

Monografia apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para obtenção do título de Bacharel em Ciência da Computação

Análise e visualização de dados do transporte público de ônibus da cidade de São Paulo

Janeiro 2016 São Paulo, Brasil

Orientando: André Kenji Yai

Orientador: Roberto Hirata Junior

Resumo

Este trabalho usa uma combinação de dados públicos disponibilizados pela SPTrans com outros dados (históricos) fornecidos, também pela SPTrans, especificamente para este projeto. Foram feitas análises e visualizações destes dados para que seja possível avaliar a qualidade dos mesmos e do serviço do transporte público de ônibus da cidade de São Paulo.

Parte deste trabalho foi desenvolvido com colaboração da Scipopulis em um projeto para a SPTrans. A parte técnica, objeto do trabalho aqui descrito, consistiu do desenvolvimento de uma metodologia para manipulação dos dados, estruturação do banco de dados, exploração e análise dos mesmos utilizando principalmente o GTFS ("*General Transit Feed Specification*") e o AVL ("*Automatic Vehicle Location*").

Palavras Chaves: Análise de Dados, AVL, Dados Abertos, GTFS, Transporte Público, São Paulo, Visualização de Dados

Abstract

This project used a combination of public data from SPTrans and historic data (also from SPTrans) provided specifically for this work. Analysis and visualizations were done out of this data to evaluate them as well as the quality of service of São Paulo's city public buses transportation service.

Part of this project was developed in collaboration with Scipopulis in a work requested by SPTrans. The technical part of the work hereby described, was the development of a methodology for data manipulation, database structure, data exploration and analysis, mainly using GTFS (*General Transit Feed Specification*) and AVL (*Automatic Vehicle Location*) datasets.

Keywords: AVL, Data Analysis, Data Visualization, GTFS, Open Data, Public Transportation, São Paulo

Conteúdo

I	Parte Objetiva	5
1	Introdução	6
1.1	Dados Abertos	6
1.2	O Projeto	9
2	Ferramentas	10
2.1	Características das Ferramentas e seu uso no Projeto	10
3	Metodologia	12
3.1	Nomenclatura	12
3.2	Metodologia	15
3.3	Arquitetura	16
4	Estruturação do Banco de Dados	17
4.1	Organização inicial dos Dados	17
4.2	MongoDB	19
4.2.1	Modelagem do GTFS no Projeto	19
4.2.2	Modelagem do AVL no Projeto	20
5	Exploração de Dados	21
5.1	Visão Detalhada das Métricas	21
6	Análise de dados do GTFS	30
6.1	Visão Geral	30
6.2	Visão Detalhada das Métricas	36
6.2.1	Pontos duplicados	36
6.2.2	Pontos de Ida e Volta	37
6.2.3	Taxa de Confiabilidade Baixa	38
6.2.4	Conexões	39
6.3	Visualização	40
6.4	Resultados da análise da qualidade de dados do GTFS	42
6.4.1	Número de paradas distantes mais de 500m	42
6.4.2	Número de paradas distantes menos de 30m	42
6.4.3	Número de partidas diárias previstas	43

6.4.4	Linha sem frequência de partidas registradas	44
6.4.5	Total de horas de operação	44
6.4.6	Quantidade de shapes na linha	44
6.4.7	Distância total	45
6.4.8	Velocidade média prevista	45
6.4.9	Quantidade de paradas duplicadas	46
6.4.10	Quantidade de paradas únicas	46
6.4.11	Linha que passa duas vezes pela mesma parada	47
6.4.12	Paradas em comum na ida e na volta	47
6.4.13	Tempo previsto de viagem	47
6.4.14	Frequência média por hora	48
6.4.15	Intervalo médio entre ônibus	48
6.4.16	Porcentagem de arestas com baixa confiabilidade para cálculo da velocidade média	49
6.4.17	Porcentagem de arestas com velocidades inconsistentes	49
6.4.18	Porcentagem de arestas compartilhadas	50
6.4.19	Maior quantidade de linhas compartilhadas por aresta	50
6.4.20	Quantidade de conexões de uma linha	51
7	Visualização dos dados do AVL	52
7.1	Visão Geral	52
7.2	Visualização do AVL	53
7.2.1	Visualização de viagens reportadas pelo AVL	55
8	Conclusão	57
II	Parte Subjetiva	59
9	Apreciação e Críticas Pessoais para o Projeto	60
9.1	Motivação Pessoal	60
9.2	Desafios	61
9.3	Matérias que me ajudaram	61
9.4	Agradecimentos	62

Parte I

Parte Objetiva

Capítulo 1

Introdução

“Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.”

Open Data Handbook [1]
<http://opendatahandbook.org/>

1.1 Dados Abertos

Governos estão bastante interessados em explorar os benefícios que uma estratégia de "Open Data"[2] pode trazer à população. Para isto estão disponibilizando cada vez mais dados relativos às cidades permitindo uma maior busca de soluções para as pessoas interessadas em temas como "smartcities".

Dentre esses dados temos os do transporte público. No decorrer da última década, governos do mundo inteiro passaram a incentivar projetos de uso de dados abertos de transporte público. Podemos citar por exemplo, as cidades de Londres [8], Nova Iorque [7], Cidade do México [3], Santiago do Chile [5] [6] e também São Paulo [9]. Uma lista de cidades que estão disponibilizando os seus dados de trânsito pode ser encontrada no site da Google Transit [10].

Conforme o relatório sobre o transporte público da Cidade do México [4], em 2005, a cidade de Portland (Oregon - EUA) [12], em parceria com a Google [11], integrou as informações da agenda do transporte público e suas rotas com o Google Maps. Durante esse período foi desenvolvido o "*General Transit Feed Specification*" (GTFS) o qual padroniza e facilita a interpretação destes dados. O GTFS consiste de um conjunto de arquivos de texto delimitados por vírgulas, onde cada arquivo contém uma característica da informação de trânsito e um conjunto de regras de como gravá-las :

seis arquivos mandatórios (agência, paradas, rotas, linhas, tempo de paradas e calendário) e sete arquivos opcionais (dias de calendário, atribuições das tarifas, regras, formato da linha, frequência, transferência e informações do fornecedor). O GTFS foi desenhado inicialmente para conter apenas informações estáticas do transporte (por exemplo: rotas, paradas e frequências de saída), mas não informações dinâmicas (por exemplo: posicionamento dos ônibus). Posteriormente foram adicionadas capacidades para informações dinâmicas. Para as informações dinâmicas normalmente utiliza-se o equipamento AVL ("Automatic Vehicle Location") [13].

Em São Paulo estas informações começaram a ser disponibilizadas em 2013 [14]. Nelas temos informações tanto do GTFS quanto alguns dados tratados do AVL.

Comparando a cidade de São Paulo com outras cidades do mundo quanto a número de linhas (figura 1.1), número de ônibus (figura 1.2) e quantidade de passageiros que utilizam o transporte público por ano (figura 1.3), observa-se que São Paulo tem a maior quantidade de linhas de ônibus, a segunda em frotas de ônibus e a segunda em número de passageiros durante o ano ¹.

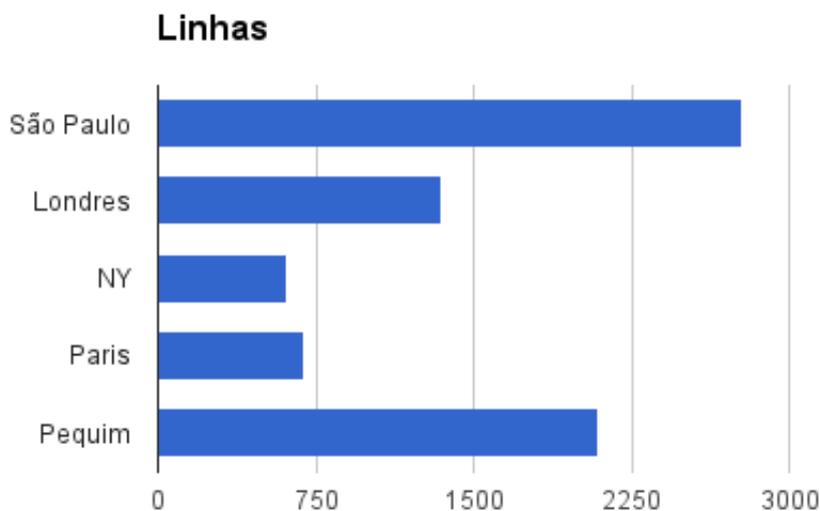


Figura 1.1: Número de linhas nas cidades.

¹Dados publicados pela SPTrans (São Paulo) [15], TFL (Londres) [16], MTA (Nova Iorque) [17], RATP (Paris) [18] e BPT e Yuntong (Pequim) [19]

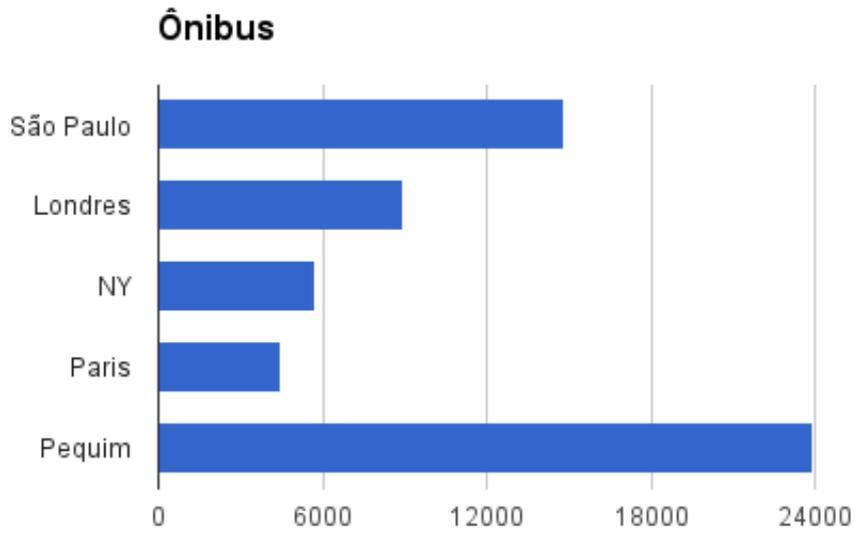


Figura 1.2: Número de ônibus em circulação.

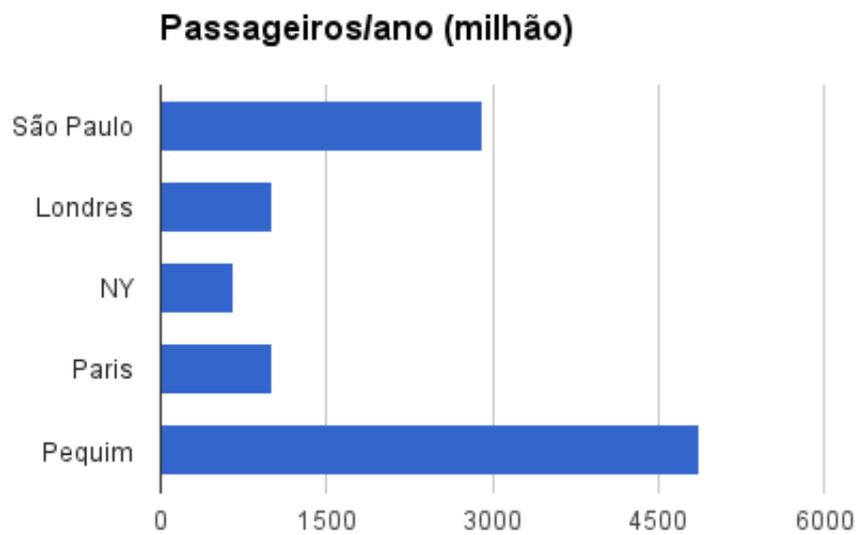


Figura 1.3: Número de passageiros que utilizam ônibus por ano.

1.2 O Projeto

O projeto apresentado neste documento tem como objetivo explorar e analisar dados para descobrir e visualizar informações do sistema de transporte público por ônibus da cidade de São Paulo.

Neste projeto iremos apresentar a exploração e análise dos dados de GTFS e AVL da cidade de São Paulo. O GTFS refere-se aos dados planejados e o AVL aos dados reais coletados dos ônibus. Para São Paulo somente os dados atuais do GTFS são dados abertos que podem ser encontrados no site do desenvolvedor da SPTrans. Já para o AVL, tivemos que nos reunir com a SPTrans para obter os dados brutos do AVL para as nossas análises. Atualmente a SPTrans também provê a API “Olho Vivo” com dados tratados do geoposicionamento dos ônibus em tempo real.

Com o GTFS e a API do *"Olho Vivo"* já foram desenvolvidos inúmeros aplicativos disponíveis no mercado, por exemplo o “Cadê o Ônibus?”, “Coletivo”, “Moovit”, “CityMapper”, e outros. Dado a sua abrangência e versatilidade, sua aplicabilidade torna-se imensa, por exemplo, identificar linhas de ônibus afetadas por um mega evento na cidade. Seu uso neste trabalho foi fundamental para descobrir anomalias e conclusões que pudessem qualificar a qualidade de serviço de ônibus na cidade de São Paulo.

Nos capítulos seguintes iremos apresentar as ferramentas, metodologia, banco de dados utilizados, assim como os resultados da exploração e análises dos dados e a conclusão.

Capítulo 2

Ferramentas

Este projeto foi desenvolvido para visualização e acesso através da Internet. Desta forma utilizamos as linguagens HTML, CSS e JavaScript. Para o ambiente do servidor foi utilizado o Node.js; para o banco de dados o MongoDB; para as visualizações o Meteor e para facilitar os cálculos, visualizações e importação dos dados foram utilizadas as bibliotecas JavaScript.

A seguir descrevemos com mais detalhes cada uma das ferramentas e como elas ajudaram o projeto.

2.1 Características das Ferramentas e seu uso no Projeto

Node.js [20] é um ambiente voltado para aplicações web do lado do servidor ("*backend*") no qual aplicações JavaScript podem ser executadas em diversas plataformas, por exemplo Linux, Windows, OS X e outros. Ele possui uma arquitetura orientada a eventos e com uma interface que não bloqueia a entrada/saída de dados, permitindo que a aplicação tenha maior vazão ("*throughput*") e escalabilidade (uma vez que ela não fica bloqueada pela API - "*Application Programming Interface*"). A escolha do ambiente Node.js foi prioritariamente selecionada porque ela permite o uso de JavaScript em todas as etapas, uniformizando o desenvolvimento das aplicações, seu "*debug*" e execução. Através deste ambiente foi possível a leitura dos dados e o desenvolvimento das métricas.

MongoDB [21] é um banco de dados "*NoSQL open-source*" orientado a documentos que provê alta performance, alta disponibilidade e escalabilidade automática. Sua escolha foi devido principalmente pelo fato de facilitar e simplificar o manuseio das informações do GTFS e do AVL em poucos conjuntos de dados (coleções) evitando o uso de múltiplas tabelas e "*queries*".

Por fim o Meteor [22] que é uma plataforma JavaScript voltada para desenvolvimento web, escrita em Node.js. Ela é de rápida prototipação e integrada com o MongoDB. Os seus principais diferenciais são:

- Modelo de programação reativa, o qual permite uma maior interação do banco de dados com o usuário;
- "*Data-on-the-wire*", no qual o servidor, ao invés de processar toda a informação, montar a página a ser mostrada e enviar o texto HTML através da linha ("wire") para o cliente, apenas envia os dados pela linha e deixa para o cliente a responsabilidade de decidir o que fazer com o mesmo;
- Integrado e flexível, onde não é necessário se preocupar com a integração e a administração dos componentes JavaScript.

Sua escolha foi porque ele tem uma resposta dinâmica, em tempo real ("*real-time*"), às ações do usuário, permitindo uma grande interatividade entre o cliente e o servidor, sua capacidade de "*data-on-the-wire*" e suporte ao Node.js.

As bibliotecas JavaScript utilizadas tiveram duas finalidades: a visualização e os cálculos.

- Para a visualização foram utilizadas bibliotecas baseadas no D3.js [23]. O D3.js é uma biblioteca JavaScript para manipulação de documentos baseado em dados e que utiliza HTML, SVG e CSS.
 - Para os gráficos, foi utilizado o C3.js [24] que é baseada na biblioteca D3.js facilitando a geração dos gráficos e sua atualização mesmo depois de criados.
 - Para os grafos foi utilizado o Alchemy.js [25] que é uma aplicação voltada para a visualização de grafos na web, baseada no D3.js.
 - Para os mapas foi utilizada a biblioteca JavaScript do Google Maps [26].
- Para os cálculos foi utilizado o Underscore.js [27], que é uma biblioteca JavaScript que provê uma série de funcionalidades que auxiliam as listas, coleções e objetos da linguagem.
- Para a importação dos dados do GTFS e do AVL foi utilizado o Fast-CSV [28] responsável pelo "*parsing*" e formatação dos dados CSV para o JavaScript.

Capítulo 3

Metodologia

Neste capítulo iremos introduzir conceitos que ajudarão a fundamentar este trabalho.

Na primeira seção temos a nomenclatura com os significados das palavras que serão utilizados durante o texto.

A seção seguinte será sobre a metodologia utilizada para o desenvolvimento do projeto. Esta foi dividida em 3 partes: (i) explorar e analisar os dados; (ii) identificar as anomalias; (iii) criar e testar as métricas para quantificar a qualidade dos dados e dos serviços.

Por fim mostraremos a arquitetura do sistema proposto de análise dos dados do transporte de ônibus da cidade de São Paulo.

3.1 Nomenclatura

Para evitar ambiguidades na leitura deste trabalho, definiremos nesta seção os principais termos que serão usados e seus respectivos significados.

AVL (*"Automatic Vehicle Location"*): Sistema de localização dos ônibus por GPS. Registra periodicamente a posição do ônibus e envia este dado a um servidor remoto;

Aresta: Representa o caminho entre dois pontos de paradas consecutivos. Veja a figura 3.1;

Frequência: Quantidade de partidas de ônibus consecutivas de uma determinada linha dentro de uma janela horária. A frequência pode mudar dependendo da janela;

GTFS (*General Transit Feed Specification*): Formato de dados que permite a especificação de uma rede de transportes de uma cidade, com suas

linhas e partidas planejadas;

Intervalo: Tempo, em minutos, entre duas partidas de ônibus consecutivas de uma determinada linha dentro de uma janela horária. O intervalo pode mudar dependendo da janela;

Linha: Uma trajetória de ida ou de volta planejadas na rede de transportes;

Parada: Um ponto de ônibus;

Partida: O início de uma viagem realizada por um ônibus sobre o trajeto de uma linha;

Pontos de ônibus: Locais de embarque ou desembarque de passageiros ao longo de uma linha;

Segmento: O trajeto formado pela conexão de dois pontos geo-localizados consecutivos (shapes). Veja a figura 3.1;

Shape: Ponto geo-localizado 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. Veja a figura 3.1;

Rota: Informações equivalente a linha de ida e a sua respectiva volta;

Terminal Primário (TP): Local de partida dos ônibus no sentido de ida de uma linha;

Viagem: O ato ou movimento de percorrer o trajeto da linha;

Terminal Secundário (TS): Local de partida dos ônibus no sentido de volta de uma linha;

Trajetória: 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 trajetória de ida, ou trajetória de volta caso o sentido seja TS -> TP;

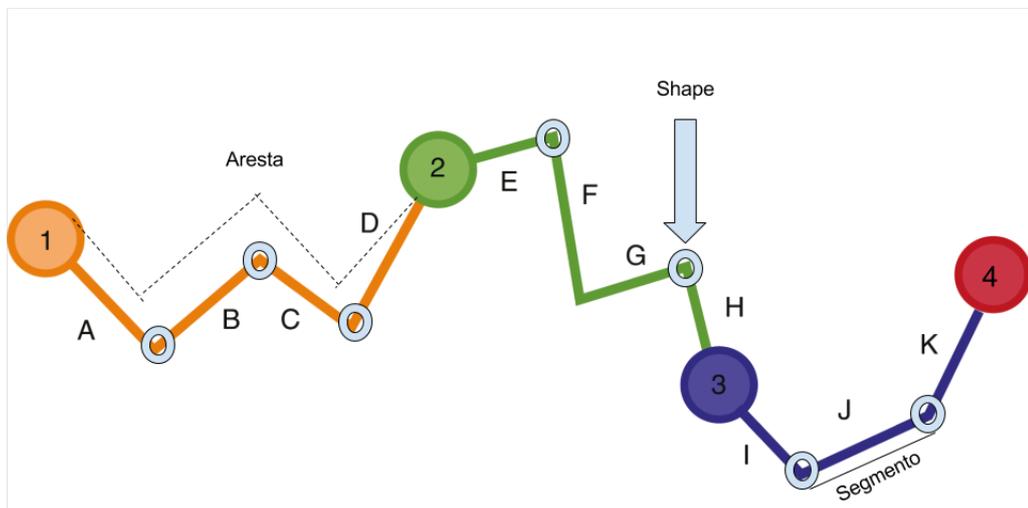


Figura 3.1: Exemplo de Paradas (círculos coloridos), Arestas (trajetória entre duas paradas), Shape (círculos azuis), Segmento (retas identificadas de A a K)

3.2 Metodologia

A metodologia utilizada neste trabalho consiste de três etapas:

- explorar e analisar os dados;
- identificar as anomalias;
- criar e testar as métricas para quantificar a qualidade dos dados e dos serviços.

A fase de exploração e análise dos dados foi importante para identificarmos as relações entre as informações, as melhores formas de organizar as estruturas de dados, os possíveis gargalos de escalabilidade da análise (considerando o volume de dados a serem analisados) e as melhores ferramentas para o tratamento dos dados apresentados.

Nesta etapa foram feitas explorações e análises de dados referente a linhas, pontos de ônibus e arestas. Dentre elas temos: quilometragem das linhas, quantidade de pontos, quantidade de segmentos, quantidade de conexões de uma linha, quantidade de pontos duplicados de uma linha, maior quantidade de linhas compartilhadas por arestas e a porcentagem de arestas com baixa confiabilidade para cálculo da velocidade média.

A etapa de identificação das anomalias foi fundamental para identificar resultados que não eram compatíveis com a realidade ou para auxiliar na compreensão dos resultados normais. É preciso saber quais linhas tem características diferentes da normal para identificarmos se um dado resultado anormal é um erro ou uma particularidade da linha.

Nessa etapa identificamos anomalias nas linhas, ponto de ônibus, arestas e segmentos. Para isso comparamos os valores medidos com os valores esperados, por exemplo: Na linha 3310-10-0 - Cidade Tiradentes, identificamos uma anomalia quanto à sua extensão. Essa linha apresenta 102 km de extensão enquanto que 95% das linhas apresenta um tamanho menor que 31 km.

Por fim as métricas, que serão detalhadas nos capítulos 6 a 8, criadas e testadas com a finalidade de quantificar a qualidade dos dados e dos serviços. Nesta etapa fizemos métricas tanto para os dados planejados (GTFS) quanto para os dados reais (AVL).

Foi analisada a semana chamada de "referência", de 18/10/2014 a 24/10/2014. Esta semana foi chamada de "semana de referência" porque utilizamos este

período para análise comparativa, uma vez que a SPTrans já possuía análise referente a esta semana de anos anteriores.

Estas análises estarão disponíveis na ferramenta de visualização de dados que faz parte deste projeto. Nos capítulos seguintes descrevemos as análises de qualidade de dados tanto do GTFS quanto do AVL.

3.3 Arquitetura

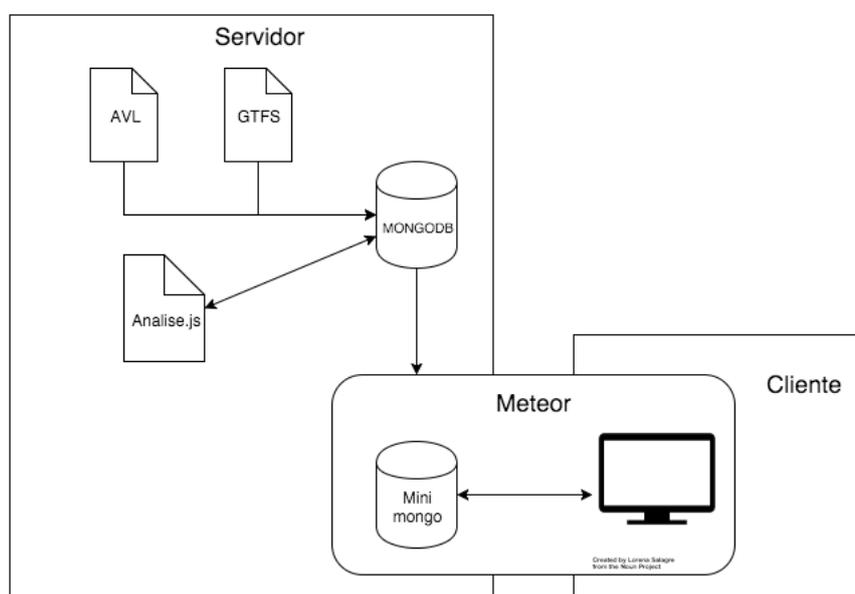


Figura 3.2: Arquitetura do sistema proposto de análise dos dados do transporte de ônibus da cidade de São Paulo.

A figura 3.2 mostra a arquitetura utilizada neste projeto. O ponto central desta arquitetura é o banco de dados MongoDB onde estão armazenadas as informações do GTFS e do AVL.

Com a utilização do Analise.js e dos dados do MongoDB, foram geradas as métricas que posteriormente foram visualizadas. Em seguida utilizamos o Meteor que nos forneceu uma plataforma de desenvolvimento e execução tanto do lado do servidor quanto do cliente. Na parte do cliente temos a nossa visualização de dados. Na parte do servidor, teremos o MiniMongo, que armazena os dados do MongoDB recentemente visualizados (*buffer*), fazendo com que o cliente acesse tais dados de forma mais rápida e sem a necessidade de acessar o MongoDB novamente.

Capítulo 4

Estruturação do Banco de Dados

Este capítulo mostra como era a organização inicial dos dados do GTFS e do AVL, a modelagem dos dados do GTFS deste trabalho e a modelagem da Scipopulis para os dados do AVL.

4.1 Organização inicial dos Dados

Neste trabalho utilizamos GTFS para indicar um conjunto de arquivos com as especificações dos dados da rede de ônibus da cidade de São Paulo.

Estes arquivos são :

- *agency.txt* Contém os dados referentes à agência. (Neste caso a SP-Trans);
- *calendar.txt* Contém os códigos dos dias de funcionamento e os seus significados;
- *fare_attributes.txt* Contém o custo de cada tipo de locomoção;
- *fare_rules.txt* Contém os identificadores das linhas e seus tipos de locomoção (ônibus, metrô, CPTM);
- *frequencies.txt* Apesar do nome frequências, este arquivo contém os intervalos, em segundos, da passagem de cada linha de acordo com o horário;
- *routes.txt* Contém as informações de uma rota, por exemplo, o nome da rota e o seu identificador;

- *shapes.txt* Contém o shape (ponto geo-localizado que indica uma mudança de direção) de cada linha;
- *stops_times.txt* Contém os pontos que cada linha passa e em que horário ela passa;
- *stops.txt* - Contém informações sobre os pontos, como a latitude e a longitude;
- *trips.txt* - Contém informações sobre as linhas. Informações como a linha a que ela pertence;

Estes arquivos estão organizados conforme abaixo.

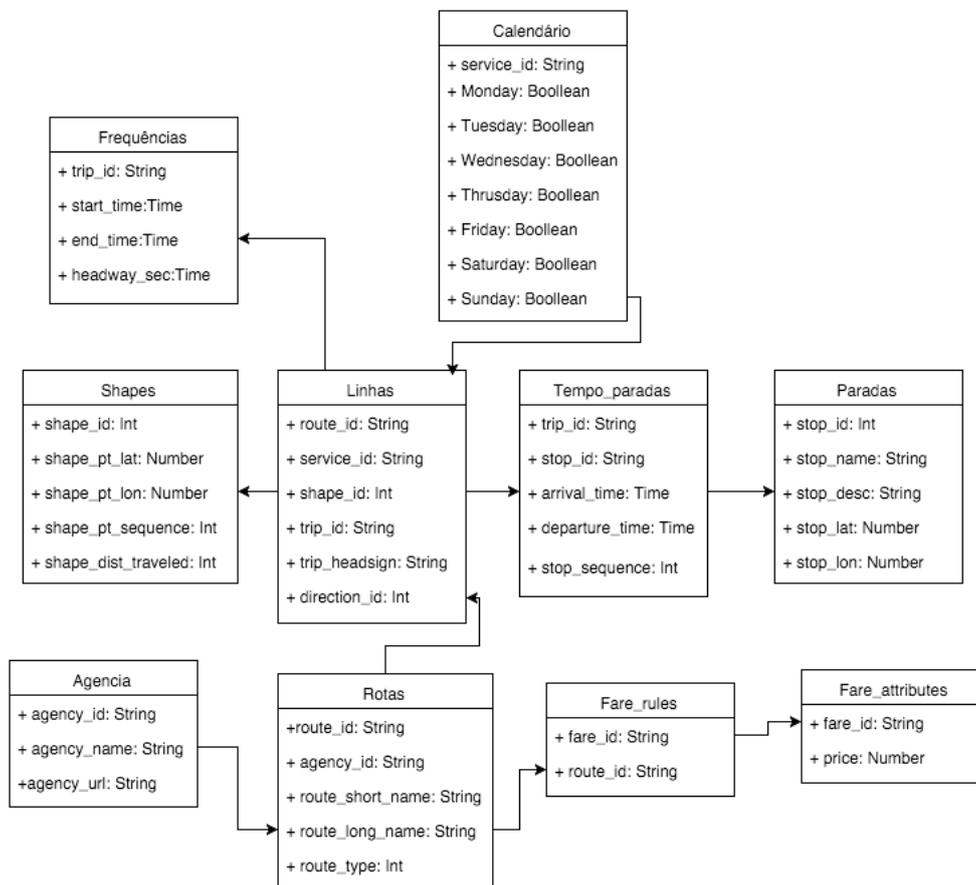


Figura 4.1: Organização prévia dos arquivos GTFS.

O AVL é um sistema de localização dos ônibus por GPS, registrando periodicamente a posição do ônibus e enviando este dado a um servidor remoto. Cada dia é gerado um AVL que é composto por 3 arquivos:

- *AL_data.txt* - Contém informações das linhas sendo monitoradas como o número da linha, complemento, código da linha e sentido da linha;
- *AV_data.txt* - Contém informações sobre o veículo como o código da linha e o código do veículo realizando o trajeto daquela linha;
- *MO_data.txt* Contém dados sobre o movimento dos veículos como o horário em que o dado foi recebido pelo servidor, o horário em que o dado foi capturado pelo equipamento, o código da linha, a latitude, a longitude e o código do equipamento;

4.2 MongoDB

O armazenamento de dados no MongoDB é organizado em coleções. Dessa forma, os arquivos do GTFS e AVL são armazenados nestas coleções.

No GTFS temos três coleções, sendo elas trips, stops e edges. No trips colocamos as informações das linhas, por exemplo, quilômetros, pontos que a linha passa e os seus segmentos. Já no stop temos as informações das paradas de ônibus, por exemplo, as linhas que passam por ela e a sua localização. Por fim o edges com informações sobre as arestas, por exemplo, quais linhas passam por ela, os paradas que compõem a aresta e a sua extensão.

No AVL temos apenas uma coleção que é a coleção dos log_avl.

4.2.1 Modelagem do GTFS no Projeto

A modelagem do GTFS para uso nesse trabalho teve como finalidade adaptar as definições genéricas da especificação às necessidades e particularidades dos dados analisados.

A importação dos arquivos do GTFS para o MongoDB é feita no ambiente node.js através da biblioteca fast-csv (versão 0.5.7).

Primeiro são lidos os arquivos stops.txt e trips.txt e a partir desta leitura são criadas duas listas com informações específicas sobre linha e parada de ônibus.

São também lidos outros arquivos, por exemplo shapes.txt, frequencies.txt e stop_times.txt, cujas informações são utilizadas para complementar as listas específicas conforme acima.

Cada elemento da lista contém todas as informações referente ao objeto em questão (se for a lista de linhas, todas as informações referentes a linha, por exemplo shapes, intervalo e nome, no caso da lista de paradas, informações como posição geográfica da parada (longitude/latitude) e nome das linhas que passam por esta parada).

Ao terminarmos esta operação, iremos verificar se uma linha (ida) tem o seu complementar (volta) ou se é uma linha circular. Para isso basta verificarmos se o identificador da linha possui seu último dígito diferente. Por exemplo: A linha 5701-10-0 - Estação Berrini possui a volta, a linha 5701-10-1 - Metrô Conceição. Já linha 8022-10-0 - Cidade Universitária é circular pois não existe a linha 8022-10-1.

Feito isso criamos uma lista com informações para as arestas (edges) onde cada um dos elementos desta lista terá informações tanto das linhas que passam por ela quanto dos pontos de ônibus que a compõem. Por exemplo, informações como : quais as paradas que compõem a aresta, o comprimento da aresta, linhas que passam pela aresta, o intervalo de tempo que as linhas percorrem a aresta.

Por fim, colocamos estas listas no MongoDB em formato de coleções. Com esta modelagem do projeto, as nossas "*queries*" ficaram otimizadas tirando a necessidade de muitas junções de tabela (que seriam necessárias no modelo original).

4.2.2 Modelagem do AVL no Projeto

Ao importar o AVL criamos o `log_avl`. Neste `log_avl` temos dados enviados pelo ônibus, como por exemplo, o horário de envio da informação, a localização do ônibus e qual linha ele estava operando.

Capítulo 5

Exploração de Dados

Neste capítulo mostramos a exploração inicial dos dados. O objetivo desta exploração inicial foi familiarizarmos com a estrutura de dados do GTFS. Foram analisados dados gerais referentes às linhas, segmentos, pontos de ônibus e arestas.

Foram definidas 7 métricas:

- Quantidade de paradas por linha;
- Comprimento de linhas;
- Comprimento de segmentos;
- Comprimento de arestas;
- Quantidade de linhas que passam por cada aresta;
- Quantidade de linhas que passam por um ponto de ônibus;
- Porcentagem de ônibus circulares;

Em adição às métricas foi também desenvolvido a visualização de arestas compartilhadas onde são enfatizadas as arestas com maior número de linhas.

5.1 Visão Detalhada das Métricas

A seguir serão apresentados detalhes sobre cada uma das métricas acima. Os dados utilizados referem-se a outubro de 2014. Como informação geral temos: 9.988.278 km transitados diariamente (dias úteis) pela frota de ônibus, 18.733 pontos de ônibus, 25.568 arestas, 2.204 linhas e 14.663 ônibus ativos.

A visualização de cada métrica é feita na forma de histograma onde são apresentadas informações sobre a mediana e desvio padrão. Além disso é também apresentado uma tabela com detalhes sobre cada métrica considerando a média, mediana, métrica (linha, aresta ou ponto de ônibus) na média e métrica anormal (fora do segundo desvio padrão de 95%).

- Quantidade de paradas por linha

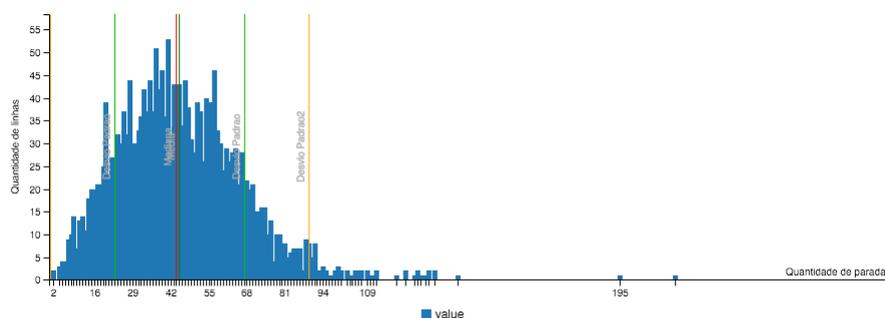


Figura 5.1: Histograma da quantidade de paradas por linha.

Média	Mediana	Linha na média	Linha anormal
45 paradas por linha	40 paradas por linha	2706-10-0 - Metrô Vila Matilde	3310-10-0 - Cidade Tiradentes (circular) - 288 paradas.

Tabela 5.1: Quantidade de paradas por linha.

Nesta métrica mostramos como é a distribuição da quantidade de pontos de ônibus por linha (fig 5.1), considerando todas as linhas de ônibus da cidade de São Paulo. Dessa forma, conseguimos determinar se uma dada linha encontra-se próxima a média ou não. Na tabela 5.1 temos os valores de forma tabular e a indicação de uma linha que esta na média e uma linha que está fora dos 95% do segundo desvio padrão.

- Quantidade de comprimentos de linhas

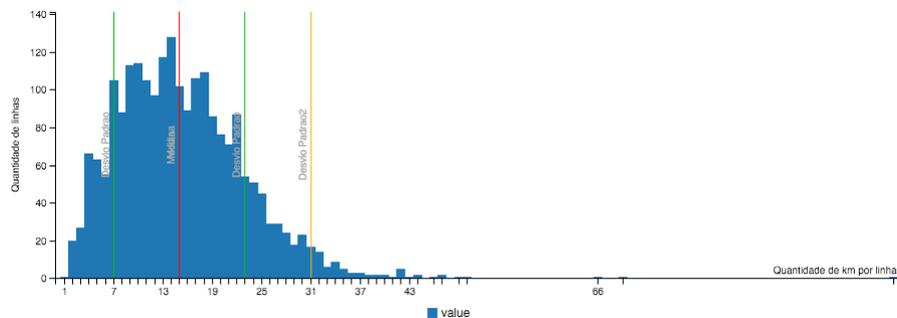


Figura 5.2: Histograma do comprimento de linhas.

Média	Mediana	Linha na média	Linha anormal
15 km por linha	15 km por linha	6074-10-1 - Jd. Gaivotas - 17 Km	3310-10-0 - Cidade Tiradentes (circular) - 102 Km

Tabela 5.2: Comprimentos de linhas.

Nessa métrica temos a distribuição da quantidade de comprimentos de linhas na cidade de São Paulo, conforme a figura 5.2. A média de comprimento de linha é de 15 quilômetros. Na tabela 5.2 temos que a linha 6074-10-1 - Jardim das Gaivotas esta perto da média com 17 quilômetros e a linha 3310-10-0 - Cidade Tiradentes esta muito acima da média, com 102 quilômetros de comprimento.

- Comprimento de segmentos

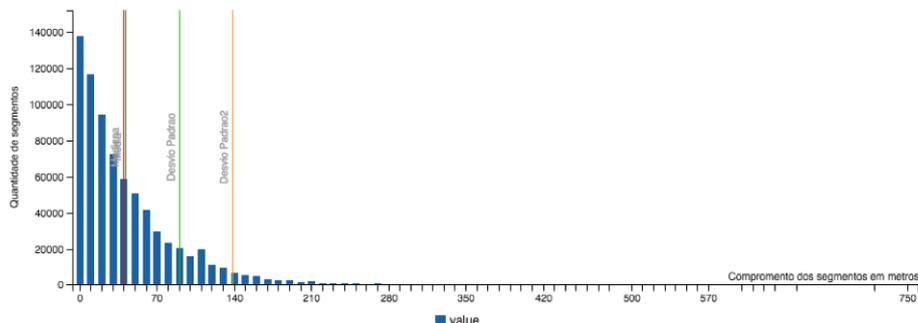


Figura 5.3: Histograma do comprimentos de segmentos.

Média	Mediana	Quantidade de segmentos da linha na média	Quantidade de segmentos da linha na anormalidade
42 metros por segmento	40 metros por segmento	3310-10-0 - Cidade Tiradentes (circular) - 2036 de 2143 dentro dos 95%	3310-10-0 - Cidade Tiradentes (circular) - 107 de 2143 fora dos 95%

Tabela 5.3: Comprimento de segmentos.

Nessa métrica temos a distribuição da quantidade de comprimentos dos segmentos das linhas de ônibus da cidade de São Paulo, conforme a figura 5.3. A média de comprimento do segmento é de 42 metros. Na tabela 5.3 temos a linha 3310-10-0 - Cidade Tiradentes (circular) onde 2036 dos 2143 segmentos tem o comprimento dentro do desvio padrão de 95%, e apenas 107 segmentos fora deste desvio padrão.

- Comprimento de arestas

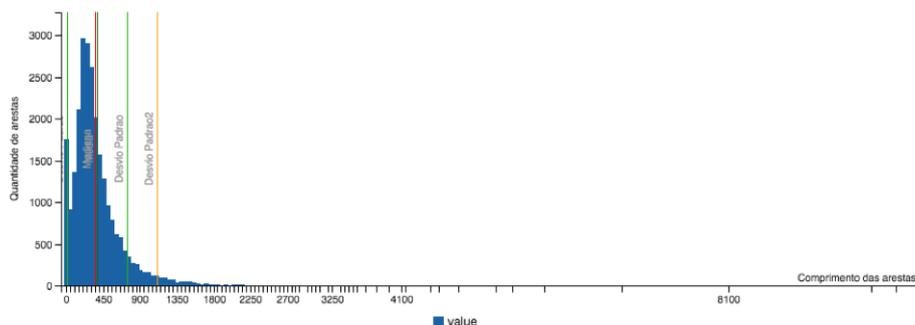


Figura 5.4: Histograma do comprimentos de arestas.

Média	Mediana	Aresta na média	Aresta anormal
374 metros por aresta	350 metros por aresta	Caminho Bela Vista / Am. Brasiliense C/B 1	Term. Parque Dom Pedro II - Plat 01 / Av. Amador Bueno Da Veiga, 236 - 10412 metros

Tabela 5.4: Comprimento de arestas.

Nessa métrica temos a distribuição da quantidade de comprimentos das arestas das linhas de ônibus da cidade de São Paulo, conforme a figura 5.4. A média de comprimento da aresta é de 374 metros. Na tabela 5.4 temos que a aresta - Caminho Bela Vista / Am. Brasiliense C/B 1 - está perto da média. Já a aresta - Term. Parque Dom Pedro II - Plat 01 / Av. Amador Bueno Da Veiga, 236 -, pertencente à linha 2022-51-1 - Jd. Dos Ipes, está fora do segundo desvio padrão, apresentando 10412 metros de extensão.

- Quantidade de linhas que passam por cada aresta

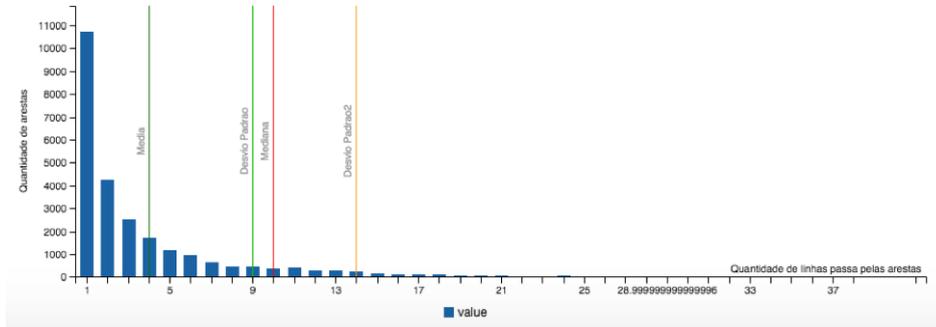


Figura 5.5: Histograma da quantidade de linhas que passam por cada aresta.

Média	Mediana	Aresta na média	Aresta anormal
4 linhas por aresta	10 linhas por aresta	Parada Ginásio do Ibirapuera / Alameda Lorena	Prudente Correia / Benedito Chaves - 41 linhas.

Tabela 5.5: Quantidade de linhas que passam por cada aresta.

Nessa métrica temos a distribuição da quantidade de linhas que passam por cada aresta das linhas de ônibus da cidade de São Paulo, conforme a figura 5.5. A média é de 4 linhas. Na tabela 5.5 temos que a aresta - Parada Ginásio do Ibirapuera / Alameda Lorena - está na média, já a aresta - Prudente Correia / Benedito Chaves - está fora do segundo desvio padrão, passando por ela 41 linhas.

- Quantidade de linhas que passam por um ponto de ônibus

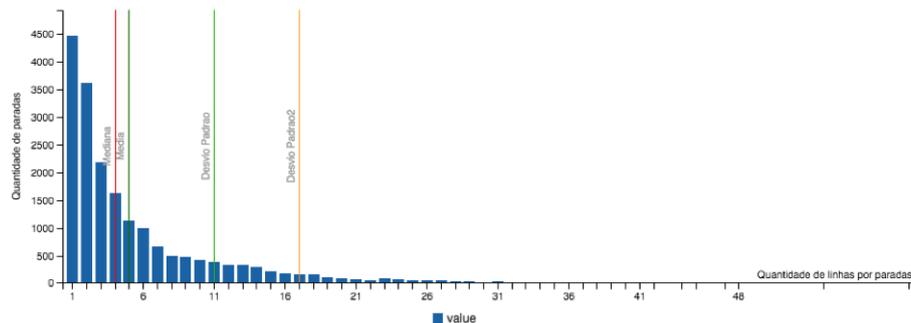


Figura 5.6: Histograma da quantidade de linhas que passam por um ponto de ônibus.

Média	Mediana	Ponto de ônibus na média	Ponto de ônibus anormal
5 linhas numa parada de ônibus	4 linhas numa parada de ônibus	940003781 - Av. Conceição, 3847	600012396-R. Rodovalho Júnior, 637 (60 linhas)

Tabela 5.6: Quantidade de linhas que passam por um ponto de ônibus.

Nessa métrica temos a distribuição da quantidade de linhas que passam por um ponto de ônibus na cidade de São Paulo, conforme a figura 5.6. A média é de 5 linhas. Na tabela 5.6 temos que o ponto - Av. Conceição, 3847 - está na média. Já o ponto - R. Rodovalho Júnior, 637 - está fora do segundo desvio padrão, passando por ela 60 linhas.

- Porcentagem de ônibus circulares

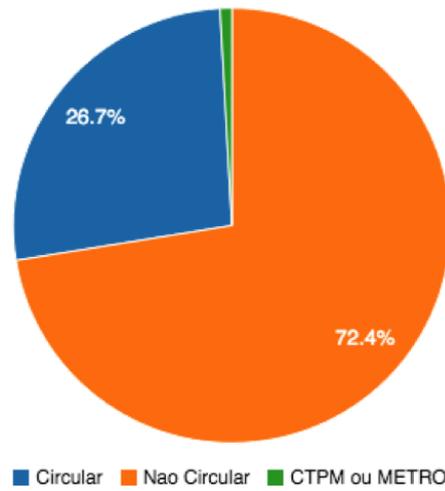


Figura 5.7: Porcentagem de linhas circulares.

Na figura 5.7 temos 26.7% são linhas circulares e 72.4% são linhas com ida e volta (não circulares). O restante (0,9%) são Metrô e CPTM.

- Compartilhamento de arestas
Além das visualizações das métricas como mostrado acima, também foi feita uma visualização para descobrir as arestas compartilhadas de uma amostra de 50 linhas.

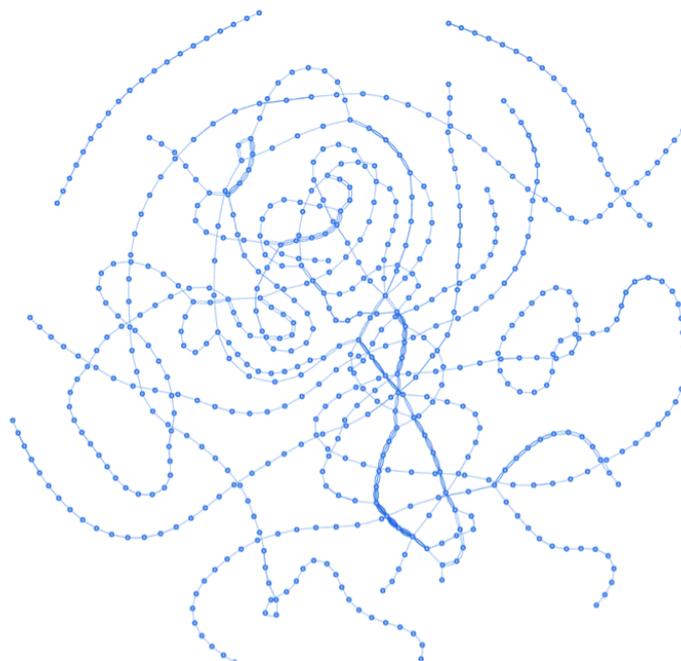


Figura 5.8: Representação de 50 linhas e onde elas compartilham arestas.

Na figura 5.8 visualiza-se partes de uma linha onde ela compartilha arestas com outras. Na figura, quanto mais espessa é uma aresta mais linhas ela compartilha. Esta imagem foi feita utilizando a biblioteca Alchemy.js.

Capítulo 6

Análise de dados do GTFS

Neste capítulo mostramos as métricas desenvolvidas para analisar os dados do GTFS e a sua visualização.

6.1 Visão Geral

No GTFS temos informações planejadas das linhas e dos pontos de ônibus. Nas linhas temos informações como o identificador da linha, o formato, o intervalo da saída dos ônibus e os horários de operação. Já sobre os pontos de ônibus temos informações como o seu identificador e sua geo-localização. A fim de saber mais sobre o planejamento das linhas e suas anomalias, foram definidas 20 métricas. Estas métricas ajudaram na análise da qualidade do serviço de ônibus da cidade de São Paulo.

Na tabela 6.1 apresentamos as informações sobre as métricas, suas descrições e seus fundamentos.

Tabela 6.1: Tabela das Métricas do GTFS.

Começo da Tabela das Métricas do GTFS		
Métrica	Descrição	Fundamento
Número de paradas distantes mais de 500m	Quantidade de paradas de ônibus que estejam distantes a mais de 500m da próxima parada.	Como o cálculo da velocidade média utiliza a distância entre duas paradas de ônibus, se houver muitas paradas que tenham distância entre si maiores que 500m, podemos ter velocidades médias fora do desvio padrão. Ou seja, com uma indicação de alta velocidade média, porém sem explicitar que isto é devido à longa distância entre paradas de ônibus.
Número de paradas distantes menos de 30m.	Quantidade de paradas de ônibus que estejam distantes a menos de 30m.	De forma similar à métrica "Número de paradas distantes mais de 500m", paradas com menos de 30m podem induzir a uma velocidade média equivocada.
Número de partidas diárias previstas.	Número total de partidas previstas no dia para uma dada linha.	Esta métrica é importante para analisar a porcentagem de cumprimento do número de partidas.
Linha sem frequência de partidas registradas.	Sinalizamos se uma dada linha não possui planejamento de partidas devidamente registrada.	Sem um planejamento de partidas de ônibus devidamente registrada não é possível o cálculo do cumprimento de partidas de uma dada linha.

Continuação da Tabela das Métricas do GTFS 6.1		
Métrica	Descrição	Fundamento
Total de horas de operação.	Quantidade de horas em que existem partidas programadas no GTFS.	As diferentes linhas da cidade de SP possuem horários de funcionamento específicos. Algumas operam apenas pela manhã, durante algumas horas, ou durante todo o dia. O número de partidas previsto é proporcional ao horário de funcionamento, e linhas com horário pequeno de funcionamento podem gerar dados de cumprimento de partida inconsistentes.
Quantidade de shapes na linha.	Número de pontos que definem o trajeto da linha sobre o viário.	Cada ponto define um segmento de reta no trajeto da linha. Muitos shapes podem indicar um caminho com muitas curvas e poucos segmentos retos, o que pode afetar a velocidade média da linha.
Distância total.	Distância do ponto inicial ao ponto final de uma viagem, em km, considerando-se o trajeto sobre o viário da cidade.	Linhas muito longas ou muito curtas podem gerar dados de velocidade média incomuns, seja por falta de dados para o cálculo, seja por acumular diversas arestas congestionadas em seu trajeto.
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.	Linhas com velocidades médias planejadas muito altas podem ter dificuldades em cumprir a quantidade planejada de partidas, enquanto linhas com velocidades médias planejadas muito baixas podem mascarar dados bons de cumprimento de partida.

Continuação da Tabela das Métricas do GTFS 6.1		
Métrica	Descrição	Fundamento
Quantidade de paradas duplicadas.	Número de paradas que aparecem mais de uma vez na sequência de paradas de uma linha.	No cálculo das velocidades médias é importante saber se uma linha passa mais de uma vez pela mesma parada, para separar o cálculo de velocidade média em cada aresta.
Linha que passa 2 vezes pela mesma parada.	Se uma linha passa mais de uma vez pela mesma parada, sinalizamos tal fato.	Se uma linha passa mais de uma vez na mesma parada em um mesmo sentido, temos que adaptar o cálculo da velocidade média para levar isso em conta.
Paradas em comum na ida e na volta.	Número de paradas com mesmo identificador na ida e na volta de uma linha.	Trajetos que possuem paradas em comum na ida e na volta podem dificultar a análise de dados do AVL gerando ambiguidade entre os trajetos.
Tempo previsto de viagem.	Tempo planejado para uma viagem de um terminal a outro.	Viagens muito rápidas ou muito demoradas exigem maior atenção no cálculo de velocidades médias.
Frequência média por hora.	Quantidade de ônibus desta linha que devem partir a cada hora, em média.	Linhas com frequência muito baixa podem dificultar o cálculo de velocidades médias.
Intervalo médio entre ônibus.	Intervalo médio entre partidas consecutivas de ônibus realizando o trajeto de uma linha.	Linhas com intervalo entre ônibus muito grande podem gerar dados de velocidade média inconsistentes.

Continuação da Tabela das Métricas do GTFS 6.1		
Métrica	Descrição	Fundamento
Porcentagem de arestas com baixa confiabilidade para cálculo da velocidade média.	Para uma dada linha, qual a porcentagem de arestas em que o intervalo de passagem de ônibus, considerando-se todas linhas que passam pela aresta, é superior a 10 minutos.	Para o cálculo da velocidade média dos ônibus em uma aresta consideramos todos os ônibus de qualquer linha passando pela aresta. Entendemos que se a frequência conjunta de todas as linhas que passam pela aresta é superior a 6 ônibus por hora, temos dados suficientes para calcular, com precisão, a velocidade média da aresta. Esta métrica nos permite calcular, para toda a linha, qual a velocidade média da linha, com precisão, considerando-se todas as arestas que ela contém.
Porcentagem de arestas com velocidades inconsistentes.	Qual a porcentagem de arestas na linha que possuem velocidades muito abaixo ou muito acima de valores típicos para um ônibus	O arquivo GTFS gerado pela SPTrans não leva em conta as velocidades reais entre paradas em uma viagem. Ao invés disso, é feita uma estimativa simplista que divide o tempo total da viagem pela quantidade de arestas para definir o tempo estimado para percorrer cada aresta. Esse cálculo pode gerar valores inconsistentes. Por exemplo : paradas muito próximas podem dar origem a velocidades esperadas muito baixas, enquanto paradas distantes podem dar origem a velocidades esperadas muito altas. Esta métrica ajuda a identificar linhas com espaçamento desigual entre os pontos.

Continuação da Tabela das Métricas do GTFS 6.1		
Métrica	Descrição	Fundamento
Porcentagem de arestas compartilhadas de uma linha.	A porcentagem de arestas de uma dada linha que é compartilhada com outras linhas.	Quantas arestas não são dependentes de uma única linha.
Maior aresta compartilhada de uma linha.	A aresta com o maior número de linhas compartilhadas.	Para identificar quais são as arestas com os maiores números de linhas que passam por ela e quantas linhas um problema nesta aresta pode afetar.
Número de conexões de uma linha.	Quantas conexões tem uma linha analisada.	A partir de uma linha quantas outras linhas consegue-se conectar.

6.2.2 Pontos de Ida e Volta

Nesta métrica verificamos se há pontos comuns de ida e volta numa linha. Uma linha tem este tipo de ponto se tanto na ida (que vai do terminal primário para o secundário) quanto na volta (que vai do terminal secundário para o primário) passa pelos mesmos pontos.

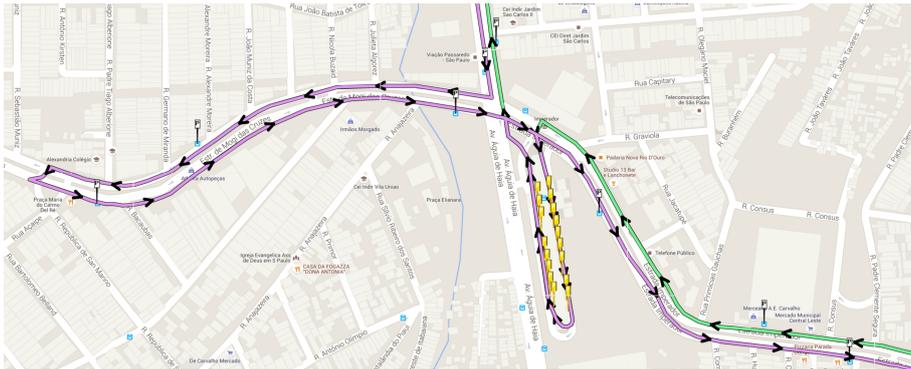


Figura 6.2: Linha: 2756-10-1 - Guaianazes (lilas) e 2756-10-1 - Metrô Patriarca (verde).

Na figura 6.2 temos as linhas 2756-10-1 - Guaianazes (lilas) e 2756-10-1 - Metrô Patriarca (verde) que compartilham os pontos assinalado em amarelo. Estes pontos são ruins pois ao fazer as análises dos dados do AVL para informações de geo-localização emitidos nesta região, poderemos ter um problema para identificar de qual linha veio o sinal.

Embora o AVL esteja associado ao ônibus, o problema de pontos de ida e volta torna-se significativo principalmente quando há situações de exceção onde faltam dados de posicionamento adjacente comprometendo a interpretação do ponto dentro de um contexto (linha). Por exemplo, um mesmo ônibus que passe por pontos comuns na ida e na volta, caso falte dados do AVL que permitam identificar de forma explícita se é ida ou volta, a coordenada do ponto de ida e volta não poderá ser utilizado uma vez que não saberemos a qual trajeto ela se refere.

6.2.3 Taxa de Confiabilidade Baixa

Esta métrica nos permite saber para uma dada linha qual é a porcentagem de arestas que passam ônibus com frequência superior a um ônibus a cada 10 minutos. Para este cálculo consideramos todos os ônibus de qualquer linha passando pela aresta. Entendemos que se a frequência conjunta de todas as linhas que passam pela aresta é superior a 6 ônibus por hora, temos dados suficientes para calcular com precisão a velocidade média da aresta. Se a frequência for inferior, temos então uma Taxa de Confiabilidade Baixa.

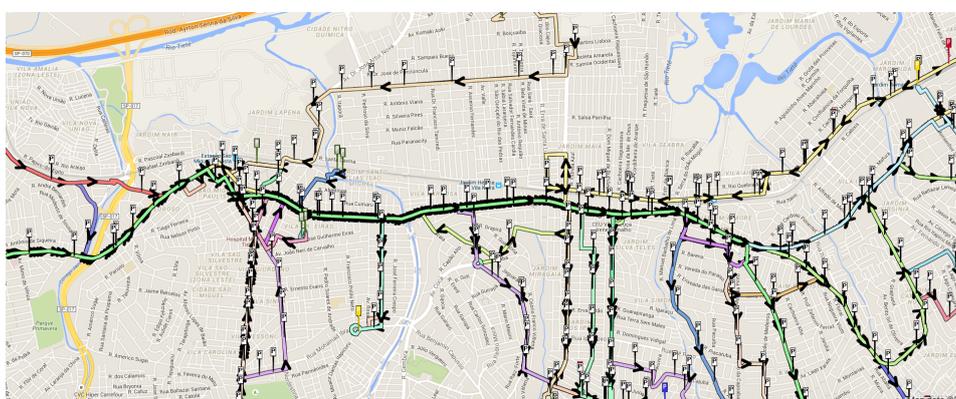


Figura 6.3: Aresta da Marechal Tito.

Observa-se na imagem 6.3 que a região de Marechal Tito tem uma confiabilidade muito boa para a sua velocidade média pois temos uma frequência de 3639 ônibus por hora pertencentes a 34 linhas.

6.2.4 Conexões

Esta métrica nos permite saber quantas outras linhas podem se conectar à linha analisada. Na métrica utilizamos informações dos pontos de ônibus que passam pela linha analisada. A cada ponto que percorremos nesta lista contamos apenas a aparição das linhas novas.

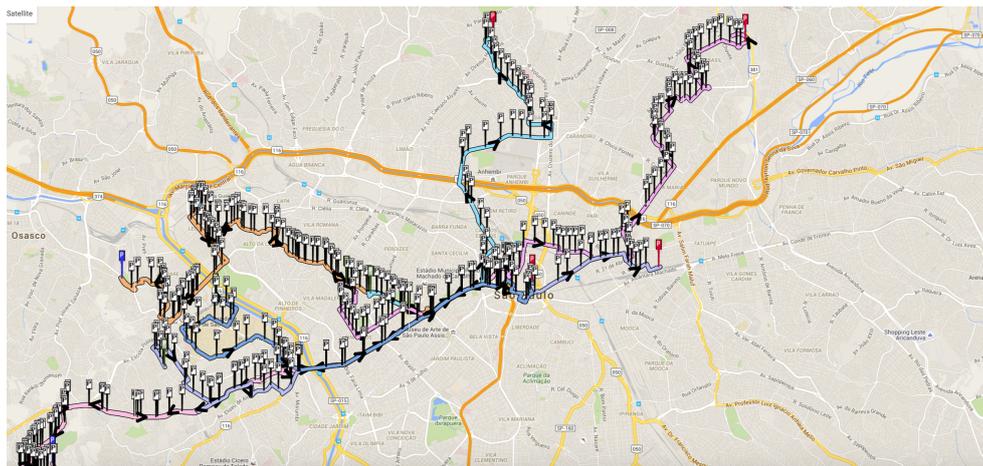


Figura 6.4: Conexões com a linha 702U-10 - Term.Pq. Dom Pedro II na região da Paulista.

Na imagem 6.4 vemos cinco das 247 linhas que se conectam com a linha 702U-10 - Term.Pq. Dom Pedro II. Isto é relevante porque quanto maior o número de conexões melhor será a malha, ou seja, opções para o passageiro chegar ao seu destino.

6.3 Visualização

Nas figuras 6.5 e 6.6 temos as visualizações feitas para o GTFS.

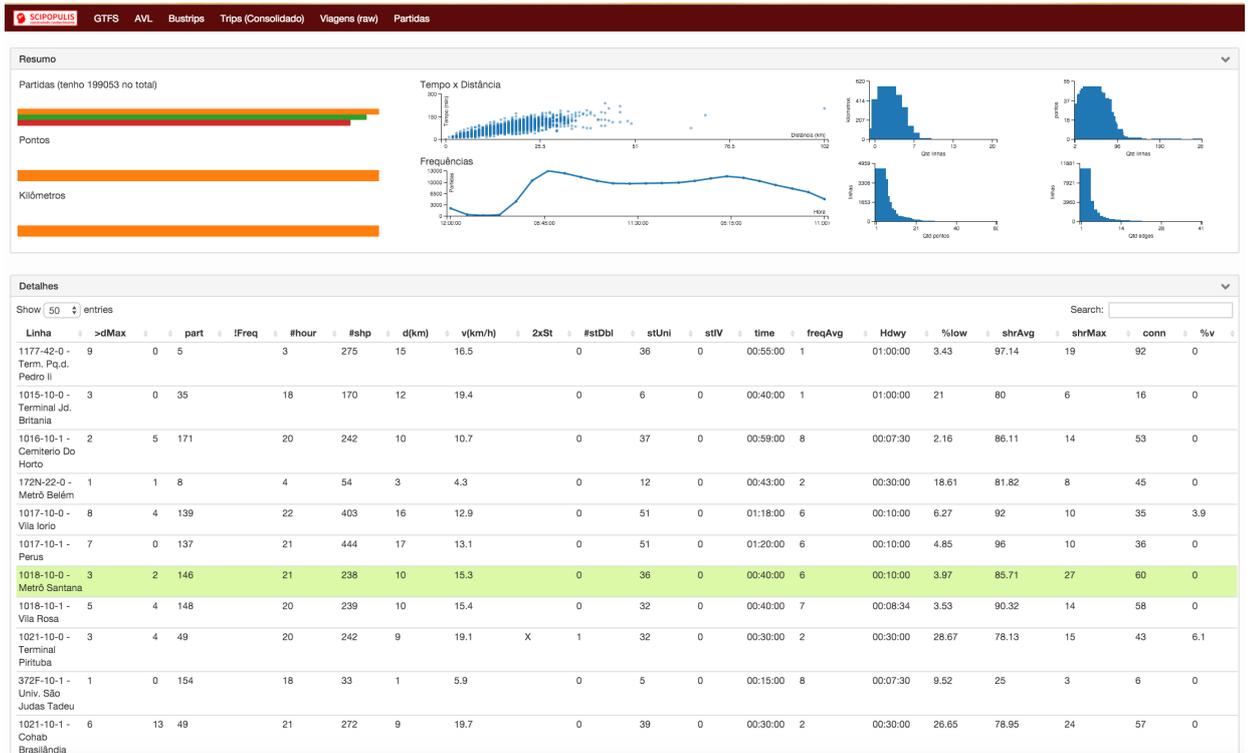


Figura 6.5: Visualização parcial do GTFS.

Na parte superior da figura 6.5 temos as informações gerais sobre as paradas de ônibus, linhas e frequências. Em seguida, temos uma tabela com todas as linhas de ônibus e suas métricas e na parte superior direita temos uma opção de busca. Além disso, temos uma funcionalidade que permite ordenar (ordem crescente ou decrescente) as linhas de acordo com uma dada métrica.

Ao selecionar uma linha desta tabela obtemos no mapa o formato desta linha e os seus pontos de ônibus, por exemplo, ao selecionarmos a linha 1018-10-0 Metrô Santana na figura 6.5, visualizamos esta linha de ônibus e seus pontos na figura 6.6. A cor da linha visualizada no mapa corresponde a cor indicada na tabela. Já para os pontos de ônibus, o sistema atribui diferentes cores para identificar pontos duplicados (azul), ida e volta (amarelo), início (verde) e fim (vermelho).

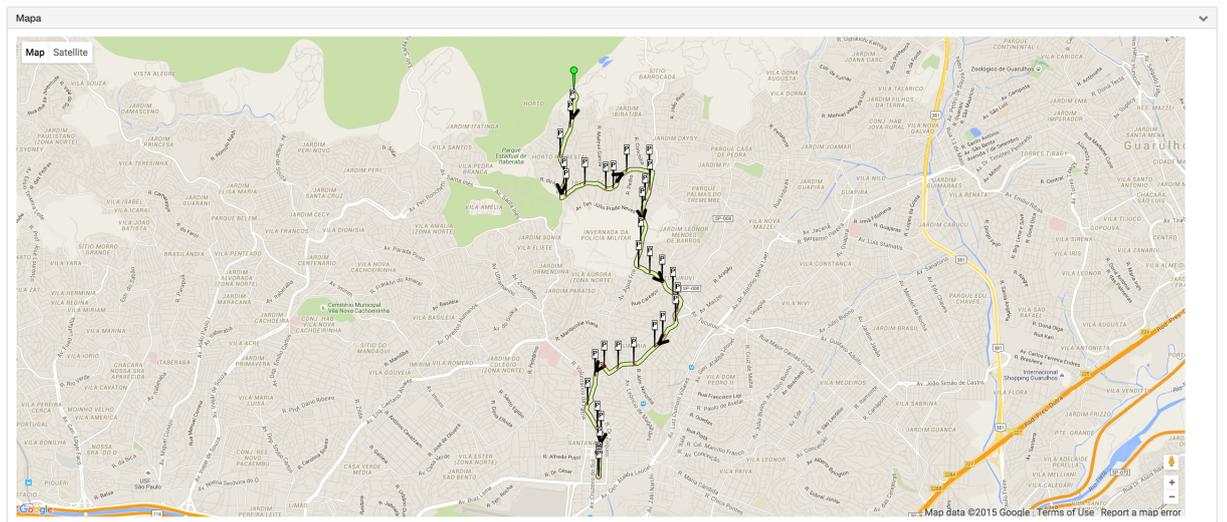


Figura 6.6: Visualização parcial do GTFS mostrando a linha 1018-10-0 Metrô Santana.

6.4 Resultados da análise da qualidade de dados do GTFS

Para cada métrica analisada, detalharemos os principais resultados para a sequência do projeto. Dependendo da métrica, destacaremos as linhas que possuem uma característica importante para a análise das partidas e frequências ou as cinco linhas com maiores e menores valores.

6.4.1 Número de paradas distantes mais de 500m

- Maiores : Linhas com o maior número de paradas distantes mais de 500m são:
 - 3310-10 - Cidade Tiradentes (circular), 56 paradas
 - 6913-21 - Itaim Bibi (circular), 42 paradas
 - 393H-10 - Jd. Sto. André, 33 paradas
 - 7903-10 - Pça. Ramos de Azevedo, 33 paradas
 - 8013-10 - Lapa, 33 paradas

- Menores : Linhas com menor número de paradas distantes mais de 500m são:
 - 372F-10 - Metrô Bresser, zero paradas
 - 1741-10 - Vila Dionísia, zero paradas
 - 1756-10 - Metrô Santana, zero paradas
 - 178Y-10 - Vila Amélia, zero paradas
 - 178Y-21 - Conj. dos Bancários, zero paradas
 - e mais 16 linhas com zero paradas distantes mais de 500m uma da outra.

6.4.2 Número de paradas distantes menos de 30m

- Maiores : Linhas com o maior número de paradas distantes menos de 30m são:
 - 273R-11 - Jd. Robru, 28 paradas
 - 1177-51 - Term. A. E. Carvalho, 27 paradas

- 2756-10 - Guaianazes, 24 paradas
 - 2435-51 - Term. A. E. Carvalho, 24 paradas
 - 3310-10 - Cidade Tiradentes (circular), 22 paradas
- Menores : Linhas com o menor número de paradas distantes menos de 30m são:
 - 372F-10 - Metrô Bresser, zero paradas
 - 178Y-21 - Conj. Dos Bancarios, zero paradas
 - 178Y-10 - Vila Amelia, zero paradas
 - 2709-10 - Metrô Patriarca, zero paradas
 - 2716-10 - Metrô Penha, zero paradas
 - e mais 374 linhas com zero paradas a menos de 30m uma da outra

6.4.3 Número de partidas diárias previstas

- Maiores : Linhas com o maior número de partidas previstas por dia são:
 - 5031-10 Vila Arapuã, 474 partidas
 - 5031-10 Term. Sacomã, 461 partidas
 - 5105-10 Term. Sacomã, 362 partidas
 - 2202-10 Term. Mercado, 359 partidas
 - 2728-10 Metrô Penha, 319 partidas
- Menores : Linhas com o menor número de partidas previstas por dia são:
 - 6000-23 Ama, 1 partida
 - 6000-24 Term. Santo Amaro, 1 partida
 - 6013-21 Jd. Angela, 1 partida
 - 9008-21 T.t.v.n.cachoeirinha, 1 partida
 - 6000-21 Cidade Dutra, 1 partida
 - e mais 16 outras linhas com apenas 1 partida

6.4.4 Linha sem frequência de partidas registradas

- A única linha que não tem frequência documentada é a 7040-21 Paraisópolis.

6.4.5 Total de horas de operação

- Maiores : Linhas com maiores faixas horárias de operação são:
 - 8700-10 Terminal Campo Limpo, 24 horas
 - 8700-10 Pça. Ramos de Azevedo, 24 horas
 - 7001-10 Jd. Angela, 24 horas
 - 7001-10 Vila Gilda, 24 horas
 - 6726-10 Terminal Grajaú, 24 horas
 - e mais outras 44 linhas com 24 horas de operação
- Menores : Linhas com menores faixas horárias de operação são:
 - 6000-23 Ama, 1 hora
 - 6000-24 Term. Santo Amaro, 1 hora
 - 6013-21 Jd. Angela, 1 hora
 - 9008-21 T.t.v.n.cachoeirinha, 1 hora
 - 6000-21 Cidade Dutra, 1 hora
 - e outras 16 linhas que funcionam por apenas 1 hora

6.4.6 Quantidade de shapes na linha

- Maiores : Linhas com mais shapes em sua geometria são:
 - 3310-10 Cidade Tiradentes, 2149 shapes
 - 393H-10 Jd. Sto André, 1652 shapes
 - 393C-10 Cohab II (circular), 1389 shapes
 - 1783-52 Pça do Correio, 1148 shapes
 - 5194-10 Pq. Ibirapuera, 1111 shapes
- Menores : Linhas com menos shapes em sua geometria são:

- 372F-10 Metrô Bresser, 32 shapes
- 372F-10 Univ. São Judas Tadeu, 33 shapes
- 5020-10 Term. Sacomã, 49 shapes
- 172N-22 Metrô Belém, 54 shapes
- 6813-10 Jd. Lidia, 54 shapes

6.4.7 Distância total

- Maiores : Linhas mais longas da cidade são:

- 3310-10 Cidade Tiradentes, 102 km
- 393H-10 Jd. Sto. André, 70 km
- 393C-10 Cohab II (circular), 67 km
- 8013-10 Lapa, 50 km
- 6913-21 Itaim Bibi (circular), 49 km

- Menores : Linhas mais curtas da cidade são:

- 372F-10 Univ. São Judas Tadeu, 1 km
- 372F-10 Metrô Bresser, 1 km
- 5020-10 Term. Sacomã, 2 km
- 6813-10 Jd. Lidia, 2 km
- 5013-21 Vila Cde. do Pinhal, 2 km
- e outras 6 linhas com 2 km de extensão

6.4.8 Velocidade média prevista

- Maiores : Linhas com maiores velocidades médias planejadas são:

- 393C-10 Cohab II (circular), 53,4 km/h
- 6L01-22 Terminal Varginha, 42,3 km/h
- 3746-31 Jd. Imperador, 38,3 km/h
- 6L01-23 Terminal Varginha, 37,9 km/h
- 8050-32 Morro Doce, 36,3 km/h

- Menores : Linhas com menores velocidades médias planejadas são:

- 172N-22 Metrô Belém, 4,3 km/h
- 172N-21 Metrô Belém, 4,5 km/h
- 6806-10 Term. João Dias, 5 km/h
- 6805-10 Term. João Dias, 5,1 km/h
- 1721-21 Metrô Carandiru, 5,3 km/h

6.4.9 Quantidade de paradas duplicadas

- Maiores : Linhas com mais paradas duplicadas são:

- 7082-10 Jd. Horizonte Azul (circular), 15 paradas de 90 no total
- 3310-10 Cidade Tiradentes, 11 paradas de 283 no total
- 273R-11 Jd. Robru, 9 paradas de 84 no total
- 393H-10 Jd. Sto. André, 8 paradas de 218 no total
- 6829-10 Jd. Vaz de Lima, 8 paradas de 66 no total
- e mais 2 linhas com 8 paradas duplicadas

- Menores : Linhas com menos paradas duplicadas são:

- 172N-22 Metrô Belém, zero paradas
- 172N-21 Metrô Belém, zero paradas
- 6806-10 Term. João Dias, zero paradas
- 6085-10 Term. João Dias, zero paradas
- 1721-21 Metrô Carandiru, zero paradas
- e mais 1991 linhas sem paradas duplicadas

6.4.10 Quantidade de paradas únicas

- Maiores : Linhas com a maior quantidade de paradas são:

- 3310-10 Cidade Tiradentes (circular), 272 paradas
- 393H-10 Jd. Sto. André, 210 paradas
- 393C-10 Cohab II (circular), 186 paradas
- 5106-10 Lgo. São Francisco (circular), 140 paradas
- 1783-52 Pca. do Correio, 134 paradas

- Menores : Linhas com a menor quantidade de paradas são:

- 6000-23 Ama, 2 paradas

- 6000-24 Term. Santo Amaro, 2 paradas
- 372F-10 Univ. São Judas Tadeu, 4 paradas
- 6000-21 Cidade Dutra, 4 paradas
- 6000-23 Term. Parelheiros, 4 paradas
- e mais 1 linha com 4 paradas

6.4.11 Linha que passa duas vezes pela mesma parada

- 185 linhas passam duas vezes por pelo menos uma parada do trajeto.

6.4.12 Paradas em comum na ida e na volta

- Maiores : Linhas com o maior número de paradas em comum na ida e na volta são:
 - 2756-10 Metrô Patriarca/Guaianazes, 24 paradas em comum
 - 273R-10 Metrô Artur Alvim/Jd. Robru, 16 paradas em comum
 - 2726-10 Metrô Penha/Limoeiro, 15 paradas em comum
 - 2722-10 Metrô Guilhermina/Jd. Veronia, 12 paradas em comum
 - 9090-10 Term. Casa Verde/Pq. Tietê, 11 paradas em comum
 - e mais 4 linhas com 11 paradas em comum
- Menores : Linhas com o menor número de paradas em comum na ida e na volta são:
 - 572 linhas não compartilham pontos entre a ida e a volta.

6.4.13 Tempo previsto de viagem

- Maiores : As viagens mais demoradas pelo tempo planejado são:
 - 7282-10 Pça Ramos de Azevedo, 3 h e 59 min
 - 5106-10 Lgo. São Francisco (circular), 3 h e 40 min
 - 502J-10 Metrô Santa Cruz (circular), 3 h e 25 min
 - 3310-10 Cidade Tiradentes (circular), 3 h e 25 min
 - 5106-21 Shopping Ibirapuera (circular), 3 h e 12 min
- Menores : As viagens mais rápidas pelo tempo planejado são:
 - 736I-21 Metrô Vila das Belezas, 15 min
 - 6L01-22 Terminal Varginha, 15 min
 - 6L01-23 Terminal Varginha, 15 min
 - 2703-21 Metrô Itaquera, 15 min
 - 2703-22 Metrô Itaquera, 15 min
 - e mais 28 linhas com tempo de viagem planejado de 15 min.

6.4.14 Frequência média por hora

- Maiores : Linhas com maior número de partidas previstas por hora em média são:
 - 5031-10 Vila Arapuã, 21 partidas por hora em média
 - 5031-10 Term. Sacomã, 20 partidas por hora em média
 - 5105-10 Term. Mercado, 17 partidas por hora em média
 - 5105-10 Term. Sacomã, 17 partidas por hora em média
 - 2728-10 Metrô Penha, 15 partidas por hora em média
- Menores : Linhas com menor número de partidas previstas por hora em média são:
 - 6L01-23 Terminal Varginha, 1 partida por hora em média
 - 8015-21 Estação de Perus, 1 partida por hora em média
 - 8015-21 Cemitério de Perus, 1 partida por hora em média
 - 6000-23 Term. Parelheiros, 1 partida por hora em média
 - 6000-21 Cidade Dutra, 1 partida por hora em média
 - e mais 220 linhas com 1 partida por hora em média

6.4.15 Intervalo médio entre ônibus

- Maiores : Linhas com maior intervalo médio planejado entre partidas de ônibus são:
 - 3310-10 Cidade Tiradentes, 1 hora
 - 5612-10 Parque Ibirapuera (circular), 1 hora
 - 6040-10 Itaim Bibi, 1 hora
 - 6091-21 Terminal Santo Amaro, 1 hora
 - 393H-10 Jd. Sto André, 1 hora
 - e mais 220 linhas com intervalo médio de 1 hora
- Menores : Linhas com menores intervalo médio planejado entre partidas são:
 - 5031-10 Vila Arapua, 2 minutos
 - 5105-10 Term. Sacomã, 3 minutos
 - 5105-10 Term. Mercado, 3 minutos
 - 5031-10 Term. Sacomã, 3 minutos
 - 8000-10 Pça. Ramos de Azevedo, 4 minutos
 - e mais 13 linhas com intervalo médio de 4 minutos

6.4.16 Porcentagem de arestas com baixa confiabilidade para cálculo da velocidade média

- Maiores : Linhas com maior quantidade de arestas não confiáveis para cálculo da velocidade média são:
 - 6L10-41 Messiânica, 94% das arestas
 - 6072-23 Balneário São José (circular), 88% das arestas
 - 6L10-41 Terminal Varginha, 81% das arestas
 - 6L11-10 Ilha do Bororé, 68% das arestas
 - 7085-10 Jd. Universal (circular), 67% das arestas
- Menores : Linhas com menor quantidade de arestas não confiáveis para cálculo da velocidade média são:
 - 1016-10 Jd. São Bernardo, 0% das arestas
 - 1016-10 Cemitério do Horto, 0% das arestas
 - 1018-10 Metrô Santana, 0% das arestas
 - 1018-10 Vila Rosa, 0% das arestas
 - 1156-10 Pça. do Correio, 0% das arestas
 - e mais 720 linhas com 0% das arestas

6.4.17 Porcentagem de arestas com velocidades inconsistentes

- Maiores: Linhas com maior porcentagem de arestas onde a velocidade planejada é incompatível com a velocidade de um ônibus são:
 - 4006-10 Circular, 42,9% das arestas
 - 273R-10 Jd. Robru, 29% das arestas
 - 737G-10 Term. Guarapiranga, 28,6% das arestas
 - 3064-41 Santa Etelvina II B6, 25% das arestas
 - 6000-23 Term. Parelheiros, 25% das arestas
 - e mais 1 linha com 25% das arestas com velocidades incompatíveis à velocidade de um ônibus
- Menores : Linhas com menor porcentagem de arestas onde a velocidade planejada é incompatível com a velocidade de um ônibus são:
 - 7053-10 Jd. Macedônia, 0% das arestas
 - 7053-21 Jd. Maria Sampaio, 0% das arestas
 - 7053-22 Jd. Maria Sampaio, 0% das arestas

- 3574-41 E. T. Itaquera, 0% das arestas
- 6031-10 Terminal Grajaú, 0% das arestas
- e mais 677 linhas com 0% de arestas com velocidades incompatíveis

6.4.18 Porcentagem de arestas compartilhadas

- Maiores: Linhas com maior porcentagem de arestas onde compartilha mais arestas com outras linhas são:
 - 1178-31-0 - Pca. Do Correio, 100% das arestas
 - 118C-22-0 - Jd.Pery Alto, 100% das arestas
 - 118C-24-0 - Jd. Pery Alto, 100% das arestas
 - 1741-21-0 - Metrô Santana, 100% das arestas
 - 1745-10-1 - Vl.nova Cachoeirinha, 100% das arestas
 - e mais 99 linhas com 100% das arestas compartilhadas
- Menores : Linhas com menor porcentagem de arestas onde compartilha mais arestas com outras linhas são:
 - 372F-10-0 - Metrô Bresser, 0% das arestas
 - 1783-51-0 - Cachoeira-dib, 0% das arestas
 - 6041-21-0 - Jd. Monte Azul- Circular, 0% das arestas
 - 6L10-41-1 - Messiânica, 5% das arestas
 - 4724-10-0 - V. Nhocune, 5% das arestas

6.4.19 Maior quantidade de linhas compartilhadas por aresta

- Maiores: Linhas com maior quantidade de linhas compartilhadas são:
 - 1178-10-1 - Sao Miguel, 34 linhas
 - 2006-51-0 - Cid. Kemel (circular), 34 linhas
 - 2016-10-1 - Jd. Das Oliveiras, 34 linhas
 - 2017-10-0 - Sao Miguel (circular), 34 linhas
 - 2022-10-1 - Jd. Dos Ipes, 34 linhas
 - e mais 29 linhas que tem como compartilhamento máximo do seu trajeto com 34 linhas.
- Menores : Linhas com menor porcentagem de arestas onde compartilha mais arestas com outras linhas são:
 - 372F-10-0 - Metrô Bresser, 0 linhas

- 1783-51-0 - Cachoeira-dib, 0 linhas
- 6041-21-0 - Jd. Monte Azul- Circular, 0 linhas
- 6L08-10-1 - Jd. São Nicolau, 2 linhas
- 3725-41-0 - Vila Aricanduva, 2 linhas
- e mais 9 linhas com 0

6.4.20 Quantidade de conexões de uma linha

- Maiores: Linhas com maiores número de conexão são:
 - 3310-10-0 Cidade Tiradentes, 360 conexões
 - 393C-10-0 - Cohab II, 328 conexões
 - 393H-10 - Jd. Sto André , 309 conexões
 - 701U-10-1 - Metrô Santana, 247 conexões
 - 1767-51-0 -Term. Bandeira,245 conexões
- Menores : Linhas que possui menores números de conexões são:
 - 5105-10-1 Terminal Sacomã, 3 conexões
 - 3734-41-0 - Vila Aricanduva, 3 conexões
 - 8015-21-0 - Estação De Perus, 4 conexões
 - 1783-51-0 - Cachoeira-dib, 4 conexões
 - 5031-21-1 - Vila Cde. Do Pinhal, 5 conexões

Capítulo 7

Visualização dos dados do AVL

Nesta seção mostramos o trabalho feito para a visualização do AVL.

7.1 Visão Geral

No AVL (*Automatic Vehicle Location*) temos informações obtidas a partir das viagens efetivamente realizadas pelos ônibus.

Os dados de posicionamento dos ônibus são relevantes para a análise principalmente quando considerada a frequência que os mesmos são coletados e enviados para o servidor. Nos arquivos históricos considerados, notamos que esta frequência de atualização começou em 90 segundos e diminuiu para aproximadamente 45 segundos gerando, portanto, dados mais precisos.

Conforme vimos no capítulo 4 de banco de dados temos que o AVL é formado por três arquivos:

- AL_[data].txt: contém informações das linhas sendo monitoradas,
- AV_[data].txt: contém informações sobre o veículo,
- MO_[data].txt: contém dados sobre o movimento dos veículos.

Ainda conforme o capítulo 4, temos que estes arquivos foram organizados e estruturados na coleção *logs_avls* no MongoDB.

Um arquivo de AVL começa a coletar a posição dos ônibus às 4:00h do dia em questão e se encerra às 4:00h da manhã do dia seguinte. Esta organização privilegia o estudo dos ônibus diurnos, mas acaba dificultando o estudo dos ônibus da madrugada que podem ter as suas viagens começando às 3:00h de um dia e se encerrando depois das 4:00h da manhã. A figura 7.1 ilustra este problema. Com isto, as viagens da madrugada, com início próximo das

04:00 h poderão ter os dados truncados (divididos em 2 arquivos).

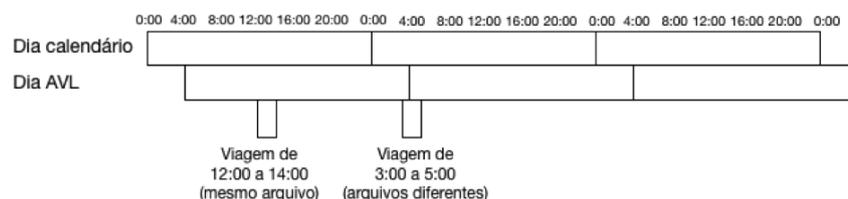


Figura 7.1: Exemplo de arquivo AVL com dados coletados ao longo do dia.

O equipamento de AVL é basicamente formado por:

- Um GPS que envia a latitude e longitude do local onde se encontra o equipamento;
- Uma interface de conexão à rede de telefonia celular para envio dos dados;
- Uma memória finita para armazenamento temporário dos dados antes do envio. Cada um destes componentes é sensível a falhas, podendo portanto enviar, ou não, dados que podem comprometer a qualidade da informação sobre a localização do ônibus em um dado instante.

Para que o AVL consiga emitir corretamente a sua posição, é preciso que o sensor de GPS esteja visível aos satélites. Desta forma o GPS funciona muito bem em áreas abertas e planas mas não tão bem em áreas com grande densidade de prédios altos. Por isso, o GPS tem dificuldade de funcionar em algumas partes da cidade passando a reportar valores imprecisos do seu posicionamento. A região da Paulista é um exemplo deste caso, onde é difícil o monitoramento através do GPS.

7.2 Visualização do AVL

Nesta parte do projeto criamos a visualização do AVL. A finalidade desta visualização foi testar as métricas para verificar se havia algum problema no algoritmo de análise ou nos dados enviados pelos AVLS.

Foram criadas as seguintes funcionalidades :

- Busca por identificador da linha;

- Busca por identificador do AVL;
- Gráfico das distâncias percorridas pelos ônibus no tempo;
- Mapa mostrando o posicionamento dos pontos emitidos pelo AVL num determinado intervalo de tempo;
- Tabela com os indicadores das viagens.

Tais funcionalidades podem ser visualizadas na figura 7.2.

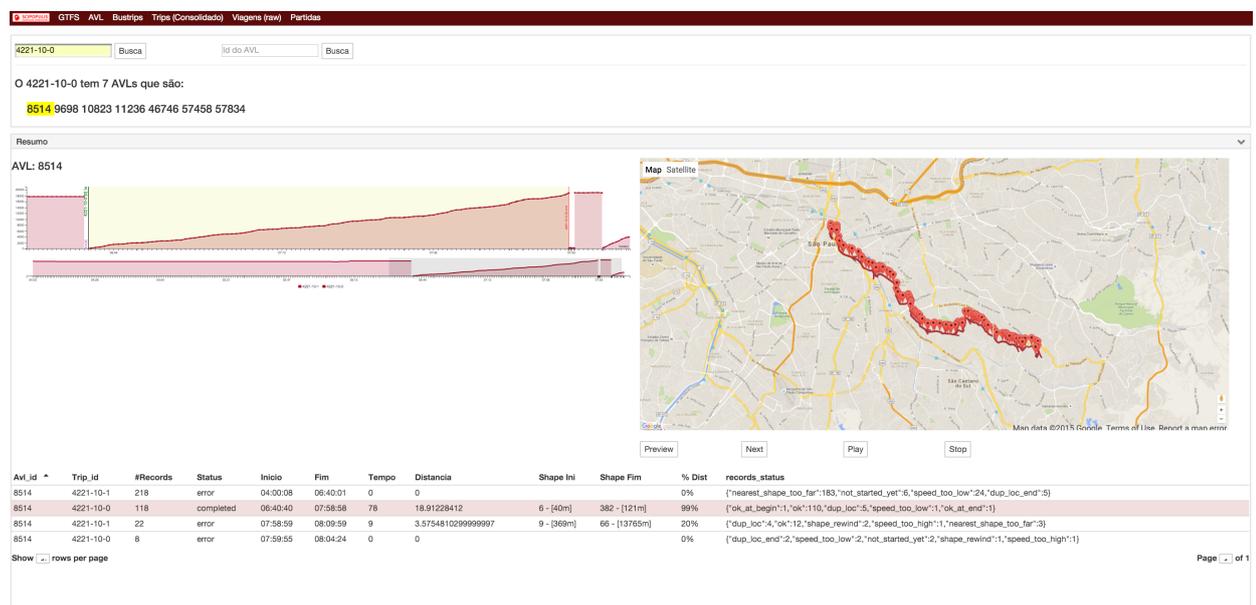


Figura 7.2: Ilustra uma viagem feita pelo AVL 8514 na linha 4221-10-0.

Para melhor entender o funcionamento desta visualização criamos um diagrama com os seus estados. Vide a figura 7.3.

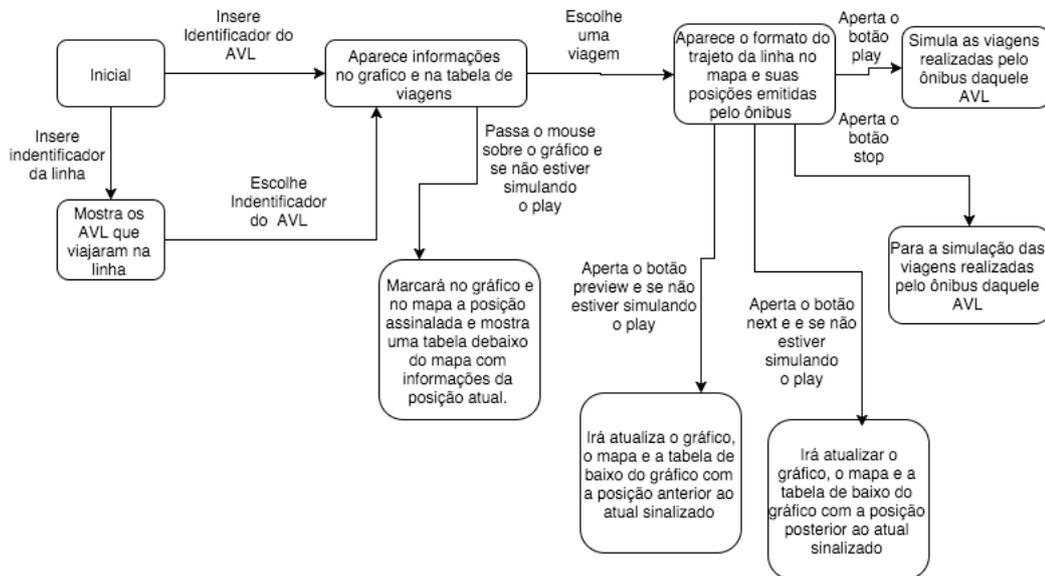


Figura 7.3: Diagrama de funcionamento da visualização do gráfico no mapa.

A partir do ponto inicial do diagrama, pode-se inserir o identificador do AVL ou o identificador da linha (neste caso é necessário, depois, selecionar o identificador do AVL) para que as informações referentes a viagem do AVL sejam mostradas no gráfico e na tabela. Ao selecionar uma linha desta tabela, irá aparecer no mapa identificadores mostrando todas as localizações que foram registradas enquanto o ônibus estava fazendo a viagem selecionada (figura 7.3).

Uma parte muito importante da visualização é a interação entre o mapa e o gráfico. Ao passarmos o mouse no gráfico, um ponto irá aparecer no mapa mostrando exatamente a localização do ônibus naquele momento. Isto só ocorrerá se o botão play não estiver rodando. Se o botão play estiver rodando, este irá simular o caminho feito pelo ônibus. Durante esta simulação só é permitido apertar o botão stop. Se o botão play não estiver rodando, pode-se realizar todas as operações relatadas anteriormente e também utilizar os botões preview e next.

7.2.1 Visualização de viagens reportadas pelo AVL

A figura 7.4 mostra um exemplo de viagem realizada dentro do padrão esperado, pois no gráfico da distância pelo tempo ela apresenta uma linha contínua e crescente.

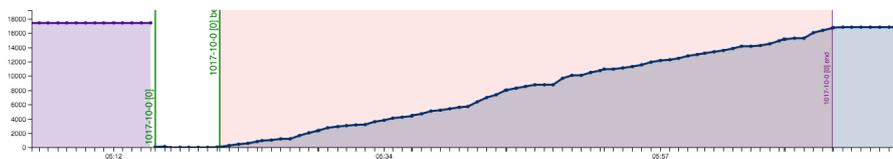


Figura 7.4: Viagem realizada dentro do padrão esperado.

Mas também podem haver problemas nos dados do AVL. Abaixo temos a figura 7.5 que demonstra esta situação com os seguintes problemas :

1. Interrupção da posição do AVL;
2. Trajeto não planejado para a linha especificada pelo identificador do AVL.

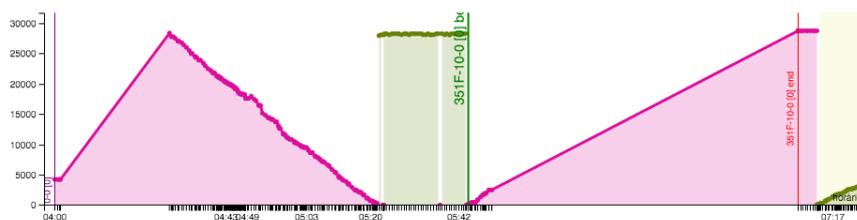


Figura 7.5: Viagem feita pelo AVL com problemas.

O problema (1) pode ser notado na figura 7.5 durante os períodos crescentes das linhas (entre aproximadamente 04:00 h e 04:30h e das 05:45 h às 07:00 h).

Já o problema (2) pode ser notado durante o período decrescente da linha (entre aproximadamente 04:30 h e 05:20 h).

Capítulo 8

Conclusão

Neste capítulo mostraremos as conclusões alcançadas neste projeto na análise do GTFS e visualização do AVL.

O objetivo deste trabalho foi explorar e analisar dados do sistema de transporte de ônibus urbano da cidade de São Paulo para descobrir e visualizar informações relevantes e a qualidade destas informações no período da semana de 18/10/2014 a 24/10/2014.

A exploração e análise dos dados foram feitas principalmente através do Node.js, MongoDB e bibliotecas JavaScript. Para a visualização de dados foi utilizado primariamente o Meteor e bibliotecas JavaScript.

Ao analisarmos os dados da última semana de Outubro de 2014, descobrimos fatos relevantes referentes ao sistema de ônibus na cidade de São Paulo, com destaques para :

- 2.204 linhas de ônibus;
- 9.988.278 km planejados para circulação da frota de ônibus diariamente (dias úteis);
- 18.733 pontos de ônibus;
- 14.663 ônibus ativos;

Constatamos alguns problemas nos dados do GTFS que de maneira geral não comprometem o seu uso, dos quais citamos :

- Linha sem registro de frequência de saída (linha 7040-21 Paraisópolis);
- Todas as linhas com tempo previsto de viagem constante, independentemente do horário (desconsidera possível congestionamento em horário de pico);

- Todas as velocidades média por aresta com valores que não consideram uma distância variável entre arestas (no GTFS este valor é calculado através de uma divisão simples do tempo total da viagem pela quantidade de paradas da linha);

Notamos também algumas características peculiares de algumas linhas, por exemplo :

- A linha mais longa da cidade tem 102 Km (3310-10 Cidade Tiradentes);
- A linha com a maior velocidade média planejada é de 53,4 km/h (393C-10 Cohab II);
- A maior parte das linhas mais extensas, demoradas, com pontos duplicados e com maior quantidade de pontos são linhas circulares;
- 15 linhas com dados insuficientes para o cálculo da velocidade média da aresta porque possuem mais de 50% das arestas onde a frequência de todas as linhas que passam pela aresta é inferior a 1 ônibus a cada 10 minutos;

Através da visualização do AVL foi possível identificar de maneira prática, rápida e simples as linhas com comportamento padrão (normal) e aquelas cujos dados do AVL apresentavam potenciais problemas tais como interrupção na transmissão da localização do ônibus ou trajeto não planejado para a linha especificada pelo identificador do AVL.

Com isso, este trabalho nos permite concluir que os dados históricos fornecidos pela SPTrans com relação ao GTFS e AVL são confiáveis e permite uma análise detalhada e crítica da qualidade do sistema de transporte público de ônibus da cidade de São Paulo.

Parte II

Parte Subjetiva

Capítulo 9

Apreciação e Críticas Pessoais para o Projeto

Este capítulo contém as razões e motivações que me levaram à escolha e realização deste trabalho, desafios, matérias que me ajudaram, agradecimentos.

9.1 Motivação Pessoal

A idéia de realizar um projeto na área de transporte público surgiu devido a dois fatores:

- Meu intercâmbio pelo Ciências Sem Fronteiras para a Karlsruhe Institute of Technology (na Alemanha);
- As manifestações do Brasil em 2013;

Na Alemanha, convivendo o dia-a-dia com uma organizada, eficiente, pontual e ampla malha de transporte de ônibus, percebi o nível de deficiência que nos encontramos no Brasil, e o quanto podemos melhorar principalmente na cidade de São Paulo.

Já nas manifestações de 2013, um dos primeiros alicerces foi a discussão sobre a tarifa de ônibus e conseqüentemente a necessidade de um transporte público de melhor qualidade. Com isto percebi que um trabalho nesta área seria desafiador e poderia ajudar em uma área crítica, polêmica e importante nos dias de hoje. Isto levou-me a pesquisar sobre o transporte público de ônibus e motivou-me a desenvolver este trabalho de conclusão de curso.

9.2 Desafios

Este projeto foi o meu primeiro grande trabalho e tive que interagir com profissionais e empresas fora do círculo acadêmico. Isto já foi algo novo e desafiador uma vez que tive que buscar apoio, contatos e formas de viabilizar tais objetivos. Em adição houveram também desafios técnicos que tiveram que ser superados para alcançar os resultados esperados.

Um dos desafios foi a obtenção dos dados históricos do GTFS e do AVL. Apesar da SPTrans disponibilizar o acesso ao GTFS, somente se tem acesso aos dados atuais. Já para o AVL, é apenas disponibilizada uma versão parcial, filtrada, e não aquela que contém os dados “puros” das quais necessitávamos para este projeto. Com isto, para a realização das análises tivemos que esperar pela concessão dos dados históricos e complementares.

Outro desafio, foi a enorme quantidade de dados. Como exemplo podemos citar os dados do GTFS de aproximadamente 500 MB a cada geração, e os dados do AVL que são de aproximadamente 2 GB por dia. Para reduzirmos um pouco a necessidade de memória consumida na leitura destes arquivos pelo sistemas utilizamos o MongoDB 3.0 [29].

9.3 Matérias que me ajudaram

As matérias que me ajudaram na realização deste projeto foram :

MAE0121 Introdução a Probabilidade e à Estatística I

Esta matéria me ajudou no cálculo de métricas importantes para o projeto, tais como a média, mediana e desvios padrões.

MAC0323 Estrutura de Dados

Esta matéria me ajudou a melhor compreender as estruturas de dados dos algoritmos.

Princípios de Interação Humano-computador (feita na Alemanha)

Esta matéria me ajudou na criação da visualização de dados do GTFS e do AVL.

MAC0426 Sistemas de Bancos de Dados e MAC0439 Laboratório de Bancos de Dados

Estas matérias me ajudaram a entender as estruturas de banco de dados, sua organização e como realizar as *queries*.

MAC0342 Laboratório de Programação Extrema e MAC0332 Engenharia de Software

Estas matérias me ajudaram compreender as práticas de engenharia de software, como trabalhar em grupo e métricas para avaliar o código.

9.4 Agradecimentos

Gostaria de agradecer ao professor Roberto Hirata Junior ; aos membros da Scipopulis : Roberto Speicys, Julian Monteiro, Ivo Pons e Renato Arbex e a SPTrans pelos dados fornecidos. Agradeço-lhes por todo suporte fornecido, pela orientação e por tornarem este trabalho possível.

Também gostaria de agradecer à minha família e amigos por todo apoio durante estes anos de graduação.

Lista de Figuras

1.1	Número de linhas nas cidades.	7
1.2	Número de ônibus em circulação.	8
1.3	Número de passageiros que utilizam ônibus por ano.	8
3.1	Exemplo de Paradas (círculos coloridos), Arestas (trajetória entre duas paradas), Shape (círculos azuis), Segmento (retas identificadas de A a K)	14
3.2	Arquitetura do sistema proposto de análise dos dados do transporte de ônibus da cidade de São Paulo.	16
4.1	Organização prévia dos arquivos GTFS.	18
5.1	Histograma da quantidade de paradas por linha.	22
5.2	Histograma do comprimento de linhas.	23
5.3	Histograma do comprimentos de segmentos.	24
5.4	Histograma do comprimentos de arestas.	25
5.5	Histograma da quantidade de linhas que passam por cada aresta.	26
5.6	Histograma da quantidade de linhas que passam por um ponto de ônibus.	27
5.7	Porcentagem de linhas circulares.	28
5.8	Representação de 50 linhas e onde elas compartilham arestas.	29
6.1	Linha 2755-10 Metrô Penha - Ponto Duplicado.	36
6.2	Linha: 2756-10-1 - Guaianazes (lilas) e 2756-10-1 - Metrô Patriarca (verde).	37
6.3	Aresta da Marechal Tito.	38
6.4	Conexões com a linha 702U-10 - Term.Pq. Dom Pedro II na região da Paulista.	39
6.5	Visualização parcial do GTFS.	40
6.6	Visualização parcial do GTFS mostrando a linha 1018-10-0 Metrô Santana.	41
7.1	Exemplo de arquivo AVL com dados coletados ao longo do dia.	53
7.2	Ilustra uma viagem feita pelo AVL 8514 na linha 4221-10-0.	54

7.3	Diagrama de funcionamento da visualização do gráfico no mapa.	55
7.4	Viagem realizada dentro do padrão esperado.	56
7.5	Viagem feita pelo AVL com problemas.	56

Lista de Tabelas

5.1	Quantidade de paradas por linha.	22
5.2	Comprimentos de linhas.	23
5.3	Comprimento de segmentos.	24
5.4	Comprimento de arestas.	25
5.5	Quantidade de linhas que passam por cada aresta.	26
5.6	Quantidade de linhas que passam por um ponto de ônibus.	27
6.1	Tabela das Métricas do GTFS.	31

Bibliografia

- [1] Definição de OpenData
http://opendatahandbook.org/guide/pt_BR/what-is-open-data/
- [2] Motivos para utilizar OpenData.
<http://opendatahandbook.org/guide/en/why-open-data>
- [3] GTFS Mexico:
<http://transitfeeds.com/p/mexico-city-federal-district-government/70>
- [4] Pesquisa da Cidade do México na area de área de transporte público:
<http://siteresources.worldbank.org/INTURBANTRANSPORT/Resources/340136-1395424136020/GTFS-Mexico.pdf>
- [5] GTFS Santiago do Chile:
<http://datos.gob.cl/dataset/1587>
- [6] Pesquisa da cidade de Santiago do Chile área de transporte público:
<https://goo.gl/DuH08Q>
- [7] Iniciativa em Nova Iorque:
<https://nycopendata.socrata.com/>
- [8] Iniciativa em Londres:
<http://www.parliament.uk/briefing-papers/POST-PN-472.pdf>
- [9] São Paulo Dados Abertos
<http://www.sptrans.com.br/desenvolvedores/>
- [10] Lista de cidades com dados transporte publico abertos
<https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds>
- [11] GoogleTransit
<https://developers.google.com/transit/gtfs/?hl=en>

-
- [12] Origem do GTFS: <http://beyondtransparency.org/chapters/part-2/pioneering-open-data-standards-the-gtfs-story/>
- [13] AVL
<http://www.itre.ncsu.edu/pupil/STG/documents/news/GIS-GPS-AVL.pdf>
- [14] Hackathon SpTrans: Começo dos dados abertos de São Paulo [2013]
<http://www.saopauloaberta.com.br/hackatona-do-onibus/>
- [15] Indicadores da SPTrans
<http://www.sptrans.com.br/indicadores/>
- [16] Indicadores de Londres
<https://tfl.gov.uk/corporate/about-tfl/what-we-do/buses?intcmp=2587>
- [17] Indicadores de Nova Iorque
<http://web.mta.info/nyct/facts/ffbus.htm>
- [18] Indicadores de Paris
http://www.ratp.fr/en/ratp/c_5044/bus/
- [19] Indicadores de Pequim
https://en.wikipedia.org/wiki/Beijing_Bus
- [20] Node.js
<https://nodejs.org/en/>
- [21] MongoDB
<https://docs.mongodb.org/manual/reference/database-references/>
- [22] MeteorJS
<https://www.meteor.com/>
- [23] D3.JS
<http://d3js.org/>
- [24] C3.js
<http://c3js.org/>
- [25] Alchemist.Js
<http://graphalchemist.github.io/Alchemy/>
- [26] Google Maps Js
<https://developers.google.com/maps/documentation/javascript/>

- [27] Underscore.JS
<http://underscorejs.org/>
- [28] Fast-CSV
<https://github.com/C2F0/fast-csv>
- [29] MongoDB 3.0
<https://www.mongodb.com/blog/post/new-compression-options-mongodb-30>