UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES

MANOEL BARBOSA ALBUQUERQUE NETO

PROPOSIÇÃO DE UM SISTEMA DE BANCO DE DADOS DINÂMICO PARA AUXÍLIO ANALÍTICO E ESPACIAL À OPERAÇÃO DO TRANSPORTE PÚBLICO

FORTALEZA
2018

MANOEL BARBOSA ALBUQUERQUE NETO

PROPOSIÇÃO DE UM SISTEMA DE BANCO DE DADOS DINÂMICO PARA AUXÍLIO ANALÍTICO E ESPACIAL À OPERAÇÃO DO TRANSPORTE PÚBLICO

Monografia apresentada ao Curso de Engenharia Civil da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Civil.

Orientador: Prof. Dr. Mário Angelo Nunes de Azevedo Filho.

FORTALEZA

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A311p Albuquerque Neto, Manoel Barbosa.
Proposição de um sistema de banco de dados dinâmico para auxílio analítico e espacial à operação do transporte público / Manoel Barbosa Albuquerque Neto. – 2018.
55 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Civil, Fortaleza, 2018.
Orientação: Prof. Dr. Mário Angelo Nunes de Azevedo Filho.

1. Banco de dados. 2. Bilhetagem eletrônica. 3. Rastreamento de frota. 4. Transporte público. 5. Planejamento. I. Título.

CDD 620

MANOEL BARBOSA ALBUQUERQUE NETO

PROPOSIÇÃO DE UM SISTEMA DE BANCO DE DADOS DINÂMICO PARA AUXÍLIO ANALÍTICO E ESPACIAL À OPERAÇÃO DO TRANSPORTE PÚBLICO

Monografia apresentada ao Curso de Engenharia Civil da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Civil.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

_____
Prof. Mário Angelo Nunes de Azevedo Filho, D.Sc. (Orientador)
Universidade Federal do Ceará (UFC)

_____
Profa. Arielle Elias Arantes, D.Sc.
Universidade Federal do Ceará (UFC)

_____
Eng. José Nauri Cazuza de Sousa Junior, M.Sc.
Agência Reguladora de Serviços Públicos Delegados do Estado do Ceará (ARCE)

Aos meus pais.

AGRADECIMENTOS

Aos meus pais, com imenso carinho, pela dedicação incansável que sempre tiveram comigo e com meus irmãos em nossa formação como pessoas boas e, principalmente, pelo estímulo à nossa educação. Mesmo de longe, cada um ao seu jeito me inspira a ser alguém melhor e a dar sempre o melhor de mim.

Aos meus sobrinhos, por serem o motivo de buscar ser um exemplo como pessoa e lutar por um mundo mais justo, digno e humano.

Aos meus irmãos, André e Luana, pelo aprendizado sobre o que é a vida que proporcionam. Em especial à minha irmã, por ser sempre presente, mesmo que não fisicamente.

À minha tia Liduina, por ser a melhor segunda mãe que eu poderia ter, sempre companheira, atenciosa, afetuosa e disposta a ajudar em qualquer momento de aflição. À minha tia Lúcia, por todo o acolhimento e carinho ao longo desse árduo caminho na Universidade.

À Neidinha, responsável por transformar conversas triviais e despretensiosas em intermináveis momentos de descontração e risadas, sempre tão revigorantes ao dia a dia.

Aos meus amigos, da vida e àqueles conquistados ao longo do curso, por terem sido uma fortaleza nas adversidades e por terem compartilhados tantos momentos que tornaram a caminhada mais agradável e enriquecedora. Em especial, Rodrigo e Matheus.

Ao Geovanny, pelo apoio, companheirismo e paciência.

Aos amigos incríveis que o intercâmbio em Leeds me proporcionou conhecer.

Ao Prof. Mário pela atenção e dedicação no desenvolvimento deste trabalho. Aos demais professores do curso de Engenharia Civil que participaram de alguma forma do meu crescimento pessoal.

A vocês, meu muito obrigado!

“Δεν υπάρχει τίποτα μόνιμο εκτός από την αλλαγή.”

“Não há nada permanente, exceto a mudança”.

(Heráclito de Éfeso)

RESUMO

A utilização de dados referentes aos Sistemas de Tecnologia de Informações (ITS, na sigla em inglês) tem se desenvolvido de forma rápida e consistente nos últimos anos, sendo observadas diversas fontes desses dados, como redes sociais, mecanismos de buscas *online* e sistemas embarcados em veículos do transporte público. Estes últimos são o foco deste trabalho, a partir do sistema de bilhetagem eletrônica (SBE) e sistema de rastreamento de frota (AVL, em inglês). Pelo fato de possuírem origens e finalidades distintas, o desafio recai sobre a necessidade de associar tais dados de forma integrada com identificadores (chaves) únicos. Assim, técnicas de *big data* são utilizadas neste tipo de análise por serem capazes de tratar um elevado volume de informações numa frequência alta de observações e com possibilidade de variabilidade nesses dados. Desse modo, as particularidades dos sistemas utilizados precisam ser tratadas de forma específica para viabilizar a metodologia proposta. Objetiva-se, portanto, desenvolver um sistema dinâmico, georreferenciado, que agrega distintas bases de dados que possuem interface entre si e que possibilite o estudo dos dados para agregação de valor, um dos outros conceitos envolvidos com *big data*. Desse modo, as oportunidades para o planejamento dos Sistemas de Transporte Público de Passageiros (STPP) vem da possibilidade de realizar uma análise tanto comparativa, estudando series históricas para se estimar e avaliar cenários futuros, quanto descritiva, para entender o comportamento de alguma variável em estudo. Além disso, o ganho referente à visualização da informação torna a tomada de decisão mais assertiva para os gestores e operadores do STPP, resultando num benefício tanto para o poder público, quanto para as empresas da iniciativa privada e para os usuários do sistema.

Palavras chave: banco de dados, bilhetagem eletrônica, rastreamento de frota, transporte público, planejamento.

ABSTRACT

The use of Information Technology Systems (ITS) data has developed rapidly and consistently over the past few years, with various sources of such data being observed, such as social networks, online search engines, and embedded systems applied to public transport vehicles. The latter are the focus of this work, from the automated fare collection (AFC) and automated vehicle location (AVL). Because they have distinct origins and purposes, the challenge lies in the need to associate such data in an integrated way with unique identifiers (keys). Thus, big data techniques are used in this type of analysis because they can treat a high volume of information in a high frequency of observations and with possibility of variability in these data. In this way, the particularities of the systems used need to be treated in a specific way to make feasible the proposed methodology. The goal is to develop a dynamic, georeferenced system that aggregates different databases that have interfaces between them and that allows the study of data for value aggregation, one of the other concepts involved with big data. In this way, the opportunities for planning Public Passenger Transport Systems (STPP) come from the possibility of performing a comparative analysis, studying historical series to estimate and evaluate future scenarios, and descriptive, to understand the behavior of some variable in study. In addition, the gain in information visualization makes decision making more assertive for STPP managers and operators, resulting in a benefit both for public authorities and for private companies and for users of the system.

Keywords: database, automated fare collection, automated vehicle location, public transport, planning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AVL	*Automated vehicle location*
DB	*Database*
DET	Departamento de Engenharia de Transportes
Etufor	Empresa de Transporte Urbano de Fortaleza
GPS	*Global Positioning System*
GTFS	*General Transit Feed Specification*
RAM	*Random Access Memory*
SBE	Sistema de Bilhetagem Eletrônica
SC	*Smartcard*
SIG	Sistema de Informação Geográfica
Sindiônibus	Sindicato das Empresas de Transporte de Passageiros do Estado do Ceará
SQL	*Structured Query Language*
STPP	Sistema de Transporte Público de Passageiros
UFC	Universidade Federal do Ceará

SUMÁRIO

1 INTRODUÇÃO

A partir do contexto da urbanização, onde os espaços começaram a ser mais intensamente ocupados (VRIES, 1990), as relações sociais se acentuaram e tornaram-se mais complexas envolvendo interações fortes ou fracas com diversas finalidades (SATO; ZENOU, 2015): o trabalho, o convívio social, a educação, entre outros. Daí, tornou-se cada vez mais necessária a existência de um sistema de transportes que viabilizasse as viagens das pessoas, por diferentes motivos, de forma democrática. Nesse contexto, como alternativa à utilização de um meio de locomoção particular e individual, surgiu o Sistema de Transporte Público de Passageiros (STPP).

Além disso, por estar inserido em um sistema dinâmico como as cidades, a necessidade de um planejamento para implantação ou correção do STPP se fez cada vez mais presente. Assim, esperava-se sair de uma fase empírica para chegar num momento onde são conduzidos estudos que propiciem uma tomada de decisão mais assertiva, a partir de modelos já existentes. Onde o planejamento será dito bem-sucedido ao atingir a finalidade de atender social e economicamente o desenvolvimento da região (SOARES, 2014).

O surgimento de novas tecnologias de informação e comunicação, e a sua aplicação aos sistemas de transportes, viabilizou o surgimento de várias fontes de dados (OECD/ITF, 2015). Em seguida, a automação do processo se deu com o uso do Sistema de Bilhetagem Eletrônica (SBE) que gera diversas informações sobre os usuários (PELLETIER; TRÉPANIER; MORENCY, 2011) e sua utilização do serviço. Atualmente, para fins de planejamento, o SBE tem sido utilizado juntamente com informações espaciais, geradas pela ferramenta de localização automática da frota (Automated Vehicle Location – AVL, em inglês) (GSCHWENDER; MUNIZAGA; SIMONETTI, 2016).

O crescente uso de todas essas ferramentas tem gerado um grande volume de dados que precisam ser organizados e disponibilizados de modo que sirvam adequadamente ao processo de planejamento.

1.1 Problematização

Para alcançar o objetivo de ter um planejamento eficiente do sistema de transporte público é imprescindível que os dados utilizados para fazer as análises e caracterização dos elementos constituintes sejam confiáveis (SOARES, 2014). Assim, a origem desses dados

precisa ser de tal forma padronizada para que sempre se receba e sejam tratados de forma consistente.

Uma alternativa poderia ser a coleta manual por meio de pesquisa de campo, porém esta mostra-se bastante custosa e onerosa. Outra possibilidade, que elimina do processo a coleta manual dos dados e tem custo relativamente baixo, é o uso das informações provenientes do SBE dos STPP. Entretanto, é preciso ponderar que esse custo pode variar a depender das características dos sistemas, se já há uma estrutura pra aquisição de informações que será aproveitada ou se será desenvolvida desde o início, por exemplo. No caso dos sistemas utilizados no STPP, por registrar dados com frequência relativamente alta (no SBE, a cada usuário que passe pela catraca; no AVL, geralmente com precisão de alguns poucos segundos), se torna uma opção que necessita de um tratamento diferenciado para os dados.

Isto deve-se ao fato de que, para cada ônibus de toda a frota, ao longo do dia os dados provenientes das duas ferramentas (SBE e AVL) são armazenados e formam bancos de dados de expressivo tamanho. Ao analisar a quantidade de informações disponível, de forma que, muitas vezes, é necessário um tratamento para corrigir ou eliminar falhas, o conceito de *Big Data* se faz presente, sendo mais um complicador da proposição de análise para o STPP.

Ao compor um banco de dados consistente, integrado e dinâmico, ainda é preciso torná-lo acessível aos usuários e operadores do STPP, do contrário, não seria uma ferramenta útil e facilitadora do processo de tomada de decisão. Assim, o entendimento de métodos de tratamento de dados e de referenciamento espacial com o uso de ferramentas de Sistema de Informações Geográficas (SIG), compõe o desafio do estudo.

1.2 Justificativa

Passada a dificuldade inicial de se implementar um sistema integrado que forneça diversos tipos de informações para usuários e operadores do STPP, o produto proposto se insere de forma benéfica para todos os envolvidos no planejamento. A disponibilidade de informações centralizadas e integradas ajuda a reduzir o tempo envolvido nas diversas etapas como a de implementar ou alterar rotas de ônibus, por exemplo. Além disso, análises específicas relacionadas ao número de usuários, tipo de pagamento, ocorrência ou não de integração viabilizam maiores oportunidades de acompanhamento das tarifas adotadas e da receita gerada.

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo geral desta monografia é georreferenciar os dados provenientes do Sistema de Bilhetagem Eletrônica (SBE), utilizando como referência o Sistema de Localização Automática da Frota (AVL). Além disso, desenvolver um *script* que aceite consultas de informações de acordo com os dados inseridos, a fim auxiliar os operadores e planejadores do STPP de Fortaleza.

1.3.2 Objetivos específicos

Os seguintes objetivos específicos foram delimitados para se alcançar o geral:

- Caracterizar os dados provenientes de cada um dos sistemas a serem utilizados (SBE e AVL);
- Verificar a estrutura e validar as bases utilizadas como entrada no processo de criação do baco de dados;
- Criar uma metodologia de combinação e automação das duas bases a partir de identificadores únicos;
- Otimizar o retorno de informações por meio de busca iterativa;
- Analisar o banco de dados final em dia aleatório para validar o processo;
- Aplicar técnicas de georreferenciamento para localização das informações espaciais dos veículos do STPP;

2 REVISÃO BIBLIOGRÁFICA

Nesta seção discutir-se-á, inicialmente, os conceitos relacionados aos dois sistemas (SBE e AVL) que servirão de base para os dados analisados em seções posteriores. Continuando, será feita uma breve revisão dos conceitos de banco de dados, discorrendo também sobre a utilização desta ferramenta no planejamento de transportes. Por último, será comentado brevemente sobre o uso de ferramentas do Sistema de Informação Geográfica (SIG) para o georreferenciamento dos dados.

2.1 Sistema de Bilhetagem Eletrônica

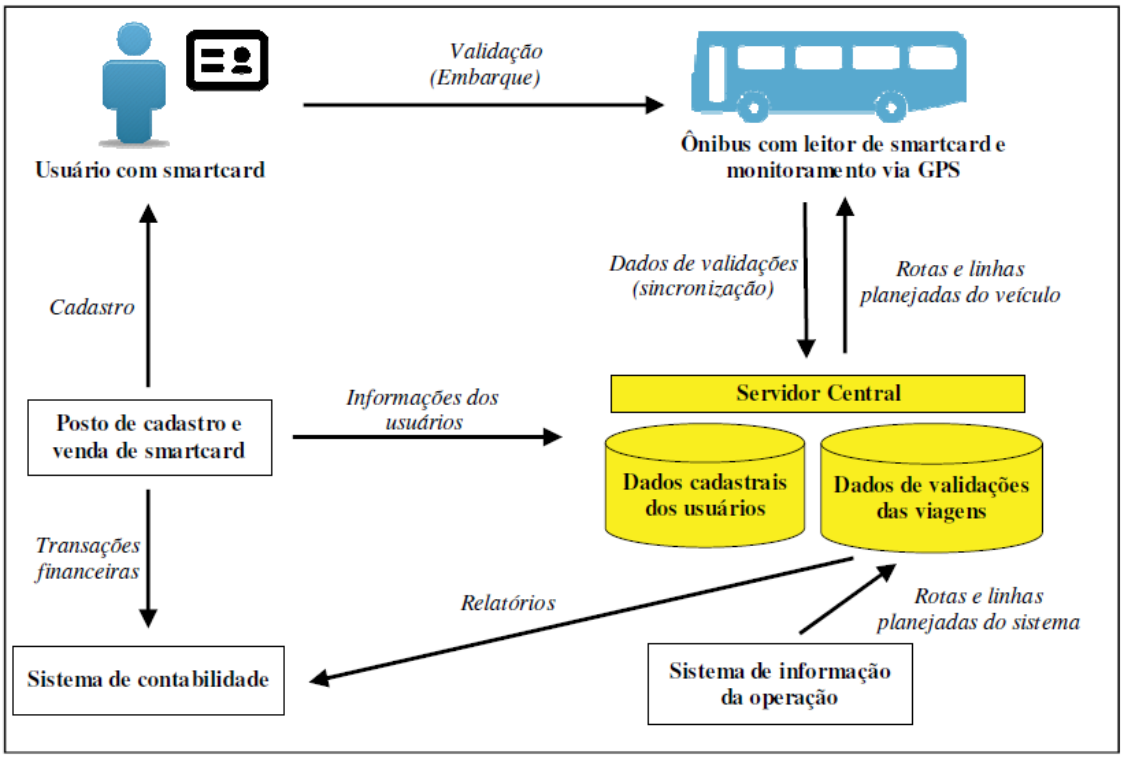
A origem do Sistema de Bilhetagem Eletrônica (SBE) está relacionada ao surgimento da tecnologia dos "cartões inteligentes" ou *smartcards* (SC), que foram desenvolvidos inicialmente no ano de 1968 pelos alemães Dethloff e Grotrupp (SHELFER; PROCACCINO, 2002). Genericamente, são cartões de plástico com um circuito integrado em um chip capazes de realizar funções pré-programadas de acordo com a programação do circuito. Em 1970, a então companhia Motorola, hoje Lenovo, aprimorou a tecnologia dos SC com o desenvolvimento de um microcontrolador capaz de proteger transações realizadas pelos cartões (PELLETIER; TRÉPANIER; MORENCY, 2011), disponibilizando comercialmente em 1977.

Com o avanço das tecnologias de comunicação, os estudos relacionados às transmissões de sinais por meio de ondas eletromagnéticas propiciaram o surgimento de uma nova tecnologia para os SC, chamada de "*contactless"* (BLYTHE, 2004). Com isso, seu uso tornou-se mais difundido e aceito nas mais diversas finalidades, sejam elas para identificação do usuário ou para transações financeiras. A Figura 1 abaixo esquematiza as etapas envolvidas no uso desta tecnologia para o sistema de bilhetagem eletrônica (SBE).

Como oportunidade para aplicação desta tecnologia no STPP, diversas cidades implementaram e fizeram testes para estudar a viabilidade da bilhetagem eletrônica (EOM, SONG, MOON, 2015; GSCHWENDER, MUNIZAGA, SIMONETTI, 2016; LI et al., 2018; PAU, 2013). Segundo Deakin e Kim (2001), a introdução de novas ferramentas no sistema de transporte é um processo complexo e que possui uma aceitabilidade lenta por parte dos usuários, além de possuir alto custo de implementação e risco associado ao investimento. Entretanto, percebeu-se benefícios significativos para todos os envolvidos no STPP.

Para os usuários, é possível adquirir passes com antecedência e usufruir de vantagens nas políticas tarifárias relacionadas a uma maior flexibilidade de acordo com categorias de usuário (usuários de vale transporte, estudantes e idosos) (MARTINELLI, J.; AROUCHA, 2012). Além disso, a bilhetagem eletrônica pode garantir um benefício social associado à possibilidade de se realizar integração tarifária, resultando em menores custos para pessoas de baixa renda.

Figura 1 – Esquema do fluxo de operação do Sistema de Bilhetagem Eletrônica (SBE).



Fonte: Adaptado de Pelletier et al., 2011.

Para o STPP, o aumento do fluxo de passageiros pelas catracas utilizando cartões eletrônicos gera redução no tempo de embarque (CHIARA-CHAVALA; COIFMAN, 1996), aumentando a velocidade de operação do sistema. Ademais, a redução do transporte de altas quantidades de dinheiro nos veículos melhora a sensação de segurança e, com o uso de SC, também há uma redução significativa em fraudes devido à extinção de passes de papel (MARTINELLI, J.; AROUCHA, 2012).

Para os gestores, a partir dos dados gerados pelo SBE, o controle da oferta do sistema tornou-se menos empírico, pela disponibilidade de informações sobre o passageiro e o destino da viagem, além da frequência realizada, por exemplo. Dessa forma, com a utilização de ferramentas para análise de alterações propostas, a operação do sistema é aprimorada entregando um melhor nível de serviço aos usuários (MCDONALD, 2000).

De acordo com Kurauchi e Schmöcker, (2017), o SBE tem se tornado cada vez mais presente nos sistemas de transporte público das grandes cidades a partir da constatação dos benefícios percebidos para os usuários, operadores e órgãos gestores. Visando o aumento da agilidade no embarque de passageiros, a adoção de linhas de ônibus sem cobrador é uma realidade em alguns países (Argentina, Canadá, Estados Unidos, entre outros) e com testes sendo realizados em cidades brasileiras, como Goiânia, São Paulo, Belo Horizonte e Fortaleza. Essa alternativa ao sistema usual de operação do transporte coletivo, com motorista e cobrador embarcados, só é possível devido às tecnologias como SBE.

2.2 Sistema de Localização Automática da Frota

Segundo Okunieff (1997), o Sistema de Localização Automática da Frota (AVL) permite que sejam utilizados dados para melhoria da performance do STPP. Isto ocorre devido à coleta, processamento e comunicação de dados associando informações de tempo e de localização espacial. Desde então, diversos estudos têm sido realizados utilizando conceitos relacionado ao AVL para avaliar e propor alternativas à operação e ao planejamento do STPP.

Os efeitos da coordenação temporal de linhas de ônibus ao longo de toda ou parte da rota representaram um importante esforço inicial desses estudos (VUCHIC, CLARKE, MOLINERO, 1981; ABKOWITZ et al., 1987; LEE, SCHONFELD, 1992; SHIH, MAHMASSANI, BAAJ, 1997). Buscou-se compreender qual era a relação entre *headways* coordenados nas linhas de ônibus e o nível de serviço em relação à confiabilidade nos horários indicados no cronograma da operação. Além disso, outros autores analisaram o uso de dados provenientes do AVL na integração com o modal ferroviário. Hall (1985) examinou essas informações para desenvolver fórmulas para intervalos ótimos entre o desembarque em um ônibus e embarque num trem, e vice-versa.

Utilizando a base de dados da TriMet, agência de transporte público de Portland, Oregon (Estados Unidos), Berkow et al. (2007) conduziram um estudo para avaliar a performance daquele STPP. A partir da automação advinda do uso do AVL, a TriMet passou a coletar e armazenar um grande volume de informações dos ônibus e linhas em operação. Com isso, os autores utilizaram tais series de dados para avaliar e aprimorar o nível de serviço e, em seguida, sugeriram que estudos futuros utilizassem as informações disponíveis para centralizar as informações em um banco de dados operacional.

Continuando na linha que este trabalho seguirá, Arbex e Cunha (2016) ressaltam a importância dos dados de localização dos veículos por meio do sistema de rastreamento da frota (AVL). Além disto, se diferenciam dos demais estudos citados por ressaltarem a vantagem que a disponibilidade de series históricas representa para a comparação entre cenários anterior e posteriores a mudanças operacionais (por exemplo, a implantação de faixas exclusivas e alteração nos ciclos semafóricos).

2.3 *Big data* e o planejamento de transportes

Segundo Khoury e Ioannidis (2014), a definição de *big data* refere-se à um grande volume de informação que pode estar conectado a diversas partes de uma estrutura complexa de dados. Assim, com o avanço tecnológico das últimas décadas, a crescente disponibilidade de informações de fontes e sistemas diversos estimulou a produção acadêmica e o interesse em *big data*, como visto pelo aumento significativo de publicações nos últimos anos (LIU et al., 2016). Redes sociais, mecanismos *online* de buscas, registros de ligações telefônicas, sistema de posicionamento global (*Global Positioning System*, GPS), AVL, SBE, são exemplos de fontes de informação utilizados em estudos relacionadas à *big data*.

De forma geral, a literatura conceitua o que é *big data* a partir do modelo dos "Vs". Inicialmente, Laney (2001) utilizou os 3Vs para essa conceituação: volume (quantidade e abrangência dos dados), velocidade (celeridade que as informações podem ser geradas ou observadas) e variedade (tipo e natureza do que está sendo coletado). Em seguida, Assunção et al. (2014) incluíram dois Vs ao modelo anterior: valor (ganhos e ideias derivados da análise dos dados) e veracidade (confiabilidade das informações utilizadas). Posteriormente, outros estudos propuseram a inclusão de novos termos, tais como: variabilidade (possibilitando a adaptação e utilização dos bancos de dados em modificações pontuais), visualização (como auxílio à análise dos dados) e validade (concordância entre informações utilizadas e a finalidade da ferramenta) (KHAN, UDDIN, GUPTA, 2014; MCNULTY, 2014; PETTIT et al., 2018).

Dos conceitos indicados anteriormente, *big data* é capaz de agregar grande valor aos projetos e negócios que consigam extrair seus benefícios. Fosso Wamba et al. (2015) indicaram cinco dimensões a essa agregação de valor: (i) criar transparência no negócio, (ii) possibilitar testes para melhorar a performance, (iii) customizar ações para populações segmentadas, (iv) auxiliar tomadas de decisão com o uso de algoritmos e (v) inovação em produtos e serviços. Partindo desses cinco aspectos, é possível estabelecer um paralelo com o

planejamento de transporte públicos que, apesar de não estar integralmente com a iniciativa privada, precisa ser rentável, de qualidade e confiável.

Apesar dos benefícios citados anteriormente, Milne e Watling (2018) citam algumas dificuldades que o uso de *big data* pode trazer na fase de planejamento de transportes devido à complexidade do método e ferramentas utilizados. Para que atenda ao conceito dos Vs, principalmente veracidade e validade, é necessário que os dados sejam continuamente monitorados, garantindo que não ocorram falhas ou *gaps* temporais na coleta. Um outro fator que pode impactar a viabilidade do uso de grandes sistemas de dados refere-se à posse dos dados utilizados pois é comum que os usuários ou analistas dos dados não sejam os que possuem legalmente a posse dessas informações. Ademais, um desafio recorrente é a inexistência de uma base de dados contendo exatamente o conteúdo que se deseja analisar, desse modo há a necessidade de se adaptar essas bases para a finalidade desejada.

2.4 Sistema de informação geográfica (SIG*)*

A utilização da expressão "Sistema de informação geográfica" surgiu na década de 1960 e, desde então, evoluiu de diversas formas para ser desde um procedimento para análise de informações no espaço geográfico até ser amplamente utilizada em *softwares*. Por essa variedade em representar fenômenos diversos de forma espacial, alguns estudiosos se referem ao conceito de SIG não como um sistema mas sim uma ciência (GOODCHILD, 2010). Estes autores defendem que a manipulação geográfica de informações para representar outros fenômenos precisa ser estudado e desenvolvido em paralelo com o surgimento de novas áreas, não apenas meramente como ferramenta espacial destas.

Na área da ciência de transportes, a nomenclatura inicial foi adaptada de SIG para SIG-T, indicando que esse segmento se tornou uma das principais aplicações do SIG (SHAW, 2011). A distinção das aplicações e usos se tornou mais especializada, sendo desenvolvidas metodologias e técnicas próprias, procedimentos de análise e adaptações para abordar as particularidades dos fenômenos de transportes. Com isso, não demorou para que o planejamento urbano de transportes, tão importante com a expansão da população, fosse visto a partir da ótica do SIG-T (MALIENE et al., 2011). A representação espacial utilizando a composição de múltiplas camadas como malha viária, delimitações de uso do solo, número de habitantes, entre outros, tornou o planejamento urbano mais efetivo e próximo à realidade estudada.

3 METODOLOGIA

O método adotado para se alcançar os objetivos delimitados anteriormente baseia-se na descrição dos tipos de dados utilizados e, posteriormente, na explanação das possibilidades e utilidades do sistema de banco de dados finalizado. Em primeiro lugar, após a aquisição das bases dos dados, delimita-se quais destas serão utilizadas. Em seguida, é necessário definir qual classe ou campo de informações servirá de "chave" para fazer a correspondência entre os dados de diferentes origens e possibilitar a espacialização dos dados. Posteriormente, tais as informações provenientes do sistema de bilhetagem eletrônica e do sistema de rastreamento da frota podem então ser visualizadas e analisadas de forma a ajudar o planejamento e a operação dos sistemas de transporte público. Por último, lista-se sucintamente os *softwares* e o computador utilizado no estudo.

3.1 Formato dos dados

A estrutura do SBE e do AVL compreende, dentre outros, os seguintes dados:

- Horário;
- Código do cartão;
- Identificação do veículo;
- Identificação da linha;
- Identificação do sentido;
- Identificação da viagem;
- Identificação de integração de viagem;
- Latitude;
- Longitude;

Um dos principais desafios na utilização de bancos de dados, principalmente quando são de origens ou fontes distintas, é a adequação das informações disponíveis para aquilo que se deseja estudar ou analisar. Desse modo, este trabalho possui, dentre outras finalidades, a compatibilização e adequação das informações dessas duas fontes de dados. Isto é necessário porque, o *modus operandi* dos sistemas do SBE e do AVL difere na forma de registrar os dados. Por exemplo, enquanto, para uma determinada viagem, serão registrados os acessos pela catraca em um dado ônibus de uma linha, esse mesmo ônibus pode, ao longo do dia, trocar de linha em operação, porém continuará com um mesmo GPS associado. Assim, é

preciso identificar e associar o número do ônibus ao código utilizado pelo sistema de rastreamento de forma a evitar anomalias na análise dos dados.

Este procedimento ocorre com a utilização de uma base de dados que contém pares de códigos relacionando o número do ônibus para o rastreamento e o número de identificação junto às empresas de ônibus (o mesmo número indicado na lataria dos veículos). Essa base, chamada de "DICIO", tem importância fundamental para o desenvolvimento da ferramenta proposta neste trabalho. Será discutido no decorrer do texto quais os impactos negativos ao se utilizar esses dados defasados em relação às datas das demais bases.

Para complementar esse processo de pareamento dos dados, também se utiliza de informações disponibilizadas pelos STPP no formato GTFS (*General Transit Feed Specification*). São dados provenientes do GTFS: empresas de transporte público, pontos de parada, rotas, horários previstos de parada, calendarização de operação das linhas, atributos relacionados ao preço das passagens, frequência e localização de operação da linha. Esta estrutura auxilia na identificação de elementos individuais de cada linha para que seja feita uma análise desagregada em relação ao conjunto dos dados.

3.2 Espacialização dos dados

O processo de associação dos dados do SBE e do AVL descrito no item anterior possibilita que se agreguem informações distintas de modo a complementar a qualidade e a eficácia dos dados a serem utilizados na análise. A partir disto, torná-los referenciados geograficamente representa um ganho relativo à espacialização dos dados de forma a facilitar e auxiliar o processo de tomada de decisão do planejamento do STPP. O georreferenciamento possibilita também a avaliação de cenários atuais, assim como a proposição de situações futuras. Para isso, os *softwares* SIG são ferramentas importantes no tratamento espacial desses dados.

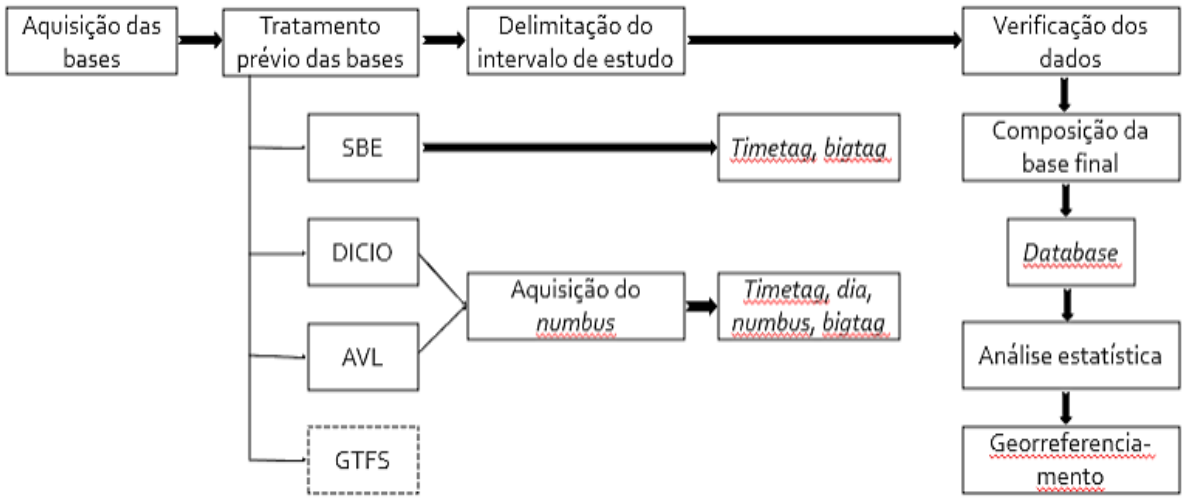
Em cartografia e demais ciências que buscam representar a superfície da Terra, é fundamental que dados de origens diferentes utilizem o mesmo *datum*. Este termo refere-se ao modelo matemático de representação da superfície tridimensional da Terra, ao nível do mar, em uma projeção bidimensional. Desse modo, o uso de coordenadas em um mesmo *datum* evita a ocorrência de distorções e de grande discrepância entre a localização real de um ponto e a sua projeção em outro sistema de coordenadas.

O sistema de referência espacial adotado utiliza a projeção SIRGAS 2000 (*Sistema de Referencia Geocentrico para las Americas*) como modelo cartográfico. Por possuir uma área mais reduzida e focada na América do Sul e Central, evita-se maiores distorções geométricas nas camadas analisadas. Esta escolha está em conformidade com a Resolução nº 1/2005 do Instituto Brasileiro de Geografia e Estatística (IBGE), baseado no decreto presidencial nº 5334/2005, que institui o SIRGAS 2000 como sistema de referência oficial.

3.3 Visualização das informações

Um recurso que deve ser explorado quando se utilizam dados de sistemas de transporte e de localização espacial é a visualização dessas informações na forma de mapas temáticos, para fins de monitoramento ou avaliação do STPP, ou rotas, para os usuários. Assim, a possibilidade de se traduzir o georreferenciamento dos dados de rastreamento com *softwares* SIG é uma etapa importante da metodologia proposta neste trabalho. A Figura 2 esquematiza de forma simplificada as etapas adotadas ao longo do trabalho como metodologia.

Figura 2 – Esquema da metodologia utilizada



Fonte: Elaborado pelo autor.

A introdução dessa abordagem de visualização dos dados coletados também viabiliza e serve de ponto de partida para técnicas mais sofisticadas como a da estatística espacial. Tem-se, por exemplo, a possibilidade de conduzir estudos mais avançados considerando a forma como os dados estão distribuídos espacialmente, visando identificar

correlação espacial com o objeto de estudo; estimar variáveis no espaço por meio de amostras bidimensionais; simulação do comportamento diferencial em fenômenos de transporte; etc. Apesar da análise espacial avançada não ser um dos focos deste estudo, o uso de tal abordagem analítica é importante pois permite que os indicadores dos STPP sejam explorados de forma espacial.

3.4 Ferramentas utilizadas

Para fins de especificação técnica e comparações posteriores de processamento, foram utilizados:

- Notebook Intel® Core ™ i5-6200U CPU @ 2,3GHz. Memória RAM 4GB DDR3. Sistema operacional de 64 bits, Windows 10 Home
- Python 3.7
- PostgreSQL 10
- QGIS Desktop 3.4.1

4 COLETA E DADOS UTILIZADOS

Conforme mencionado nas seções anteriores, a tipologia dos dados utilizados no estudo é inerentemente de grande quantidade de informações e esses dados ocupam espaços consideráveis no armazenamento do computador. No caso do AVL, possui registro periódico de mais de 2 mil ônibus, em uma frequência de aproximadamente 30s. Para o SBE, contém informação sobre cada validação de passageiros nas catracas dos ônibus. Os arquivos não poderiam, portanto, ser gravados em mídias digitais convencionais (CD, DVD ou DVD *dual layer*, *pen drives*), pois esbarrariam na limitação física de memória disponível. A utilização de mídias mais modernas, como *Blu-ray*, supriria a limitação anterior, com ressalvas a serem esclarecidas em seguida, mas trariam a limitação de requerer tecnologias mais avançadas para leitura de tais discos.

4.1 Definição dos dados

Apesar de possuírem fornecedores distintos, os dados referentes aos diversos sistemas relacionados ao transporte público de passageiros são centralizados na Etufor e Sindiônibus e, a partir daí, algumas bases foram repassadas para utilização nas pesquisas no Departamento de Engenharia de Transportes (DET) da UFC. No objetivo de propor uma ferramenta para facilitar o estudo e a análise dessas informações, aplicou-se as etapas metodológicas apenas para um mês de dados,, visto que esta pode ser replicada para quaisquer outros meses do ano desejado. Para este trabalho, as bases inicialmente utilizadas foram as de bilhetagem eletrônica (i) e de rastreamento (ii) ambas de setembro de 2014 e o dicionário de veículos de junho de 2017 (iii).

No início deste capítulo, foi abordada a característica do tamanho dos arquivos a serem utilizados. Para se ter uma noção quantitativa, em (i) cada dia útil representa um arquivo de tamanho médio de 96 Megabytes (MB), enquanto sábados e domingos, 64 e 22MB, respectivamente. Em (ii), para cada mês, o arquivo possui em média 23 Gigabytes (GB). Para se registrar um ano de dados relativos aos dois sistemas, seriam necessários aproximadamente 300GB. Limitando ao intervalo de um mês, os dados foram transportados em disco rígido (HD) externo, o que reduziu a possibilidade de fragmentação e consequente perda de dados, assegurando a integridade das informações a serem tratadas.

O arquivo indicado em (iii) pode ser considerado como chave para possibilitar a associação das duas bases (i e ii). Nele, são feitos os devidos vínculos entre o código de um ônibus no sistema de rastreamento e qual seu respectivo código na base de bilhetagem. Posteriormente verificou-se que, devido à importância que esta relação representa, este arquivo evidencia um fator fundamental nas análises almejadas por este estudo.

4.2 Tratamento das bases

Um dos principais objetivos de fazer manipulações nas bases é otimizar o tempo de processamento reduzindo dados não desejados ou que podem ser obtidos de uma forma simplificada. Além disso, também é preciso fazer modificações na estrutura para extrair informações adicionais que não estão explicitamente indicadas. A nomenclatura adotada para indicar ou fazer referência aos componentes das bases será similar ao da linguagem SQL (*Structured Query Language*) que será utilizada em uma das etapas finais. A saber:

- Banco de dados: *database*
- Bases ou arquivos utilizados: tabelas ou *tables*
- Colunas: campos ou *fields*
- Linhas: registros ou *rows*

4.2.1 SBE

A validação de usuários nos ônibus do sistema público de transportes guarda informações diversas conforme citado em seções anteriores (dados do *smartcard*, caso seja utilizado, da linha e do horário de embarque). São nove campos que possuem tipos distintos em cada um deles, como: números inteiros, *string* (cadeias de texto) e data. A remoção dos campos redundantes ou que não estão diretamente relacionados à análise dos dados serve para reduzir o espaço ocupado no HD e agilizar as etapas posteriores. Em seguida, são criados dois novos campos para inserir informações úteis para o tratamento da tabela: *timetag* e *bigtag*. A Figura 3 abaixo ilustra a estrutura inicial e final após os procedimentos de manipulação da base.

O *timetag* é uma variável criada para representar o horário da validação em intervalos de 30s. Isto é, dividir os 86400s do dia em trechos de 30s, conforme equação 1 abaixo.

$$\text{timetag [s]} = \frac{hora * 3600 + minuto * 60 + segundo}{30} \qquad \text{Equação 1}$$

Já o segundo campo criado, *bigtag*, é uma composição (concatenação de termos), conforme equação 2, que visa estabelecer um modelo de chave para a junção com a base de rastreamento. Essa relação pode ser unitária ou não, isto é, em um intervalo de 30s, para um mesmo ônibus, pode ter ocorrido apenas uma validação ou mais de uma. No primeiro caso, será uma chave única e no segundo, uma compartilhada, pois as validações compartilham um mesmo *timetag*.

bigtag = concat('1' & dia & mês & ano & timetag & numbus) Equação 2

Para exemplificar a equação 2 e como o *bigtag* é criado, o registro a seguir pode ser considerado: um veículo qualquer de número 14002, trafegando no dia 07/09/2014 às 13:41:26. Inicialmente, o primeiro termo "1" é inserido à cadeia de caracteres para evitar que informações sejam perdidas (no caso de uma sequência inicial de zeros, por exemplo) em processamentos futuros. Em seguida, toma-se o dia e mês com dois caracteres e ano, com quatro. A etapa seguinte é transformar o horário em *timetag* conforme equação 1. Por último, concatena-se o número do ônibus. Assim, no exemplo fornecido, o *bigtag* seria a união dos termos: "1", "07", "09", "2014", "1643" e "14002". Resultando no termo: 107092014164314002.

Figura 3 – Mudança na estrutura das bases do SBE antes e pós modificação da estrutura.

SBE$_{INÍCIO}$			**SBE$_{FIM}$**	
sigom	Número do *smartcard* (se utilizado)		**bigtag**	**Equação 2**
nlin	Número da linha		**sigom**	-
nomelin	Nome da linha	✖	**nlin**	-
nbus	Número do ônibus		**nbus**	-
datahora	Data e hora		**datahora**	-
tipcar	Código do tipo de pagamento		**tipcar**	-
desc_pgto	Descrição do tipo de pagamento	✖	**sent**	-
sent	Sentido da linha		**int**	-
int	Realização de integração temporal		**timetag**	**Equação 1**

Fonte: Elaborado pelo autor.

Os arquivos de SBE que anteriormente possuíam em média 96MB por dia de coleta, sofrem uma redução de aproximadamente 21%, já considerando a configuração final (SBE$_{FIM}$). Como pôde ser visto na Figura 3 anterior, os campos excluídos da tabela final eram cadeias de caracteres que descreviam o nome da linha representado pelo código (cod_lin) e a descrição do

pagamento a partir do código do tipo (tip_pgto). Por serem frequentes em todos os dias do mês, esses dois campos foram transformados em *tables* independentes: “codpgto” e “codlin”. Assim, é possível consultar a quais descrições cada código representa, além de tornar mais eficiente a aquisição de informação devido à redução do tamanho do arquivo principal.

4.2.2 AVL

A tabela de rastreamento possui um tamanho que inviabiliza que sua leitura seja feita normalmente (a partir de *softwares* de planilhas como o Microsoft Excel, por exemplo) utilizando a memória primária (RAM) do computador. Isto significa que, para nenhum computador convencional, haverá memória RAM suficiente que possibilite a leitura completa e simultânea de um arquivo contendo um mês de registros, aproximadamente 23GB. Dessa forma, os dois objetivos anteriores de tratamento das bases se tornam insuficientes, sendo necessário aplicar técnicas mais avançadas de tratamento nos dados.

Para contornar a limitação de memória dos computadores convencionais (como o utilizado neste trabalho), algumas linguagens de programação oferecem soluções e ferramentas mais apropriadas para cada caso. Do ponto de vista de memória RAM disponível, um fator comum para que seja superada essa restrição de armazenamento e leitura de dados, é a utilização de blocos (*chunks*) menores que consigam ser operados e tratados individualmente. Esta seria a etapa anterior às descritas no tópico 4.2.1. Desse modo, a memória só é comprometida até um limite determinado pelo usuário, onde pode-se ajustar, por exemplo, blocos maiores utilizando um tempo de processamento maior ou *chunks* menores que conseguem ser lidos de forma mais rápida. Pelos testes realizados no trabalho, este último é mais indicado em etapas iniciais, que requerem agilidade na identificação e tratamento de erros. O primeiro é mais recomendado quando já se possui um código robusto que irá tratar corretamente as informações, fazendo um menor número de pausas para fragmentar em blocos.

Esta base e, consequentemente, cada bloco, é composta inicialmente por 43 campos. Possuindo os seguintes tipos de dados: a parte principal para este estudo possui uma maioria numérica (inteiros ou decimais), um campo contendo o registro de data e hora, uma cadeia de caracteres que registra um campo geométrico especial (a ser melhor discutido na seção pertinente aos dados espaciais) e inúmeros campos contendo caracteres que podem ser ignorados. Estes não agregarão qualquer valor à análise por conterem caracteres relacionados a funcionalidades do equipamento de rastreamento que não estão ativas no caso de Fortaleza.

Para cada bloco, são efetuados processos iterativos para remover os campos que não serão utilizados e, em seguida, adicionar novos dados. Serão mantidas 10 das 43 colunas iniciais e adicionadas 4 em duas etapas distintas descritas a seguir. A primeira etapa refere-se à inclusão do *timetag* conforme descrito pela equação 1 e de uma nova coluna que indica o dia do registro de cada rastreamento. A segunda etapa aborda uma importante operação para a continuidade da análise, fazendo a primeira junção (*merge*) de duas bases distintas: o dicionário de veículos indicado em 4.1 (iii) e o rastreamento de veículos, utilizando como chave o "vehicleid". Este procedimento é fundamental para que seja possível identificar corretamente cada ônibus, ou seja, possibilita que cada registro de rastreamento carregue o código do ônibus utilizado pelas empresas de transporte (Sindiônibus, Etufor, prestadoras de serviços) e não apenas o código do aparelho individual de rastreamento (da empresa M2M Solutions). Para finalizar a segunda etapa de tratamento da base AVL, é inserido o campo contendo o *bigtag*, descrito no item anterior (equação 2). A Figura 4 abaixo ilustra os procedimentos mencionados.

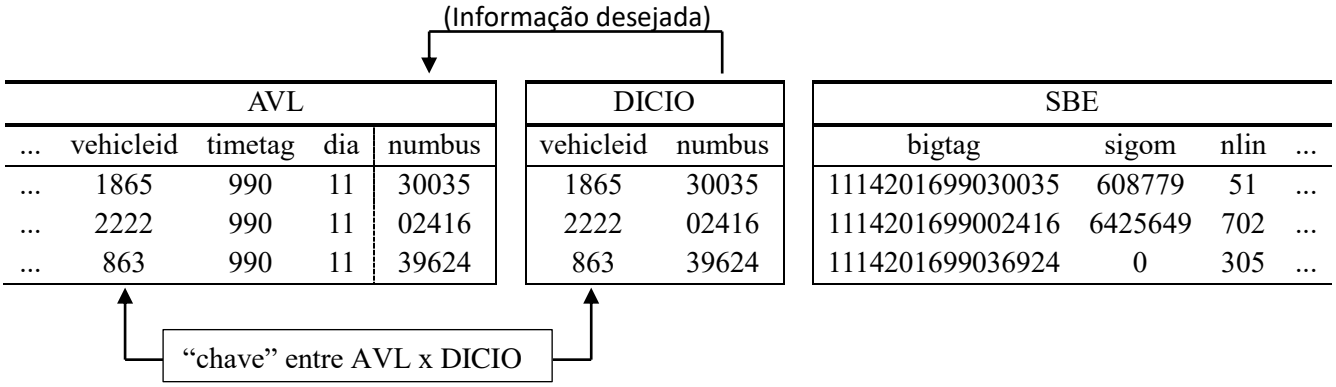
Figura 4 – Pseudocódigo ilustrando o código escrito para tratar a base AVL

```
PARA CADA arquivo_AVL:
    LER ARQUIVO EM BLOCOS (tamanho = 524288 linhas)
    PARA CADA BLOCO:
        DELETAR COLUNAS INDESEJADAS
        CRIAR COLUNA TIMETAG (equação 1)        } Etapa 1
        CRIAR COLUNA DIA

        FAZER JUNÇÂO COM A TABLE DICIO
        CRIAR COLUNA BIGTAG (equação 2)         } Etapa 2

        FAZER JUNÇÃO COM A TABLE SBE*

        SALVAR A TABELA RESULTANTE
    FIM PARA
FIM PARA
```

Fonte: Elaborado pelo autor. A junção com a base SBE (*) será descrita no próximo item.

A compatibilização mencionada está ilustrada no exemplo da Figura 5 abaixo. O campo "vehicleid" presente tanto na base AVL quanto no DICIO, faz com que seja possível identificar o "numbus" de cada registro no arquivo de rastreamento. Após o tratamento da tabela de AVL, a redução no tamanho do arquivo é de aproximadamente 40%, reduzindo em 10GB no mês observado. Em seguida, com todos os termos do *bigtag* explicitados, é possível fazer uma segunda junção, dessa vez entre a base AVL e a do SBE, a ser discutida a seguir.

Figura 5 – Relação de identidade entre campos de diferentes *tables*.



Fonte: Elaborado pelo autor.

4.3 União dos arquivos após tratamento

Após os arquivos terem sidos formatados e modificados individualmente, a etapa seguinte se trata da junção dos dados de rastreamento e de bilhetagem eletrônica. O produto final deste procedimento é um banco de dados contendo dois tipos de informações distintas que não seriam obtidas de forma separada. A chave nesse caso foi o campo criado em ambas as bases: o *bigtag*. O tipo de união realizado pode resultar em estruturas bastante distintas entre si, por isso, é importante compreender como se deu a escolha do *output* desejado. Uma das opções seria mesclar as duas tabelas mantendo uma ou outra como padrão, conservando sua estrutura (colunas). Nesse caso, devido à natureza dos dados de rastreamento (alta frequência de gravação para cada um dos mais de 2 mil ônibus), existe um número bastante superior de registros de rastreamento em detrimento dos dados de bilhetagem. Dessa forma, uma fração pequena do campo *bigtag* seria combinada, gerando muita informação incompleta. Outra possibilidade, parcialmente empregada neste estudo, toma apenas os dados comuns em ambas as *tables* a partir da chave adotada. Com isso, a obtenção dos dados é completa por associar com êxito a validação do usuário no transporte público com uma localização geográfica e é também concisa, pois reduz o tamanho do banco de dados final.

A delimitação do intervalo de estudo mencionada no início do capítulo, referiu-se à possibilidade de se replicar a mesma abordagem para trechos quaisquer ao longo de um ano, contanto que se tenha disponível todas as bases necessárias. O período escolhido para fazer o estudo foi o mês de setembro de 2014, por conter dados de bilhetagem eletrônica de todos os dias, assim como toda a rastreabilidade dos veículos no mês. O que aconteceu, entretanto, não correspondeu ao esperado por não ter retornado uma quantidade significativa de pontos. Em

outras palavras, enquanto era esperado que houvesse um registro no banco de dados para cada usuário de ônibus no STPP de Fortaleza no referido mês, esse comportamento não aconteceu.

Alguns questionamentos foram então levantados para compreender as possíveis causas desse problema. Um dos maiores desafios presentes desde o início do estudo, foi a visualização de cada um dos registros. O próprio conceito de *big data* se refere a um conjunto de dados muitas vezes incapaz de ser analisado pontualmente e quase sempre inviável de se verificar por *softwares* como o Microsoft Excel. Mesmo com a utilização de ferramentas de processamento e manipulação de dados, utilizando bibliotecas específicas de programação, se mostrou uma tarefa laboriosa devido ao tempo de computação exigido para cada verificação. Após análise das etapas de verificação em cada processo mencionado anteriormente, concluiu-se que a falha estava no momento de fazer a última junção entre as bases, ou seja, estas estavam incompletas ou as chaves utilizadas não estavam encontrando correspondência. Finalmente, foi verificado que, na base de rastreamento, muitos ônibus não estavam indicando corretamente o campo "numbus", ou seja, a base DICIO utilizada não contemplava todos os ônibus que estavam em circulação no momento do rastreamento.

A distinção temporal entre as datas do dicionário (junho de 2017) e dos dados obtidos de rastreamento (setembro de 2014) tornou inviável o processo de identificação entre os campos "vehicleid" e "numbus", impossibilitando a continuidade da análise do modo como planejado. Cada base, separadamente, possui informações que propicia, como será visto no capítulo 5, algum tipo de estudo sobre o transporte público de passageiros, como por exemplo: AVL – distribuição geográfica de ônibus, velocidades exercidas, distribuição e frequência temporal dos veículos; SBE – análise das horas com maior número de validação, linhas mais utilizadas, formas de pagamento mais adotadas, entre outros. Entretanto, como proposta de ferramenta para uso no planejamento e operação do STPP, este trabalho busca a integração destes dados. Assim, para demonstrar a eficácia do método proposto, deu-se prosseguimento ao estudo empregando as mesmas operações citadas anteriormente, porém utilizando os dados do mês de abril de 2016. A estrutura do banco de dados final está indicada na Tabela 1.

Tabela 1 – Estrutura final do banco de dados.

Campo	Descrição
metricid	Identificador único para o AVL
calcspeed	Velocidade instantânea do veículo
direction	Direção em graus do veículo
lat	Latitude
long	Longitude

Campo	Descrição
timestamp	Data e hora (AVL)
odometer	Odômetro
the_geom	Campo geométrico
vehicleid	Código do veículo para o rastreamento
timetag	Campo adicionado (equação 1)
dia	Campo adicionado – dia do registro
numbus	Número do ônibus para a frota (5 dígitos)
bigtag	Campo adicionado (equação 2)
sigom	Código do *smartcard* utilizado (se utilizado)
nlin	Número da linha em operação
datahora	Data e hora (SBE)
tipcar	Tipo de pagamento efetuado
sent	Sentido da linha
integr	Realização de integração temporal

Fonte: Elaborado pelo autor.

5 ANÁLISE DOS DADOS

Neste capítulo será inicialmente explicitado como foi realizada a determinação da amostra dos dados. Em seguida, para cada base utilizada, serão analisadas informações geradas provenientes dos dados disponíveis em sua forma bruta, ou seja, antes da realização do tratamento das bases (Capítulo 4). Posteriormente, com o auxílio da linguagem SQL para tratamento de banco de dados com campos geográficos e do *software* QGIS, serão apresentadas as análises realizadas a partir do produto final do capítulo anterior.

5.1 Delimitação da amostra analisada

Após identificação da falha que impossibilitou a utilização dos dados de setembro de 2014, foram utilizados arquivos alternativos de outros períodos que também estavam à disposição para estudo. Essa mudança não exigiu alteração nas linhas de programação porque já foram escritas para considerar o *input* de arquivos distintos e que podem ser alterados a qualquer momento pelos usuários. Procedeu-se, entretanto, a mais uma análise das bases que foram tratadas e geradas conforme descrito no capítulo anterior. Essa verificação serviu para assegurar que a mudança dos arquivos, por serem de anos distintos, por exemplo, possuíssem a mesma estrutura, não resultando em erros para o estudo. Uma vez concluídas as etapas de verificação de erros, cada base tratada poderia seguir para a geração dos bancos de dados.

Como mencionado anteriormente, a princípio foi considerado avaliar integralmente um intervalo de um mês de informações disponibilizadas para o DET. Porém, a abordagem adotada objetivou trazer resultados mais próximos daqueles a serem futuramente abordados por quaisquer usuários ou operadores que desejam analisar bancos de dados gerados a partir da união dessas duas fontes de dados (AVL e SBE). Isto é, para cada uma dessas bases originais é possível extrair conhecimentos distintos e, ainda mais, após a junção delas. De forma resumida, foram exploradas características gerais da utilização do sistema de transporte público para, em seguida, limitar os horizontes de estudo da pesquisa.

A restrição identificada na base “DICIO”, três anos mais recente que os dados utilizados a princípio, direcionou a escolha para que se optasse por utilizar outras bases que tivessem a data mais próxima possível daquela da base limitante (“DICIO”). Entretanto, em busca de otimizar o tempo de computação do código e dar continuidade ao trabalho, enquanto eram processadas as novas informações, foram realizadas a delimitação e a filtragem da amostra

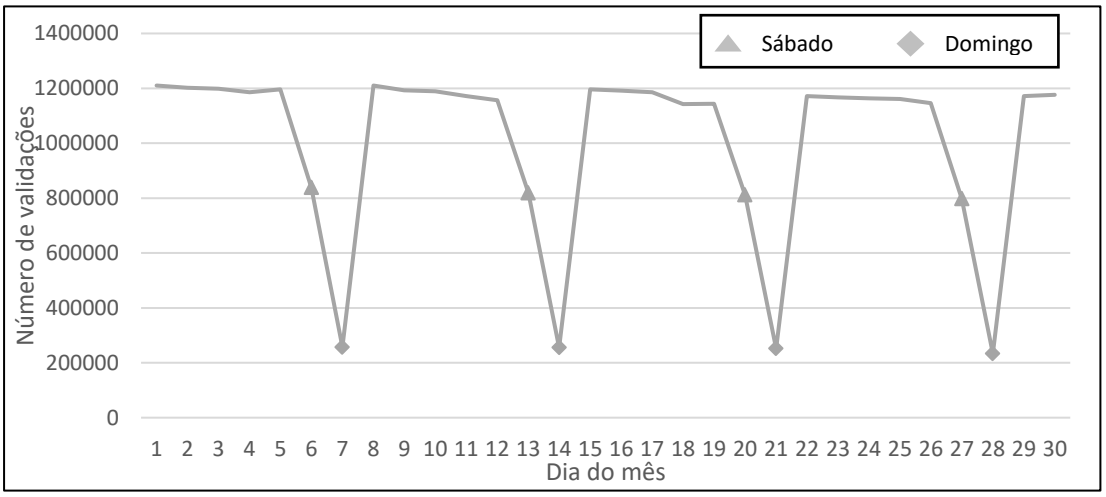
(com as bases anteriores). Por se tratar da utilização do STPP, determinados períodos do mês e do dia possuem maior representatividade no contexto geral do transporte coletivo urbano. Foi definido, assim, o comportamento da utilização de ônibus ao longo do mês observado, dias da semana de maior e menor número de usuários, intervalos do dia com picos de validação, linhas de maior relevância para o sistema, modalidade de pagamento utilizado, ocorrência de validações provenientes de integração temporal, entre outros.

5.2 Validações de usuários – SBE

A análise quantitativa dos dados de bilhetagem eletrônica fornece embasamento prático para os operadores do STPP. Isto se deve ao fato destes proporcionarem um maior conhecimento e familiaridade com os hábitos de utilização do sistema no qual operam. Dessa forma, o planejamento se torna mais assertivo e as tomadas de decisões são mais realistas, por considerarem um comportamento observado e mensurado quantitativamente.

Neste estudo, buscou-se fazer a caracterização do comportamento dos usuários no STPP para o mês de setembro de 2014. Neste mês, foram 30.249.248 usuários que tiveram seu embarque registrados no SBE, seja nos ônibus ou terminais de passageiros, tanto para passagens de integração temporal quanto novas validações. Esse valor resulta em uma média diária de 1.008.308 usuários. Porém, observando a utilização dos veículos do transporte público, é evidente que em determinados dias da semana esse comportamento se distancia do valor médio. A Figura 6 abaixo ilustra o número de validações por dia, ao longo do mês.

Figura 6 – Número de validações por dia (setembro/2014)

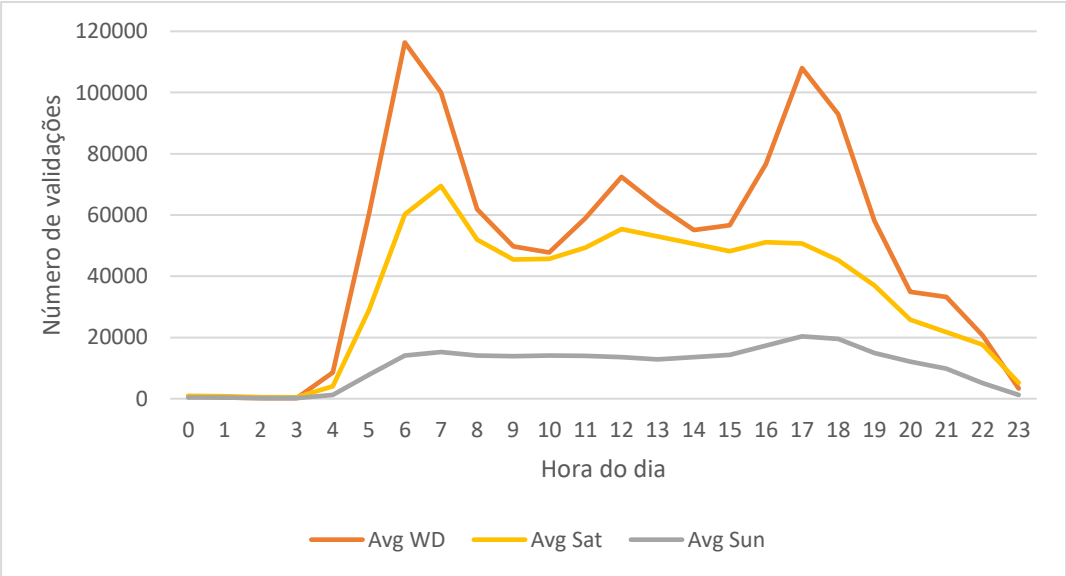


Fonte: Elaborado pelo autor.

A primeira característica identificada se refere ao número de usuários aos finais de semana. Há uma redução média de 30,6% e 78,8% aos sábados e aos domingos, respectivamente, em relação aos dias úteis. Essa redução é confirmada, ao fazer a comparação individualmente, semana a semana. Além disso, há uma tendência de que, ao longo da semana e ao longo do mês, o número de validação decresça. Isto é, no início da semana há um número maior de usuários, assim como no início do mês (em relação ao final). Este comportamento pode ser estudado em maior profundidade em trabalhos futuros mas se sugere abordar, como ponto de partida, as datas em que os pagamentos dos usuários que possuem empregos sãos realizada, por exemplo.

Ao buscar caracterizar a forma como os usuários utilizam o transporte público de ônibus ao longo de cada dia, os resultados mostraram mais uma vez que há um comportamento preponderante em relação às horas de maior número de validações. A Figura 7 mostra a média de usuários por hora do dia. No entanto, já incluindo a distinção verificada no parágrafo anterior, é preciso discernir entre os dias da semana (*weekdays* – WD), que possuem semelhanças entre si, e os dias sábado e domingo.

Figura 7 – Número médio de validações por hora, por categoria (WD, Sábado e Domingo)

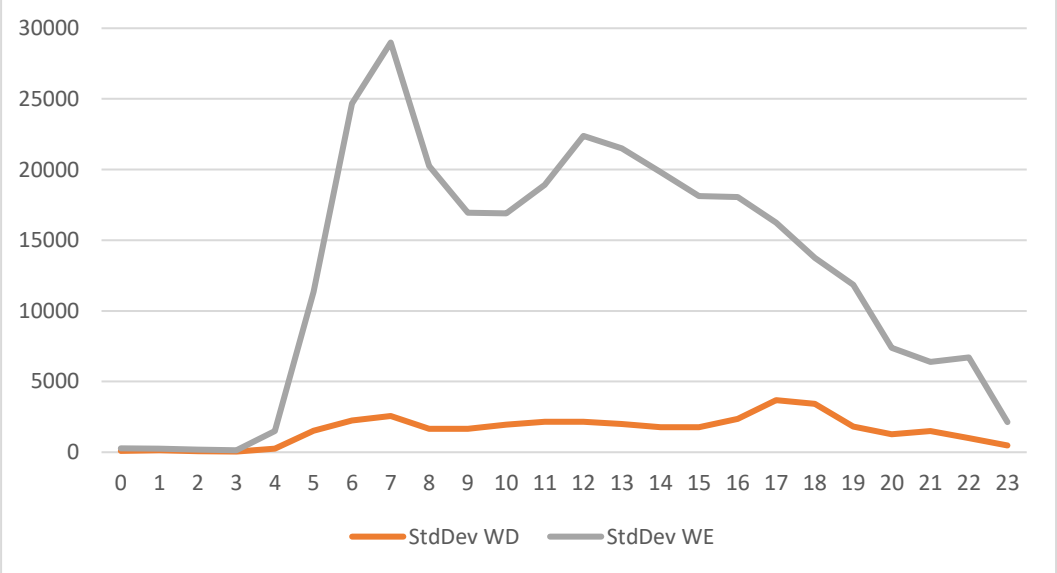


Fonte: Elaborado pelo autor.

O gráfico acima permite que inúmeras informações sejam extraídas da análise do mês em questão. A primeira delas se refere à média (diária) bastante distinta entre cada um dos períodos considerados: 1.180.417 usuários em dias da semana (número superior à média semanal), 819.462 usuários aos sábados e 250.558 usuários aos domingos. Considerando a

média do número de usuários ao longo de 24 horas temos: 49.184 usuários em dias de semana, 34.144 usuários aos sábados e 10.440 usuários aos domingos. Estes números não agregam valor à análise pois desconsideram a dinâmica social do uso de transporte público dos usuários e a variabilidade temporal das validações, ou seja, padronizam um fato que não acontece na realidade. Analisando os desvios-padrões por hora, estes são baixos para os dias da semana e elevados no final de semana (*weekends* – WE) como ilustrado na Figura 8. O coeficiente de variação, o qual mostra a dispersão dos dados em relação à média, para WE é igual a 60%, enquanto para WD é de apenas 3%. Este fato representa que, embora possuam valores relativamente distintos de acordo com a hora do dia, no caso dos dias da semana é possível haver uma maior previsão do número de usuários, o que é bastante relevante para que os operadores do STPP façam o planejamento de frota, por exemplo.

Figura 8 – Desvio-padrão por horário, por categoria (WD, WE)



Fonte: Elaborado pelo autor.

Para concluir a definição de quais horários serão considerados como principais para o estudo, considerou-se três intervalos em dias da semana onde há picos de validação de usuários. São os horários picos da manhã (HP_1) de 5 a 8h, do almoço (HP_2) de 11 a 14h e da tarde (HP_3) de 16 a 18h. Esses trechos, por sua vez, possuem uma hora (comum a todos os dias da semana) que concentra o maior número de validações, sendo: HP_1 – 6 a 6h59, HP_2 – 12 a 12h59 e HP_3 – 17 a 17h59. Aos sábados a hora com maior número de usuários é de 7 a 7h59 e, aos domingos, de 17 a 17h59.

Dando continuidade às análises possíveis com as informações do SBE, é possível também analisar quantitativamente as linhas de ônibus e identificar quais delas têm maior impacto na operação dos ônibus de Fortaleza. Apesar de também conter identificadores para os terminais de ônibus, as validações indicadas para cada um deles só contabilizam os usuários que estão chegando diretamente às suas catracas, ou seja, que farão sua primeira viagem ali. Logo, não é possível, por esse campo da base de bilhetagem eletrônica, indicar quantas e quais integrações ou viagens ocorreram em cada um dos terminais. Contudo, a identificação das principais linhas de ônibus ocorre sem maiores prejuízos, possibilitando priorizar quaisquer opções desejadas. Foram observadas 308 linhas distintas em operação (regular ou especial) no mês dos dados utilizados. Dessas, inicialmente, focou-se na identificação das 10 linhas com maior número de validações, conforme a Tabela 2 abaixo. Do total de 29.053.672 validações (excluindo as realizadas na entrada dos terminais), aproximadamente 4 milhões (3.991.049) ocorreram somente nessas 10 linhas indicadas, o que corresponde a 13,7% do total de usuários.

Tabela 2 – Resumo das dez linhas maior número de validações no mês analisado (set/2014).

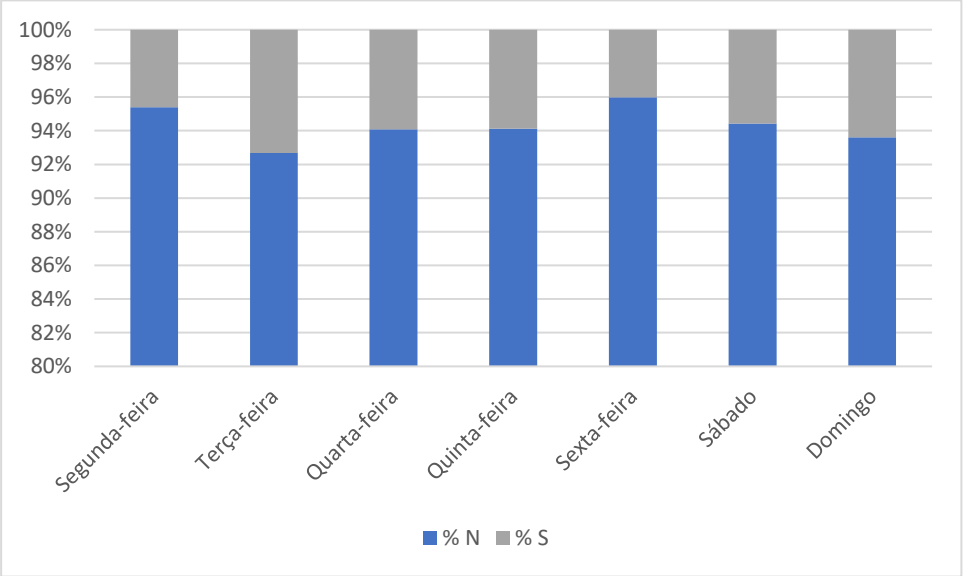
RK	Linha	Ônibus	Soma	Avg	Avg WD	Avg WE	StdDev	StdDev WD	StdDev WE	Total (%)
1	41	Parangaba/Oliv. Paiva	459298	15310	17759	8576	4707	766	4400	1,58%
2	45	Cj Ceará/Papicu/Mont	446336	14878	17602	7387	4902	369	3419	1,54%
3	42	Ant. Bez./Fco Sá/Papicu	434068	14469	16808	8037	4410	366	3960	1,49%
4	26	Antônio Bezerra/Messej.	430492	14350	16506	8420	4172	550	4049	1,48%
5	24	Antônio Bezerra/Lagoa	406951	13565	16217	6272	4772	372	3318	1,40%
6	75	Campus do Pici/Unifor	396853	13228	16467	4322	5731	1064	3006	1,37%
7	44	Parangaba/Papicu/Montese	357039	11901	13881	6456	3887	471	3966	1,23%
8	52	Grande Circular 2	357036	11901	12160	11190	864	372	1373	1,23%
9	76	Cj Ceará/Aldeota	354219	11807	14122	5443	4296	399	3587	1,22%
10	74	Antônio Bezerra/Unifor	348757	11625	14374	4066	4857	884	2520	1,20%

Fonte: Elaborado pelo autor.

A utilização dos terminais de integração de passageiros como porta de entrada segue um comportamento diferente ao se verificar, dentre as 10 principais linhas, quais terminais fazem parte dessas linhas. Considerando apenas a proporção de embarques entre os terminais, naquele mês, os três mais utilizados foram Parangaba (33%), Antônio Bezerra (17%) e Siqueira (13%), enquanto a porcentagem de usuários de forma geral (que também realizaram integração temporal, por exemplo) é: Papicu (25%), Antônio Bezerra (20%) e Parangaba (17%). Conforme visto na tabela acima, das 10 linhas mais utilizadas, 5 delas possuem o terminal do Papicu como destino inicial ou final, o que corrobora a afirmação anterior. Além da integração nos terminais

físicos, o quantitativo de integrações temporais realizadas também pode ser extraído, indicando o cenário de utilização desse benefício, que, dentre outras possibilidades, fornece insumo para o cálculo de ajustes tarifários, por exemplo. A composição média do mês foi de 94,4% validações sem integrações temporais e de 5,6% integrações realizadas. A Figura 9 ilustra como ocorreu essa distribuição em dias da semana.

Figura 9 – Porcentagem do número de integrações temporais por dia da semana



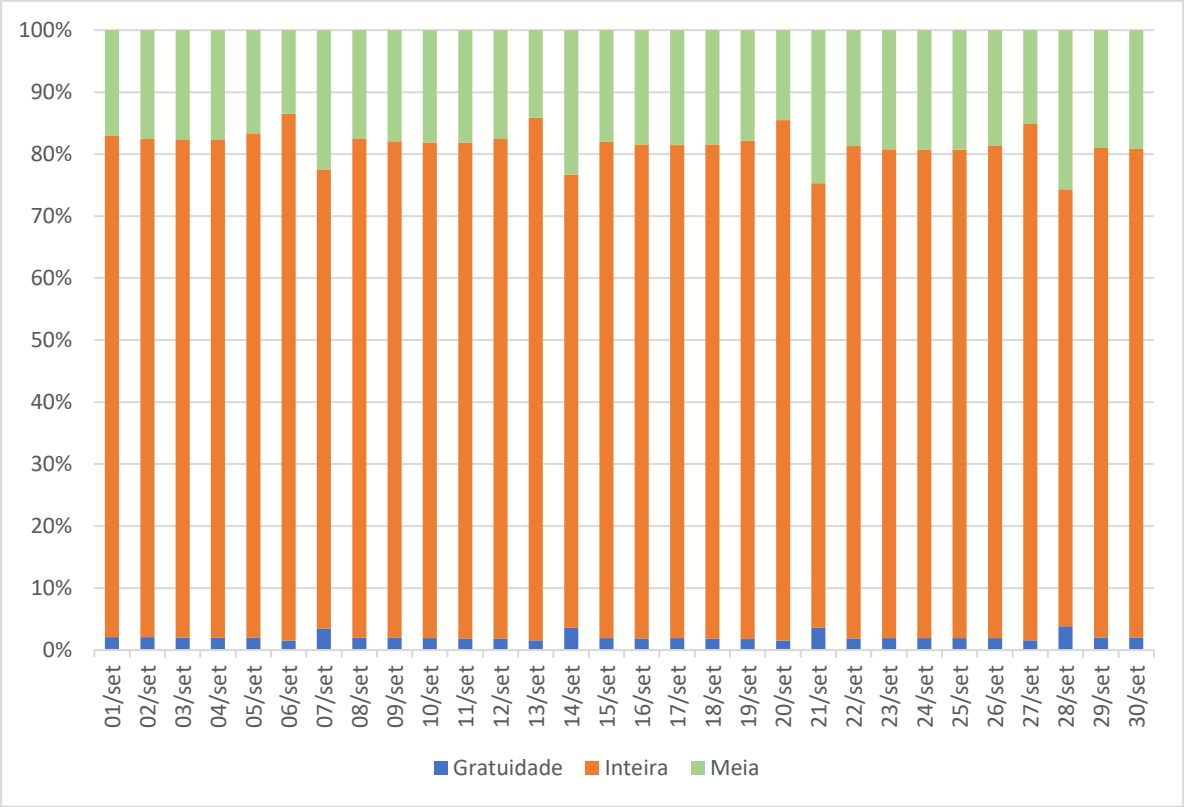
Fonte: Elaborado pelo autor.

Para finalizar a caracterização do sistema, é preciso analisar as modalidades de pagamento utilizadas. Estas informações proporcionam aos tomadores de decisão do STPP maior seguridade para um planejamento mais eficaz de como está o retorno do econômico do sistema. Isto é, associado aos dados de integração temporal, a composição de usuários pagantes (passagens inteiras e meias) e dos que gozam de algum benefício que conceda gratuidade, é utilizado como base na construção de políticas tarifárias: seja do ponto de vista público (por meio dos subsídios para manutenção do transporte público) ou da iniciativa privada. A Figura 10 abaixo ilustra a proporção que cada categoria de pagamento representa ao longo do mês.

Em seguida, conforme esclarecido no início deste tópico, para muitas análises, o comportamento de utilização do transporte público é muito característico (dependendo do dia considerado) para que se agreguem todos os dias de forma singular. A qualidade das informações resultantes depende do reconhecimento de tais características e da restrição de cada intervalo considerado. Logo, a Tabela 3 indica algumas medidas estatísticas (média aritmética – avg, desvio-padrão – std, coeficiente de variação – CV) utilizando a distinção por

categoria de cada modalidade de pagamento (gratuidade, inteira e meia) e dia da semana (TT – total, WD – dias úteis e WE – finais de semana).

Figura 10 – Modalidades de validações por modalidade (gratuidade, inteira e meia)



Fonte: Elaborado pelo autor.

No mês de setembro, em média 80% das validações foram de passagens do tipo inteira (vale avulso, passagem em dinheiro, vale transporte), 18% foram do tipo meia (cartão recarregado ou pagamento em dinheiro) e 1,9% do tipo gratuitas. Este último valor, entretanto, deve ser maior, pois nem todos os usuários com direito à gratuidade fazem a validação utilizando os cartões eletrônicos, embarcando apenas com documentos de identidade com foto, por exemplo. O coeficiente de variação dos dados tem valor mediano quando tomados de forma agregada, porém se tornam pouco variáveis (de 2 a 6%) ao se analisar somente os dias da semana. As validações que ocorrem nos finais de semana, do tipo inteira e meia, são as que possuem maior CV: 31 e 17%, respectivamente.

Tabela 3 – Tendência central e dispersão por modalidade de validade e período da semana

	Gratuidade	Inteira	Meia
avg TT	19.629	807.677	181.002
std TT	5.522	260.702	57.961

	Gratuidade	Inteira	Meia	
CV1	28,1%	32,3%	32,0%	CV1 = std TT / avg TT (%)
avg WD	22.852	942.976	214.590	
std WD	1.277	21.935	7.470	
CV2	6%	2%	3%	CV2 = std WD / avg WD (%)
avg WE	21.537	871.213	177.270	
std WE	1.850	272.200	30.567	
CV3	9%	31%	17%	CV_3 = std WE / avg WE (%)
Somatório	588.883	24.230.314	5.430.051	
Proporção	1,9%	80,1%	18%	

Fonte: Elaborado pelo autor.

De forma geral, a caracterização elaborada neste tópico servirá de fundamento para as etapas finais deste trabalho. A partir dos dados de validação ao longo do dia, observou-se que o comportamento de utilização do sistema tende a ser maior na primeira quinzena do mês e, em seguida, apresenta uma leve redução. Analisando um dia típico, três horários de pico foram constatados como maiores contribuintes para a utilização do transporte público. Após, explorando as validações ocorridas em cada linha, identificou-se as dez principais que possuem um maior número de usuários. Entre estas algumas serão selecionadas para serem analisadas em conjunto com as outras bases do banco de dados, devido à presença de particularidades da rota e do público que atendem.

É importante ressaltar que, embora se trate do estudo de um mês de dados, as características aqui identificadas podem não ser representativas ao se analisar outros períodos. Cada mês possui uma dinâmica complexa de fatores que pode alterar os pontos aqui determinados. Meses distintos do mesmo ano e, principalmente, de anos posteriores, demandam estudos independentes de acordo com o cenário ao qual se deseja avaliar. A ferramenta aqui constituiu no georreferenciamento do SBE com base no AVL. Neste sentido, esta é útil para a tomada de decisão por proporcionar uma espacialização dos dados de bilhetagem.

5.3 Identificação dos veículos – AVL (base tratada)

O serviço de rastreamento prestado pela M2M Solutions é importante para que sejam monitorados os mais de 2 mil veículos da frota de ônibus. Os dados registrados a partir daí possuem informações tanto de interesse direto das empresas proprietárias dos veículos, quanto para os planejadores do STPP. Alguns dos exemplos da utilização desses dados são: a

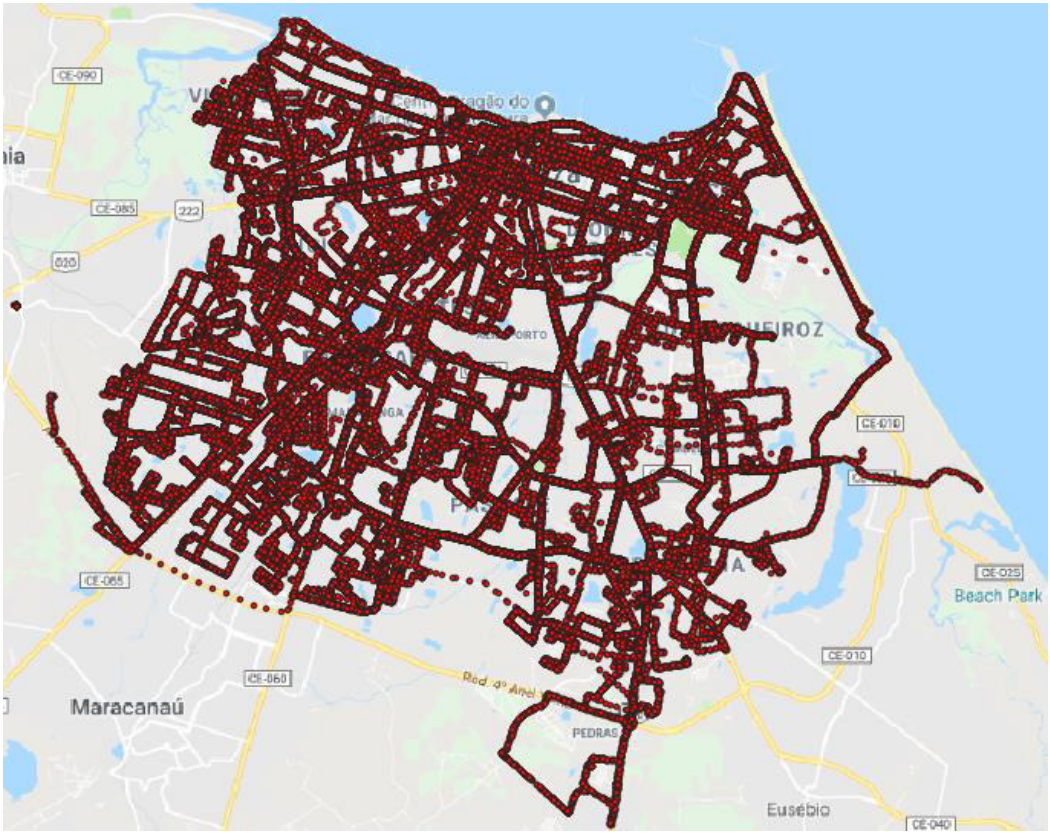
situação de quilometragem dos veículos a partir do campo “odômetro”; a localização dos veículos para averiguação em casos de sinistros por meio dos campos “latitude” e “longitude”; no “calcspeed”, as velocidades instantâneas desenvolvidas ao longo de determinados trechos das vias; também sobre localização, a aderência de trajeto dos ônibus às rotas estabelecidas para um período; a partir do campo “data_hora” e da localização é possível monitorar o cumprimento dos tempos determinados de início e fim da operação; utilizando *softwares* de informação geográfica, o campo “the_geom” permite um georreferenciamento mais preciso por possuir a característica de ser um dado registrado no formato espacial desde a origem, evitando truncamento e arredondamento dos valores de coordenadas geográficas, por exemplo.

Neste trabalho, a aplicação desta base de rastreamento se restringirá à correta identificação do veículo, às coordenadas de latitude e longitude obtidas a partir do “the_geom”, às velocidades registradas e à data e hora de cada observação. Para exemplificar a amplitude do número de observações registradas, a Figura 11 abaixo ilustra cada uma delas em seu devido ponto geográfico. São 524.288 pontos entre 0 e 6h de um dia típico qualquer. Considerando as discussões passadas, esse é o período do dia que há menor utilização dos ônibus, portanto, para estimar número total de um dia não é possível extrapolar linearmente esse número de pontos.

Essa primeira representação do que é a distribuição dos veículos pela cidade, indica que, de modo geral, as linhas de ônibus abrangem uma área satisfatória dos espaços urbanos. Apesar de haver concentração em pontos como no bairro Centro, Parangaba e Papicu, as linhas circulam pelo demais bairros da cidade não deixando grandes áreas desprovidas de acesso ao sistema público de ônibus. As exceções são áreas onde o uso e a ocupação do solo são restritos, como nas proximidades da foz do Rio Cocó e do Aeroporto Pinto Martins.

A princípio, no tratamento da base de dados, não foram excluídos pontos considerados externos à rota regular dos veículos, ou seja, inclui tanto o percurso usual de cada rota, quanto o trajeto entre as garagens e também algum eventual desvio realizado. Optando por isso, as informações são preservadas no banco de dados e podem ser utilizadas de acordo com a necessidade do usuário.

Figura 11 – Pontos de rastreamento de veículos de um dia típico (set/2018 – 0 a 6h)



Fonte: Elaborado pelo autor.

O fato citado anteriormente não afetará, entretanto, a análise utilizando os dados de bilhetagem eletrônica, pois só são mantidos os pontos comuns entre as bases, ou seja, com validações reais. Posterior à etapa de junção das duas bases, o rastreamento se torna mais valioso como objeto de estudo pois, além das coordenadas de cada ponto, é possível identificar qual ônibus aquele ponto representa e, principalmente, o número da linha em operação.

5.4 Identificação espacial da base de bilhetagem (SBE+AVL)

Conforme mencionado no início do capítulo, após identificar que o uso da base DICIO não retornava resultados numa ordem de grandeza esperada, buscou-se utilizar bases alternativas mais recentes. A caracterização realizada ao longo do estudo auxiliou que se priorizasse intervalos temporais que retornassem um maior número de validações em um mês. Desse modo, independente das variações sazonais próprias do mês escolhido, buscavam-se bases de rastreamento e de bilhetagem eletrônica com datas próximas ao da base DICIO (jun/17) e que fossem da primeira quinzena do mês escolhido. Assim, optou-se por utilizar as bases AVL e SBE de abril de 2016. Embora estivessem defasadas 14 meses do período desejado,

estas bases identificaram 96% dos “vehicleid”, campo chave para a tradução do código do aparelho *GPS* nos ônibus para o “numbus” (número utilizado na operação).

A ferramenta utilizada para gerar o banco de dados foi desenvolvida de modo a permitir que fossem inseridos dados dessas bases de quaisquer períodos. Após ter concluído de forma exitosa todos os processos anteriores, o banco de dados resulta em alguns elementos básicos à sua operação. São eles: base SBE modificada, base AVL modificada, base DICIO, duas bases auxiliares (opcionais) que traduzem o número da linha e o código do tipo de pagamento em seus respectivos nomes e, finalmente, uma base (*table*) integrada contendo a localização geográfica aproximada de cada validação de bilhetagem.

O foco deste tópico será a última base mencionada e que, deste momento em diante, devido à sua complexidade e à quantidade de informações contidas, poderá ser mencionada no texto como “banco de dados”. Conceitualmente, entretanto, este termo se refere à uma estrutura de dados organizados em *tables* individuais e que estão à disposição de usuários finais de forma prática e dinâmica (Christensson, 2009).

A migração da base final para o banco de dados ocorre para garantir que ocorra a disponibilização dos recursos da linguagem SQL a favor da manipulação, registro e consulta das informações de forma mais eficiente, sem comprometer inteiramente a memória do computador. Uma das limitações iniciais é a criação da estrutura das tabelas de forma exata aos dos dados que serão inseridos, por isso a importância da caracterização descrita no capítulo 4. A ferramenta proposta neste trabalho cria de forma automatizada a estrutura necessária para que os dados sejam, em seguida, inseridos. O fluxograma presente na Figura 12 representa de forma simplificada os passos nesta etapa do processo.

Figura 12 – Pseudocódigo ilustrando a criação da estrutura em SQL e importação do banco de dados

```
HABILITAR BANCO DE DADOS
PARA CADA base:
    LER ARQUIVO
    IDENTIFICAR ESTRUTURA
    CRIAR TABLES NO DATABASE
        SE table <- AVL:
            LER ARQUIVO EM BLOCOS
            IMPORTAR ARQUIVO
        SE NÃO:
            IMPORTAR ARQUIVO
FIM PARA
```

Fonte: Elaborado pelo autor.

A possibilidade de se utilizar *softwares* que analisam e processam dados na forma de um sistema de informações geográficas (SIG) é mais uma característica positiva do *output* do estudo. Para demonstrar a utilização de alguma dessas ferramentas, foram selecionadas três linhas que possuem particularidades distintas entre si. A primeira delas, a linha "52 – Grande Circular 2" é uma das maiores linhas em extensão do sistema de transporte coletivo urbano de Fortaleza; circula por bairros de diferentes situações socioeconômicas; possui um trajeto radial em volta do perímetro urbano; das linhas selecionadas, é a que possui uma proporção de validações do tipo "gratuidade" (1,2%) mais próxima à média mensal. A segunda linha, "26 – Antônio Bezerra/Messejana", trafega entre dois importantes terminais de integração (de mesmo nome da linha), além de ter parte do seu trajeto no bairro Centro da cidade; sua rota possui orientação que pode ser resumida como "Norte-Sul"; das três linhas citadas aqui, é a que possui maior proporção de validações do tipo "inteira" (83%). Finalmente a última linha, "75 – Campus do Pici/Unifor" liga as duas maiores universidade da cidade, passando por três *campi*; com exceção do trecho final, a orientação geográfica da rota é predominantemente "Leste-Oeste"; no período analisado, a proporção de validações do tipo "meia" (42%) foi superior ao dobro da média mensal para esta categoria. A Tabela 4 abaixo resume estas características. É importante reforçar que as três linhas fazem parte das dez mais utilizadas ao longo do mês.

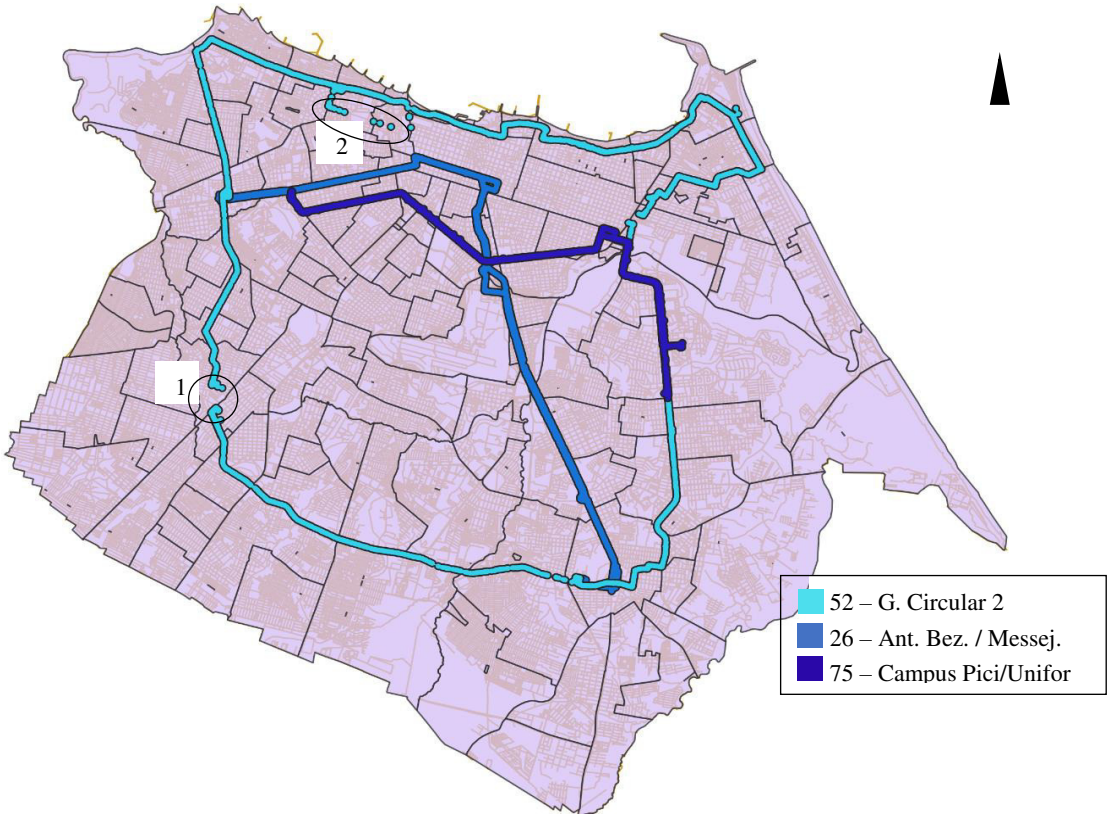
Tabela 4 – Resumo das características das linhas 52, 26 e 75

	Linhas de ônibus		
	52	26	75
Orientação predominante	Radial	Norte-Sul	Leste-Oeste
Número de terminais	4	2	0
Validações: Gratuidade	93 (1,2%)	47 (0,5%)	39 (0,4%)
Validações: Inteira	6322 (79%)	7136 (83%)	6147 (58%)
Validações: Meia	1556 (20%)	1400 (16%)	4438 (42%)
Perfil do usuário por tipo de validação	Misto	Inteira	Estudantes

Fonte: Elaborado pelo autor.

A Figura 13 abaixo mostra a rota realizada por essas linhas num dia útil típico.

Figura 13 – Representação do trajeto das rotas 52, 26 e 75



Fonte: Elaborado pelo autor.

Devido à quantidade de observações de cada linha ser elevada, os pontos utilizados para desenho do trajeto são bastantes próximos ao que seria a rota de fato. Na ilustração acima, apenas a linha 52 apresenta alguns pontos de descontinuidade, conforme indicado nas duas regiões demarcada. Em 1, se encontra o terminal de ônibus do Siqueira, as validações, portanto, ocorrem pouco antes da entrada ou na parada seguinte, interrompendo o fluxo das validações neste trecho. Na região demarcada 2, são observadas poucas validações nesse trecho paralelo à Av. Leste Oeste. Esses registros, por serem em quantidade insignificante em relação ao total, podem indicar algum desvio de fluxo que ocorreu ao longo do dia por motivos corriqueiros como a ocorrência de algum acidente de trânsito. A precisão das coordenadas geográficas registradas pelo GPS dos sistemas de rastreamento é adequada à condução de estudos do tipo, pois de forma geral não foram observados pontos muito destoantes das rotas.

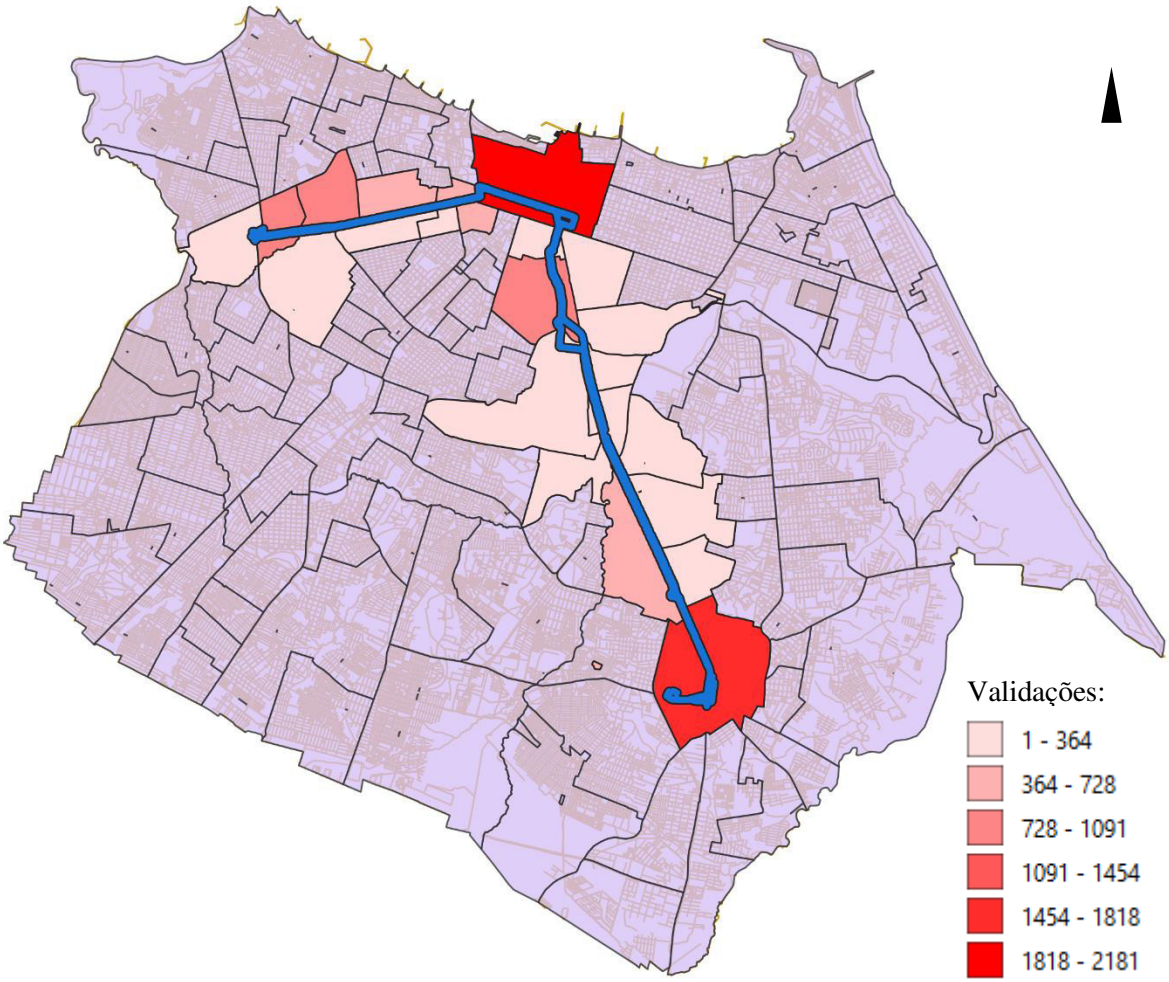
A associação de informações sobre a utilização dos transportes públicos com dados demográficos permite que sejam analisadas alternativas diversificadas sobre a forma como a circulação de pessoas ocorre em grandes centros. Apesar de não ser um dos objetivos propostos nesse trabalho, uma forma primária de se avaliar áreas com maior número de geração de viagens é identificar a posição geográfica das validações em relação a essas zonas. Não se trata de uma

matriz de viagens (O/D) pois os fluxos nos arcos não podem ser determinados a partir disto, mas, de forma preliminar, mostra como é distribuído espacialmente esses movimentos.

Ao se determinar quais são os principais pontos que contribuem para que determinado trecho seja mais ou menos utilizado, é facilitado o estudo de alternativas ao atendimento de transporte público da região em questão. Seja por meio de alternativas de integração modal, pela alocação de um número maior de veículo ou a criação de novas rotas. Desse modo, a Figura 14 ilustra como esse tipo de informação pode ser visualizada, exemplificando como as validações ocorreram, em cada bairro, na linha 26. Conforme esperado, zonas que possuem um maior número de habitantes tendem a possuir maior número de embarques (considerando que o planejamento inicial da linha tenha ocorrido de forma correta). Ainda na figura, as cores mais intensas são no bairro Centro e Messejana, que possuíram 2181 e 1487 validações, respectivamente. Esse comportamento é esperado pois, no caso do Centro, possui elevada concentração de habitantes e de atividade comercial. O bairro Messejana, embora possua terminal próprio, possui também elevado número de habitantes, o que explica o segundo maior número de validações. Comportamento oposto ao verificado nas proximidades do terminal Antônio Bezerra que, embora seja uma zona muito populosa, os embarques se concentram no terminal, não contabilizando nessas validações indicadas no mapa.

Além de informações sobre os usuários, alguns dados puderam ser extraídos com a utilização dos dados do GTFS. Conforme citado anteriormente, são um conjunto de arquivos fornecidos pelas agências de transporte público, é possível extrair diversas informações que podem se tornar recursos para desenvolvimento de novos serviços e produtos. A Tabela 5 expõe os elementos do formato padrão que contêm informações sobre o sistema de transporte público.

Figura 14 – Número de validações nos bairros da linha 26



Fonte: Elaborado pelo autor.

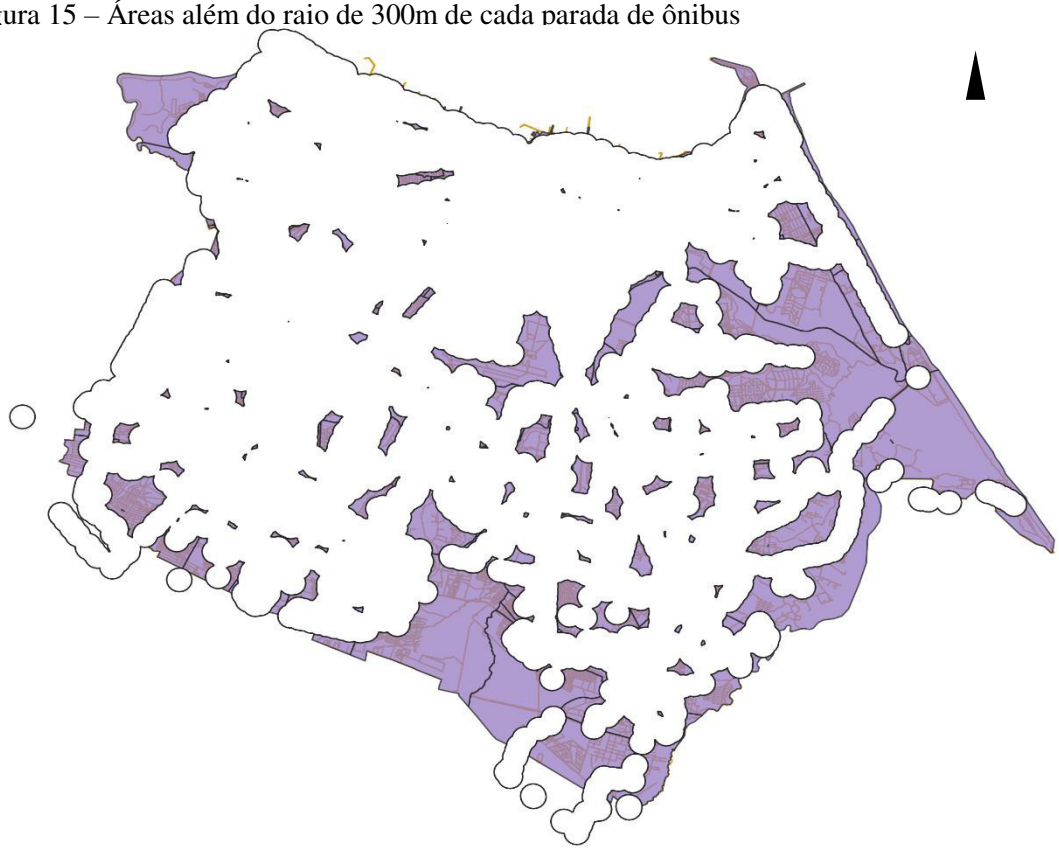
Tabela 5 – Arquivos obrigatórios à estrutura do GTFS

Arquivo	Descrição
Agency.txt	Agências de transporte público que operam o serviço e fornecem os dados do GTFS
Stops.txt	Pontos de embarque ou desembarque de passageiros
Routes.txt	Rota utilizada pelos veículos
Trips.txt	Sequência de paradas ao longo de um período considerado
Stop_times.txt	Hora de chegada e saída dos veículos em pontos do trajeto
Calendar.txt	Delimita os dias e horários que o serviço estará disponível para cada linha

Fonte: Elaborado pelo autor.

O arquivo *stop*.txt foi utilizado para verificar a acessibilidade do transporte público em relação à distância entre as paradas. Segundo Ferraz e Torres (2004) são recomendados de 200 a 400m entre cada parada. A proximidade desses pontos influencia na velocidade operacional dos ônibus e, portanto, não podem ser muito próximos. Já distâncias elevadas entre paradas é prejudicial do ponto de vista da acessibilidade e no conforto proporcionado aos usuários. A Figura 15 representa a cidade de Fortaleza em relação à distância entre as paradas, sendo considerado um raio de 300m como aceitável. As áreas escuras representam as regiões mais desprovidas de acesso ao sistema público de transportes, com ênfase, os bairros Edson Queiroz, Manuel Dias Branco, Prefeito José Walter, de Lourdes.

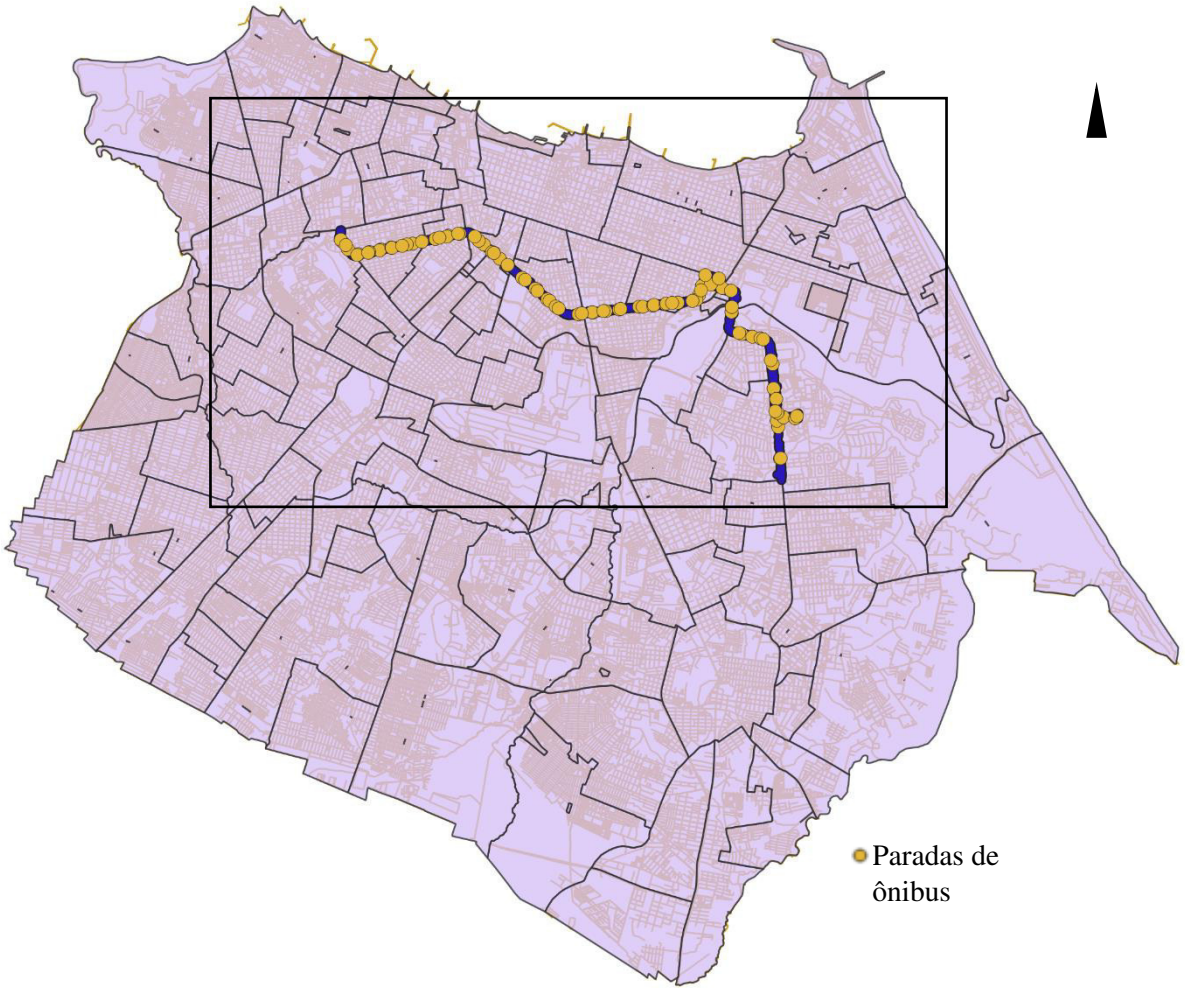
Figura 15 – Áreas além do raio de 300m de cada parada de ônibus



Fonte: Elaborado pelo autor.

Buscando analisar como as validações dos usuários ocorrem ao longo do trajeto da linha 75 (Campus do Pici/Unifor), as paradas dessa linha foram destacadas conforme indicado na Figura 16. São 85 paradas distribuídas em aproximadamente 31Km de extensão total da rota. A partir da divisão simples entre a extensão e o número de paradas, calcula-se o distanciamento médio de 365m, valor próximo ao utilizado na comparação anterior (300m) e dentro do recomendado na literatura citada.

Figura 16 – Distribuição de pontos de paradas de ônibus da linha 75



Fonte: Elaborado pelo autor.

A decisão dos usuários de passar pela catraca do veículo imediatamente após embarcar não é facilmente determinado apenas com as características da linha e do trecho. Elas envolvem desde aspectos dos usuários como maior conforto ao buscar fazer a viagem sentado, maior sensação de segurança após a catraca, proximidade com o destino final, etc. Além de características dos ônibus como total de assentos disponíveis, distribuição de assentos antes e após a catraca, quantidade de portas para desembarque, entre outros. Como forma de avaliar

simplificadamente os embarques que ocorrem próximos às paradas, buscou-se traduzir o tempo de embarque numa distância considerada razoável para que os usuários façam a validação na catraca. O valor foi calculado a partir de três parâmetros: (i) aceleração média de um ônibus, (ii) velocidade média desenvolvida (para limitar o cálculo da distância), (iii) tempo estimado para validação considerando a existência de filas (para passageiros que desejam passar pela catraca logo ao embarcar).

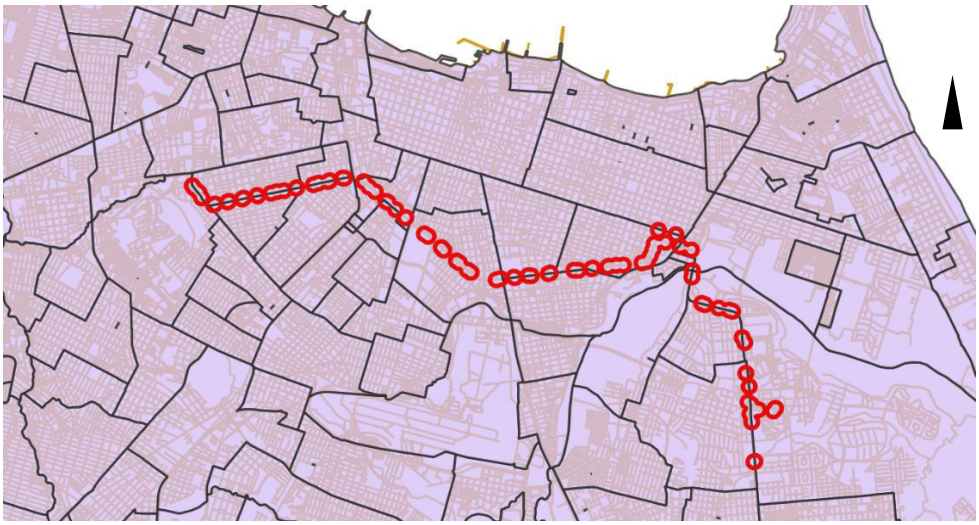
A aceleração considerada foi de 0,8m/s^2, valor estimado para preservar o conforto e segurança dos usuários. A partir da observação de velocidades em um dia útil, durante um intervalo de 5h (abrangendo registros antes e depois do horário-pico), determinou-se, conforme indicado na Tabela 6 abaixo, a velocidade média: 20,20Km/h. Finalmente, o tempo considerado foi de 30s. Assim, a distância calculada para delimitar o espaço percorrido pelos usuários que pretendem completar a validação nas catracas logo ao embarcar é 120m, ilustrado na Figura 17.

Tabela 6 – Descrição estatística de observações de velocidade (5 a 10h) de um dia útil

	5h	6h	7h	8h	9h	10h
Contagem	55395	103692	84428	51550	36968	31856
Valor mínimo	0,19	0,19	0,19	0,19	0,19	0,19
Valor máximo	79,45	77,04	68,71	74,27	79,27	78,34
Valor médio	21,73	19,03	18,28	20,55	21,05	20,55
Mediana	20,37	17,59	16,67	19,26	20,00	19,63
Desvio padrão	15,77	15,11	14,86	15,56	15,63	15,20
CV	0,73	0,79	0,81	0,76	0,74	0,74

Fonte: Elaborado pelo autor.

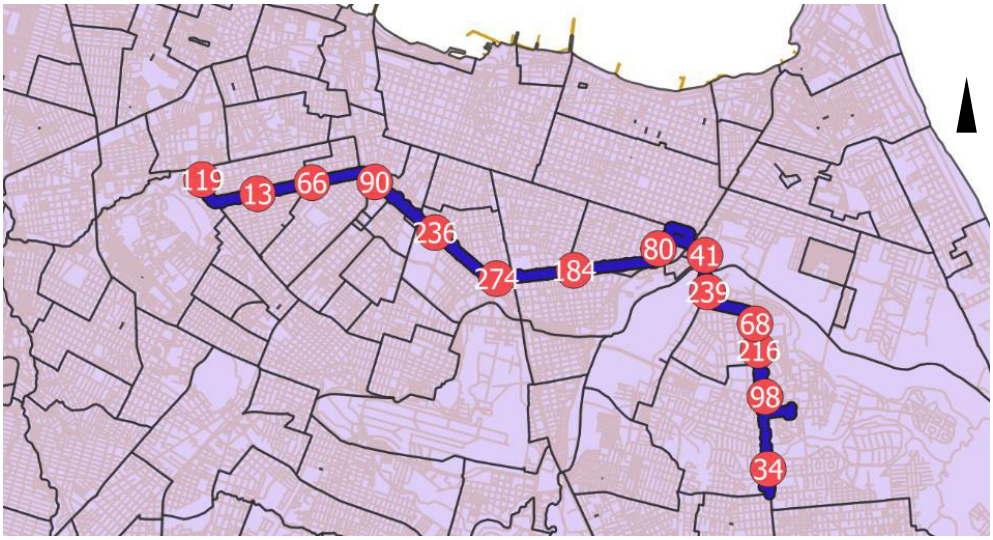
Figura 17 – Representação das paradas de ônibus com delimitação de área com 120m de raio



Fonte: Elaborado pelo autor.

Utilizando novamente ferramentas de geoprocessamento do *software* SIG, extraiu-se as validações que ocorreram externamente às áreas demarcadas na figura anterior. Agregando cada conjunto de observações em regiões por proximidade, é possível avaliar onde estão concentrados os maiores números de validação distantes mais de 120m de paradas de ônibus. Na Figura 18 abaixo estão indicados esses valores, de forma agrupada por trecho.

Figura 18 – Número agregado de validações que ocorreram fora do raio de 120m das paradas



Fonte: Elaborado pelo autor.

Uma das possíveis interpretações para cada um desses valores agrupados pode ser a ocorrência de filas nas catracas antes da validação dos usuários ou superlotação nos ônibus, impedindo que o fluxo se desloque para a parte anterior do veículo. Entre outras possibilidades, é possível relacionar este número com os trechos de maior demanda de passageiros pois estão situados em trechos posteriores às zonas de elevado número de embarque /desembarque. No caso da linha do exemplo, esses trechos se situam principalmente nas proximidades do Campus do Pici, Shopping Benfica, reitoria da UFC, shopping Iguatemi e Unifor.

6 CONSIDERAÇÕES FINAIS

A utilização de sistemas de dados cada vez maiores é uma realidade nos tempos atuais e tende a continuar em constante evolução devido à conectividade da vida moderna, com geração de informação a todo instante. Este trabalho teve como objetivo o desenvolvimento de uma ferramenta para criação de um banco de dados capaz de abordar esse tipo de situação, onde o volume disponível de dados requer métodos auxiliares para que seja gerado informações úteis a partir de dados brutos. Além disso, foi preciso considerar a natureza dinâmica e ininterrupta do surgimento e registro de novas informações para serem inseridas no banco de dados.

O planejamento para desenvolvimento do código iniciou com a averiguação da estrutura de cada base utilizada e, em seguida, o tratamento de limitações que poderiam surgir na utilização dessas bases. Desse modo, aumentou-se a robustez da ferramenta ao considerar e antever quais tipos de erro poderiam invalidar a análise, propondo pontos de verificação ao longo do processo. Por utilizar conceitos de banco de dados relacionais, ou seja, percebidos pelo usuário como partes integrantes de estruturas individuais que possuem relações entre si, é imprescindível que os termos utilizados como chave possuam correspondentes. Neste trabalho, dois atributos foram utilizados como chave: *vehicleid* e *bigtag*. O primeiro deles é fundamental para a ferramenta apresentada pois a taxa de sucesso, isto é, existência de pares correspondentes entre as bases, representa a proporção de registros que serão tratados no decorrer da análise. A limitação presente na ferramenta se deve à aquisição de bases de dados contendo informações de um período próximo para que se amenize a perda de registros no pareamento das bases.

Para o planejamento urbano, com foco em transportes, a utilização de bases como as do sistema de bilhetagem eletrônica (SBE) e do rastreamento da frota (AVL), existentes de forma separada, representa uma vantagem ainda maior quando feita de forma conjunta. Mostrou-se que informações desagregadas de validações de passagens e de localização destas permitem diversos tipos de análise, fornecendo insumos para o planejamento tanto estratégico quanto operacional de governos, agências de trânsito e empresas de transporte de passageiros.

De forma geral, a ferramenta se mostrou útil no armazenamento das informações em um banco de dados que pode ser direcionado para servidores paralelos, externos às limitações de computadores domésticos, além de permitir a constante alimentação e atualização dos registros guardados. Ademais, está estruturado para ser visualizado em *softwares* que reconhecem bancos de dados e projetam as informações desejadas espacialmente. Portanto, facilitando análises estatísticas básicas ou espaciais, a geração de mapas temáticos, entre outros.

7 SUGESTÕES PARA TRABALHOS FUTUROS

As motivações que guiaram o interesse pela pesquisa sobre este tema referem-se às diversas oportunidades advindas da aplicação de um sistema dinâmico como suporte à tomada de decisão do STPP. A partir da literatura discutida ao longo deste texto, percebe-se que este é um tema atual e promissor para o desenvolvimento de novos estudos que utilizem uma abordagem semelhante de aquisição, tratamento e análise de dados. Como forma de sugestão, a partir deste trabalho é possível indicar algumas pesquisas futuras como:

- A utilização de dados de velocidade e de localização geográfica (tipo de via urbana, elevação, tipo de uso e ocupação do solo, horário) para se estimar o consumo médio de combustível;
- Estimativas de poluição a partir da emissão de gases poluentes emitidos pelos ônibus, relacionando o código e o modelo do veículo em conjunto com informações tais como as citadas no ponto anterior;
- Aprimoramento de matrizes de viagem que considerem o número de validações dos usuários, a população das zonas, o tipo de uso e ocupação do solo e previsões de demanda;
- Redução do intervalo de tempo do campo *timetag* para avaliar o impacto na velocidade de processamento das informações e a possível redução na perda de informações devido a falhas no pareamento das bases;
- A verificação de um gradiente de descolamento utilizando informações dos registros temporais e do sentido que os veículos adotam instantaneamente em cada ponto de observação da base de rastreamento;
- Aprimorar a determinação da rota ao verificar se o trecho em questão faz parte do caminho determinado para a linha em questão, considerando possíveis pontos de desvio e o horário de operação do veículo;
- Estudo econômico dos ativos empregados no STPP por meio do cruzamento de dados sobre os veículos e as características da operação nas rotas.

REFERÊNCIAS

ABKOWITZ, M. et al. Operational Feasibility of Timed Transfer in Transit Systems. **Journal of Transportation Engineering**, v. 113, n. 2, p. 168–177, 1987.

ARBEX, R. O.; DA CUNHA, C. B. Avaliação das mudanças nas velocidades das linhas de ônibus da cidade de São Paulo após a implantação de faixas exclusivas através da análise de dados de GPS. **Transportes**, v. 24, n. 4, p. 21, 2016.

ASSUNÇÃO, M. D. et al. Big data computing and clouds: Trends and future directions Marcos. **Journal of Parallel and Distributed Computing**, p. 2–44, 2014.

BERKOW, M. et al. Transit performance measurement and arterial travel time estimation using archived AVL data. **ITE District**, v. 6, n. July, 2007.

BLYTHE, P. T. Improving public transport ticketing through smart cards. **Proceedings of the Institution of Civil Engineers**, n. July 1998, p. 47–54, 2004.

CHIARA-CHAVALA, T.; COIFMAN, B. Effects of Smart Cards on Transit Operators. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1521, n. Transportation Research Board of the National Academies, p. 84–90, 1996.

CHRISTENSSON, P. "Database Definition." **TechTerms**. Sharpened Productions, 27 de outubro de 2009. Internet. Acesso em: 30 de novembro de 2018. <https://techterms.com/definition/database>.

DEAKIN, E.; KIM, S. **Transportation Technologies: Implications for Planning**. [s.l.] UC Berkeley: University of California Transportation Center, 2001.

EOM, J. K.; SONG, J. Y.; MOON, D. S. Analysis of public transit service performance using transit smart card data in Seoul. **KSCE Journal of Civil Engineering**, v. 19, n. 5, p. 1530–1537, 2015.

FERRAZ, A.C.P.; TORRES, I. G. E. **Transporte Público Urbano**. 2ª Edição. ed. São Carlos: Rima, 2004.

FOSSO WAMBA, S. et al. How “big data” can make big impact: Findings from a systematic review and a longitudinal case study. **International Journal of Production Economics**, v. 165, p. 234–246, 2015.

GOODCHILD, M. F. Twenty years of progress: GIScience in 2010. **Journal of Spatial Information Science**, v. 1, n. 1, p. 3–20, 2010.

GSCHWENDER, A.; MUNIZAGA, M.; SIMONETTI, C. Using smart card and GPS data for policy and planning: The case of Transantiago. **Research in Transportation Economics**, v. 59, p. 242–249, 2016.

HALL, R. Vehicle scheduling at a transportation terminal with random delays en route. **Transportation Science**, v. 19, p. 308–320, 1985.

KHAN, M. A. U. D.; UDDIN, M. F.; GUPTA, N. Seven V's of Big Data understanding Big Data to extract value. **Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education - "Engineering Education: Industry Involvement and Interdisciplinary Trends", ASEE Zone 1 2014**, 2014.

KHOURY, M. J.; IOANNIDIS, J. P. A. Big data meets public health. **Science**, v. 346, n. 6213, p. 1054–1055, 2014.

KURAUCHI, F.; SCHMOCKER, J.-D. **Public Transport Planning with Smart Card Data**. [s.l.] CRC Press, 2017.

LANEY, D. META Delta. **Application Delivery Strategies**, v. 949, n. February 2001, p. 4, 2001.

LEE, K. K. T.; SCHONFELD, P. Optimal Slack Time for Timed Transfers at a Transit Terminal. **Journal of Advanced Transportation**. v. 25, n. 3, p. 281–308, 1992.

LI, T. et al. Smart Card Data Mining of Public Transport Destination: A Literature Review. **Information**, v. 9, n. 1, p. 18, 2018.

LIU, J. et al. ISPRS Journal of Photogrammetry and Remote Sensing Rethinking big data : A review on the data quality and usage issues. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 115, p. 134–142, 2016.

MALIENE, V. et al. Geographic information system: Old principles with new capabilities. **URBAN DESIGN International**, v. 16, n. 1, p. 1–6, 2011.

MARTINELLI, J; AROUCHA, M. Fase Atual da Bilhetagem Elerônica. **Sistemas Inteligentes de Transportes**, Series Cadernos Técnicos, vol. 8,. p. 76–99, 2012.

MCDONALD, N. Multipurpose Smart Cards in Transportation: Benefits and Barriers to Use. **Spring** 630, 8 de dezembro de 2000. Internet. Disponível em: <http://www.uctc.net/research/papers/630.pdf>. Acesso em: 9 jun. 2018.

MCNULTY, E. **Understanding Big Data: The Seven V's**. Dataconomu, 22 de maio de 2014. Disponível em: <http:// dataconomy.com/2014/05/seven-vs-big-data/>. Acesso em: 9 jun. 2018.

MILNE, D.; WATLING, D. Big data and understanding change in the context of planning transport systems. **Journal of Transport Geography**, n. November, p. 0–1, 2018.

OECD/ITF. Big Data and Transport. **OECD/ITF**, p. 66, 2015.

OKUNIEFF, P. E. TRCP Synthesis of Transit Practice 24: AVL Systems for Bus Transit. **Transportation Research Board**, National Research Council, Washington, D.C.. 1997.

PAU, S. A. Using Smart Card Technologies to Measure Public Transport Performance : Data

Capture and Analysis. 2013.

PELLETIER, M. P.; TRÉPANIER, M.; MORENCY, C. Smart card data use in public transit: A literature review. **Transportation Research Part C: Emerging Technologies**, v. 19, n. 4, p. 557–568, 2011.
PETTIT, C. et al. Planning support systems for smart cities. **City, Culture and Society**, v. 12, n. November 2017, p. 13–24, 2018.

SATO, Y.; ZENOU, Y. How urbanization affect employment and social interactions. **European Economic Review**, v. 75, p. 131–155, 2015.

SHAW, S. L. Geographic information systems for transportation - An introduction. **Journal of Transport Geography**, v. 19, n. 3, p. 377–378, 2011.

SHELFER, K. M.; PROCACCINO, J. D. Smart card evolution. **Communications of the ACM**, v. 45, n. 7, p. 83–88, 2002.

SHIH, M.-C.; MAHMASSANI, H.; BAAJ, M. Trip Assignment Model for Timed-Transfer Transit Systems. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1571, n. 971011, p. 24–30, 1997.

VRIES, J. DE. Studying Cities in their Context. **193 Urban History Review**, v. XVIII, 1990.
VUCHIC, V. R.; CLARKE, R.; MOLINERO, A. Timed Transfer System Planning , Design and Operation. n. October, 1981.