

20 Volume

SÉRIE CADERNOS TÉCNICOS

***Big Data* para análise de métricas de qualidade de transporte: metodologia e aplicação**



SÉRIE CADERNOS TÉCNICOS
volume 20

***Big Data* para análise de métricas de
qualidade de transporte: metodologia e
aplicação**



Coordenador

Luiz Carlos Mantovani Néspoli

Equipe Técnica Scipopulis

André Yai
Igor Gonçalves
Ivo Pons
Julian Monteiro
Nicole Escobar
Renato Arbex
Roberto Speicys

Equipe ANTP

Geraldo José Calmon de Moura
Marcos Pimentel Bicalho

ISBN: 978-85-5815-000-2

MONTEIRO, J.; PONS, I.; SPEICYS, R. Big Data para análise de métricas de qualidade de transporte: metodologia e aplicação, São Paulo, v. 20, 2015.

SÉRIE CADERNOS TÉCNICOS

volume 20

***Big Data* para análise de métricas de
qualidade de transporte: metodologia e
aplicação**

dezembro 2015

Elaboração

**Scipopulis Desenvolvimento e Análise
de Dados Ltda.**

Ivo Pons
Julian Monteiro
Roberto Speicys



1

Apresentação ANTP

A utilização das tecnologias de informação para controle dos sistemas de transporte ainda é um assunto pouco explorado e praticamente não utilizado pelos gestores públicos ou pelos operadores privados dos serviços de transporte coletivo no Brasil, apesar de boa parte da frota em operação nas cidades brasileiras já disporem de recursos para isso, mais especificamente sistemas de bilhetagem eletrônica e sistemas de localização veicular por satélite (GPS).

Apesar da disponibilidade desses equipamentos, a gestão cotidiana dos serviços pouco os utiliza e as atividades rotineiras de planejamento operacional, controle e fiscalização continuam sendo feitas – quando são feitas – com base em pesquisa de campo sobre amostras bastante reduzidas e apontamentos manuais, com previsíveis limitações de abrangência e de confiabilidade.

Este projeto desenvolvido para a Cidade de São Paulo, onde as dimensões do sistema de transporte coletivo são, no mínimo, desafiadoras, mostra que este salto na gestão dos serviços operados por ônibus tem todas as condições de ser dado, com as condições tecnológicas disponíveis no mercado e a custos acessíveis.

Por outro lado, o estudo denuncia que, se a tecnologia ajuda, ela não é uma panaceia para solução de todos os problemas e permanece dependente da qualidade da atuação dos órgãos gestores. No projeto desenvolvido, uma análise preliminar foi necessária para verificar a qualidade das informações disponíveis a respeito dos itinerários, dos quadros de partidas programadas para as linhas, das velocidades médias utilizadas para dimensionamento ou a localização dos pontos de parada. Infelizmente muitas prefeituras, mesmo de cidades de médio e grande portes, não dispõem desses dados, ou não os têm de forma confiável.

Essas fragilidades, porém, não atrapalharam o desenvolvimento exploratório destas ferramentas para o Município de São Paulo e os seus resultados e a sua divulgação podem servir como um estímulo para que outras cidades pretendam também utilizá-las, ainda que, para isso, precisem investir na estruturação e na capacitação dos seus órgãos gestores.

Luiz Carlos Mantovani Néspoli

Superintendente da ANTP

2

Apresentação SPTrans

A Cidade de São Paulo possui hoje cerca de 15.000 ônibus gerenciados pela SPTrans, que circulam em cerca de 1300 linhas, operadas por 37 empresas consorciadas e que desenvolvem quase 200 mil viagens diárias transportando mais de 9 milhões de pessoas por dia. Números grandiosos para uma cidade estado, de mesma proporção.

A totalidade dos ônibus é rastreada através de um sistema de geolocalização e esses dados estão abertos para a população na plataforma “Olho Vivo” da SPTrans, primeira capital brasileira a abrir esse tipo de informação, mostrando mais uma vez o pioneirismo da cidade. Disponibilizar os dados para os cidadãos é fundamental para ampliar a transparência do governo, fortalecer o protagonismo do cidadão na fiscalização e a evolução da qualidade do sistema de transporte como um todo, o que é por si só uma ruptura de paradigma.

Para seguir vencendo os complexos desafios contemporâneos no transporte público, e em especial do transporte de ônibus, em grandes cidades é preciso avançar ainda mais na busca e no desenvolvimento de tecnologias

que possam contribuir para a gestão planejamento e controle. Nesse âmbito a parceria com o Banco Mundial, através da ANTP soma esforços e adiciona mais uma etapa na constante busca pela melhoria e ampliação da qualidade.

Este caderno técnico da ANTP, “Big Data para análise de métricas de qualidade de transporte: metodologia e aplicação” é um marco na utilização dos dados de rastreamento dos ônibus, seja do ponto de vista metodológico, seja do ponto de vista prático. Os exemplos aqui apresentados usam como base os dados coletados pela SPTrans no âmbito da Cidade de São Paulo, porém, contribui para além desse horizonte, para o desenvolvimento metodológico em nível nacional e para a aplicação em qualquer cidade que sofra com problemas de transporte dado que os códigos desse desenvolvimento, em linha com a filosofia da cidade, também estarão abertos ao público.

O grande problema do modelo atual de concessão dos serviços são os parâmetros da mesma. Há 3 formas essenciais de se delegar os serviços de transporte público. A primeira alternativa, usual em países desenvolvidos, é remunerar de maneira fixa pelo serviço prestado. Assim, estima-se o custo do serviço prestado e divide-se por algum parâmetro como, por exemplo, o total de quilômetros rodados ou o número de veículos da frota. A segunda alternativa, usual em países em desenvolvimento, é remunerar pelo número de passageiros transportados. A terceira alternativa é remunerar em função de parâmetros operacionais. Seria como se remunerasse por partida completa, regularidade no intervalo entre partidas, etc.

A terceira alternativa, que seja do meu conhecimento, nunca foi aplicada diretamente. De fato, diversos contratos misturam aspectos dos três grandes modelos de concessão listados acima. Originalmente, tanto a remuneração fixa quanto a remuneração por passageiro procurava controlar o nível do serviço com fiscais multando pelo descumprimento. Esse sistema preponderou em praticamente todos os exemplos conhecidos até recentemente quando algumas cidades começaram a realizar ajustes nos desembolsos a partir de parâmetros operacionais.

Assim, o controle de parâmetros operacionais é hoje um elemento fundamental para uma concessão de serviços de transporte público. As novas tecnologias discutidas nesse relatório e implementadas no serviço web gerado dentro desse estudo mostram a capacidade que temos hoje em dia de aferição em um serviço tão descentralizado como o transporte coletivo sobre pneus. Uma aferição precisa permite que se precifique cada um dos parâmetros de qualidade de um contrato como esse e, portanto, garanta o seu cumprimento por estímulos econômicos e não por dependência de um controle que não é possível de se ter. Em um modelo no qual se pretende operar por comando e não por controle, essa medida é fundamental para se transitar entre os dois modelos de políticas públicas.

O produto gerado nesse estudo permite que se analise o cumprimento de parâmetros básicos com o cuidado de decompor as “falhas” entre o viário e a empresa operadora. A hipótese básica é que o operador não pode se responsabilizar pelo viário. O operador não pode se responsabilizar com um aumento no trânsito que certamente deve impedir o cumprimento dos parâmetros operacionais. Poderemos ver nesse estudo o grau de adesão ao planejado em dois parâmetros fundamentais: cumprimento de partidas e intervalo entre os mesmos decompondo a falta de adesão entre o trânsito acima do previsto e outros fatores.

A quantidade de estudos aplicados que se pode gerar a partir dessa nova forma de visualização e organização da informação massiva gerada por esse trabalho é incontável. É possível ver onde o trânsito impede o planejamento e onde as empresas estão falhando mais. Cada um desses problemas requer tratamentos distintos. O cruzamento dessa informação com dados de passageiros seria a última etapa em uma avaliação profunda do sistema de transportes corrente. Nos casos em que o planejado se cumpre porém a linha está andando com um excesso de passageiros ou com grande ociosidade estão mal planejadas. Esse raio X do sistema aplicado a um mês específico, quando generalizado para o incrível acúmulo de dados históricos realizado pela SPTrans deve gerar um conhecimento indispensável para a melhoria no planejamento dos transportes bem como o aprimoramento dos contratos de concessão dos serviços. Esse trabalho é o primeiro esforço nesse sentido utilizando ferramentas de big data e

computação em nuvem na Cidade de São Paulo e talvez um pioneiro no Brasil. A sua leitura deve deixar claro o que se pode avançar nas próximas décadas em termos de conhecimento dos sistemas de transporte descentralizados.

Ciro Biderman

Chefe de Gabinete da SPTrans

Sumário

Conteúdo

1 - Introdução ao caderno técnico _____	pág. 11
2 - Definições do caderno técnico _____	pág. 13
3 - Aplicações do Big Data à gestão e ao planejamento do transporte público _____	pág. 16
4 - Descrição da metodologia aplicada _____	pág. 23
5 - Etapa preliminar do trabalho _____	pág. 34
6 - Cálculo de valores de base _____	pág. 46
7 - Análise do cumprimento de partidas _____	pág. 61
8 - Análise de intervalo de passagem _____	pág. 67
9 - Conclusões sobre o trabalho desenvolvido _____	pág. 72

Introdução



Introdução ao caderno técnico

A recente evolução das tecnologias de armazenamento e processamento de bases de dados massivas - popularmente conhecidos como *Big Data* - pode trazer uma mudança de paradigma na gestão e planejamento dos sistemas de transporte público.

Redes de transporte público são sistemas extremamente complexos e dinâmicos, cujo desempenho depende de diversos parâmetros e de suas variações. Por esta razão, a área de transportes sempre procurou obter dados sobre diversos indicadores que afetam o sistema como número de passageiros, origens e destinos, demanda, viagens realizadas, entre outros.

Tradicionalmente, estas informações são obtidas através de pesquisas de campo e estudos técnicos realizados a partir de amostras representativas da população analisada, como por exemplo pesquisas de Origem e Destino ou pesquisas de embarque/desembarque em ônibus. Desta forma, para se determinar a demanda de viagens, por exemplo, entrevista-se uma parcela relativamente pequena de domicílios de uma região e extrapola-se o resultado para o total da população. Para determinar-se a velocidade de um ônibus ao longo de uma linha, uma equipe técnica segue alguns ônibus ao longo do trajeto tomando notas, e aproxima-se o resultado para o restante das partidas.

Pesquisas e estudos são fundamentais para a melhoria do transporte público, mas possuem algumas desvantagens. A aquisição e tabulação dos dados, por exemplo, é extremamente cara pois requer um grande esforço humano. Pesquisas são criadas com o propósito de responder a uma série de perguntas relevantes no momento de sua criação, mas os dados obtidos dificilmente podem ser reaproveitados para responder perguntas a posteriori.

Nos últimos anos, a informatização dos sistemas de transporte se transformou em realidade. Um subproduto destes sistemas é a digitalização de diversas transações características do transporte público, como por exemplo o uso dos bilhetes de transporte (através do sistema de bilhetagem) ou o posicionamento dos ônibus (através do sistema de rastreamento). Estas informações, que até recentemente eram analógicas, assíncronas e possuíam alto custo de obtenção e armazenamento, sob forma digital são recuperadas automaticamente, em tempo real e podem ser armazenadas por longos períodos de tempo com um custo irrisório.

Estes dados são extremamente ricos e podem fornecer informações valiosas sem o esforço de coleta de dados requerido pelas informações analógicas. Além disso, a armazenagem histórica destes dados facilita a elaboração de análises ao longo do tempo e a caracterização de tendências relacionadas à evolução do sistema de transporte com maior precisão e flexibilidade.

Neste caderno técnico apresentamos uma metodologia para a análise de grandes massas de dados para determinação de indicadores de qualidade do sistema de transporte público. Esta metodologia pode ser aplicada a dados atuais e antigos para avaliação da evolução dos indicadores ao longo do tempo. Utilizamos os dados de rastreamento dos veículos e informações sobre a rede de transportes para identificar índices de cumprimento de partida em relação ao tráfego real da cidade e a variação no intervalo de tempo entre passagens de cada linha. Esta metodologia foi aplicada à cidade de São Paulo com o objetivo de levantar indicadores de qualidade que eventualmente podem ser também empregados em novas licitações de transporte público da cidade.

Definições

2

Definições do caderno técnico

A fim de evitar ambiguidades na leitura deste caderno técnico, definiremos nesta seção os principais termos que serão utilizados ao longo do documento e a forma em que nós os utilizamos e como eles devem ser interpretados pelo leitor:

AVL Automatic Vehicle Location, sistema de localização dos ônibus por GPS⁹, registra periodicamente a posição do ônibus e envia este dado a um servidor remoto.

Frequência quantidade de viagens realizadas em uma determinada linha dentro de uma janela horária. A frequência de uma linha pode variar conforme a hora do dia

GTFS General Transit Feed Specification, formato de dados que permite a especificação de uma rede de transportes de uma cidade, com suas linhas e

⁹ GPS é a sigla de “Global Positioning System” que significa sistema de posicionamento global, em português.

partidas planejadas. Neste documento utilizaremos o acrônimo GTFS para identificar também o conjunto de arquivos contendo a especificação da rede no formato GTFS.

Headway o *headway* é o inverso da frequência, ou seja, é o intervalo em minutos entre duas partidas da linha.

Linha identificação de uma viagem de ida e sua respectiva volta planejadas na rede de transportes.

Parada um ponto de ônibus.

Partida uma viagem realizada por um ônibus sobre o itinerário previsto de uma linha.

Ponto de ônibus locais de embarque ou desembarque de passageiros ao longo de uma linha.

Segmento a linha formada pela conexão de dois shapes consecutivos.

Shape ponto geolocalizado que define o vértice de um dos segmentos que formam o trajeto de uma linha considerando-se a geometria das vias por onde trafegam os ônibus. Conectando-se os shapes consecutivos é possível obter o trajeto completo da linha no viário da cidade.

TP terminal primário, local de saída dos ônibus no sentido de ida de uma linha.

Trecho sequência de shapes consecutivos contidos no conjunto de shapes que definem o trajeto de uma linha.

Trip uma das viagens (ida ou volta) que constituem a linha.

TS terminal secundário, local de saída dos ônibus no sentido de volta de uma linha.

Viagem sequência de pontos, a partir de um TP ou um TS, em ordem crescente de distância do ponto inicial, com destino ao TS ou TP correspondente. Se o sentido da viagem é TP -> TS ela é chamada de viagem de ida, ou viagem de volta caso o sentido seja TS -> TP.

3

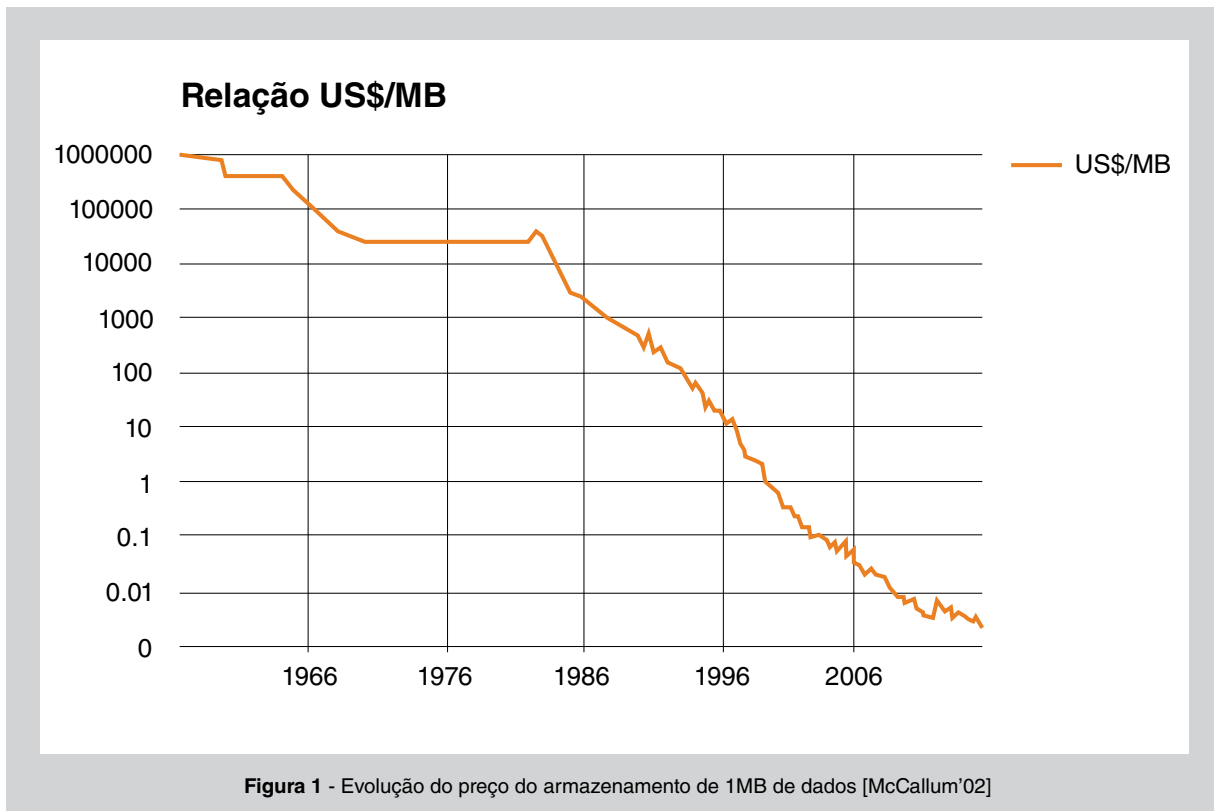
Aplicações

Aplicações do Big Data à gestão e ao planejamento do transporte público

A queda dos custos relacionados ao poder computacional e ao armazenamento de dados não são uma novidade. A Lei de Moore, uma previsão feita em 1965 pelo Dr. Gordon Moore, um dos fundadores da Intel, dizia que a capacidade computacional dos processadores dobraria a cada ano [Moore'65] (alguns anos depois a previsão foi atualizada para 2 anos). Esta lei vem se confirmando nos últimos 50 anos apesar dos limites físicos estarem próximos de serem atingidos.

Em relação ao armazenamento de dados, o custo de estocagem vem caindo exponencialmente, como mostra a Figura 1, onde a escala do eixo das abscissas é logarítmica.

Esta redução nos custos tanto de processamento como de armazenamento propiciou a criação de diversas tecnologias que estão revolucionando a maneira como as pessoas se relacionam, como as empresas fazem negócios e como os governos aumentam sua eficiência. *Smartphones*, computação em nuvem e *Big Data* são algumas das novas tecnologias viabilizadas por esta redução nos custos de processamento e armazenamento de dados.



O termo *Big Data*, em particular, vem sendo muito utilizado ultimamente em diversas áreas que empregam a análise de grandes massas de dados para a obtenção de informações que anteriormente eram desconhecidas. Entretanto, e talvez até por isso, o termo *Big Data* não tem uma definição clara: um levantamento recente [Ward'13] encontrou pelo menos 6 definições diferentes, incluindo uma para cada grande empresa de tecnologia (Microsoft, Intel, Oracle e IBM). Em [Ward'13] é proposta uma nova definição que mantém as ideias principais das definições anteriores e que iremos adotar neste documento: “*Big Data* descreve o armazenamento e a análise de conjuntos de dados grandes e complexos usando técnicas como NoSQL, MapReduce e Aprendizagem Computacional”.

Outra tendência tecnológica recente é a da abertura de dados, ou *Open Data*. Segundo a *Open Knowledge Foundation* [okfn] dado aberto é “todo aquele que pode ser livremente utilizado, compartilhado e transformado por qualquer pessoa, em qualquer lugar e por qualquer motivo”. O princípio dos dados abertos se aplica a qualquer dado gerado em qualquer área da sociedade, mas ele se tornou particularmente importante nas áreas governamentais devido a três fatores. Por um lado, os dados governamentais são gerados a partir de sistemas financiados pela sociedade. Argumenta-se que, por este motivo, a sociedade tem o direito de ter

acesso a estes dados. Por outro lado, os órgãos governamentais tem dificuldade em inovar, tanto por falta de recursos técnicos e financeiros como por terem uma estrutura organizacional que desestimula a inovação e o risco. A abertura de dados governamentais, neste caso, permite que a inovação na resolução de problemas do governo venha da sociedade civil, e que o governo colha os frutos desta inovação. Finalmente, a abertura dos dados aumenta a transparência do governo e estimula a participação dos cidadãos na fiscalização e nas decisões governamentais, dificultando desvios e aumentando a eficiência da máquina pública.

O processo de abertura do governo à sociedade civil é chamado de *Open Government*, ou Governo Aberto. O Governo Aberto é baseado na ideia de que “as pessoas tem o direito de acessar os documentos e os procedimentos do governo” [Lathrop’10]. Diversas cidades e países ao redor do mundo estão adotando a doutrina do Governo Aberto para guiar suas ações. Nos Estados Unidos, logo que assumiu, o presidente Barack Obama redigiu um memorando sobre transparência e governo aberto [Obama’09] que define diretrizes governamentais que favorecem a transparência e a abertura de dados do governo. Cidades como Boston⁹ e Chicago¹⁰ divulgam diversos dados do governo e mantém estas bases atualizadas. Na área específica do transporte, a RATP¹¹ em Paris e a TfL¹² de Londres são duas empresas que não apenas publicam os dados em formato aberto como também fornecem APIs (*Application Programming Interface*) para acesso a diversos dados dinâmicos.

No Brasil, a Lei de Acesso à Informação [Lei12527] foi um marco na abertura de dados e na transparência do governo. Através desta lei, qualquer cidadão pode requisitar acesso à qualquer informação do governo e, caso não existam restrições ao acesso dos dados, eles são liberados ao cidadão. A lei determina os mecanismos, prazos e princípios para a liberação dos dados governamentais em todas esferas públicas. No Brasil, diversas cidades estão adotando o modelo de governo aberto, como por exemplo, São Paulo¹³. No caso específico do transporte público, entre as cidades que fornecem dados abertos à sociedade civil como forma de aumentar a fiscalização e estimular a inovação, estão o Rio de Janeiro¹⁴ e São Paulo¹⁵.

A área de planejamento urbano, e em particular a área de transportes, tradicionalmente consome grandes massas de dados. A pesquisa de origem/destino

9 <http://www.cityofboston.gov/open>

10 <https://data.cityofchicago.org>

11 <http://data.ratp.fr>

12 <https://tfl.gov.uk/info-for/open-data-users>

13 <http://saopauloaberta.prefeitura.sp.gov.br/>

14 <http://data.rio>

15 http://www.prefeitura.sp.gov.br/cidade/secretarias/desenvolvimento_urbano/dados_abertos

realizada pelo Metrô de São Paulo em 2007, por exemplo, abrangeu uma região com população estimada de 19,5 milhões de pessoas através de uma amostra de 30 mil domicílios. Apesar de ter sido feita em 2007, seus resultados foram divulgados apenas em dezembro/2008.

Essa demora na divulgação dos resultados se justifica por diversos fatores. Os dados são obtidos manualmente, e devem ser digitalizados antes de serem processados. Além disso, os dados também devem ser georreferenciados manualmente para que as estatísticas apresentadas possam ser relacionadas a localizações geográficas na cidade. Em todo o processo de coleta e análise de dados para este tipo de estudo, existem diversos procedimentos humanos que retardam e dificultam a análise dos dados.

Até pouco tempo atrás, portanto, essa era a realidade dos engenheiros e responsáveis pelo planejamento de transportes das grandes cidades brasileiras: as decisões eram tomadas a partir de dados coletados em estudos extensos, dispendiosos, com tempo de execução relativamente longo e que deveriam ser cuidadosamente planejados.

Estamos hoje no meio de uma mudança de paradigma. A evolução do poder de processamento dos computadores e a redução do preço do armazenamento de dados, que está mudando a sociedade, também vem gerando impactos na gestão e no planejamento de transportes, como não poderia deixar de ser. A principal mudança foi a popularização dos ITS (do inglês *Intelligent Transport Systems*). ITS combinam alta tecnologia, e avanços dos sistemas de informação, comunicação, sensores e métodos matemáticos avançados com o mundo convencional da infraestrutura de transporte [Sussman'96] para a resolução de problemas de transporte como congestionamentos, segurança, eficiência dos transportes e impacto no meio-ambiente [Figueiredo'01].

O conceito de ITS não é recente. As discussões originais sobre o uso da tecnologia para melhorar a eficiência dos sistemas de transporte datam da década de 30, com os primeiros semáforos elétricos [Figueiredo'01]. Desde então a tecnologia vem se tornando cada vez mais importante na gestão de trânsito e no planejamento de transporte com a introdução de diversos outros sistemas que hoje fazem parte do dia-a-dia de diversas cidades, como a cobrança automatizada de pedágios, parquímetros eletrônicos, semáforos sincronizados e radares de detecção de alta velocidade.

Este processo se intensificou a partir da metade da década de 90 quando

Smart Motorway



Figura 2 - Velocidade dinâmica em estradas inglesas [Highways England]

a tecnologia começou a se tornar mais presente e houve uma padronização dos sistemas e das soluções. Hoje em dia, sistemas de rastreamento de veículos, bilhetagem eletrônica e de navegação por GPS são comuns e utilizados por bilhões de pessoas diariamente.

A popularização dos ITS teve como consequência a digitalização automática de diversas transações que antes eram analógicas. Processos de coleta e consolidação de dados que antes eram caros e praticamente inviáveis como a contabilização do número de passagens utilizadas diariamente em uma grande cidade, passaram a ser viáveis e baratos com a introdução, por exemplo, dos sistemas de bilhetagem eletrônica. A disponibilidade quase imediata de dados deste tipo e a facilidade de utilizá-los nos processos de operação, gestão e planejamento da rede de transportes tem o potencial de revolucionar a forma como as cidades administram suas malhas de transporte.

Diversas cidades estão desenvolvendo iniciativas de uso das tecnologias de *Big Data* no processamento das imensas bases de dados geradas pelos ITS em combinação com dados provenientes de outras fontes e aplicando os resultados na gestão do transporte. Um dos países pioneiros no uso do *Big Data* para gestão do transporte é o Reino Unido.

Uma das iniciativas interessantes são as *Smart Motorways* [*SmartMotorways*]. São rodovias que utilizam dados vindos de uma série de sensores como câmeras e sensores de tráfego instalados no piso da rodovia para monitorar o tráfego e alterar dinamicamente o comportamento da pista de acordo com as informações coletadas. Os limites de velocidade, por exemplo, são alterados conforme a quantidade de veículos (maiores quando há menos veículos e menores à medida que a quantidade de veículos aumenta), pistas são fechadas em caso de acidente e o acostamento é liberado para tráfego, se conveniente.

Outro uso de Big Data no Reino Unido é para determinação dos locais de embarque, desembarque e conexão dos passageiros do sistema de transporte público. A TfL - *Transport For London* participou de um projeto [Gordon'12] de análise de dados em conjunto com o MIT - *Massachusetts Institute of Technology*. Este projeto combina dados de tarifação dos *Oyster Cards* - os cartões do sistema de bilhetagem eletrônica - com dados de posicionamento dos ônibus para inferir as viagens realizadas pelos passageiros. Estas informações são importantes para que a TfL compreenda melhor as necessidades de mobilidade dos passageiros e possa adaptar frequências de passagem, trajetos e pontos de ônibus a estas necessidades.

Finalmente, uma última aplicação interessante de *Big Data* na gestão de transportes de Londres é a caracterização dos passageiros a partir de dados do *Oyster Card* [Ortega-Tong'13]. Este estudo estende a metodologia para a inferência de viagens proposta em [Gordon'12] para classificar os passageiros em perfis de acordo com a sua movimentação pela cidade. Foram identificados dois grandes grupos de passageiros: passageiros frequentes e passageiros ocasionais. Estes grupos foram subdivididos conforme sua frequência de viagem e modal utilizado, resultando em oito grupos de passageiros. As informações sobre os padrões de viagem destes grupos e a comparação entre estes grupos de usuários anônimos e os padrões de viagem dos grupos de usuários registrados no sistema ajudaram a TfL a planejar melhor a rede e a quantificar o viés das análises que utilizam apenas dados dos usuários registrados.

A cidade de Santiago no Chile também aplica as técnicas de *Big Data* para análise da sua rede de transportes. A Transantiago, como é chamada a rede, foi criada em 2007 e possui bilhetagem eletrônica e rastreamento GPS nos veículos. Estes dados foram utilizados para analisar a qualidade do serviço de transporte oferecido [Núñez'13]. Foram estudados diversos aspectos como (i) tempo de viagem, (ii) velocidade de viagem dos usuários, (iii) razão distância na rota/distância euclidiana, (iv) conexões por viagem, (v) velocidade do ônibus x velocidade dos carros e (vi) conexões críticas. Estes valores foram usados para sugerir mudanças na rede de transporte como por exemplo a construção de corredores ou a criação de novas linhas para substituir uma linha muito longa.

O projeto detalhado neste caderno técnico se enquadra na mesma orientação das iniciativas descritas anteriormente. Utilizamos *Big Data*, dados abertos e técnicas de análise de dados para obtenção de informações qualitativas sobre o funcionamento da rede de transportes. Descrevemos uma metodologia para o cálculo de métricas de qualidade de serviço da rede de ônibus de uma cidade, a partir de bases de dados massivas de deslocamento dos veículos e de dados abertos. Esta metodologia foi utilizada na cidade de São Paulo a pedido da SPTrans, o órgão responsável pela gestão da rede de transportes da cidade, a fim de calcular os valores de duas métricas principais: o índice de cumprimento de partidas por linha e o índice de respeito às frequências de passagem planejadas por linha e por ponto. Estes valores seriam utilizados na redação da nova licitação do sistema de transportes da cidade de São Paulo a fim de estabelecer níveis de qualidade de serviço superiores ao serviço prestado até então.

4

Metodologia

Descrição da metodologia aplicada

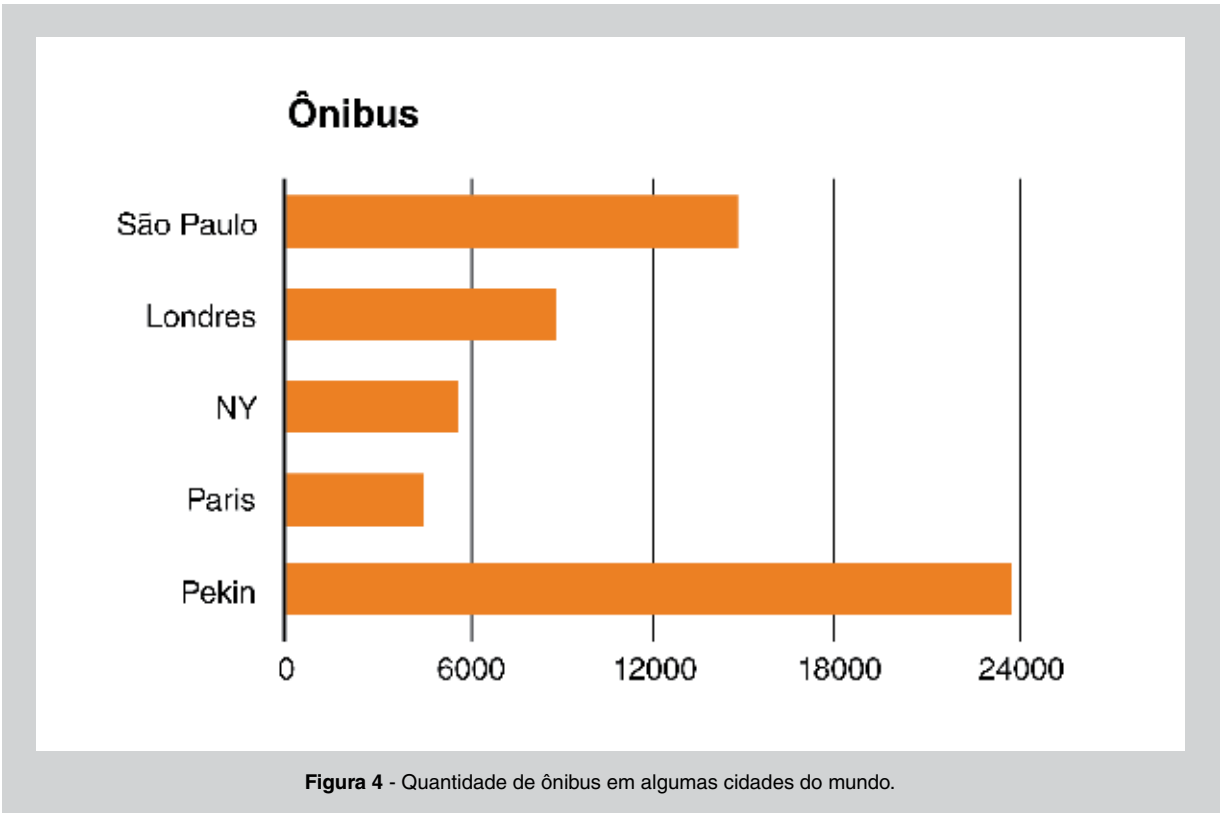
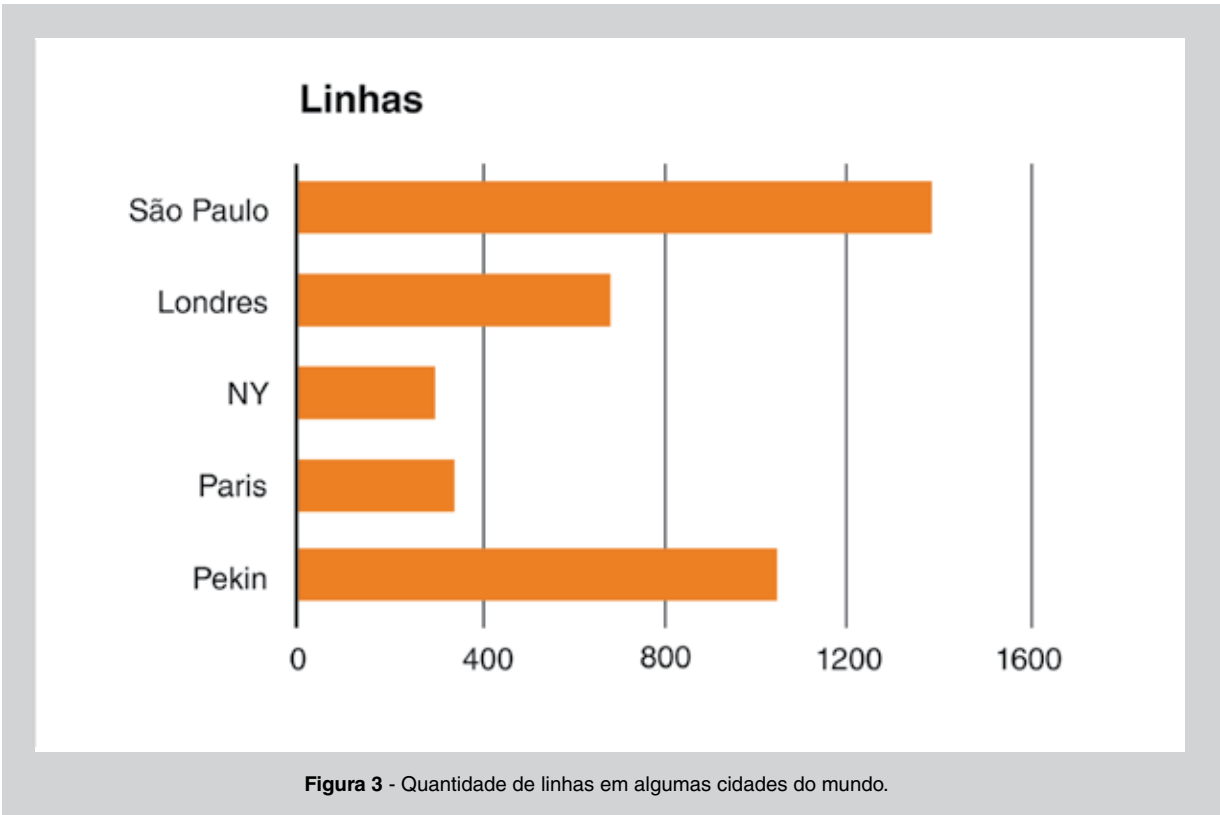
A metodologia que descreveremos a seguir foi aplicada à rede de transportes de São Paulo. O sistema de transporte público sobre rodas de São Paulo é um dos mais complexos do mundo. Com cerca de 15.000 veículos e mais de 1.300 linhas, a rede de ônibus da cidade de São Paulo é mais complexa que a da maioria das cidades européias e é uma das maiores do mundo, como pode ser visto nas figuras 03, 04 e 05.

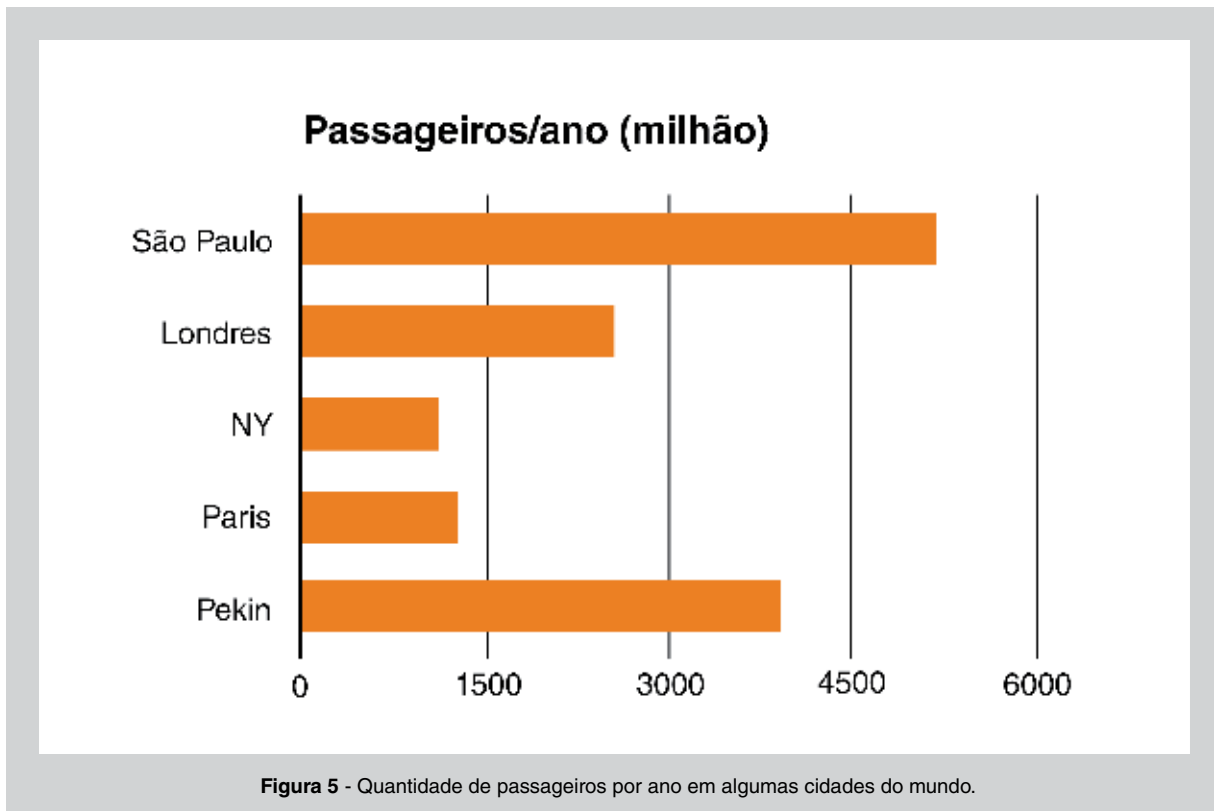
Nossa metodologia se baseia na análise conjunta das seguintes bases de dados:

1. GTFS

2. AVL

A seguir detalharemos o conteúdo e formato de cada uma dessas bases de dados.





GTFS

GTFS é o acrônimo de *General Transit Feed Specification*. É um formato padrão de troca de dados de transporte, incluindo dados de programação e dados geográficos⁹. É um formato de dados que foi desenvolvido originalmente em uma parceria entre o Google e a TriMet, a agência de transportes de Portland, Oregon. O formato foi inicialmente chamado de *Google Transit Feed Specification* e depois renomeado para que ficasse claro que se tratava de um formato aberto e não de um formato do Google.

Uma base GTFS contém diversos arquivos compactados em um único arquivo com formato ZIP. Cada arquivo contém dados no formato CSV, ou seja, com valores separados por vírgulas.

A especificação GTFS define 6 arquivos obrigatórios e 7 arquivos opcionais. São obrigatórios os seguintes arquivos:

⁹ <https://developers.google.com/transit/gtfs/>

o **agency.txt** contém informações sobre a(s) agência(s) de trânsito que fornecem dados para esta base

o **stops.txt** cada local onde os passageiros embarcam e desembarcam dos veículos

o **routes.txt** um serviço de transporte, formado por uma ou mais viagens

o **trips.txt** as viagens do sistema. Cada viagem é uma sequência de duas ou mais paradas que acontece em um determinado horário

o **stop_times.txt** horários em que os veículos chegam e saem de cada parada para cada viagem

o **calendar.txt** datas de operação de cada serviço. Define quando um serviço começa, termina e em quais dias da semana está disponível

Já os seguintes arquivos são opcionais:

o **calendar_dates.txt** exceções dos serviços em algumas datas do calendário (por exemplo feriados)

o **fare_attributes.txt** informações sobre valores de passagens e validade de tickets

o **fare_rules.txt** regras sobre como aplicar as tarifas nos serviços definidos

o **shapes.txt** percurso geográfico das viagens definidas pela agência

o **frequencies.txt** intervalo entre saídas dos serviços com frequências variáveis

o **transfers.txt** regras para fazer conexões e transferências entre serviços

o **feed_info.txt** informações adicionais sobre os dados como por exemplo versão e data de expiração

A figura 6 resume graficamente a relação entre os arquivos de uma base GTFS.

Em resumo, o arquivo GTFS fornece informações **estáticas** sobre a rede de transportes de uma cidade. A partir do GTFS é possível saber quais são os trajetos dos ônibus, onde se localizam os pontos de ônibus e qual a frequência de passagem de cada linha.

É difícil determinar quantas agências publicam informações sobre suas redes de transporte no formato GTFS, mas no site *GTFS Data Exchange*¹⁰ estão catalogadas aproximadamente mil agências de transporte ao redor do mundo que publicam seus dados neste formato.

Existe também um formato de dados para a publicação de informações em tempo real, chamado *GTFS-realtime*¹¹. Ao contrário do GTFS, este formato é bem menos popular e poucas agências o utilizam para anunciar mudanças das informações estáticas (por exemplo o cancelamento de viagens em razão de um acidente) ou a posição dos ônibus em tempo real.

As redes de transporte das grandes cidades são extremamente complexas. Existem serviços que operam apenas durante algumas horas de alguns dias da semana, conexões entre serviços que são gratuitas em alguns casos mas que são pagas em outros, as velocidades dos veículos variam durante o dia e os horários de paradas nos pontos de ônibus também, entre outras dificuldades. Apesar do formato GTFS suportar a definição de redes de transporte com alto nível de complexidade, a autoridade de transporte pode ter dificuldades em criar uma descrição precisa contendo todos os detalhes do sistema. Nossa metodologia inicialmente avalia a qualidade do arquivo GTFS para determinar a validade dos resultados obtidos posteriormente. A análise será mais consistente quanto mais preciso forem os dados do GTFS.

AVL é o acrônimo de *Automatic Vehicle Location*. É o nome genérico que se dá a todos os sistemas de rastreamento de veículos utilizados pelas empresas

10 <http://www.gtfs-data-exchange.com/agencies>

11 <https://developers.google.com/transit/gtfs-realtime>

GTFS: General Transit Feed Specification

Um guia para o Sistema de Trânsito

Nomes das tabelas dos arquivos .txt mostrados na **caixa laranja**.
E os campos conectando as tabelas em **texto laranja**.

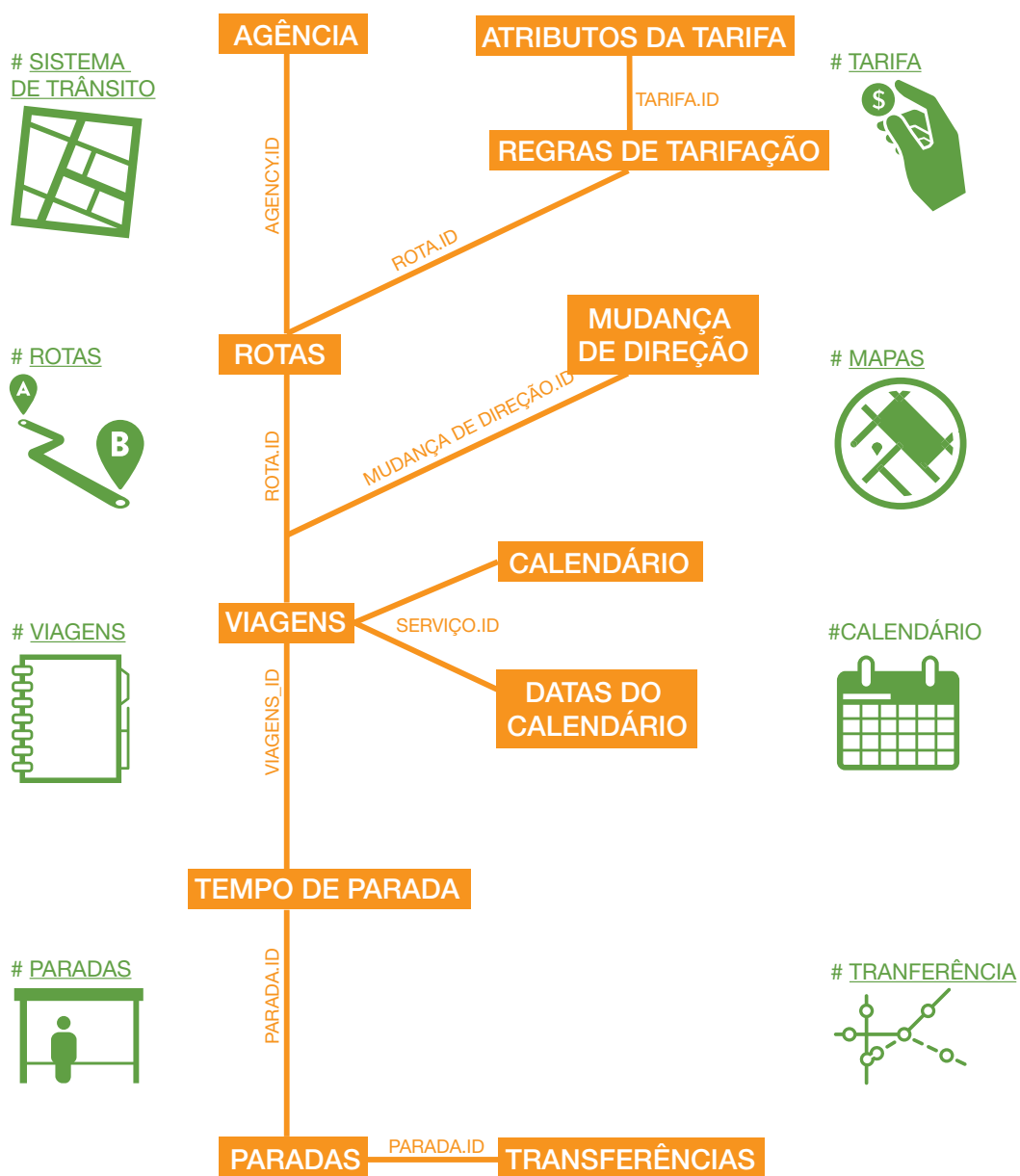
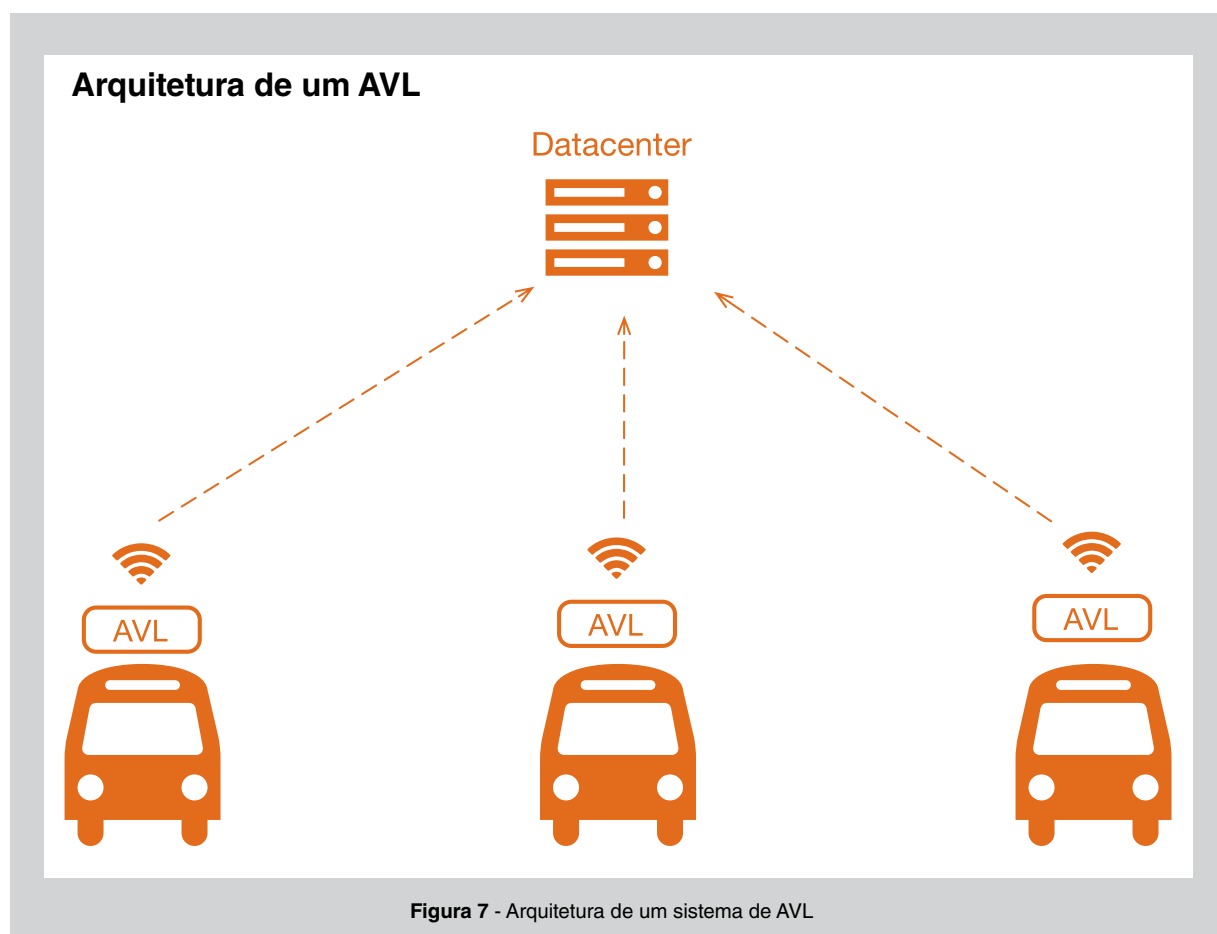


Figura 6 - Relação entre arquivos GTFS

AVL

de ônibus para localizarem seus carros na cidade. Estes sistemas normalmente são baseados em um equipamento dotado de GPS e uma conexão sem fio de dados que é instalado em cada ônibus. Ao longo do dia, os equipamentos enviam a posição detectada pelo GPS por meio de uma conexão de dados 3G ou GPRS a um servidor remoto, onde eles são armazenados. A figura 7 descreve a arquitetura de um sistema de AVL.

O formato dos dados de um sistema de AVL não é padronizado. Cada provedor



de sistemas de rastreamento define seus próprios protocolos de transmissão de dados proprietários e seus formatos de dados particulares. Em geral, os dados transmitidos pelo AVL contém as coordenadas (latitude e longitude), horário de registro da coordenada e identificador do equipamento.

Como exemplo, o sistema de AVL da cidade de São Paulo gera os seguintes arquivos a cada dia de operação:

o **AL_[data].txt** arquivo no formato CSV (com dados separados por vírgula) contendo as seguintes informações: número da linha, complemento, código da linha e sentido da viagem

o **AV_[data].txt** arquivo no formato CSV (com dados separados por vírgula) contendo as seguintes informações: código da linha e o código do veículo realizando o trajeto daquela linha

o **MO_[data].txt** arquivo no formato CSV (com dados separados por vírgula) contendo as informações de movimento dos veículos: horário em que o dado foi recebido pelo servidor, horário em que o dado foi capturado pelo equipamento, código da linha, latitude, longitude e o código do equipamento

Destes arquivos, os que contém a maior quantidade de dados são os arquivos **MO**. Em São Paulo, estes arquivos compreendem mais de 28 milhões de registros atingindo mais de 2GB de dados para apenas um dia de operação da rede de transportes sem nenhum processamento.

A própria arquitetura do sistema de rastreamento, entretanto, faz com que uma parte importante dos dados coletados não sejam exploráveis. O sistema de GPS tem dificuldades de operar em cidades, em particular em regiões com prédios altos que formam “vales urbanos”. Nestas situações o sistema de GPS não tem visibilidade para detectar os satélites de GPS no céu e não consegue localizar o ônibus com precisão. A conexão 3G também é uma fonte de problemas: a cobertura das operadoras de telefonia 3G na cidade varia de região para região, e locais com grande concentração de pessoas concorrendo pela conexão de telefonia celular ou locais muito periféricos com cobertura de telefonia insuficiente dificultam o envio de informações pelos equipamentos de AVL.

Uma parte do processo de funcionamento dos AVLS envolve operação manual, o que é uma outra fonte de imprecisão dos dados. O equipamento deve ser ajustado para a linha na qual o veículo está operando e esta configuração é feita manualmente pela equipe de operação. Muitas vezes o ônibus faz o trajeto de volta com o equipamento de AVL configurado para a ida, ou o ônibus é realocado para outro trajeto mas o equipamento não é reconfigurado. Estes erros de operação também geram dados incorretos que diminuem a precisão das informações coletadas pelo

AVL.

Assim sendo, antes de se analisar os dados provenientes dos equipamentos de AVL é preciso verificar sua qualidade e eliminar dados incorretos ou muito imprecisos.

Metodologia

O objetivo da metodologia é determinar os índices de cumprimento de partidas e de frequência de passagem de todos os ônibus circulando por todas as linhas da cidade. Este cálculo leva em consideração não somente a quantidade de partidas planejadas para a linha ou a frequência de passagem estimada, mas também considera a velocidade média das vias durante o dia para estimar a quantidade de partidas possível de ser realizada levando-se em conta a situação do trânsito da cidade. Com isso é possível determinar o impacto da fluidez das vias ao longo dos dias nos índices de cumprimento de partidas e no respeito aos intervalos de passagem.

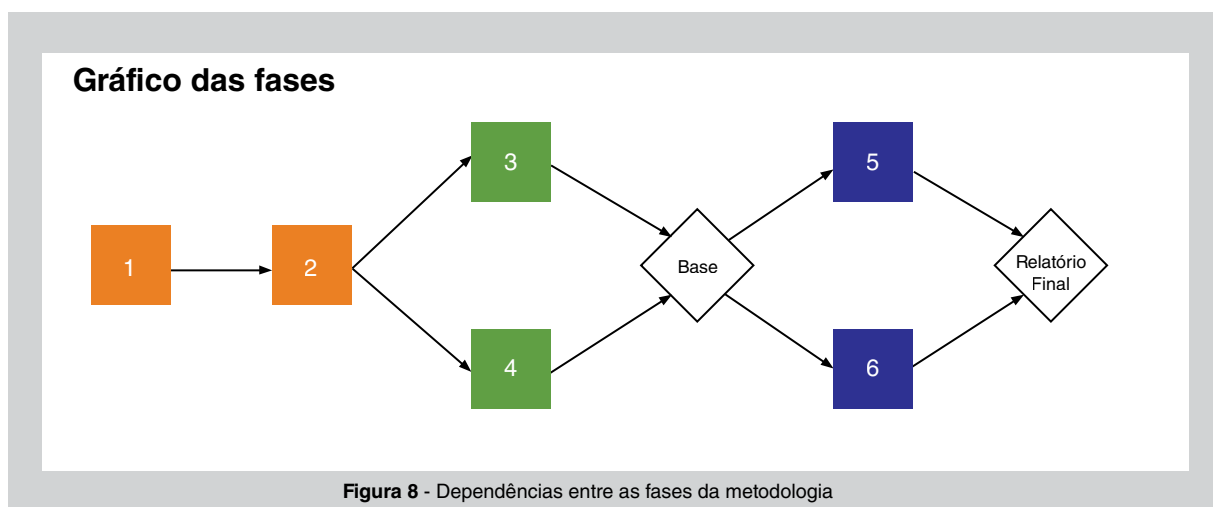
Para isso, analisamos em conjunto a base de dados GTFS, que contém os valores planejados do funcionamento do sistema de transporte, com os dados de AVL, que representam as viagens efetivamente realizadas na cidade durante o dia.

Uma outra fonte de dados que pode ser usada nesta análise são os dados de bilhetagem. Estes dados são coletados pelos validadores de bilhete instalados em cada ônibus de cidades que contam com um sistema de bilhetagem digital. Os dados de validação do bilhete podem ser associados à localização geográfica naquele instante do ônibus em que ocorreu o embarque, permitindo identificar em quais posições os passageiros efetuaram o pagamento da tarifa. Por conterem informações pessoais identificáveis, entretanto, a análise deste tipo de base de dados traz implicações de privacidade evidentes. Neste trabalho decidimos não utilizar, portanto, os dados da bilhetagem.

Nossa metodologia contém 6 etapas:

1. Exploração dos dados
2. Análise da qualidade dos dados
3. Detecção das viagens
4. Construção do mapa de velocidades
5. Cálculo dos índices de cumprimento de partidas
6. Cálculo dos desvios da frequência de passagem

As duas primeiras etapas fazem parte de uma fase preliminar de familiarização dos analistas com os dados da cidade e definição da confiabilidade dos resultados



finais que serão obtidos. As etapas 3 e 4 consistem no cálculo de valores de base para velocidades e viagens que serão utilizados nas etapas seguintes. Finalmente, as etapas 5 e 6 se referem à análise de dados propriamente dita. Levamos em conta as viagens detectadas e as velocidades médias calculadas nas etapas anteriores para determinar os cumprimentos de partida e os desvios das frequências de passagem em relação ao esperado.

Nas próximas seções, detalharemos cada etapa da metodologia acima,

exemplificando com dados do sistema de transportes de São Paulo onde a metodologia foi aplicada. Na Seção 5 descreveremos o processo de exploração dos dados e mostraremos alguns dos resultados para a cidade de São Paulo. Explicaremos também o processo de análise da qualidade dos dados da base GTFS e da base de AVLs. Na Seção 6 mostraremos como detectamos as viagens realizadas através dos dados de AVL e as dificuldades na utilização dos dados brutos para este fim. Também detalharemos o cálculo das velocidades médias das vias a partir dos dados do AVL. Finalmente na Seção 7 explicaremos como realizamos a avaliação do cumprimento de partidas e da frequência de passagens dos ônibus a partir dos dados analisados.

5

Etapa preliminar

Etapa preliminar do trabalho

Nesta etapa os analistas de dados procuram se familiarizar com a estrutura da rede de transportes da cidade analisada. Cada cidade tem particularidades que são refletidas na estrutura da malha, na disposição dos pontos de ônibus e na programação das linhas. A compreensão destas particularidades auxilia a identificação nas etapas seguintes de erros de cálculo ou inconsistências nos valores obtidos.

Também é estimada a fidelidade com que os dados fornecidos descrevem a rede de transportes da cidade. Os diversos dados são cruzados, comparados e analisados em busca de inconsistências, valores incompatíveis com a realidade ou simplesmente informações erradas.

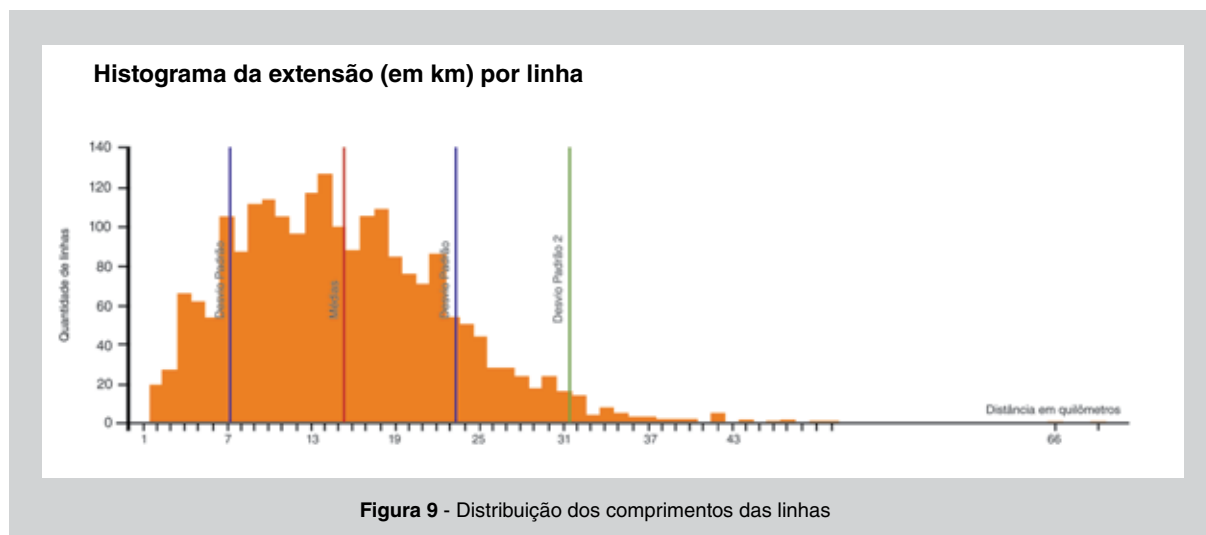
Nesta seção descrevemos os objetivos e alguns resultados das duas fases que compõem a etapa preliminar: a fase de exploração dos dados e a fase de análise da qualidade dos dados.

Exploração dos dados

De posse dos dados fornecidos para execução desta análise, é realizada uma exploração inicial para compreender a estrutura dos dados e como ela se reflete na organização da rede de transportes da cidade. Nesta fase procura-se entender como os elementos da rede são distribuídos em relações a métricas do mundo real: tempos, distâncias, quantidades. Também é analisada a rede de transportes e como ela está interconectada.

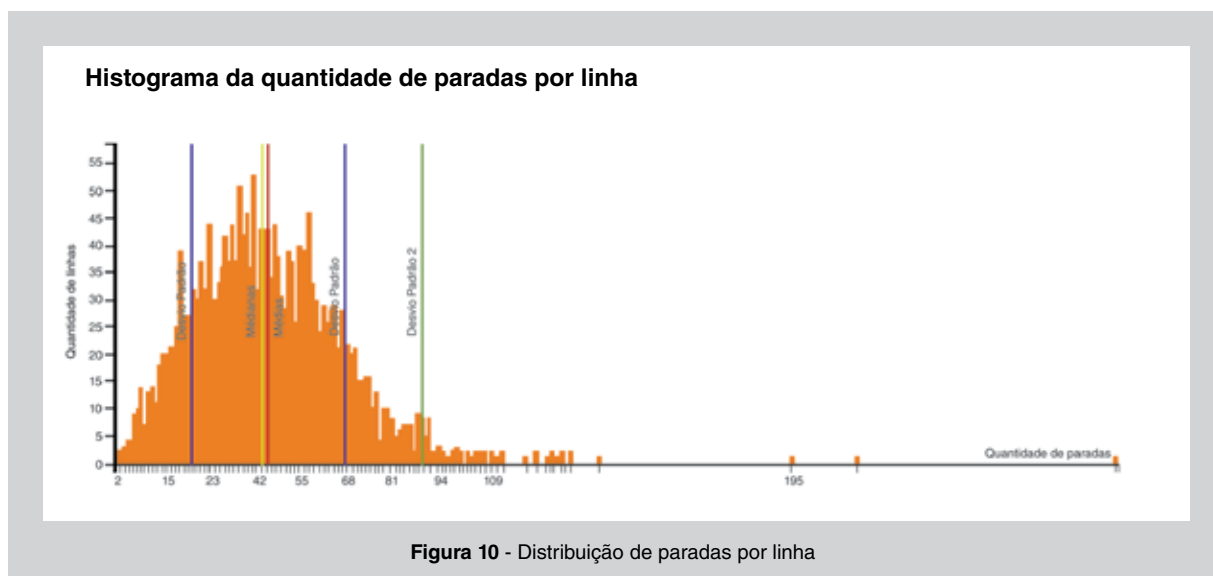
Durante esta fase algumas visualizações de dados simplificadas são criadas para auxiliar na compreensão dos dados. O objetivo destas visualizações não é o de se tornarem ferramentas de análise, ou de possibilitar a comparação de diversas fontes de dados diferentes a fim de evidenciar algumas relações. O objetivo destas visualizações é apenas de permitir aos analistas enxergarem a enorme massa de dados de forma estruturada.

No contexto da cidade de São Paulo desenvolvemos algumas visualizações exploratórias que mostramos a seguir. Estudamos, entre outros, as distribuições do comprimento das linhas, das paradas em cada linha e da quantidade de linhas que passa em cada ponto.



Estas visualizações nos ajudaram a identificar casos extremos da rede de transportes de São Paulo. Por exemplo, apesar da maioria das linhas ter comprimento ao redor de 14km, existem linhas com mais de 40 km de extensão, o que distorce os valores de quantidade de partidas ou velocidade média dessas linhas. Em relação à quantidade de paradas, apesar da maioria das linhas terem ao redor de 40 paradas,

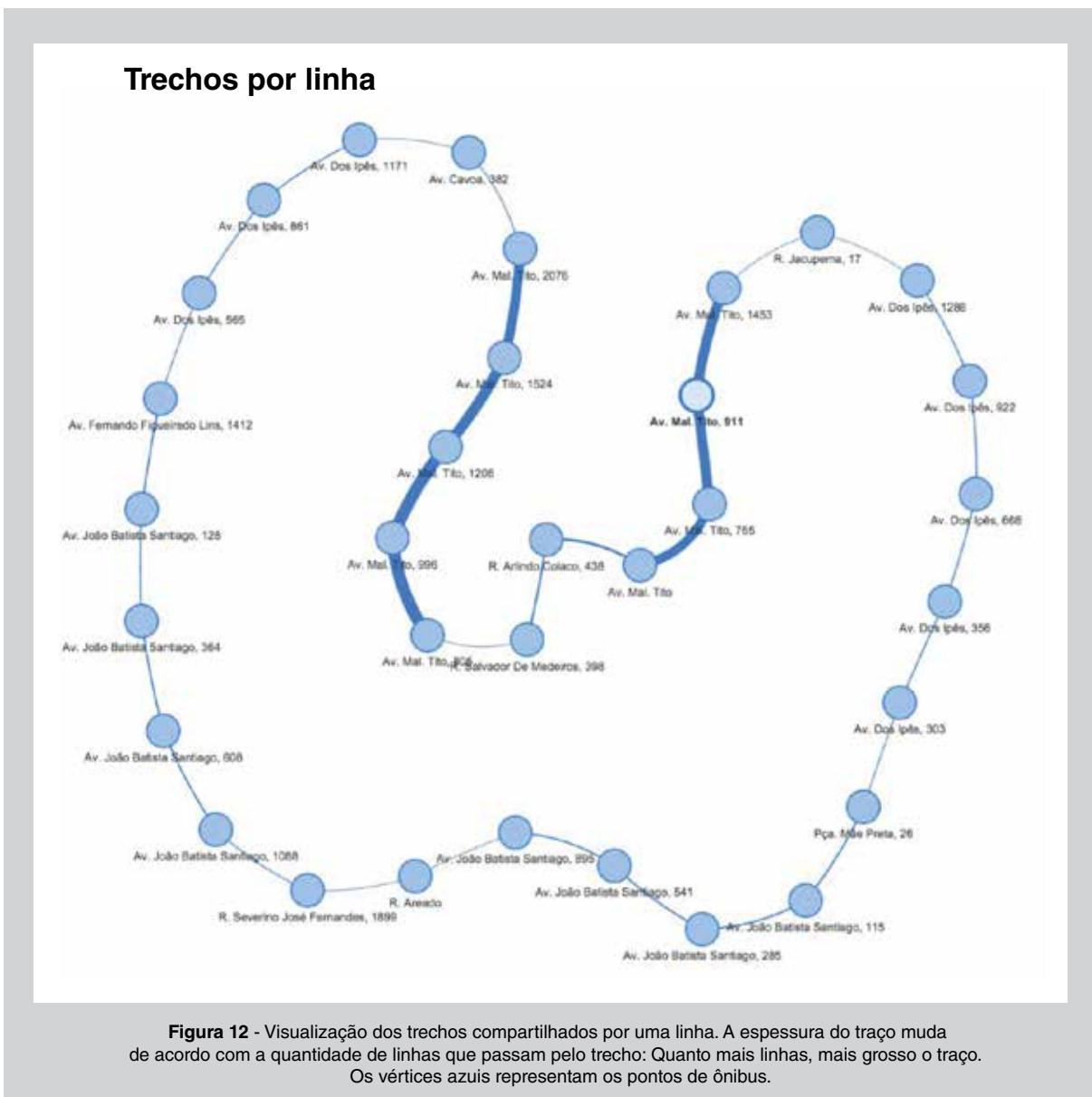
algumas chegam a ter mais de 100, enquanto outras tem menos de 10 paradas. Finalmente, na maioria dos pontos de ônibus da cidade passa apenas uma linha de ônibus, o que dificulta a conexão com outras partes do sistema. Ao mesmo tempo, existem pontos onde passam mais de 40 linhas diferentes.



Evidentemente, estes casos representam uma pequena parcela do total de linhas analisadas. Entretanto, o conhecimento destes extremos ajuda a compreender a razão de alguns resultados anormais na sequência da análise.

Criamos também algumas visualizações para compreender melhor como os ônibus compartilham a estrutura viária da cidade. Procuramos entender em quais vias existem mais ônibus simultaneamente, qual parte do trecho de uma linha é compartilhado com ônibus de outras linhas ou quantos ônibus de linhas diferentes passam pela mesma via. Para isso, construímos um grafo onde cada vértice é um ponto de ônibus e onde a espessura da aresta corresponde ao número de linhas diferentes que passam por aquela aresta. Algumas das visualizações são apresentadas nas figuras 12 e 13.

Mais uma vez, o objetivo destas visualizações é permitir aos analistas terem uma ideia geral dos dados que serão analisados. Devido à quantidade de informações, é impossível avaliar os dados individualmente. As visualizações desta



Trechos 50 linha

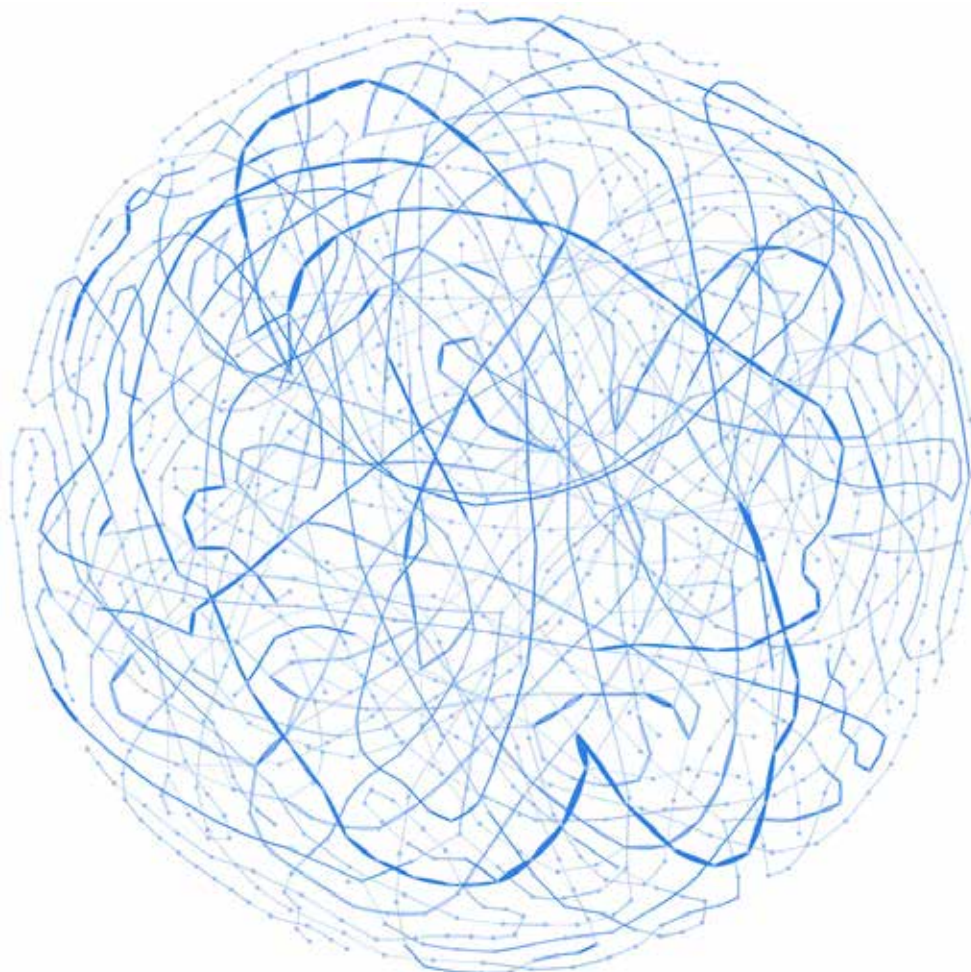


Figura 13 - Visualização dos trechos compartilhados entre 50 linhas, sendo a espessura do traço representando a densidade de linhas no trecho e os vértices representando os pontos de ônibus.

fase permitem que os analistas identifiquem partes problemáticas dos dados e se aprofundem em um subconjunto dos dados.

Análise da qualidade dos dados

Após a exploração inicial dos dados, é possível definir algumas métricas para avaliar a qualidade dos dados fornecidos e a confiança dos resultados obtidos ao fim do projeto. Analisamos separadamente a qualidade dos dados do GTFS e dos dados de AVL.

Para o GTFS, foram determinadas 19 métricas de qualidade que descrevemos abaixo:

1 Número de paradas distantes mais de 500m quantidade de pontos de ônibus distantes a mais de 500m do ponto de ônibus seguinte

2 Número de paradas distantes menos de 30m quantidade de pontos de ônibus consecutivos distantes a menos de 30m

3 Partidas diárias previstas total de partidas previstas no dia para a linha

4 Linha sem frequência de partidas documentada caso a linha não possua planejamento de partidas, ela é sinalizada

5 Total de horas de operação quantidade de horas em que existem partidas programadas no GTFS

6 Quantidade de shapes na linha número de pontos que definem o trajeto da linha sobre o viário

7 Distância total distância do ponto inicial ao ponto final de uma viagem, em km, considerando-se o trajeto sobre o viário da cidade

8 Velocidade média prevista velocidade média planejada para a linha, levando-se em conta seu comprimento e os horários previstos de parada em cada ponto de ônibus

9 Quantidade de paradas duplicadas número de paradas que aparecem mais de uma vez na sequência de paradas de uma linha

10 Quantidade de paradas únicas número de paradas únicas que aparecem ao longo da linha

Trajeto linha Metrô Parada Inglesa

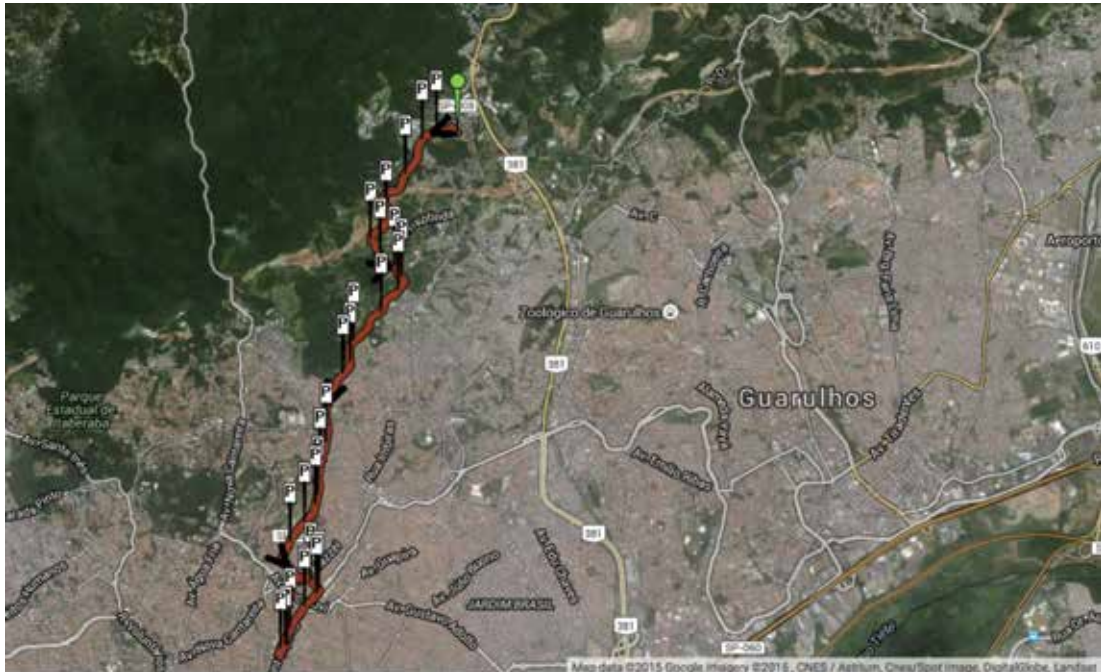


Figura 14 - Trajeto identificado no mapa da linha Metrô Parada Inglesa, com respectivos pontos

11 Linha que passa 2 vezes pela mesma parada se uma linha passa mais de uma vez pela mesma parada, sinalizamos

12 Paradas em comum na ida e na volta número de paradas com mesmo identificador na ida e na volta de uma linha

13 Tempo previsto de viagem tempo planejado para uma viagem de um terminal a outro

14 Frequência média por hora quantidade de ônibus desta linha que devem partir a cada hora, em média

15 Intervalo médio entre ônibus intervalo médio entre partidas consecutivas de ônibus realizando o trajeto desta linha

16 Porcentagem de trechos com baixa confiabilidade para cálculo da velocidade média para uma dada linha, qual a porcentagem de trechos em que o intervalo de passagem de ônibus, considerando-se **todas linhas que passam pelo trecho**, é superior a 10 minutos

17 Porcentagem de paradas próximas ao shape qual porcentagem das paradas desta linha está a menos de 50m do shape mais próximo

18 Maior distância de uma parada ao shape maior erro entre a posição da parada de ônibus e do shape mais próximo

19 Porcentagem de trechos com velocidades inconsistentes qual a porcentagem de trechos na linha que possuem velocidades muito abaixo ou muito acima de valores típicos para um ônibus

Entre as métricas que merecem ser detalhadas está o índice de confiabilidade no cálculo da velocidade média da linha. Este índice depende da quantidade de ônibus passando pelos trechos daquela linha, incluindo aí veículos de outras linhas. Na Seção 8 apresentamos a metodologia para cálculo da velocidade média, mas apenas como intuição, quanto mais ônibus passam por uma rua mais confiável a velocidade média calculada para aquela rua pois os valores são atualizados frequentemente (a cada passagem de um ônibus). Pode ser que aquela linha tenha alta frequência de passagem e os ônibus daquela linha, independente das outras, são suficientes para que o cálculo da velocidade média seja preciso. Pode ser que a linha tenha baixa frequência, mas ela compartilha grande parte do trajeto com outras linhas de forma que a frequência conjunta de todos os trechos do trajeto seja alta.

Definimos que, para cada trecho, se a frequência média conjunta de todos os ônibus que passam pelo trecho é superior a um ônibus a cada 10 minutos para 50% dos trechos de uma linha, então esta linha tem alta confiabilidade no cálculo da sua velocidade média.

Com relação a outras métricas, na cidade de São Paulo, por exemplo, a linha 3310-10 Cidade Tiradentes tem mais de 56 pontos com distância superior a 500 m entre eles. A linha 7082-10 Jd. Horizonte Azul tem 15 pontos de ônibus duplicados no trajeto. A linha 719P-10 Metrô Armênia possui apenas 31,3% dos pontos de paradas a menos de 50 m do shape do trajeto. Finalmente, apenas 0,6% das linhas analisadas possuem mais de 50% dos trechos com pouca confiabilidade no cálculo

da velocidade média.

Para os dados de AVL, foram definidas 16 métricas a fim de determinar a quantidade de dados que podem ser explorados nas etapas seguintes da análise. Esta análise é importante pois, como detalhamos na Seção 6, a própria arquitetura do sistema de AVL propicia a geração de dados incorretos. As 16 métricas são descritas abaixo.

1 Localização fora da área de análise dado de localização se encontra fora da área da cidade de São Paulo

2 Diferença superior a 24h entre tempo de coleta e de recebimento a diferença entre o horário de coleta do dado e da recepção do dado pelo servidor é superior a 24h

3 Tempo de coleta no passado dado de AVL foi coletado no dia anterior ao dia analisado

4 Tempo de coleta no futuro dado de AVL foi coletado no dia seguinte ao dia analisado

5 Código da linha não encontrado código da linha descrito no arquivo MO não se encontra no arquivo **AL** correspondente

6 Linha não encontrada linha não existe no arquivo GTFS

7 Shape não encontrado não foi possível associar a latitude e longitude do dado com um shape da linha

8 Shape mais próximo muito distante o shape mais próximo desta latitude e longitude está a mais de 300m de distância

9 Horário idêntico ao anterior horário desta coordenada é idêntico ao da coordenada anterior

10 Localização idêntica à anterior - ponto inicial latitude e longitude desta coordenada são idênticos aos da coordenada anterior e próximos ao ponto inicial

11 Localização idêntica à anterior - ponto final latitude e longitude desta coordenada são idênticos aos da coordenada anterior e próximos ao ponto final

12 Localização idêntica à anterior latitude e longitude desta coordenada são idênticos aos da coordenada anterior em locais intermediários do trajeto

13 Velocidade extremamente alta velocidade entre esta coordenada e a anterior é superior a 80km/h

14 Velocidade extremamente baixa velocidade entre esta coordenada e a anterior é inferior a 0,1km/h

15 Sentido invertido o shape associado a esta coordenada é anterior ao shape associado à coordenada anterior, como se o ônibus “andasse para trás”

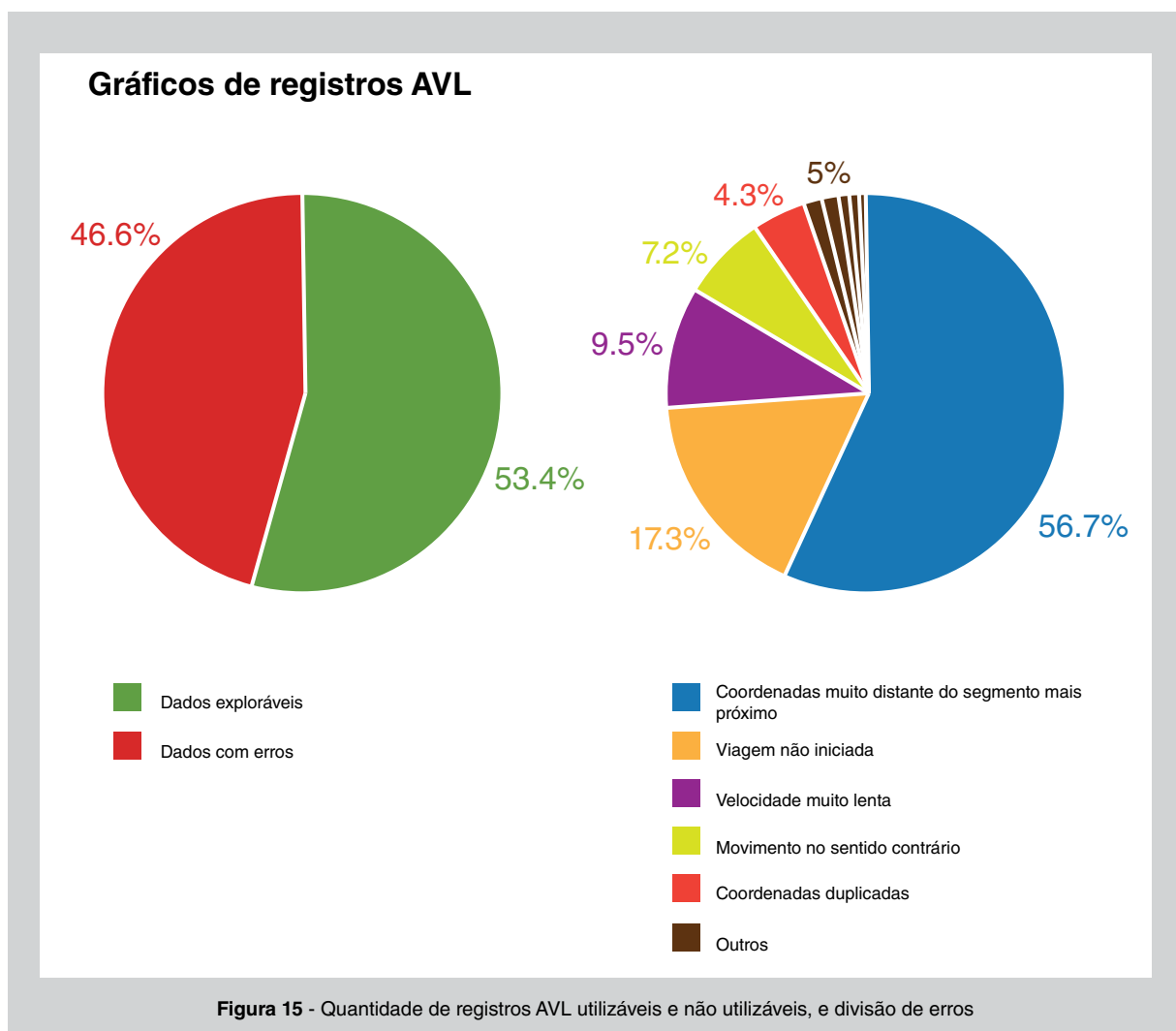
16 Localização em nuvem se após 3 coordenadas a diferença entre duas coordenadas consecutivas é sempre inferior a 50m, as coordenadas estão dentro de uma nuvem de valores similares

Considerando-se alguns valores para a cidade de São Paulo, 0,4% dos registros pertencem a linhas não cadastradas. Isto é reflexo de uma inconsistência entre os dados de AVL e os dados do GTFS em relação a linhas circulares: o GTFS documenta as linhas circulares como linhas de um único sentido, enquanto os equipamentos de AVL dividem as linhas circulares em dois sentidos (ida e volta), criando uma linha adicional virtual não documentada.

Em relação às posições fora da área de análise, a linha 1783-41 Metrô Parada Inglesa teve 18,9% das posições fora da região metropolitana da cidade de São Paulo. Conforme a figura 14, esta linha se encontra no extremo norte da capital, perto do limite da cidade, e passa pela Serra da Cantareira, um local onde o sistema de GPS funciona mal. Finalmente, a linha 3787-10 Metrô Itaquera possui 9,8% dos registros indicando uma velocidade do ônibus fora do normal, também por problemas com o sistema de GPS.

Resumindo, do total de registros analisados, 46,6% dos dados não podem ser utilizados para detecção de viagens por conta de inconsistências na captura dos dados. Entre os erros detectados, o mais frequente (56,7%) foi um erro indicando que o trajeto da linha se encontra muito longe da coordenada enviada. Este erro acontece por diversos motivos: ônibus que estão na garagem mantêm o equipamento de AVL ligado sinalizando sua posição embora não estejam realizando o trajeto programado, ônibus se deslocam da garagem ao ponto inicial da linha com o equipamento AVL ligado, entre outros.

O segundo maior motivo de erros de AVL acontece quando um ônibus fica reportando sua posição a partir de um terminal de chegada antes de reconfigurar o AVL para o sentido da volta. Neste caso é como se o ônibus estivesse esperando a viagem começar no terminal errado. Estes dados também não acrescentam nada na detecção de viagens e podem ser eliminados.



Das 2.182 linhas existentes nas bases de dados, apenas uma não possuía frequências documentadas e não pode ser estudada. Além disso, não existiam registros de AVL para uma segunda linha, então não foi possível estudar as viagens desta linha. Portanto, apenas 2 linhas de um total de 2.182, ou seja, 0,09% das linhas da cidade não puderam ser analisadas através desta metodologia.

6

Valores de base

Cálculo de valores de base

Uma vez que a estrutura dos dados está clara e a análise de qualidade dos dados identificou formas de limpar os dados e remover o ruído das bases sendo analisadas, podemos começar a inferir as informações que servirão de apoio para o estudo de cumprimento de partidas e intervalos de passagem.

Duas informações são necessárias para entender o comportamento do sistema de transportes em relação a partidas e frequência de passagem. Primeiro é necessário calcular a matriz de velocidades médias das vias da cidade. A partir desta matriz podemos estimar o tempo necessário para que um ônibus realize uma viagem em um dado horário do dia. Através desta matriz descobrimos se o tempo da viagem é igual, inferior ou superior ao tempo planejado. Em segundo lugar é preciso descobrir as viagens realizadas: quando elas foram feitas, quanto tempo demoraram, se foram completas ou não, entre outros. A partir da detecção das viagens podemos passar a calcular o cumprimento de partidas e a frequência de passagens.

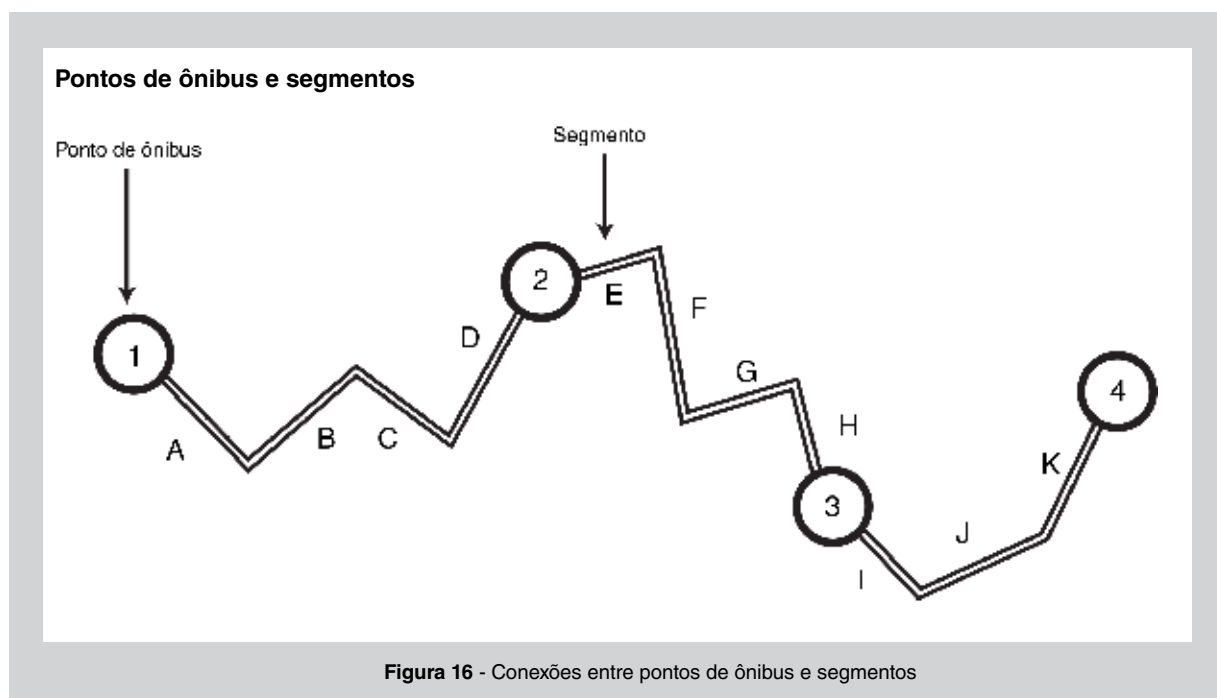
Os valores de velocidade média ajudam a compreender melhor as causas de problemas de cumprimento de partida e de frequência de passagem. Se as condições da via impossibilitam o operador de cumprir a programação de partidas, ou se um acidente bloqueou a via impedindo a passagem dos ônibus, os desvios do

previsto são compreensíveis. Caso contrário, eles devem ser coibidos.

Nas próximas seções explicaremos a metodologia para cálculo da velocidade média nas vias e o processo de detecção de viagens a partir dos dados de AVL.

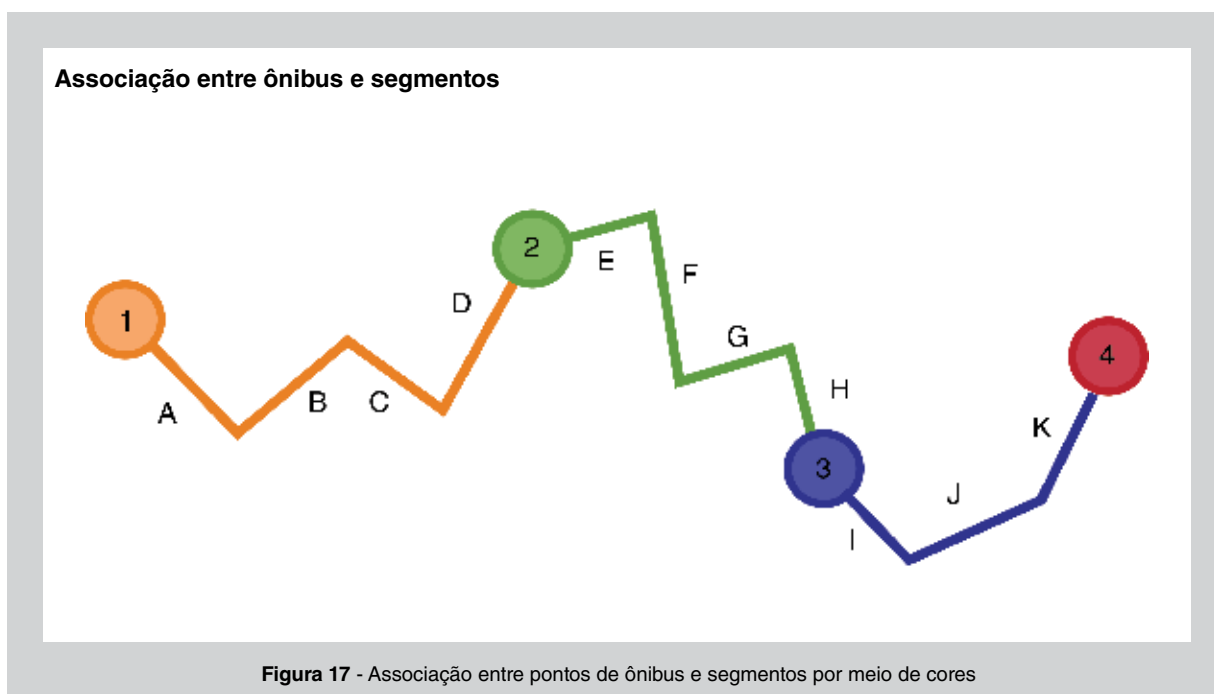
Cálculo de velocidades médias

É preciso entender inicialmente como estão estruturadas as informações da rede de transportes em relação à realidade. A figura 16 mostra as conexões entre pontos de ônibus e segmentos, que são as entidades documentadas nos arquivos GTFS.



Na figura 16 existem quatro pontos de ônibus (numerados de 1 a 4) e 11 segmentos (letras A a K). Os pontos de ônibus e os segmentos do trajeto são documentados em arquivos diferentes, e não necessariamente se encontram. Ou seja, é possível que o trajeto do ônibus definido pelos segmentos não passe através dos pontos de ônibus ou nem mesmo próximo aos pontos de ônibus.

Utiliza-se como unidade para cálculo da velocidade média o trecho entre 2 pontos de ônibus. Este trecho pode ser formado por um ou mais segmentos do arquivo GTFS. O conjunto de segmentos é então associado ao ponto de ônibus de origem. A próxima figura mostra através de cores quais segmentos são associados a quais pontos de ônibus.



A velocidade média é calculada para a distância entre os pontos de ônibus pois isso facilita o cálculo da velocidade média para uma linha qualquer e a estimativa do tempo de chegada de um ônibus ao seu destino. Entretanto, os equipamentos de AVL não fornecem sua posição ao passarem exatamente em frente a um ponto de ônibus. Os equipamentos fornecem sua posição em intervalos de tempo regulares, independente da localização, o que dificulta o cálculo da velocidade média como mostra a figura 18.

Na figura 18, um ônibus reportou sua posição entre os segmentos B e C às 9:30 AM e em seguida reportou sua posição entre os segmentos G e H às 9:42 AM. Sabendo-se que a distância entre estes dois pontos é de 4 km, é fácil calcular a velocidade entre as duas posições reportadas pelo ônibus como sendo de 20 km/h. Entretanto, isso não diz nada sobre a velocidade média entre os pontos 1 e 2 nem sobre a velocidade média entre os pontos 2 e 3.

Estimativa da velocidade média da via

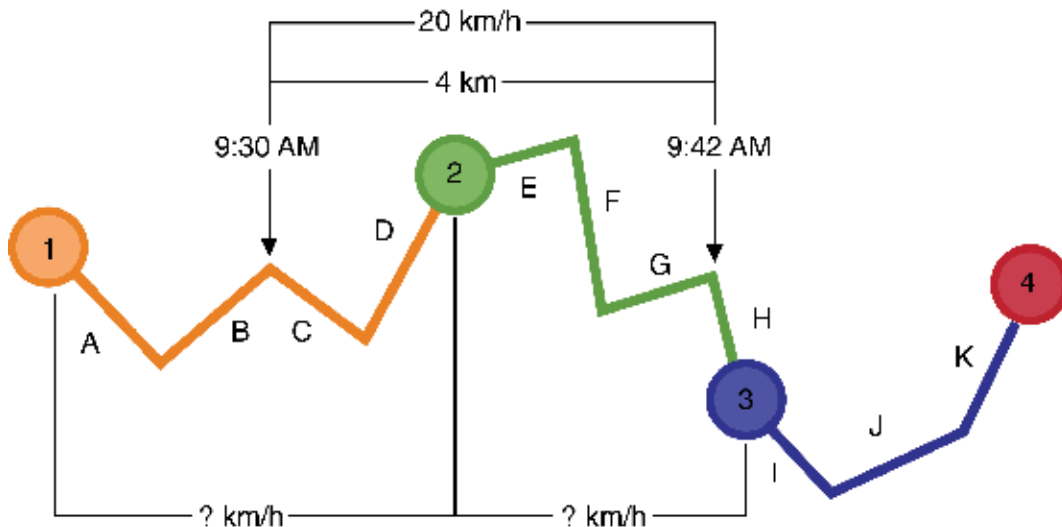


Figura 18 - Estimativa da velocidade média de trecho por meio da relação entre a posição do veículo e o horário

Para encontrar estes valores é necessário interpolar as posições reportadas até as posições dos pontos de ônibus e aproximar as velocidades médias proporcionalmente à distância do ponto de ônibus às posições reportadas pelo equipamento de AVL. Considera-se não a distância euclidiana entre os pontos mas sim a distância seguindo-se o trajeto documentado.

Além disso, qualquer ônibus trafegando pelo trecho é utilizado para atualizar a velocidade média do trecho, independente da linha que ele esteja servindo. Com isso pode-se prever com precisão os tempos de viagem mesmo para linhas com baixa frequência.

Ao final do processo, é criado um mapa com as velocidades médias de todas as vias por onde passam ônibus ao longo do tempo, similar ao da figura 19. Este mapa é utilizado posteriormente para estimar os tempos de viagens de todas as linhas ao longo do dia e para estimativa das frotas ideais para atender as frequências planejadas.

Velocidades médias

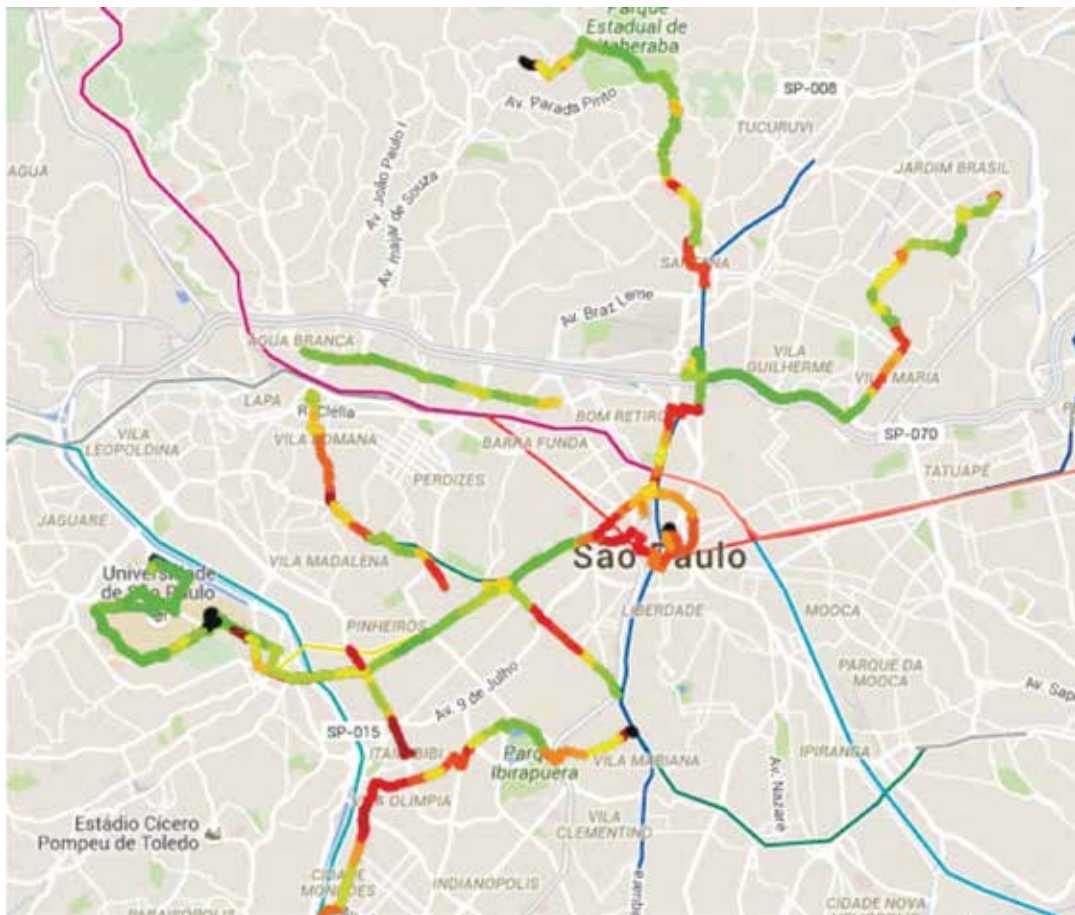


Figura 19 - Exemplo de mapa com velocidades médias. A velocidade é demonstrada pela escala de cor, sendo verde bom e vermelho ruim.

Determinação das viagens realizadas

A fase anterior de análise de qualidade dos dados permite que sejam eliminados dados de AVL que não fazem sentido no contexto do transporte da cidade. Ainda assim, é preciso identificar se os dados fazem sentido no contexto das viagens sendo analisadas. A figura 20 mostra a dispersão dos pontos de um equipamento de AVL na cidade de São Paulo.

Na imagem 20 podemos ver claramente um eixo que concentra as coordenadas reportadas pelo equipamento de AVL. Este eixo provavelmente representa o trajeto

da linha servida pelo ônibus. Mas repare que existem diversas posições reportadas pelo AVL que estão fora deste eixo principal. No alto, à esquerda, uma grande nuvem de pontos coincide com uma garagem de ônibus. Outras posições à esquerda deste eixo principal provavelmente representam o caminho do ônibus da garagem ao ponto inicial da linha e o caminho do ponto final de volta à garagem.

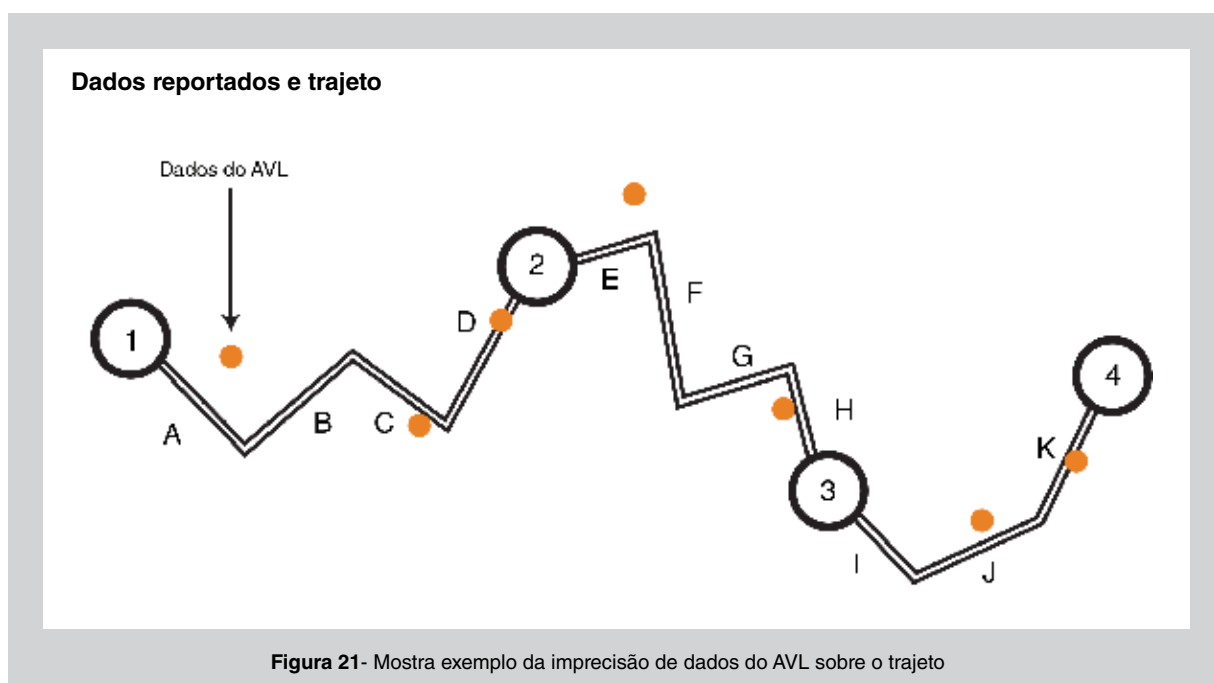


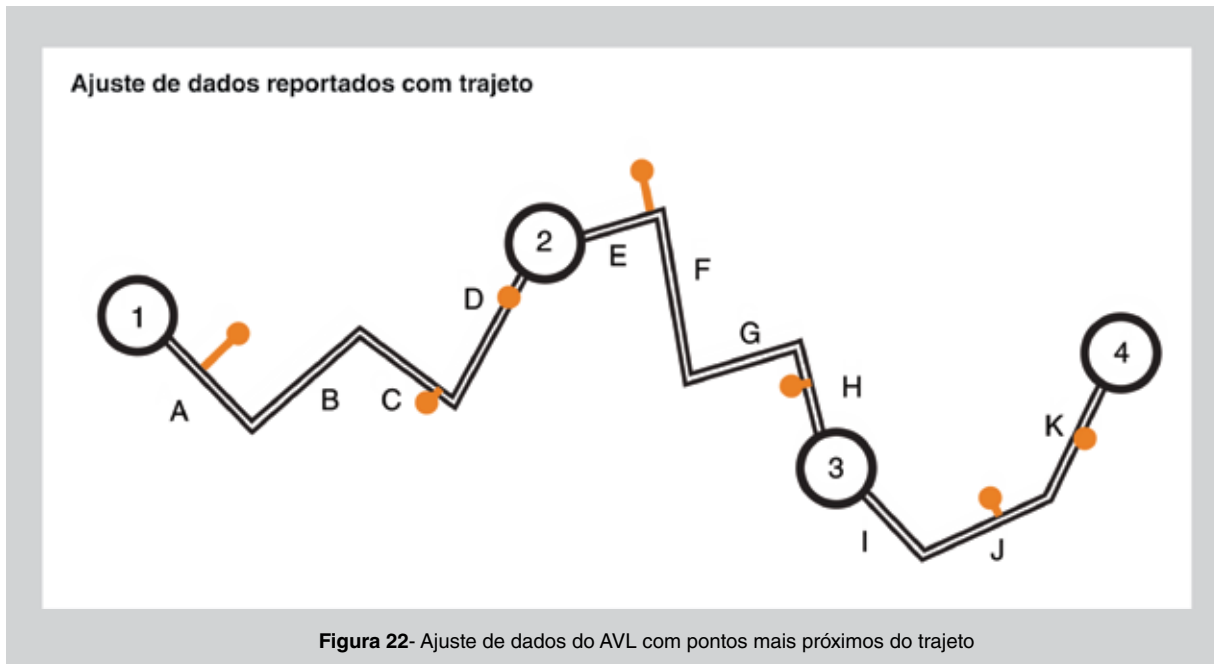
É importante notar que estas posições reportadas pelo equipamento AVL são perfeitamente válidas quando analisadas individualmente (foram capturadas durante o período de análise, se encontram na região metropolitana de São Paulo, etc.), mas se revelam inconsistentes quando analisadas em relação a outras posições reportadas pelo AVL e quando cruzadas com o trajeto da linha configurada no equipamento.

Existem três categorias de problemas com os dados de AVL que dificultam a identificação de viagens a partir dos dados do equipamento. A primeira categoria são os dados que fazem sentido quando analisados isoladamente mas no contexto de uma viagem não são consistentes com os demais dados da viagem. Por exemplo, uma localização do ônibus em um terminal da linha que ele está operando é um dado a princípio válido. Mas se este dado aparece em uma sequência de dados que mostra o ônibus em movimento longe do terminal, provavelmente essa localização próxima ao terminal se trata de um erro.

A segunda categoria de problemas que dificultam a identificação das viagens a partir dos registros de AVL é a falta de dados críticos ao longo da viagem. Em muitos casos, existem dados para restituir o trecho inicial da viagem mas não existem dados sobre o trecho final. Neste caso, não podemos afirmar, por exemplo, que a viagem foi concluída. Em outros casos, os dados cobrem, proporcionalmente, um trecho considerável da viagem mas a distância entre cada posição reportada é muito grande. Nestas situações, não podemos afirmar nada sobre longos trechos das viagens e a viagem restituída pode não ser explorável.

Finalmente, a terceira categoria de problemas diz respeito a erros na coleta de dados que podem ser corrigidos durante a fase de análise. Por um lado, alguns erros na coleta de dados não podem ser corrigidos a posteriori, como por exemplo nos casos em que o GPS detecta uma coordenada incorreta. Por outro lado, alguns erros ficam evidentes ao analisar os dados, como erros na configuração do AVL que associam um conjunto de coordenadas à uma linha no sentido contrário, por





exemplo. Nestes casos, é possível identificar a viagem e corrigir a associação entre a viagem e a linha durante a análise para que a viagem possa ser contabilizada no sentido correto.

A figura 21 mostra um exemplo de como os dados reportados pelo AVL se relacionam ao trajeto de um ônibus. Algumas posições coincidem com o trajeto, enquanto outras podem estar distantes do trajeto por imprecisão do equipamento de GPS.

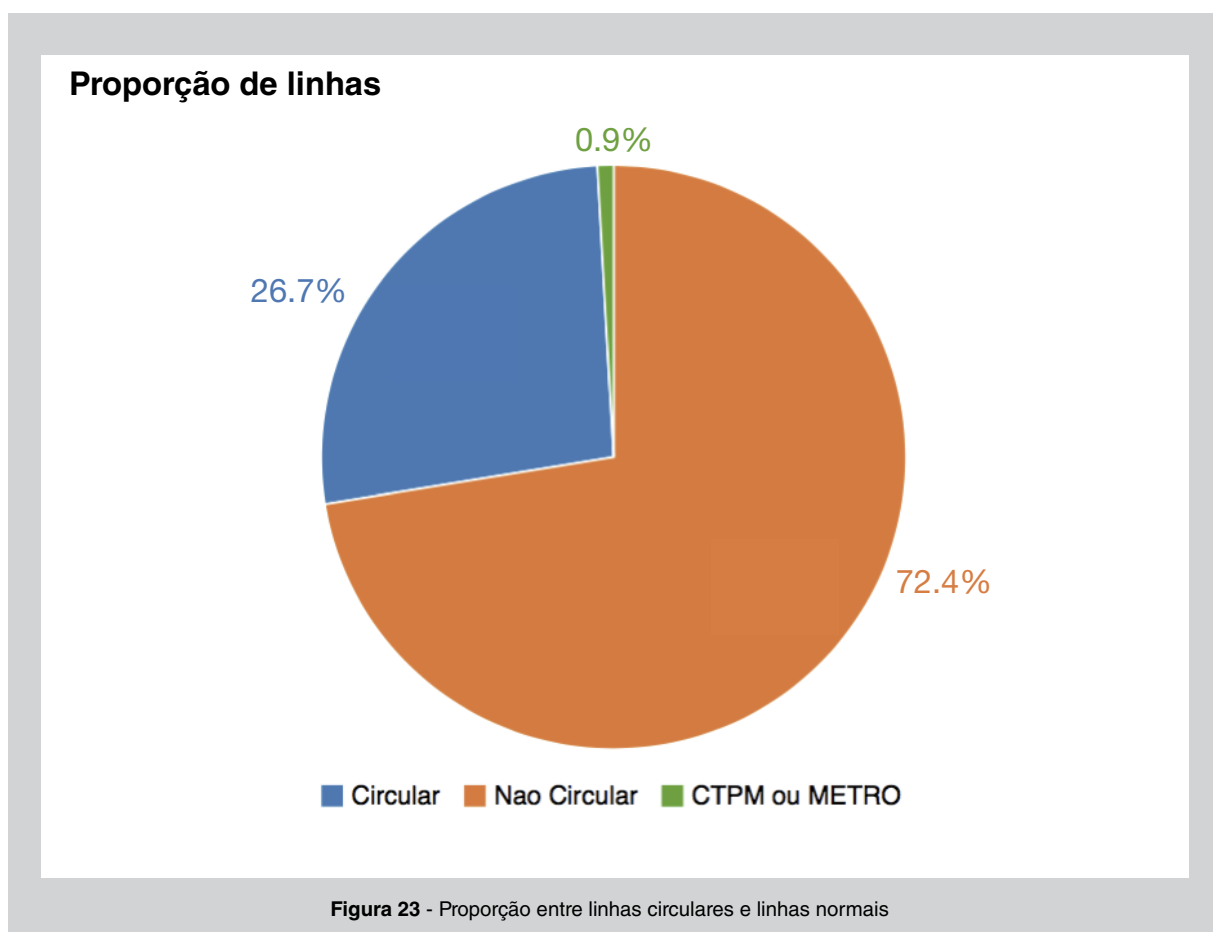
A identificação e reconstituição das viagens é feita da seguinte forma. Para cada posição reportada pelo equipamento, a coordenada é normalizada ao associá-la ao segmento mais próximo no trajeto da linha configurada para o equipamento. Em seguida a sequência de segmentos é analisada para determinar-se a cobertura dos registros identificados e se a viagem reconstituída teve sua origem no ponto inicial e término no ponto final. Finalmente, se o sentido documentado pelo equipamento está invertido, ele é corrigido.

Esta técnica funciona muito bem na maioria dos casos, mas tem problemas na análise de linhas circulares. Quando uma linha é circular, ela passa por trechos muito próximos com o mesmo identificador de viagem. Ao associar-se um registro AVL ao segmento mais próximo, algumas vezes o algoritmo acaba selecionando o segmento do sentido de volta enquanto o ônibus ainda está no sentido de ida. Quando isso acontece, a viagem detectada é inconsistente e supõe que o ônibus realizou um trajeto que na prática é impossível. As linhas circulares são tratadas

de forma especial, considerando-se não apenas o segmento mais próximo mas também a ordem dos segmentos já detectados para decidir entre dois ou mais candidatos a segmentos mais próximos. A figura 23 mostra a proporção entre linhas circulares e linhas normais na base de dados analisada (além das linhas de trem e metrô que não foram consideradas na análise).

A figura 24 mostra um AVL com dados de posicionamento completos e de boa qualidade. No eixo X estão representados os horários de coleta dos dados enquanto no eixo Y estão representadas as distâncias da coordenada AVL, normalizada para o trajeto da linha, a partir do ponto inicial da viagem. Nesta imagem podemos ver que as viagens se sucedem alternando ida e volta da mesma linha. Além disso, em cada viagem a distância do ponto inicial aumenta progressivamente até atingir o ponto final. Neste registro é possível identificar facilmente ao menos 7 viagens da linha 107T-10, sendo 4 no sentido de ida, e 3 no sentido de volta.

Infelizmente, grande parte dos registros capturados pelo AVL não são tão claros como no caso acima. Na figura W mostramos os registros AVL de um



Registros AVL

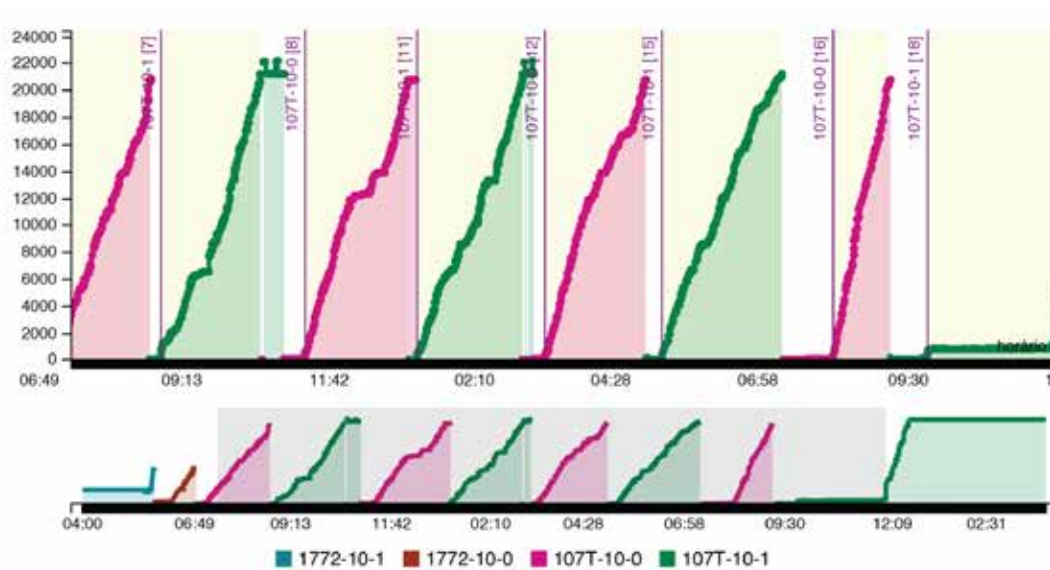


Figura 24 - Registros AVL com problemas de qualidade nos dados

equipamento com diversos problemas de qualidade nos dados enviados. Por volta de 4:30 PM as coordenadas enviadas pelo equipamento representam uma viagem que tem início a cerca de 7000 m do ponto inicial. Apesar da viagem ter atingido o ponto final, durante os 50% iniciais do percurso não temos nenhuma informação.

Por volta das 6:30 PM uma nova viagem com problemas é reportada pelo equipamento de AVL. A viagem começa normalmente do ponto inicial mas ao atingir 4000 m o equipamento passa a reportar exatamente a mesma coordenada durante uma hora. Quando o equipamento passa a reportar uma posição diferente, as distâncias começam a diminuir em relação ao ponto inicial. Isso acontece porque o equipamento está configurado para um sentido da viagem mas o ônibus está viajando no sentido contrário. Além disso, essa viagem invertida começou a 8000 m do ponto inicial em uma linha com trajeto de 14000 m. Mais uma vez, não temos informações a respeito de cerca de 43% do trajeto. Mesmo corrigindo o erro de configuração do equipamento para o sentido correto da viagem, a viagem reconstituída teria apenas 57% dos dados e não poderia ser explorada.

Registros AVL

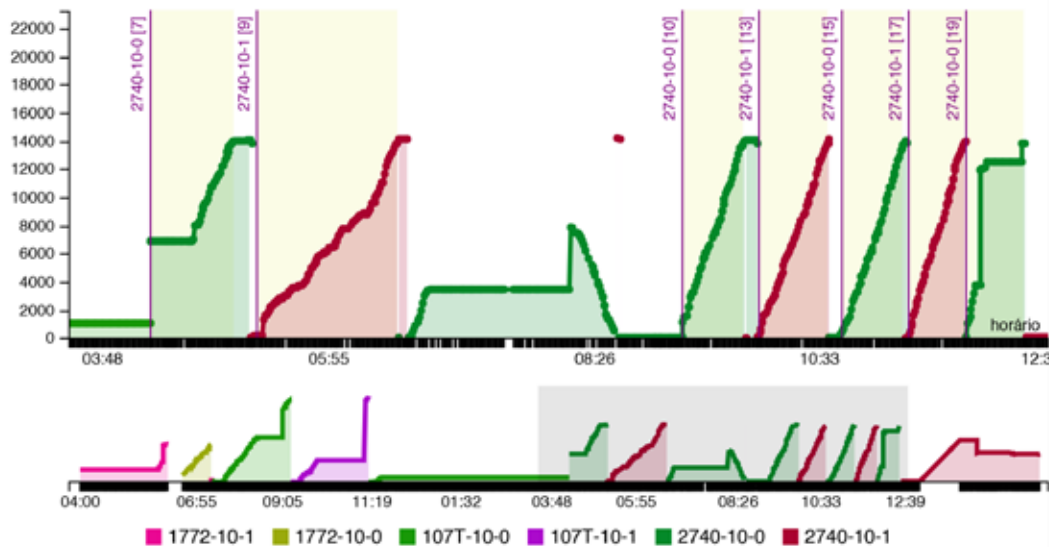


Figura 25 - Registros AVL com problemas de qualidade nos dados

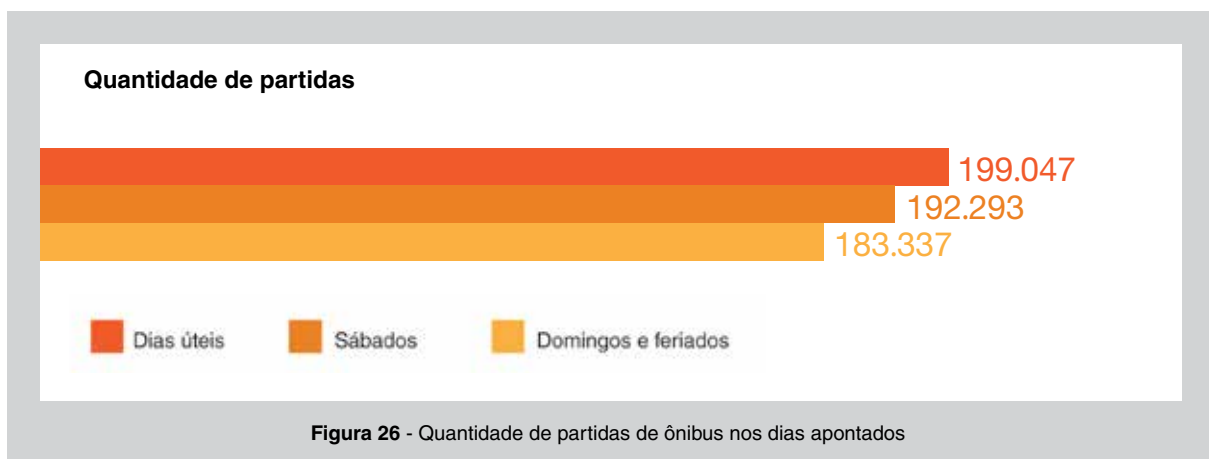
Finalmente, por volta das 8:40 PM o ônibus aparentemente mudou sua posição bruscamente de um terminal para outro (representados por dois pontos rosas no alto do gráfico). O que aconteceu neste momento é que a viagem descendente da linha 2740-10-0 na verdade deveria ser uma viagem ascendente no sentido contrário, 2740-10-1. O ponto inicial do sentido 2740-10-0 é na verdade o ponto final do sentido 2740-10-1. A linha descendente azul deveria ser uma linha ascendente rosa. O momento em que aparecem os pontos rosas é o momento em que o sistema é reconfigurado corretamente, após a falha de 6:30 PM.

É importante notar que esses problemas não aconteceram porque este equipamento tem algum defeito nem porque este trajeto passa por locais com dificuldades em acessar a rede de dados ou de obter precisão no GPS. Repare que a partir de 9:20 PM o equipamento registra 4 viagens nos dois sentidos perfeitamente. Erros como esses acontecem frequentemente independentemente do equipamento ou do trajeto por falhas momentâneas da tecnologia.

A imagem 26 mostra a quantidade total de partidas de ônibus em dias úteis, sábados e domingos para a cidade de São Paulo, conforme documentado e previstas no GTFS. Das 199.047 partidas programadas para um dia em São Paulo, conseguimos detectar usando esta metodologia 179.733 mil, ou seja, aproximadamente 90,3%.

Os gráficos 27 e 28 mostram a distribuição das partidas detectadas. O primeiro gráfico [figura 27] mostra a porcentagem total de partidas detectadas em cada linha. Nele podemos ver que grande parte das linhas teve mais de 80% do total de partidas identificadas e que diversas linhas tiveram próximo de 100% das partidas identificadas. Os valores ultrapassam 100% porque algumas linhas realizaram mais partidas que o planejado.

O segundo gráfico [figura 28] mostra a distribuição acumulada das



porcentagens de partidas detectadas. Nele podemos ver que quase 100% das linhas tiveram mais de 50% das partidas detectadas. Apesar de apenas 55% das linhas terem 100% ou mais de partidas detectadas, 91% das linhas tiveram 70% ou mais de partidas detectadas. Levando-se em consideração que diversas partidas planejadas não são dadas ao longo do dia, temos uma amostra grande das viagens realizadas e as próximas etapas da análise possuem dados suficientes para serem conclusivas.

O mapa ao lado [figura 29] mostra a distribuição geográfica das linhas em relação às partidas detectadas na cidade de São Paulo. As linhas com mais de 80% das partidas detectadas foram marcadas em verde, as que tiveram entre 50% e 80% foram marcadas em amarelo e as que tiveram menos de 50% marcadas em vermelho. Podemos perceber que a maioria das linhas com baixo número de

Partidas e linhas

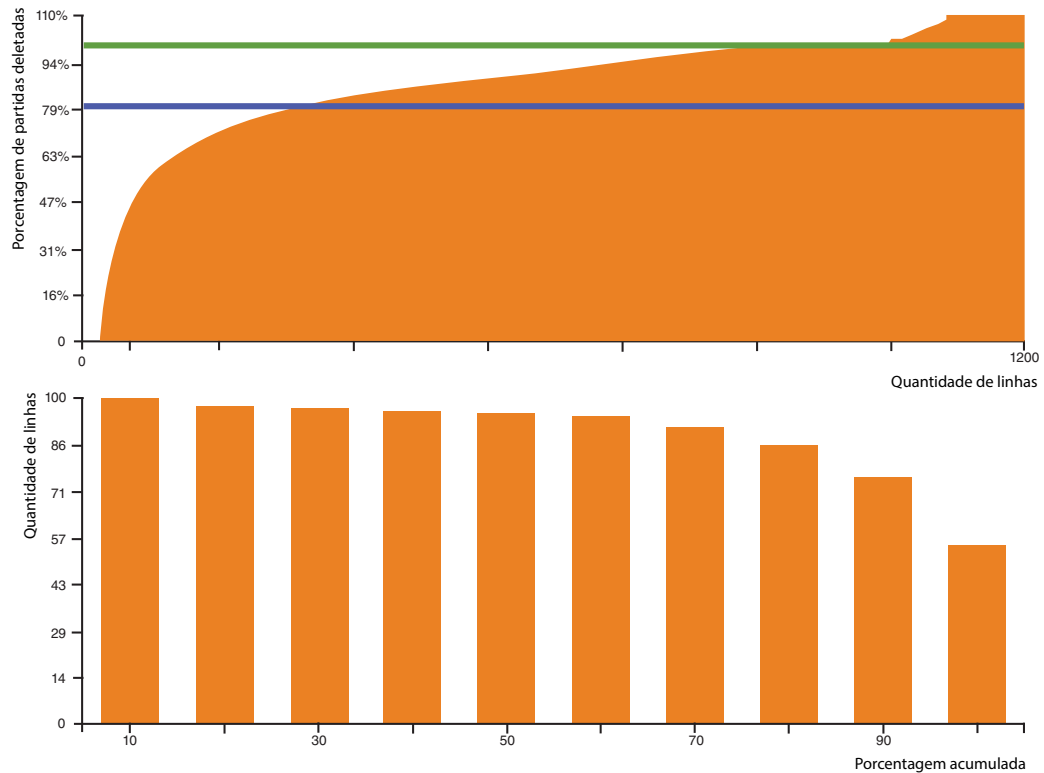


Figura 27 e 28 - Mostram respectivamente a porcentagem de partidas detectadas por linha e a distribuição acumulada das partidas detectadas.

viagens detectadas se encontra na periferia. Entretanto, na mesma região onde se encontram linhas com baixo número de partidas detectadas existem também linhas com alto número de partidas detectadas. O extremo sul da cidade é a região na qual a qualidade dos dados foi a menor do estudo. por ser uma região remota, é possível que os problemas estejam ligados ao funcionamento da tecnologia de rastreamento de veículos nesta região.

A figura 30 mostra um comparativo entre o número de partidas planejado e o número de partidas detectado para rotas com menos de 10 partidas programadas, 20, 30 e assim por diante. Neste gráfico considera-se o número de partidas da rota completa, ou seja, a quantidade de partidas planejadas no sentido de ida mais as partidas planejadas no sentido de volta. A linha azul representa 100% enquanto a linha laranja representa 80%.

Mapa de distribuição de linhas

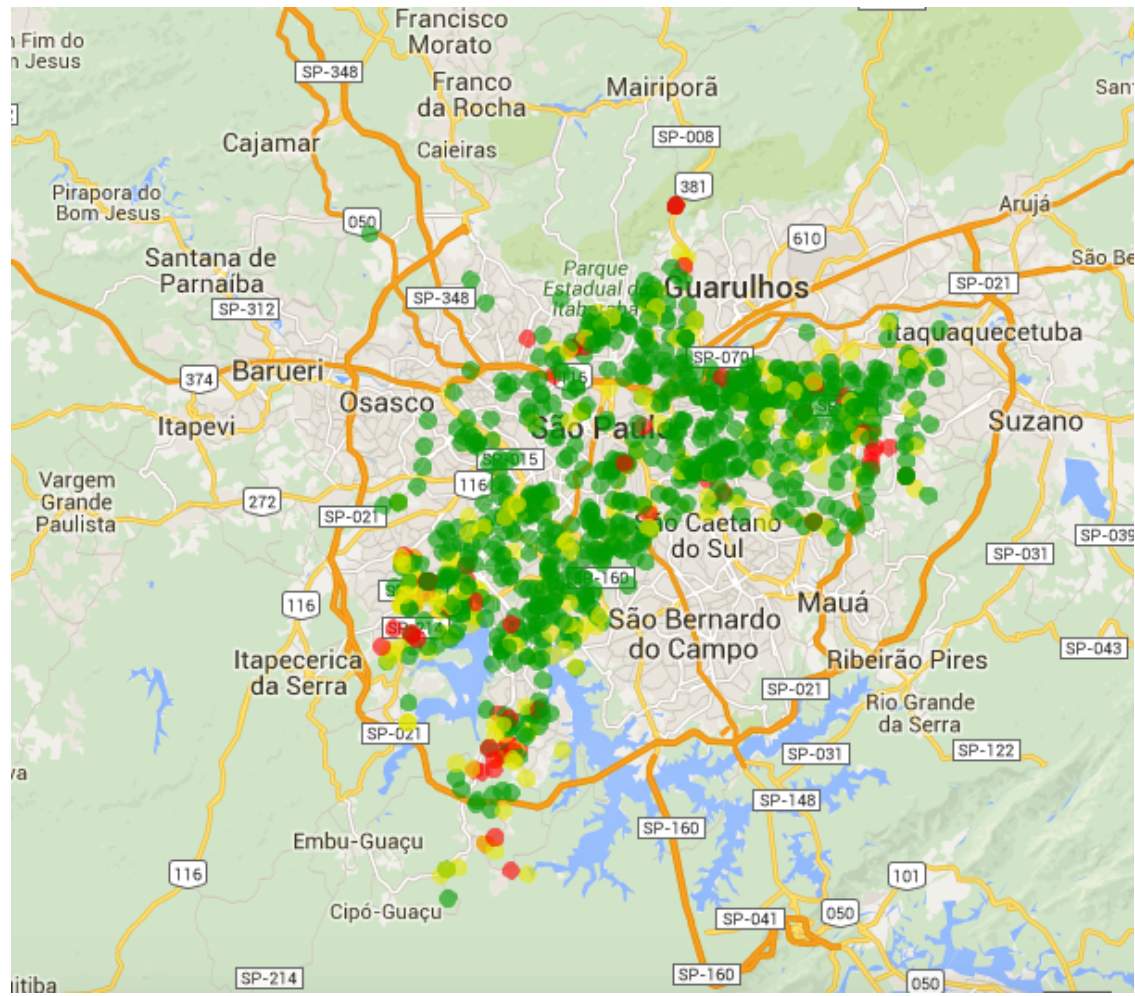


Figura 29 - Distribuição geográfica das linhas na cidade de São Paulo. As linhas com mais de 80% de partidas estão sinalizadas de verde. Entre 80% e 50% estão de amarelo, e as linhas com menos de 50% de partidas detectadas encontram-se sinalizadas de vermelho.

Podemos ver que para a grande maioria das linhas a metodologia identificou entre 80% e 100% das partidas. É importante notar também que mesmo linhas com poucas partidas planejadas tiveram um índice de detecção relativamente alto. A não-detecção de viagens, portanto, está mais associada à posição geográfica da linha e a algumas linhas com problemas na geração dos dados que com a quantidade de partidas planejadas durante o dia.

Partidas planejadas e detectadas

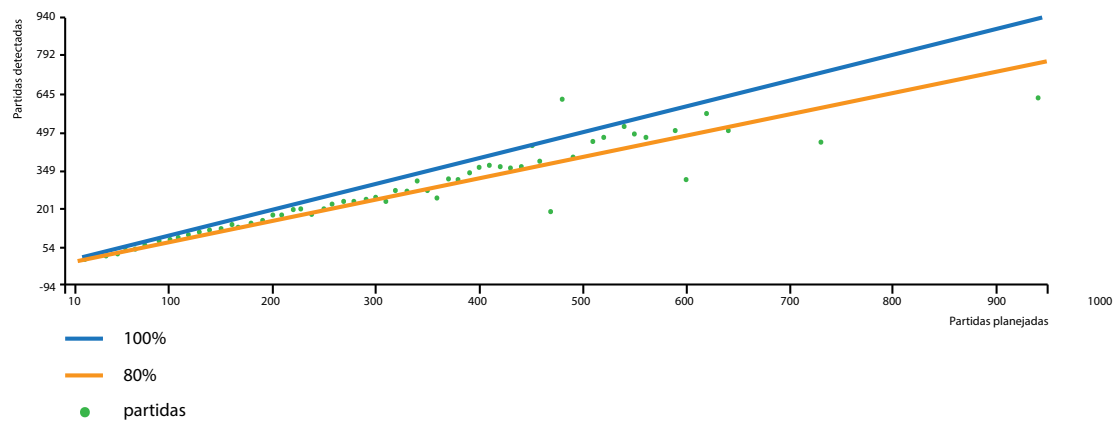


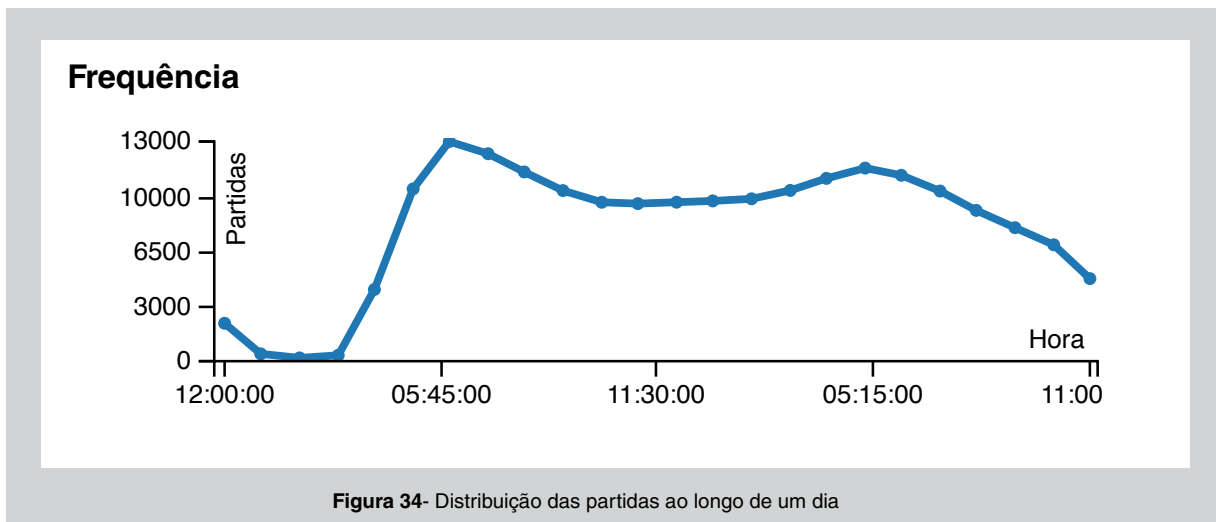
Figura 30 - Gráfico comparando número de partidas planejadas com o número de partidas detectadas

Cumprimento de partidas



Análise do cumprimento de partidas

O objetivo da análise de cumprimento de partidas é o de determinar a porcentagem de viagens planejadas para a linha que foi efetivamente cumprida pelo prestador de serviço daquela linha. A quantidade de partidas varia a cada hora de acordo com a programação da linha. Esta programação é definida conforme a demanda de passageiros estimada ao longo do dia para aquele trajeto. A figura 31 mostra a distribuição da frequência de partidas na cidade de São Paulo ao longo do dia.



Mas apenas contar a quantidade de viagens realizadas ao longo dos diferentes horários não é suficiente: é possível que a frota toda planejada para atender uma linha esteja circulando, mas as condições de tráfego do momento impedem que o prestador de serviços cumpra o planejamento.

O índice de cumprimento de partidas, portanto, não está ligado apenas à quantidade de viagens realizadas. Ele depende também da variação no tempo de percurso do trajeto ao longo do dia e da disponibilidade da frota alocada para aquela linha. Caso a velocidade das vias caia, o tempo de trajeto aumente e a frota prevista da linha já esteja sendo toda utilizada, as eventuais partidas que deixarem de ser feitas devem ser tratadas de outra forma.

Portanto, para compreender melhor a análise de cumprimento de partidas, precisamos definir 5 novos conceitos que utilizaremos ao longo desta seção.

Tempo de trajeto planejado é o tempo de trajeto documentado no arquivo GTFS para cada linha. No caso da cidade de São Paulo, este valor é único ao longo do dia pois só existe a especificação dos tempos de viagem de uma única partida no arquivo do GTFS. Isso quer dizer que supõe-se que o tempo de trajeto de uma linha às 6:00 AM é o mesmo que às 6:00 PM. Por este motivo, o tempo de trajeto planejado é uma aproximação ruim para os tempos de trajeto reais.

Tempo de trajeto realizado é o tempo que um ônibus servindo aquela linha efetivamente levou para chegar ao ponto final a partir das viagens restituídas dos arquivos de AVL e dos cálculos de velocidade média dos trechos quando a viagem não pode ser restituída. Este valor é exatamente o que aconteceu durante um trajeto, ou uma boa aproximação calculada a partir dos dados de velocidade média daquele horário.

Frota planejada com base no tempo de trajeto planejado e nas frequências de partida planejadas a cada hora, calcula-se a frota necessária ideal para que as partidas de cada hora sejam cumpridas supondo-se o tempo de trajeto planejado. Mais uma vez, como a frota planejada é baseada no tempo de trajeto planejado, que é uma aproximação ruim para o tempo de trajeto real, a frota planejada também é uma aproximação ruim para a frota realmente necessária para que as partidas planejadas sejam cumpridas.

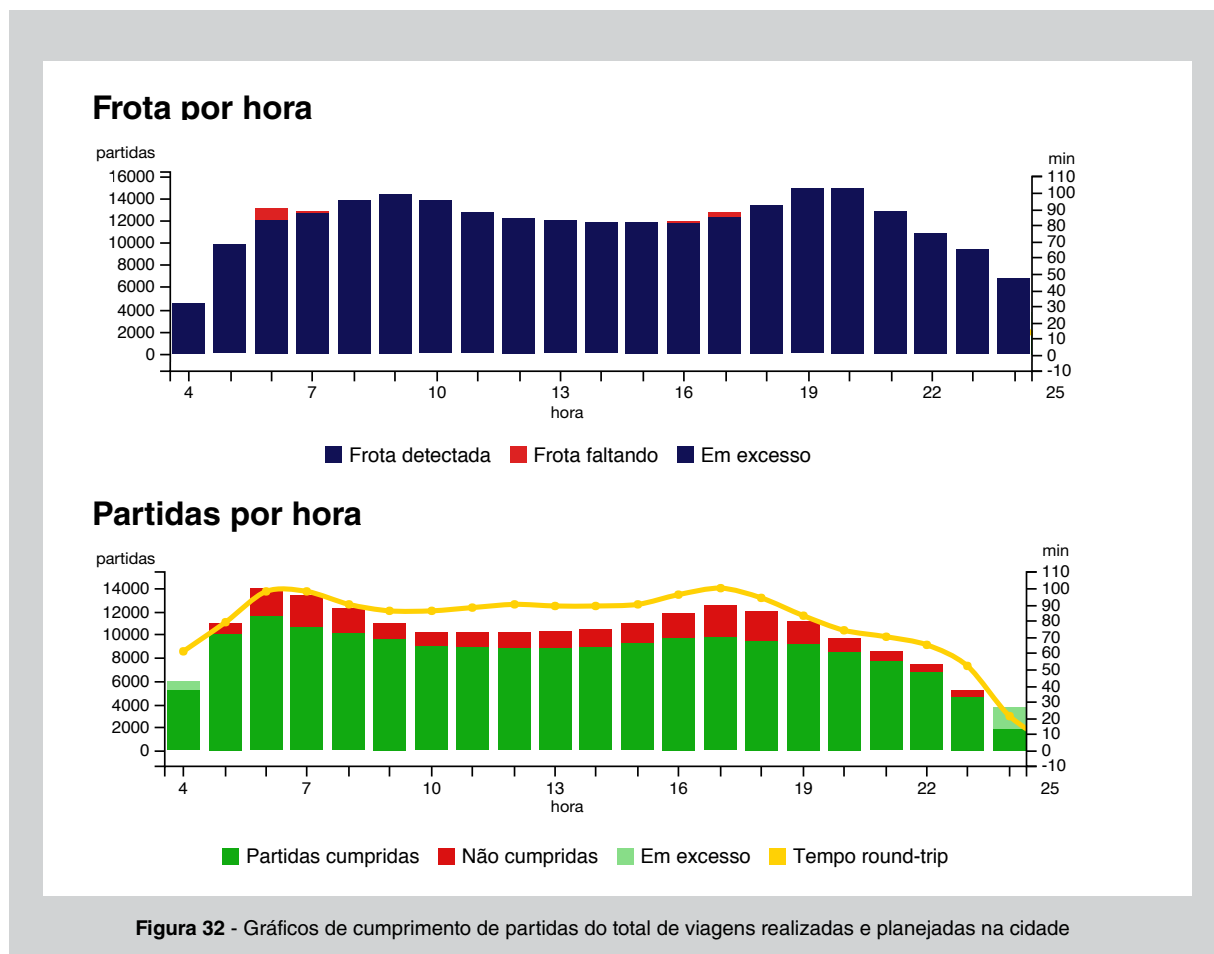
Frota requerida a frota requerida é calculada a partir do tempo de trajeto real detectado e das frequências de partida planejadas a cada hora. Em outras palavras,

dado os tempos de trajeto reais verificados para os ônibus daquela linha, a frota requerida é a menor quantidade de ônibus necessária operando nos dois sentidos da linha a cada hora para cumprir as partidas ao longo do dia. Como este cálculo utiliza o tempo de trajeto real, este valor é muito próximo da frota real.

Frota detectada é a frota verificada através da análise das viagens reconstituídas ao longo do dia. Este valor representa a quantidade de ônibus efetivamente em operação para aquela linha naquele dia.

Para a análise do cumprimento de partidas, portanto, comparamos não apenas as partidas realizadas com as previstas, mas também a frota detectada com as frotas requeridas e planejadas. Calculamos então todas as frotas hora a hora e comparamos os valores verificados com requeridos.

Para cálculo das frotas utilizamos um algoritmo baseado em [Ceder'07]. O algoritmo foi adaptado da seguinte forma: para o cálculo da frota requerida, o tempo



de *round-trip* considerado é a média dos tempos de *round-trip* para aquela linha durante a hora. Esse tempo de *round-trip* é o chamado tempo de ciclo da linha, que é a soma do tempo de ida da linha, tempo de volta e os tempos parados nos terminais até a próxima saída. Ou seja, é o tempo que um mesmo veículo leva para atender um mesmo ponto duas vezes. Para cada linha, também variou-se a frequência de acordo com as frequências documentadas no GTFS.

A figura 32 mostra um gráfico de cumprimento de partidas que considera o total de viagens realizadas e planejadas para toda a cidade, média de todas as linhas.

A parte superior do gráfico 32 mostra informações relativas à **frota**. Para cada hora do dia, calcula-se a frota detectada naquela hora e compara-se com a frota requerida para cumprir as partidas planejadas para cada linha naquele horário. Caso a frota detectada seja inferior à frota requerida, a quantidade de ônibus faltando aparece em vermelho. Caso sejam detectados veículos além do necessário para cumprir as partidas planejadas, eles aparecem em azul-claro.

A parte inferior do gráfico 32 mostra informações relativas ao **cumprimento de partidas**. A cada hora do dia calcula-se o número de partidas planejadas para cada linha e compara-se com a quantidade de partidas detectadas naquela hora. Caso o número de partidas detectadas seja inferior ao número de partidas planejadas, as partidas ausentes são mostradas em vermelho. Caso sejam detectadas mais partidas que o planejado para aquela hora, as partidas excedentes são mostradas em verde-claro.

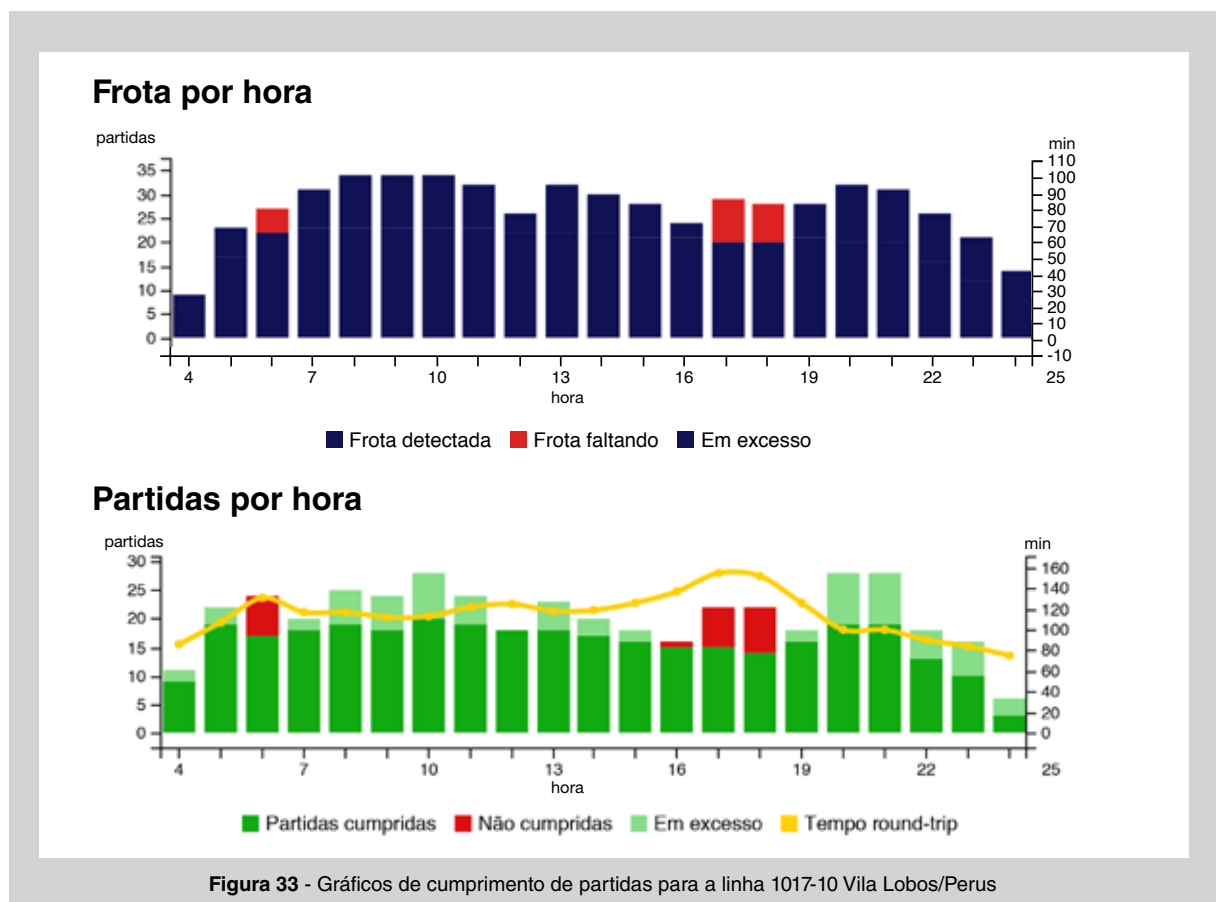
Esse gráfico também mostra uma linha amarela representando o tempo real de *round-trip* calculado para a rota durante aquela hora. Quanto maior o tempo de *round-trip*, mais tempo os ônibus demoram para chegar em um dos terminais e maior a dificuldade em se cumprir as partidas. O tempo de *round-trip* ajuda a avaliar se o não cumprimento das partidas deveu-se a um aumento no tráfego nas vias percorridas pela linha ou à negligência do operador.

Podemos ver que, no caso da cidade de São Paulo, a frota ativa na cidade é inferior à frota requerida principalmente nos horários de pico. Às 6:00 AM faltam 1045 carros e às 7:00 AM faltam 162 carros. Já no final da tarde, faltam 384 carros às 17:00. O tempo de viagem médio durante o pico da manhã é apenas 9% mais lento que o tempo médio fora do horário de pico, enquanto no pico da tarde esse tempo é apenas 11% mais lento que o tempo médio fora do horário de pico. Entretanto, durante o pico da manhã não são cumpridas 18% das viagens previstas, enquanto

no pico da tarde não são cumpridas 25% das viagens.

A metodologia também obtém estes valores para cada linha da cidade ao longo do dia. A imagem 33 mostra os resultados para a linha 1017-10 Vila Lório/Perus. Podemos ver claramente que a partir de 4:00 PM o tempo de round-trip da linha aumenta, e como consequência a linha começa a perder partidas, pois, devido à baixa velocidade, os veículos alocados não conseguem retornar aos pontos iniciais e finais para manter o cumprimento de partidas. Neste momento os veículos estavam provavelmente parados nos engarrafamentos e seriam necessários muitos outros veículos para atender a quantidade de partidas planejadas com a real velocidade da linha naquele momento.

Para cada linha, nossa metodologia calcula a quantidade de partidas não cumpridas que podem ser atribuídas a uma queda da velocidade da linha (maior tempo de *round trip*, influência do trânsito das vias da cidade percorridas pela linha) e aquelas que não podem ser explicadas por esse motivo. Para isso, utilizamos a seguinte conta: inicialmente calculamos o tempo de trajeto médio real para cada linha com base nos tempos de trajeto verificados ao longo do dia. A cada hora,



calculamos a porcentagem de lentidão da linha naquela hora em relação ao tempo de trajeto médio, ao dividir do roundtrip detectado na hora pelo roundtrip médio do dia. Definimos então a quantidade de partidas cujo não cumprimento poderia ser justificado pelo aumento no tempo de trajeto. Descontamos das partidas não cumpridas totais aquelas devidas ao aumento do tempo de viagem e calculamos um índice líquido de partidas não cumpridas, descontados os não cumprimentos devido ao tráfego enfrentado pelos ônibus.

Intervalo



Análise de intervalo de passagem

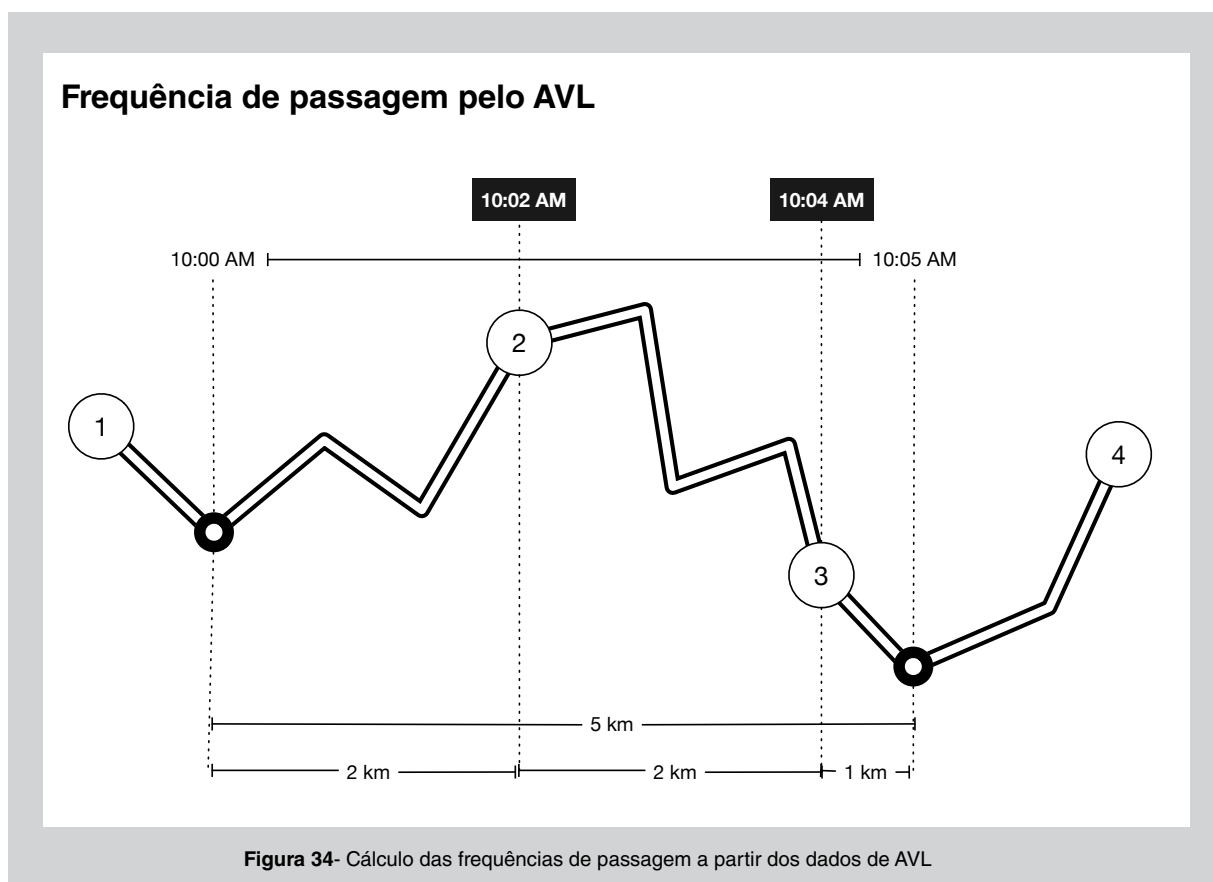
A análise do cumprimento de partidas permite avaliar um aspecto importante da operação do sistema: o respeito às partidas programadas a cada hora do dia para o bom funcionamento de cada linha, conforme as demandas de passageiros estimadas em cada ponto. Entretanto, esta análise não fornece nenhuma informação sobre outro aspecto importante para garantir que os passageiros estão sendo bem atendidos: a frequência de passagem dos ônibus em cada ponto.

Efetivamente, a programação de partidas define a quantidade de partidas na hora mas também o intervalo entre elas. Se uma linha teve 4 partidas dentro da hora, é importante saber se elas aconteceram regularmente a cada 15 minutos ou se elas foram concentradas no começo da hora. No segundo caso, além da espera dos passageiros ser maior do que o previsto para a linha, a demanda se concentraria e alguns ônibus circulariam com lotação acima do programado para a linha.

Não basta apenas verificar que os veículos estão saindo do ponto inicial a intervalos regulares. É importante verificar se esta regularidade se mantém ao longo do trajeto. Caso haja alguma retenção ao longo da linha, os passageiros após a retenção podem sofrer com a irregularidade dos ônibus e o nível de qualidade de serviço que eles obterão será menor que o planejado.

Para isso, são estimados a partir do AVL os horários de passagem de cada ônibus por cada ponto ao longo do trajeto. Como detalhado na Seção 8.2, as posições coletadas pelo sistema de AVL não coincidem com os pontos de parada dos ônibus. A figura 34 mostra como estimamos os horários de passagem a partir das posições do AVL.

A figura 34 mostra um trajeto de ônibus fictício com duas posições coletadas pelo equipamento do ônibus: uma entre os pontos 1 e 2 e outra entre os pontos 3 e 4. A primeira posição foi coletada às 10:00 AM e a segunda às 10:05 AM, com um

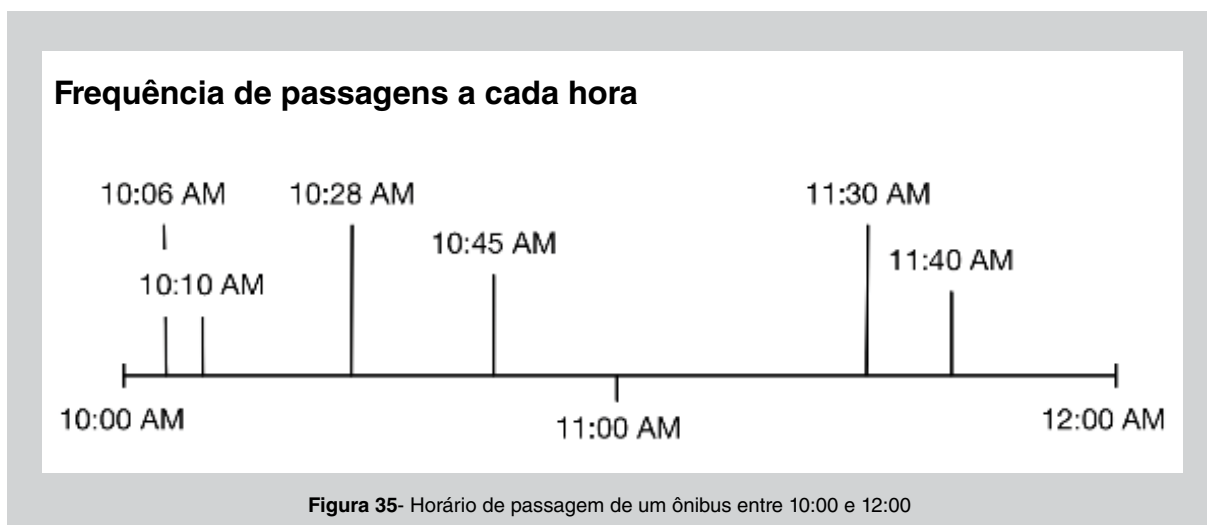


intervalo, portanto, de 5 minutos entre elas. A distância entre as duas posições **de acordo com a geometria do trajeto** (e não a distância euclidiana) é de 5 km.

Sabemos também que a distância entre a primeira posição detectada pelo equipamento de AVL e o ponto de ônibus número 2 seguindo a geometria do trajeto é de 2 km. Podemos estimar portanto o horário de passagem do ônibus no ponto 2 em 10:02 AM. Da mesma forma, a distância entre a primeira posição do AVL e o ponto de ônibus número 3 é de 4 km e o horário estimado de passada do ônibus é 10:04 AM.

A partir da matriz com os horários de passagem de todos os ônibus por todos os pontos da cidade, calcula-se a frequência de passagem dos ônibus hora a hora em cada ponto. A figura 35 mostra como são calculadas as frequências de passagem médias a cada hora.

Suponha que para um dado ponto de ônibus foram estimados os horários de passagem acima. Se forem consideradas apenas as passagens dentro de cada hora, no caso acima teríamos uma frequência média de 13 minutos (com desvio padrão de 7 minutos) entre 10:00 AM e 11:00 AM e de 10 minutos (com desvio padrão de 0) entre 11:00 AM e 12:00 AM, ao considerar apenas o intervalo entre a passagem

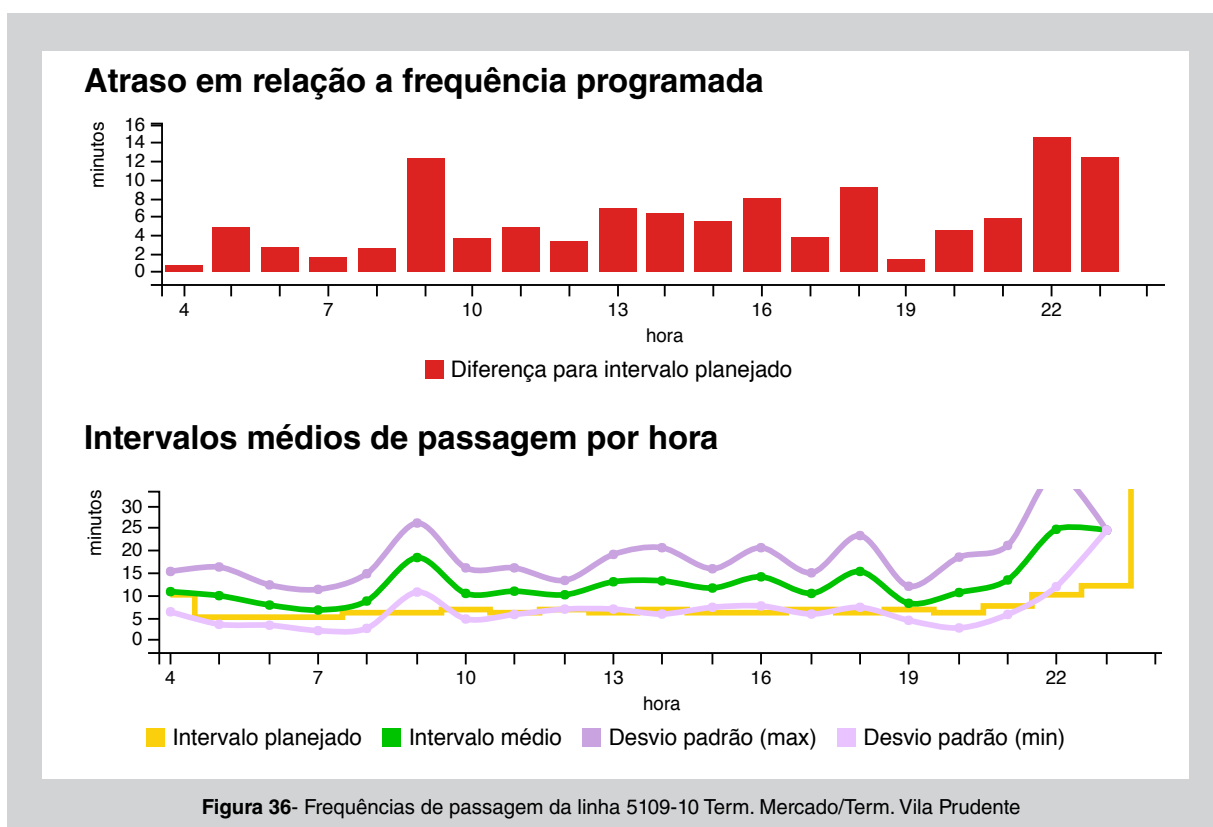


de 11:30 AM e 11:40AM, ao considerar o mesmo intervalo entre a passagem antes de 11:30 AM e 11:40 AM. Repare que o intervalo de 45 minutos entre a última passagem antes das 11:00 AM e a primeira passagem depois das 11:00 AM não seria contabilizada na média da primeira hora nem na média da segunda hora.

Por este motivo, o cálculo de frequência média considera todas passagens de ônibus dentro da hora mais a primeira passagem da hora seguinte. No caso acima, utilizando nossa metodologia, a frequência média de passagem entre 10:00 AM e 11:00 AM é de 21 minutos com desvio padrão de 17 minutos ao invés de 13 minutos e desvio padrão de 7 minutos.

A partir da frequência média de passagem horária em cada ponto da cidade para cada linha, é criado um gráfico como o da figura 36. Ele representa um sumário das variações na frequência de passagem de todos os pontos que compõem a linha, a cada hora do dia, em relação à frequência planejada.

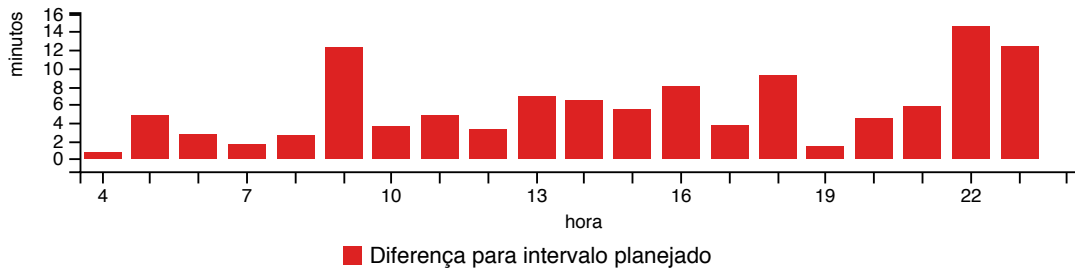
A parte superior do gráfico mostra a quantidade de minutos em média que o ônibus atrasou em relação a frequência esperada de passagem, ou seja, a diferença entre a frequência programada e a frequência calculada com os dados de AVL. A parte inferior do gráfico mostra em amarelo a frequência de passagem planejada para a linha a cada hora, e em verde a frequência média detectada a cada hora. As linhas roxas representam um desvio padrão acima e um desvio padrão abaixo das frequências médias detectadas. Neste exemplo, a partir de 9:00 AM a linha passou a atrasar consistentemente seus horários de passagem em relação ao planejado, com uma frequência real menor que a programada



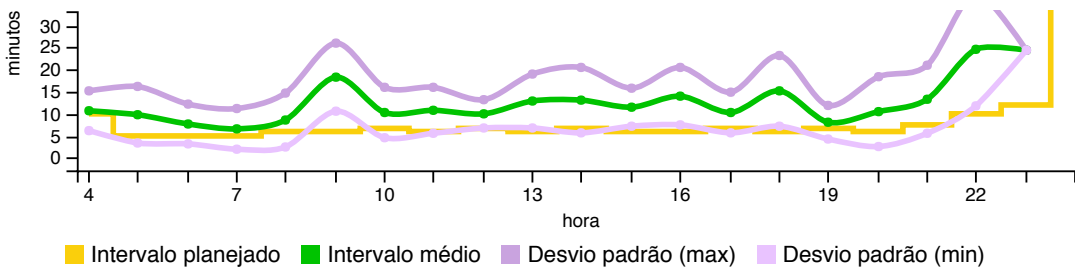
Os dados de frequência podem ser analisados de forma mais detalhada através da visualização na figura 37. Nela podemos ver todas as diferenças da frequência verificada em relação à frequência planejada, a cada hora, para todos os pontos da linha.

Na imagem acima a matriz colorida representam os pontos de ônibus em ordem crescente de baixo para cima, e as horas estão representadas no eixo das colunas, diretamente alinhadas com os horários do gráfico imediatamente superior. Nesta imagem podemos perceber que o grande atraso desta linha por volta do meio-dia ocorreu por problemas nos pontos de ônibus do primeiro terço do trajeto,

Atraso em relação a frequência programada



Intervalos médios de passagem por hora



Atraso em relação ao programado no trajeto

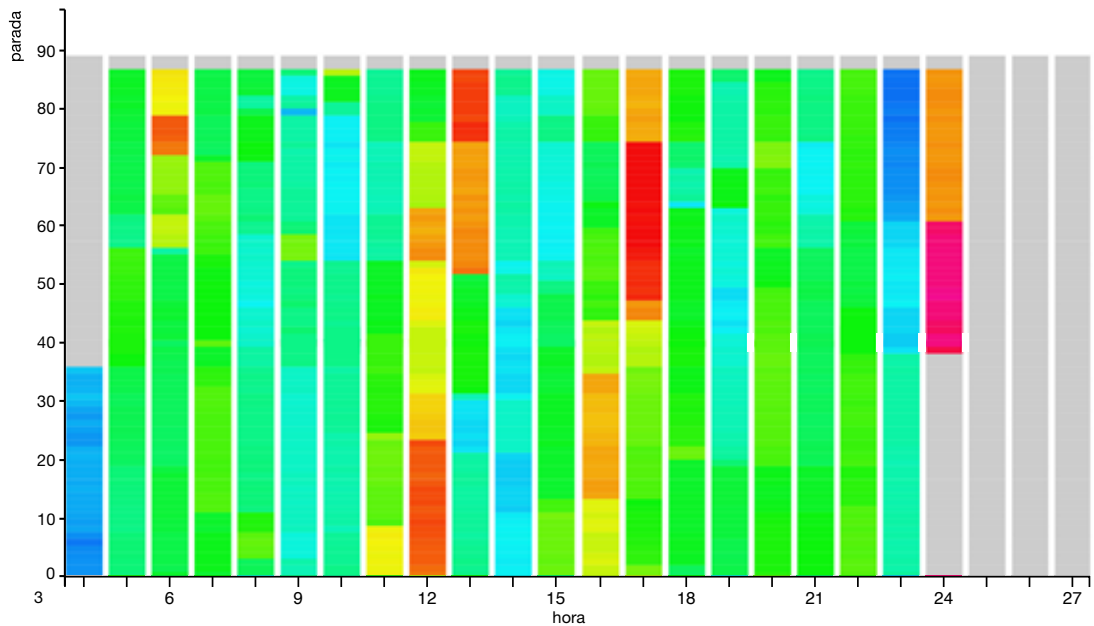


Figura 37- Diferença entre a frequência de passagem real e a planejada por ponto

enquanto os atrasos que aconteceram por volta das 17:00 se deveram a problemas nos pontos da segunda metade da linha.

9

Conclusões

Conclusões sobre o trabalho desenvolvido

Novos conceitos na área da tecnologia da informação como os Dados Abertos e o *Big Data* podem revolucionar o planejamento de transportes em megacidades. Há menos de 5 anos a metodologia típica para um estudo sobre dados de transporte envolvia invariavelmente diversos pesquisadores, realizando um extenso trabalho de coleta e de análise de dados. Os resultados demoravam para aparecer e o custo do estudo era extremamente alto.

Com a informatização cada vez maior dos sistemas de transporte, consequência natural da popularização dos ITS e da redução dos custos de armazenagem, processamento e comunicação de dados, milhares de transações que antes eram analógicas foram digitalizadas, reduzindo a complexidade da coleta de dados e dando origem a uma base de dados valiosíssima para o planejamento e para a operação de redes de transporte complexas.

Diversas experiências ao redor do mundo estão explorando as bases de dados massivas criadas pela popularização da tecnologia na gestão dos transportes com resultados animadores. Trabalhos realizados em Londres e em Santiago mostram que os dados gerados por sistemas como rastreadores de veículos e *smart cards*, quando analisados, podem suportar a tomada de decisão de gestores e planejadores de sistemas de transporte que são carentes de informações para

fundamentar suas decisões.

Neste caderno técnico apresentamos uma metodologia de análise de dados de transporte que permite estimar com precisão os índices de cumprimento de partidas e as variações nas frequências de passagem de grande parte dos ônibus circulando por uma cidade. A metodologia se apoia em duas bases de dados principais: (i) o mapa estático da rede de transportes e (ii) os históricos de movimento dos ônibus registrados pelos sistemas de rastreamento.

O mapa estático da rede de transportes contém trajetos das linhas, posições dos pontos de ônibus, frequências planejadas de partidas, entre outras informações que descrevem uma rede de transportes. Estes dados são normalmente especificados através do formato GTFS - *General Transit Feed Specification*, um formato definido pelo Google em conjunto com alguns operadores de transporte que facilita o processamento sistemático de dados estáticos de transporte e é flexível o suficiente para permitir a descrição de redes de transportes complexas, com tarifas variáveis e modais diferentes.

Os movimentos dos ônibus são obtidos através de equipamentos de AVL - *Automatic Vehicle Location* e são armazenados em formatos proprietários geralmente bem documentados. Estes dados contém a posição do ônibus em um determinado momento, o instante em que o dado foi coletado, a linha associada ao dado, entre outras informações. Eles são enviados a um servidor remoto periodicamente, normalmente em intervalos curtos, o que permite que sejam reconstituídos os trajetos e as velocidades dos ônibus ao longo dos dias.

Nossa metodologia compara os dados da base de dados de AVL, que representam as viagens efetivamente realizadas, com as informações da base GTFS, que representam o planejamento da rede, ou como se espera que a rede de transportes funcione. Essa comparação possibilita ao gestor compreender melhor os desvios da operação em relação ao planejado, identificar os principais pontos de problemas e definir medidas corretivas.

Essa análise entretanto é dificultada por três razões principais: (i) as informações do comportamento planejado da rede de transportes podem ser inconsistentes / incompletas, (ii) a qualidade dos dados obtidos pelos equipamentos de AVL é baixa, com diversas ocorrências ausentes e alguns dados corrompidos e errados e (iii) a quantidade de dados a analisar é extremamente grande, chegando a 28 milhões de registros por dia na cidade de São Paulo.

Nossa metodologia contém seis etapas que visam a eliminar estes problemas quando possível e mensurar o impacto dos problemas na massa de dados quando for impossível eliminá-los. A partir dos dados exploráveis da massa de dados analisada, nossa metodologia determina a quantidade de viagens realizadas em comparação com o planejado e a distribuição das frequências de passagens em pontos ao longo das linhas. As seis etapas são: (i) exploração dos dados, (ii) análise da qualidade dos dados, (iii) detecção das viagens, (iv) construção do mapa de velocidades, (v) cálculo dos índices de cumprimento de partidas e (vi) cálculo dos desvios da frequência de passagem.

No caso de São Paulo, apenas uma das 2.321 linhas não possuía dados de frequência documentados na base GTFS e não pode ser analisada, o que representa 0,04% das linhas de ônibus da cidade. Em relação aos dados de AVL, 46,6% dos dados não são exploráveis por diversos motivos ligados a falhas de operação, sendo os principais deles o acionamento do equipamento de AVL antes do ônibus iniciar o trajeto da linha (por exemplo entre a garagem e o ponto inicial) ou a configuração errada do equipamento (ajustado para uma linha quando o ônibus está realizando outro trajeto). Apesar da grande quantidade de registros AVL não exploráveis, apenas uma linha não possuía nenhum registro na base de dados AVL e não pode ser analisada. Considerando-se a outra linha que não possuía dados no GTFS, apenas 0,09% das linhas da cidade de São Paulo não pode ser analisada.

A partir destes dados de AVL, nossa metodologia foi capaz de restituir 85,5% das viagens da cidade. As viagens que não puderam ser recuperadas se distribuíram pela cidade independentemente da frequência planejada ou da área de operação das linhas, com uma pequena concentração no extremo sul da cidade. Essas viagens representam significativamente as diversas linhas da cidade de São Paulo: 86% das linhas tiveram mais de 80% das viagens planejadas restituídas.

Sobre o cumprimento das partidas, nossa metodologia deixa clara a relação entre as partidas realizadas e a velocidade média das vias naquele instante. É possível visualizar para cada linha a quantidade de partidas realizadas no dia, a quantidade de partidas não cumpridas, e o tempo de trajeto da linha a cada hora. Também é possível ver a frota detectada a cada hora e a frota requerida para cumprimento das partidas a cada hora.

A metodologia também define um índice de descumprimento de partidas por razões desconhecidas. Estas são as partidas que não foram cumpridas pela linha e que não podem ser explicadas pela lentidão detectada para a linha. Esta análise considera a variação da velocidade real inferida para a linha ao longo do dia

e calcula o número de partidas que poderiam ser impactadas por esta variação. As demais partidas não cumpridas são consideradas como partidas não cumpridas por razões desconhecidas.

Finalmente, a análise de frequência de passagem estima os horários de passagem de todos os ônibus da cidade em todos os pontos ao longo do dia. A partir desta estimativa, calcula-se a quantidade de passagens de ônibus a cada hora por cada ponto, e compara-se este número com o número de passagens planejadas ao longo do dia para aquela linha. É importante notar que a frequência de passagem planejada varia ao longo do dia, e a análise de frequência leva em conta essa variação no cálculo da diferença da frequência verificada para a frequência planejada.

A análise fornece como resultado, para cada linha, um gráfico sumário com a média da diferença entre a frequência real e a planejada ao longo da linha a cada hora, e a média da frequência de passagem em relação ao desvio padrão dos valores. Além disso, caso seja necessário se aprofundar na análise das razões que causaram os atrasos médios, é possível aumentar o detalhamento dos dados e visualizar, para cada ponto e a cada hora do dia, métricas como o desvio padrão da frequência de passagem ou a diferença entre a frequência de passagem e a frequência planejada. Desta forma é possível identificar exatamente em qual ponto de ônibus da linha e em qual horário do dia as frequências de passagem começaram a divergir do planejado.

Em suma, a grande massa de dados de transporte que é coletada sistematicamente pelos sistemas de transporte inteligentes das grandes cidades traz diversos desafios à sua análise por conta da enorme quantidade de dados que devem ser processados e também pelos problemas na qualidade dos dados obtidos. Por outro lado, uma vez superados os desafios na análise, estes dados podem fornecer uma visão macro do sistema de transportes que não pode ser obtida facilmente de nenhuma outra forma. Esta visão possibilita ao gestor identificar desvios da operação em relação ao planejamento, fornecendo informações importantes para suportar tomadas de decisão e definir ações corretivas e alterações na estrutura de transportes da cidade.

Bibliografia e Fotografia

[Figueiredo'01] Figueiredo, Lino, Isabel Jesus, JA Tenreiro Machado, J. Ferreira, and JL Martins De Carvalho. "Towards the development of intelligent transportation systems." In Intelligent transportation systems, vol. 88, pp. 1206-1211. 2001.

[McCallum'02] McCallum, J. "Price-performance of computer technology." The computer engineering handbook (2002): 136-153.

[Ward'13] Ward, Jonathan Stuart, and Adam Barker. "Undefined by data: a survey of big data definitions." arXiv preprint arXiv:1309.5821 (2013).

[Moore'65] Moore, Gordon. "Cramming more components onto integrated circuits." In Electronics, Volume 38, Number 8, April 19, 1965

[Sussman'96] Sussman, Joseph M. "ITS: a short history and perspective on the future." Transportation Quarterly 50.4 (1996).

[SmartMotorways] Smart Motorways <https://www.gov.uk/government/collections/smart-motorways>

[Gordon'12] Gordon, Jason Benjamin. Intermodal passenger flows on

London's public transport network: automated inference of full passenger journeys using fare-transaction and vehicle-location data. Diss. Massachusetts Institute of Technology, 2012.

[Ortega-Tong'13] Ortega-Tong, Meisy Andrea. Classification of London's public transport users using smart card data. Diss. Massachusetts Institute of Technology, 2013.

[Núñez'13] Núñez, César, Marcela Munizaga, and Antonio Gschwender. "Cálculo de indicadores de calidad de servicio del sistema de transporte público de Santiago a partir de datos pasivos." Congreso Chileno de Ingeniería de Transporte. No. 16. 2013.

[okfn] <http://blog.okfn.org/2013/10/03/defining-open-data/>

[Lathrop'10] Lathrop, Daniel, and Laurel Ruma. Open government: Collaboration, transparency, and participation in practice. " O'Reilly Media, Inc.", 2010.

[Obama'09] Obama, Barack. Memorandum on Transparency and Open Government. White House, 2009. Disponível em https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment

[Lei12527] Lei 12.527 de 18 de Novembro de 2011 da Presidência da República. Disponível em http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

[Ceder'07] Ceder, Avishai. Public transit planning and operation: theory, modeling and practice. Elsevier, Butterworth-Heinemann, 2007.

Fotografia

<http://bit.ly/22J7bwX>

Lista de Figuras

Figura 01 - Relação US\$ / MB	_____	pág. 17
Figura 02 - Smart Motoway	_____	pág. 20
Figura 03 - Linhas	_____	pág. 24
Figura 04 - Ônibus	_____	pág. 24
Figura 05 - Passageiros/ano (milhão)	_____	pág. 25
Figura 06 - GTFS	_____	pág. 28
Figura 07 - Arquitetura de um AVL	_____	pág. 29
Figura 08 - Gráfico das fases	_____	pág. 32
Figura 09 - Quantidade de linhas por extensão	_____	pág. 35

Figura 10 - QUantidade de linhas por parada _____	pág. 36
Figura 11 - Quantidade de linhas por parada_____	pág. 36
Figura 12 - Trecho por linha _____	pág. 37
Figura 13 - Trecho por 50 linhas _____	pág. 38
Figura 14 - Trajeto linha Metrô Parada Inglesa _____	pág. 40
Figura 15 - Gráficos de registros AVL_____	pág. 44
Figura 16 - Ponto de ônibus e segmentos_____	pág. 47
Figura 17 - Associação entre ônibus e segmentos_____	pág. 48
Figura 18 - Estimativa da Velocidade média da via_____	pág. 49
Figura 19 - Velocidades médias_____	pág. 50
Figura 20 - Posições reportadas por AVL _____	pág. 51
Figura 21 - Dados reportados e trjeto_____	pág. 52
Figura 22 - Ajuste de dados reportados com trajeto_____	pág. 53
Figura 23 - Proporção de linhas_____	pág 54
Figura 24 - Registros AVL_____	pág. 55
Figura 25 - Registros AVL_____	pág. 56

Figura 26 - Quantidade de partidas	_____	pág. 57
Figura 27 - Partidas e linhas	_____	pág. 58
Figura 28 - Mapa de distribuição de linhas	_____	pág. 58
Figura 29 - Mapa de distribuição de linhas	_____	pág. 59
Figura 30 - Partidas planejadas e detectadas	_____	pág. 60
Figura 31 - Frequência	_____	pág. 61
Figura 32 - Cumprimento de partidas	_____	pág. 63
Figura 33 - Cumprimento de partidas	_____	pág. 65
Figura 34 - Frequência de passagrm pelo AVL	_____	pág. 68
Figura 35 - Frequência de passagens a cada hora	_____	pág. 69
Figura 36 - Frequência de passagem	_____	pág. 70
Figura 37 - Frequência de passagem	_____	pág. 71

Prefeitura da Cidade de São Paulo

Fernando Haddad

Prefeito

Jilmar Augustinho Tatto

Secretário Municipal de Transportes – SMT

José Evaldo Gonçalo

Secretário Adjunto - SMT

Jilmar Augustinho Tatto

Diretor Presidente da São Paulo Transporte S.A. – SPTrans

Ciro Biderman

Chefe de Gabinete da Presidência da SPTrans

Conselho Diretor (biênio 2014/2015)

Ailton Brasiliense Pires | presidente

Antonio Luiz Mourão Santana (Oficina); **Claudio de Senna Frederico** (Artificium); **Jilmar Augustinho Tatto** (SMT/São Paulo); **Joaquim Lopes da Silva Jr.** (EMTU/SP); **João Gustavo Haenel Filho** (SOCICAM); **José Antonio Fernandes Martins** (SIMEFRE); **José Geraldo Baião** (AEAMESP); **Joubert Fortes Flores Filho** (OPPORTANS); **Lélis Marcos Teixeira** (Rio-Ônibus); **Leo Carlos Cruz** (CETURB-GV); **Leonardo Ceragioli** (Prodata Mobility); **Luiz Antonio Carvalho Pacheco** (Metrô/SP); **Marcos Bicalho dos Santos** (SETRABH); **Mário Manoel Seabra Bandeira** (CPTM); **Nelson Barreto C. B. de Menezes** (Grande Recife); **Oscar José Gameiro Silveira Campos** (STVP/SBC); **Otavio Vieira da Cunha Filho** (NTU); **Paulo Henrique do Nascimento Martins** (ManausTrans); **Plínio Oswaldo Assmann** (membro benemérito); **Ramon Victor Cesar** (BHTRANS); **Renato Gianolla** (URBES Sorocaba); **Roberto Gregório da Silva Junior** (URBS Curitiba); **Vanderlei Luis Cappellari** (EPTC Porto Alegre); **Vicente Abate** (ABIFER)

Membros Natos

Jurandir Ribeiro Fernando Fernandes (ex-presidente);
Rogério Belda (ex-presidente)

Suplentes

Atilio Pereira (SMTT Guarulhos); **Francisco Carlos Cavallero Colombo** (CBTU/RJ); **Humberto Kasper** (TRENURB Porto Alegre); **Julio Grilo** (TACOM); **Nazareno S. N. Stanislau Affonso** (RuaViva); **Wagner Colombini Martins** (LOGIT); **Willian Alberto de Aquino Pereira** (Sinergia)

Conselho Fiscal

Roberto Renato Scheliga (membro benemérito); **João Carlos Camilo de Souza** (SETPESP); **Carlos Alberto Batinga Chaves** (TTC)

ANTP/São Paulo

Rua Marconi, 34, 2º andar, conjs. 21 e 22, República, CEP 01047-000, São Paulo, SP | Tel.: (11) 3371.2299 | Fax: (11) 3253.8095 | E-mail: antpsp@antp.org.br | Site: www.antp.org.br

Equipe ANTP

Luiz Carlos M. Néspoli | superintendente
Nazareno Stanislau Affonso | escritório de Brasília
Eduardo Alcântara Vasconcellos | assessor técnico
Cassia Maria Terence Guimarães | administração/finanças
Valéria Aguiar | eventos

Premio ANTP de Qualidade

Alexandre Rocha Resende (Coordenador Nacional)
Miguel Sérgio Lima
João Batista de Moraes Ribeiro Neto
Paulo Afonso Lopes da Silva
Jackson Mattos da Rocha
Cássia Maria Terence Guimarães
Andréia Lopes Catharina
SCS, Q. 4, Ed. Mineiro, Bl. A, S. 506, CEP 70304-000, Brasília, DF |
Tel. e fax: (61) 3202.0899 | E-mail: antpmdt@gmail.com

Coordenadores Regionais

Regional Centro-Oeste (ANTP/CO)

Paulo Souza
psouzan@uol.com.br

Espírito Santo (ANTP/ES)

Denise de M. Cadete Gazzinelli Cruz
Av. Hugo Viola, 1.001, Bl. A, Sala 215, Mata da Praia, CEP 29060-420, Vitória, ES |
Tel. e fax: (27) 3223.9100 e-mail: denise@antp.org.br

Minas Gerais (ANTP/MG)

Ricardo Mendanha Ladeira
Rua Januária, 181 – Floresta, CEP 31110-060, Belo Horizonte, MG |
Tel: (31) 3224.0906 | e-mail: antpmg@antp.org.br

Norte (ANTP/N)

Patrícia Bittencourt Tavares das Neves
Av. Duque de Caxias, 863, apto. 301, Marco, CEP 66093-400, Belém, PA | Cel.: (91)
8804.7651 | e-mail: pbneves@ufpa.br

Nordeste (ANTP/NE)

César Cavalcanti de Oliveira GR/CTM
Cais de Santa Rita, 600 – Santo Antonio, CEP 50020-360, Recife, PE |
Tel.: (81) 3182.5609 | Fax: (81) 3182.5610 e-mail: cesar.antp@gmail.com

Paraná (ANTP/PR)

Rosângela Maria Battistella
Av. Pres. Afonso Camargo, 330, CEP 80060-090, Curitiba, PR |
Tel.: (41) 3320.3211 | e-mail: rosangela@antp.org.br

Rio de Janeiro (ANTP/RJ)

Willian Alberto de Aquino Pereira
Praia do Flamengo, 278, cj. 52, CEP 22210-030, Rio de Janeiro, RJ |
Tel. e fax: (21) 2553.3994 e-mail: sinergia@transporteideias.com.br

Conheça os outros Cadernos Técnicos editados pela ANTP

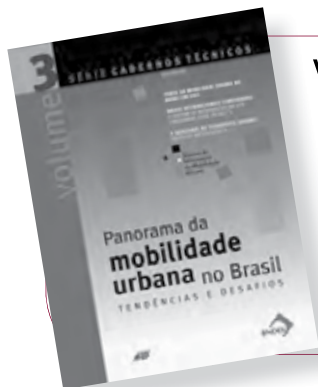


Volume 1 • Bilhetagem Automática e Gestão nos Transportes Públicos

Publicado em 2003, após a realização de um Seminário sobre o mesmo tema, o Caderno mostra o estado da arte do setor de bilhetagem eletrônica, considerada pela ANTP como um importante instrumento de ação pública e não apenas como uma ferramenta da operação privada. Os textos contidos no Caderno descrevem o panorama nacional, discutem o impacto de nova tecnologia na melhoria e nos custos dos sistemas de transporte coletivo, no emprego e na gestão pública, apontam as tendências da evolução tecnológica e relatam algumas experiências implementadas em cidades brasileiras.

Volume 2 • Transporte Metroferroviário no Brasil

Coordenado pela Comissão metroferroviária da ANTP, este Caderno apresenta o perfil dos serviços de transportes urbanos de passageiros sobre trilhos no Brasil. Os textos destacam os sistemas integrados, as oportunidades de novos projetos no setor, as perspectivas mundiais de desenvolvimento tecnológico e as condições de acessibilidade para as pessoas portadoras de deficiência. Do ponto de vista da gestão das empresas operadoras, outros textos abordam o perfil de consumo de energia, a gestão dos ativos das empresas e a gestão dos riscos.



Volume 3 • Panorama da Mobilidade Urbana no Brasil

O terceiro Caderno mostra o perfil da mobilidade urbana no Brasil, em 2003, com base na análise dos dados do Sistema de Informações da Mobilidade Urbana da ANTP. De forma sintética são apresentados os principais dados e indicadores de mobilidade, custo e produtividade nas cidades brasileiras com mais de 60 mil habitantes. O Caderno ainda apresenta alguns indicadores internacionais sistematizados pela União Internacional de transportes Públicos (UITP) e uma proposta de desenvolvimento de um Índice de Desenvolvimento do Transporte Urbano (IDT).

Volume 4 • Acessibilidade nos Transportes

O Caderno de número 4 foi produzido pelo grupo de trabalho da Acessibilidade da ANTP e traz um amplo balanço dos avanços e dos desafios que o setor vem enfrentando na construção de cidades acessíveis para todos. Em seus 17 artigos, são abordadas desde a evolução da luta dos movimentos sociais pela equiparação de oportunidades e remoção das barreiras, até um breve balanço das condições reais de uso dos diversos modos de transporte público por pessoas com dificuldades de locomoção. Também são comentados temas como: criação de espaços institucionais de gestão, desenho urbano, financiamento, construção de indicadores, entre outros.





Volume 5 • Integração nos Transportes Públicos

Os artigos apresentados neste Caderno reafirmam o princípio da integração como um atributo essencial na construção de redes de transporte coletivo urbano. Os textos foram distribuídos em quatro capítulos que tratam, respectivamente, de conceitos gerais (definição política, conceito de rede e experiência latino-americana com sistemas estruturadores), dos aspectos institucionais (gestão integrada, gestão metropolitana e premissas para financiamento pelo bNDES), da política tarifária (integração temporal e de desafios) e da integração física e operacional (bilhetagem eletrônica, terminais, iniciativa empresarial e integração com o transporte hidroviário e com o não motorizado).

Volume 6 • Transporte e Meio Ambiente

O Caderno “Transporte e Meio Ambiente” foi estruturado a partir de dois seminários realizados em São Paulo, em 2006 e 2007, que contaram também com apoio financeiro do bNDES, o que permitiu a participação de diversos especialistas, inclusive internacionais. O seu primeiro artigo e as linhas de ação da Comissão de Meio Ambiente da ANTP apresentam o conceito ampliado de sustentabilidade e os princípios que têm norteado as ações da ANTP. Os textos seguintes tratam das relações dos transportes urbanos com diversos temas presentes na discussão ambiental: o aquecimento global, as medidas de controle de emissões veiculares que estão sendo aplicadas no país, as fontes energéticas alternativas aos derivados de petróleo, o programa de eficiência energética da Petrobrás, as políticas urbanas e o processo de licenciamento ambiental entre outros textos.



Volume 7 • Transporte Ciclovitário

Este Caderno vem num momento oportuno, quando a bicicleta vive no Brasil uma fase de popularidade e transição impulsionada por uma nova consciência ecológica. Ele busca alertar que a bicicleta é um meio de transporte alternativo e viável e que a sociedade pode considerá-la uma ferramenta eficiente para melhorar a qualidade de vida urbana. Esta é a mensagem da Comissão Técnica de Bicicleta da ANTP.

Volume 8 • Sistemas Inteligentes de Transportes

Este Caderno organizado pela Comissão de I+D+I da sigla internacional para “Sistemas Inteligentes de Transportes” tem por finalidade difundir as experiências implantadas em nosso país – bilhetagem eletrônica, operação metroferroviária e de rodovias, funcionalidades para BRTs, Centrais de Controle e Monitoramento e Informações aos Usuários.





Volume 9 • Transporte por Fretamento

Este Caderno organizado pela ANTP em parceria com a Fier ESP trata deste segmento de transporte que ganha espaço na matriz da mobilidade urbana em muitas cidades brasileiras e que, segundo a Lei Federal 12.587/2012 deve ser considerado parte integrante do Sistema Nacional de Mobilidade Urbana.

Volume 10 • Excelência na Gestão do Transporte e Trânsito

Este Caderno, elaborado pela Comissão Técnica da Qualidade e Produtividade visa consolidar a experiência acumulada pela Comissão, visando estimular as organizações do setor para a adoção de modelos de gestão, clarificando os conceitos envolvidos e os caminhos de implantação.



Volume 11 • Marketing e Comunicação

O Caderno consolida o papel da Comissão de Marketing da ANTP como um fórum permanente para desenvolvimento de propostas para melhorar o desempenho e a atratividade do setor de transporte público e do trânsito urbano. Apresenta também um resumo dos resultados da Bienal de Marketing da ANTP.

Volume 12 • Inventário de Emissões e Remoções Antrópicas de Gases de Efeito Estufa no Município de São Paulo

O Inventário de Emissões e remoções de gases de Efeito Estufa no município de São Paulo, realizado no âmbito do Programa S t AQ (Sustainable Transport and Air Quality) com recursos do gEF repassados pelo banco mundial, contém o levantamento das fontes e dos sumidouros e a respectiva contabilização das emissões e remoções dos gEE resultantes das atividades humanas. O Caderno técnico ANTP nº 12 apresenta a metodologia utilizada e os seus principais resultados, pretendendo contribuir para sua aplicação em outras cidades, como uma ferramenta básica para desenvolvimento de políticas que possibilitem a redução dessas emissões.



Volume 13 • Avaliação de Vulnerabilidade Ambiental e Socioeconômica para o Município de Curitiba

O inventário apresenta uma matriz de análise de riscos ambientais para a cidade de Curitiba, a partir do mapeamento de fenômenos associados às estruturas de transporte. O trabalho permitiu a identificação de setores vulneráveis e a localização das infraestruturas e populações expostas aos riscos ambientais, permitindo orientar uma política de gerenciamento desses riscos e o desenvolvimento de planos de adaptação às mudanças climáticas, reduzindo ou mitigando tais situações. O estudo foi desenvolvido dentro do Programa "Transporte Sustentável e Qualidade do Ar - S t AQ (Sustainable Transport and Air Quality), financiado com recursos do global Environment Facility (gEF) repassados pelo banco mundial.

Volume 14 • Metodologias para Inventário, Avaliação e Gestão Ambiental do Sistema de Transporte de Baixa e Média Capacidades nas Regiões Metropolitanas de São Paulo

A metodologia foi desenvolvida para avaliar as condições de sustentabilidade de um sistema de transporte coletivo operado por ônibus, considerando instalações administrativas, garagens dos operadores, terminais urbanos e condições operacionais. Os indicadores desenvolvidos no estudo permitem avaliar tanto a adequação dos sistemas, a partir da aplicação de tecnologias e práticas sustentáveis disponíveis no mercado, quanto o desempenho das frotas e das instalações de apoio dos operadores dos serviços. Em caráter experimental, a metodologia foi aplicada para coleta de dados em uma pequena amostra dos serviços metropolitanos administrados pela EMTU/SP.



Volume 15 • Estudo Comparativo de Tecnologias Veiculares de Tração Aplicáveis a Ônibus Urbanos

O Caderno resume o resultado de projeto desenvolvido em parceria com a Empresa Metropolitana de Transportes Urbanos de São Paulo – EMTU/SP compreendendo um estudo comparativo de tecnologias de propulsão veicular para sistemas de transporte coletivo de baixa e média capacidades, isto é, para os serviços operados por ônibus. Os resultados apresentados não se limitam à caracterização das alternativas disponíveis no mercado e avançam em uma simulação de possibilidade de substituição do uso do diesel tradicional a partir de uma matriz multicritério. Este é o quarto Caderno Técnico produzido no âmbito do Programa Sítio AQ de Transporte Sustentável e Qualidade do Ar, conduzido pela ANTP com recursos do Global Environment Facility – GEF, aportados pelo Banco Mundial.



Volume 16 • Cidades a Pé

O Caderno 16 foi preparado para subsidiar a realização do seminário “Cidades a Pé” e contém contribuições de especialistas brasileiros e estrangeiros que participaram dele como palestrantes, e também textos produzidos pelos integrantes da Comissão Técnica de Mobilidade a Pé e Acessibilidade da ANTP, que promoveu o evento. Os textos e o seminário se somam no resgate da valorização do mais antigo e ainda o mais praticado modo de transporte urbano: o pedestre.



Para maiores informações acesse o site da ANTP, www.antp.org.br ou entre em contato com Luciana (11) 3371-2290 ou luciana@antp.org.br

