

Universidade de São Paulo
Instituto de Matemática e Estatística
Bacharelado em Matemática Aplicada e Computacional

Danilo Santana e Silva

Modelagem do Transporte Urbano Através de Grafos

São Paulo - SP

25 de setembro de 2018

Danilo Santana e Silva

Modelagem do Transporte Urbano Através de Grafos

Esta monografia tem por objetivo modelar, através de grafos não orientados, as viagens do transporte público de ônibus da cidade de São Paulo.

Universidade de São Paulo – USP
Instituto de Matemática e Estatística
Matemática Aplicada

Orientadora: Joyce da Silva Bevilacqua

São Paulo - SP
25 de setembro de 2018

Danilo Santana e Silva
Modelagem do Transporte Urbano Através de Grafos/ Danilo Santana e Silva.
– São Paulo - SP, 25 de setembro de 2018-
Orientadora: Joyce da Silva Bevilacqua

Monografia – Universidade de São Paulo – USP
Instituto de Matemática e Estatística
Matemática Aplicada, 25 de setembro de 2018.

1. Ônibus. 2. Transporte Público. 3. São Paulo. I. Dados abertos. II. Universidade de São Paulo. III. Instituto de Matemática e Estatística. IV. Transporte público na Cidade de São Paulo
CDU 00:000:000.0

Danilo Santana e Silva

Modelagem do Transporte Urbano Através de Grafos

Esta monografia tem por objetivo modelar, através de grafos não orientados, as viagens do transporte público de ônibus da cidade de São Paulo.

São Paulo - SP, 25 de setembro de 2018.

Joyce da Silva Bevilacqua
Orientadora

São Paulo - SP
25 de setembro de 2018

*Este trabalho é dedicado à cidade de São Paulo,
que possa ser um instrumento útil na gestão de recursos públicos.*

Agradecimentos

Meus agradecimentos são direcionados majoritariamente à minha família pelo apoio constante nestes anos de graduação. Agradecimentos especiais à minha noiva Luana Cherri de Oliveira, que motivou-me durante todo o período de graduação, com amor e compreensão nas dificuldades enfrentadas.

Agradeço também ao SENAI - Serviço de Aprendizagem Industrial pelo apoio na gestão de horas para viabilizar o desenvolvimento deste projeto e aos mantenedores dos canais de acesso aos dados públicos da SPTRANS - São Paulo Transportes S/A.



Resumo

O sistema público de transporte desloca milhões de pessoas diariamente. Em São Paulo, já foram iniciadas diversas medidas para que haja uma gestão de qualidade. Hoje se adotam diferentes tecnologias, inclusive de georreferenciamento e pagamentos com bilhetes eletrônicos, informações capazes de mapear de forma inteligente todo o fluxo de passageiros da cidade.

Em tempos de transparência e acesso à informação, os cidadãos são estimulados a consultar e buscar novas soluções para suas cidades, auditando seus gestores com base em dados abertos. Neste sentido, este trabalho apresenta um procedimento para geração de grafos não orientados dos fluxos de passageiros com base em registros eletrônicos de pagamento. Apesar de muito utilizados no estudo de transportes, hoje nenhuma companhia pública da cidade disponibiliza o acesso à informação neste formato.

Neste trabalho, coletaram-se informações do transporte público de ônibus já disponibilizadas. Destes dados identificaram-se os maiores fluxos do serviço com base no registros de pagamento eletrônico, identificando as rotas de ônibus mais utilizadas em uma dada faixa horária. Itinerários das linhas identificadas foram mapeadas e sobrepostas à mapas de regiões político-administrativas e modeladas em grafos.

Para fins de exemplificação de aplicabilidade dos uso do método, estimaram-se quais são as principais demandas de transporte em diferentes regiões da cidade.

Palavras-chaves: ônibus. transporte público. grafos. dados abertos. São Paulo.

Abstract

The public transport system carries millions of people daily. In São Paulo, several actions have already been taken to ensure quality management. Today different technologies are adopted, including georeferencing and payments with electronic tickets, information capable of intelligently mapping the entire flow of passengers in the city.

In times of transparency and access to information, citizens are encouraged to consult and seek new solutions for their cities, auditing their managers based on open data. In this sense, this paper presents a procedure for the generation of non-oriented graphs of the flows of passengers based on electronic payment records. Although widely used in the study of transport, today no public company in the São Paulo city provides access to information in this format.

In this work, public transport information was collected from already published channels. From these data, we identified the largest flows of the service based on electronic payment records, identifying the most used bus routes in a given hourly range. Itineraries of the identified lines were mapped and overlapped with maps of political-administrative regions and modeled in graphs.

For purposes of exemplifying the applicability of the method's use, it was estimated the main transportation demands in different regions of the city.

Key-words: bus. public transport. graphs. open data. São Paulo.

Lista de ilustrações

Figura 1 – Mapa da densidade populacional do estado de São Paulo.	14
Figura 2 – Boxplot - Manhã (com <i>outliers</i>).	22
Figura 3 – Boxplot - Manhã (sem <i>outliers</i>).	22
Figura 4 – Regiões Administrativas da SPTrans	24
Figura 5 – Regiões Administrativas de São Paulo	24
Figura 6 – Grafo Exemplo	27
Figura 7 – Principais Linhas da Manhã	28
Figura 8 – Principais Linhas da Manhã	28
Figura 9 – Principais Linhas M/T/N	29

Lista de códigos

Código 1 – redução.sh	34
Código 2 – top10.sh	34
Código 3 – contagem-faixas.sh	35
Código 4 – linhas-por-faixa.sh	36

Lista de tabelas

Tabela 1 – Relatório de Passageiros Transportados	18
Tabela 2 – Dados do Bilhete Eletrônico	19
Tabela 3 – Estrutura do Arquivo Referencial	19
Tabela 4 – Faixas Horárias da Manhã	22
Tabela 5 – Principais Linhas da Manhã	23

Lista de símbolos

n	Representa o número de elementos em uma coleção de dados.
Min	Representa o valor mínimo em uma coleção de dados.
q₁	Representa o primeiro quartil em uma coleção de dados.
\tilde{x}	Representa a mediana em uma coleção de dados.
\bar{x}	Representa a média em uma coleção de dados.
q₃	Representa o terceiro quartil em uma coleção de dados.
Max	Representa o valor máximo em uma coleção de dados.
s	Representa o desvio padrão amostral em uma coleção de dados.
IQR	Representa o intervalo entre os quartis de uma coleção de dados.

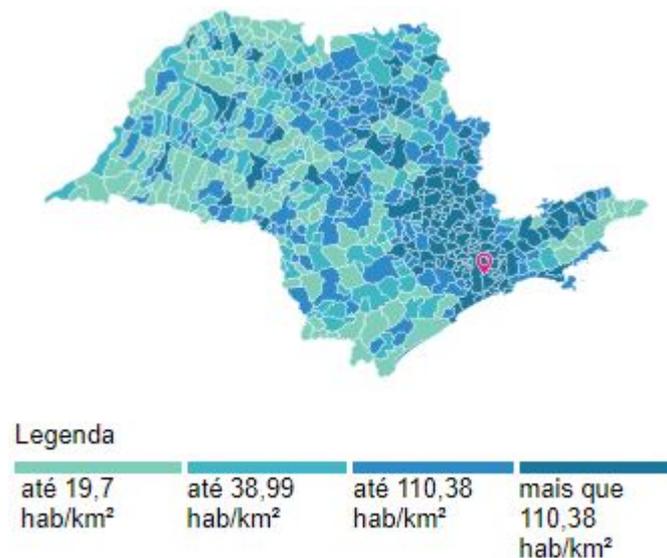
Sumário

	Introdução	14
1	MATERIAIS	17
1.1	Fontes	17
1.2	Dados do Transporte	17
1.3	Hardware e Software	20
2	MÉTODO	21
2.1	Seleção de Faixas Horárias	21
2.2	Identificação das Principais Linhas	23
2.3	Seleção do atributo qualitativo	23
2.4	Consolidação de mapas	25
2.5	Elaboração de Grafos	26
3	RESULTADOS	28
	Conclusão	30
	REFERÊNCIAS	31
	ANEXOS	33
	ANEXO A – CÓDIGOS	34

Introdução

São Paulo tem uma população estimada em mais de 12 milhões de habitantes no ano de 2018, sendo o município mais populoso do Brasil [1]. No mapa de densidade populacional, representado na Figura 1, verificamos que esta capital estadual também está cercada por diversos outros municípios com semelhante concentração demográfica, destacando-se a Região Metropolitana de São Paulo (RMSP).

Figura 1 – Mapa da densidade populacional do estado de São Paulo.



Fonte: IBGE

Estudos mostram que além dos moradores da cidade outras pessoas de diferentes municípios deslocam-se diariamente nesta capital, fenômeno denominado *Movimentos Pendulares*. O estudo destes movimentos indica que a maior concentração dos deslocamentos pendulares ocorridos no território da RMSP ocorre em sentido ao município de São Paulo. Por se tratar de uma cidade com grande representatividade, decidiu-se adotá-la como objeto para este estudo. [2]

Os sistemas públicos de transporte de São Paulo trabalham de forma integrada compondo uma malha viária de alta complexidade permitindo o constante deslocamento da população. São mais de 19 mil pontos de ônibus para 4.491 km em vias da cidade, mais de 58 estações para 65,3 km de trilhos, dividindo espaço dentre os mais 7 milhões de veículos. [3]

A problemática do transporte é agravada quando se verifica que o crescimento de vias não acompanha o número de novos veículos na cidade. Estratégias como aplicação

de rodízio no centro expandido ajudam a reduzir os congestionamentos, que atingem em média 80 km em horários de pico. [4]

Apesar de ser considerado o melhor serviço de transporte para deslocamento na capital, o sistema metroviário transporta apenas 3,7 milhões de passageiros diariamente, enquanto o "ônibus" leva em média 7,8 milhões no mesmo período, o dobro. [5]

Atualmente os sistemas de transporte, sob gestão da prefeitura, adotam um mecanismo de cobrança integrado, o Bilhete Único. Este cartão fornece ao cidadão um meio alternativo ao pagamento realizado em dinheiro e permite que a gestão pública colete informação em formato digital. Com o uso desta tecnologia como forma de pagamento, a cidade produz informações muito valiosas para seus gestores. [6]

Na intenção colaborar com a problemática do transporte urbano da cidade de São Paulo, este trabalho propõe um modelo para tratar a informação coletada dos bilhetes eletrônicos, processando-a em grafos não orientados, por faixa horária. Embora este modelo possa ser aplicado a qualquer outro modal, optou-se por estudar o transporte de ônibus metropolitano que transporta, em média 64% da população diariamente. [7]

Em 2013, o governo brasileiro criou o Plano de Dados Abertos, que aliado aos portais de transparência dos órgãos públicos garante ao cidadão o acesso a toda a informação que esteja em posse do governo. [8] É notória a importância desta ação para promover o desenvolvimento das cidades, permitindo que seja iniciada uma convergência para a era das cidades inteligentes. [9]

Consultando diferentes canais de transparência da cidade de São Paulo, verificou-se que cada secretaria apresenta um grau diferente de maturidade no que tange à abertura de dados públicos. Aparentemente ainda não há padronização nos formatos dos arquivos disponibilizados nestes canais, não estando centralizados, nem apresentando o escopo de toda a informação existente. [10]

Hoje os serviços de transporte públicos sob responsabilidade da Secretaria Municipal de Transportes de São Paulo são privatizados, sendo geridos por empresas por meio de concessões. São Paulo Transporte S/A (SPTrans) é atualmente a empresa responsável pela gestão das linhas de ônibus da cidade de São Paulo e, portanto, pela manutenção das informações de passageiros e suas viagens. [11]

Analisando respectivo portal de transparência, da SPTrans, observou-se que pouca informação está disponível, tendo em vista o potencial tecnológico já instalado através da tecnologia de bilhetagem eletrônica, desde 2010. Esperava-se que, pelo formato digital da informação, dados brutos já estivessem acessíveis à população e que a informação processada disponibilizada, como os relatórios técnicos, permitisse um estudo mais amplo e detalhado. [12]

Ao definir como tema de estudo deste trabalho a formatação da informação através

de grafos não orientados, deve-se primeiramente à ampla aplicabilidade devido à existência de vasto campo de estudo e modelos que fazem uso deste formato. Em segundo, pela condensação da informação, que permite uma saída gráfica, de fácil interpretação. ^[13] ^[14]

1 Materiais

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.

(Opendata Handbook) [9]

Nesta seção, descrevem-se os recursos e procedimentos realizados para pré-formatar os dados obtidos da agência de transporte de forma que estes possam ser convertidos em grafos não orientados, segundo a metodologia proposta no [Capítulo 2](#). Para fins de padronização, os dados originários de diferentes fontes serão denominados *Dados Brutos*, sendo que ao término do processo de pré-formatação será gerado um arquivo compilado, chamado *Arquivo Referencial*.

1.1 Fontes

Foram diversas as fontes utilizadas para prospecção de informação, dentre os dados abertos utilizados neste trabalho destacam-se os obtidos através do portal da transparência da prefeitura de São Paulo [12], da área de desenvolvedores da SPTrans [15], do portal das cidades do IBGE [1] e do portal de georreferenciamento da cidade [16].

1.2 Dados do Transporte

Hoje o cidadão que desejar obter informações sobre as viagens de ônibus na cidade de São Paulo poderá verificar os *Indicadores do Sistema de Transporte* disponibilizados no portal da transparência da prefeitura [10]. Dentre os indicadores disponibilizados encontram-se apenas relatórios diários e compilações mensais de somas do número de passageiros transportados em cada linha de ônibus, além de informações sobre a frota de ônibus contratada.

Estes relatórios são apresentados em tabelas e disponibilizados em formato de planilha eletrônica, a [Tabela 1](#) mostra quais são os campos de informação disponibilizados. Destes relatórios compilados, podemos obter informações quanto às somas de um dado período. A exemplo, somos informados que em 2017 foram transportados em média 7,8 milhões de pessoas diariamente com uma frota média de 14.500 veículos [7].

Contudo, apesar de obtermos o **nome de todas as linhas de ônibus** e o total de passageiros transportados em cada uma delas, por se tratar de informação compilada, *não temos, na [Tabela 1](#), a identificação individual de cada viagem*. Assim, como o objetivo

deste trabalho é formatar a informação, considerando períodos de interesse (horários de pico), foi necessário entrar em contato com a agência de gestão de transporte para que esta disponibilizasse acesso aos dados que serviram de base para a geração destes relatórios.

Tabela 1 – Relatório de Passageiros Transportados

Campos	Descrição
Tipo	Autorização concedida à empresa operadora linhas de ônibus.
Área	É a área numerada de abrangência da empresa.
Empresa	Nomes das empresas de transporte de ônibus envolvidas.
Linha	Nome da linha de ônibus.
Soma de Passageiros Pagtes Em Dinheiro	Soma de viagens pagas com dinheiro.
Soma de Passageiros Pagtes Comum	Soma de viagens do tipo comum.
Soma de Passageiros Pagtes Bu Comum M	Soma de viagens do tipo comum mensal.
Soma de Passageiros Pagtes Estudante	Soma de viagens do tipo estudante.
Soma de Passageiros Pagtes Bu Est Mensal	Soma de viagens do tipo estudante com cota mensal.
Soma de Passageiros Pagtes Bu Vt	Soma de viagens do tipo vale transporte que é pago pelas empresas.
Soma de Passageiros Pagtes Bu Vt Mensal	Soma de viagens do tipo vale transporte mensal.
Soma de Passageiros Pagtes Int M/CPTM	Soma de viagens que fizeram integração de metrô/-trem para ônibus.
Soma de Passageiros Pagantes	Soma de viagens de todos os passageiros pagantes.
Soma de Passageiros Pagtes Int M/CPTM M	Soma de viagens que fizeram integração de metrô/-trem para ônibus com cota mensal.
Soma de Passageiros Int Ônibus->Ônibus	Soma de viagens que fizeram integração de ônibus para ônibus.
Soma de Passageiros Com Gratuidade	Soma de viagens que usam ônibus de forma gratuita.
Soma de Passageiros Com Gratuidade Est	Soma de viagens que usam ônibus de forma gratuita sendo estudantes.
Soma de Tot Passageiros Transportados	Soma de viagens de todos os passageiros transportados.

Fonte: SPTrans

O acesso a esta informação é mais restrita, sendo possível sob solicitação feita pelo *Sistema de Informação do Cidadão e-SIC*. Através deste canal foi solicitado, à Secretaria Municipal de Transportes, o acesso à informação do histórico das viagens provenientes do bilhete eletrônico. Ao final do trâmite, foi disponibilizado acesso aos dados dos bilhetes eletrônicos, com informação diária referentes ao período de um mês de coleta.

Na [Tabela 2](#), estão descritos todos os campos de informação produzidos pelo sistema de bilhete eletrônico. Dentre os campos disponibilizados destacam-se o **horário e código de identificação da linha** no momento de pagamento da viagem. É importante destacar que esta base não contém informação suficiente para identificação dos usuários.

A elaboração de um arquivo referencial é crítico para consolidar toda a informação e permitir que seja realizada a análise temporal, ou seja, para que haja uma análise temporal precisamos essencialmente dos dados que contenham campos com os horários e a identificação das linhas, *DT_UTILIZACAO* e *NI_IDLINHA* respectivamente.

Tabela 2 – Dados do Bilhete Eletrônico

Campos	Descrição
DT_UTILIZACAO	Data da utilização.
NI_IDCARLOG	Identificação do veículo.
NI_IDCOLETA	Identificação da leitora do cartão.
NI_IDLINHA	Número da linha de ônibus atual.
NI_IDLINHAANT	Número da linha de ônibus anterior.
NI_IDPREFIXO	Identificação de prefixo da linha.
NI_IDTPCARTAO	Identificação do tipo do cartão.
NI_TPTRANSACAO	Tipo de transação.
NI_GRUPOCARTAO	Grupo do cartão.
NI_IDCARTEIRA	Identificação do número do cartão.
NI_SEQINTEGRA	Identificação de integração de passagem.
NI_SEQTRANSV	Identificação do valor existente no cartão.
NI_SENTIDOMV	Identificação do sentido da viagem.
NI_IDSAM	Identificação interna do sistema de bilhete único.
NI_VALORA	Valor gasto atual.
NI_VALORB	Valor gasto anterior.
NI_SEQTRANSV	Número de viagens feitas pelo cartão desde a última recarga.
NI_SALDOCARA	Saldo atual.
NI_SALDOCARB	Saldo anterior.

Fonte: SPTrans

No caso dos dados obtidos a partir da informação presente nos bilhetes únicos, [Tabela 2](#), foi necessário aplicar técnicas de filtragem da informação bruta, principalmente devido ao volume de dados. A filtragem foi realizada usando o [Código 1](#), cuja estrutura resultante consiste em um arquivo formatado por tabulação, representado na [Tabela 3](#).

Tabela 3 – Estrutura do Arquivo Referencial

Horário do Registro	Identificação da Linha
01/01/2018 06:30	12345
01/01/2018 07:30	67890
01/01/2018 07:35	12345

É importante destacar que, devido a uma falta de padronização das bases, a informação difere dependendo da fonte, ou seja, o código de uma determinada linha de ônibus é diferente dependendo do sistema de onde a informação foi extraída. O campo *Linha* da [Tabela 1](#) difere da informação presente em *NI_IDLINHA* da [Tabela 2](#), sendo necessário sincronizar a informação com a agência recorrentemente.

Nesta etapa, foram analisadas informações de diferentes fontes, sendo os dados de bilhete eletrônico os necessários para a elaboração do *Arquivo Referencial*. A *identificação da linha* selecionada para formar este arquivo corresponde, portanto, aos registros obtidos de [Tabela 2](#).

1.3 Hardware e Software

Ao trabalhar com os dados disponibilizados foi verificada uma dificuldade inicial em conseguir manipulá-los, principalmente devido ao volume de informação existente a partir dos bilhetes eletrônicos. Cada arquivo, diário, contém cerca de 12 milhões de registros que descompactados ocupam cerca de 1 gigabyte. Trabalhar com os dados referentes a um mês completo, foi desafiador pelas limitações de hardware e software existentes na maioria dos ambientes, sendo portanto este, o principal motivo para estabelecer um arquivo referencial, mais enxuto.

Este trabalho foi realizado em um notebook ¹, com sistema operacional *Chrome OS*, processador 1.4GHz, 2GB memória e 10GB de armazenamento.

Para manipular e filtrar a informação em grandes arquivos de texto optou-se por adotar a linguagem interpretada shell script ^[17]. Como diferencial ele foi escolhido por conseguir aplicar modificações, otimizando o uso dos recursos de hardware.

O projeto R ^[18] foi selecionado por ser uma ferramenta estatística modular, permitindo o uso de pacotes auxiliares, estendendo suas potencialidades. Dentre eles foi utilizado o pacote *reporttools* ^[19] é um pacote usado para gerar estatísticas descritivas a partir de diferentes objetos de informação. Estas estatísticas são computadas e permitem que tabelas formatadas sejam exportadas para diferentes ambientes, inclusive \LaTeX .

O *Infotrans* é um dos sistemas informatizados ^[20] para a gestão do transporte utilizados pela empresa SPTrans S.A. Os dados permitem ao cidadão obter informações quanto aos itinerários da frota de ônibus. Para desenhar, configurar e produzir mapas, mesclaram-se os dados de georreferenciamento e dos itinerários, para isso foi utilizado o *Google MyMaps*².

¹ Chromebook HP 14 - <<https://support.hp.com/us-en/document/c04030100#AbT0>>

² My Maps - <<https://www.google.com/intl/pt-BR/maps/about/mymaps/>>

2 Método

Neste capítulo, explica-se qual o processo na elaboração de grafos não orientados que representem o fluxo das viagens de ônibus a partir da importação dos dados pré-formatados.

No [Capítulo 1](#), os dados disponibilizados pelas empresas de transporte já foram utilizados. Neste [Capítulo 2](#), a informação apresentada foi produzida a partir da aplicação da metodologia que será apresentada.

A descrição do método é segmentada em etapas, são elas: seleção da faixa horária de interesse, identificação das principais linhas, seleção do atributo qualitativo, consolidação de mapas e elaboração de grafo.

2.1 Seleção de Faixas Horárias

Devido ao volume de informação dos bilhetes eletrônicos, decidiu-se filtrar a informação para que sejam estudadas faixas horárias de interesse, sendo que este trabalho tem como objetivo estabelecer e estudar os horários de pico.

Primeiramente foram identificadas as linhas que mais transportam passageiros diariamente, realizando uma amostragem *top100* para cada dia. Deste levantamento, verificou-se que independente do dia, as primeiras linhas relacionadas eram sempre as mesmas, muito distantes das identificadas nas últimas posições. Além disso, a contagem dos cerca de 30 x 11 milhões de registros apenas para este fim é muito custosa, então, na intenção de obter uma amostra mais enxuta, realizou-se um *top10*, considerando uma randomização uniforme da amostra inicial. Para extrair as 10 principais linhas diárias foi adotado o [Código 2](#).

Nos *Arquivos Referenciais*, observamos que o campo que contém os registros horários de utilização dos bilhetes apresentam informações instantâneas, isto é, trata-se de uma variável quantitativa contínua. Assim, para tabular as frequências destes registros estabeleceram-se classes ^[21].

Foram três os períodos observados em cada dia, manhã (06:00 - 11:59), tarde (12:00 - 17:59) e noite (18:00 - 23:59), sendo que cada um dos períodos foi segmentado em classes de 30 minutos. Esta classificação das faixas horárias e a contagem do número de viagens das *top10* linhas foi realizada pelo [Código 3](#). Para exemplificar os resultados, temos a [Tabela 4](#) e os gráficos na [Figura 2](#) e [Figura 3](#), do período da manhã.

Tabela 4 – Faixas Horárias da Manhã

Variáveis	n	Min	q ₁	\tilde{x}	\bar{x}	q ₃	Max	s	IQR
06:00 às 06:29	30	2931	8597.2	16343.0	44169.4	36027.5	276138	74564.9	27430.2
06:30 às 06:59	30	2939	8873.2	18891.5	52043.3	43091.5	329571	88193.6	34218.2
07:00 às 07:29	30	3056	8548.5	19306.5	53324.2	45466.5	330079	87598.5	36918.0
07:30 às 07:59	30	2886	7710.2	17879.5	49129.1	41779.0	303257	79960.4	34068.8
08:00 às 08:29	30	2613	7271.8	15072.5	42924.2	38482.0	257910	68097.4	31210.2
08:30 às 08:59	30	2486	6919.0	13497.0	37005.9	33614.8	216871	57424.8	26695.8
09:00 às 09:29	30	2436	6179.8	11080.0	32048.4	29038.5	182263	49442.2	22858.8
09:30 às 09:59	30	2710	6185.8	9952.5	28648.7	25289.5	158169	43927.1	19103.8
10:00 às 10:29	30	2796	6056.2	9566.5	26601.8	22095.0	144417	40983.7	16038.8
10:30 às 10:59	30	2663	6175.2	9970.0	26728.3	21718.2	144636	41281.5	15543.0
11:00 às 11:29	30	3042	6427.0	10514.5	28958.7	24260.0	157084	44454.4	17833.0
11:30 às 11:59	30	3178	6491.8	11175.0	31112.2	25289.5	172411	48220.7	18797.8

Figura 2 – Boxplot - Manhã (com outliers).

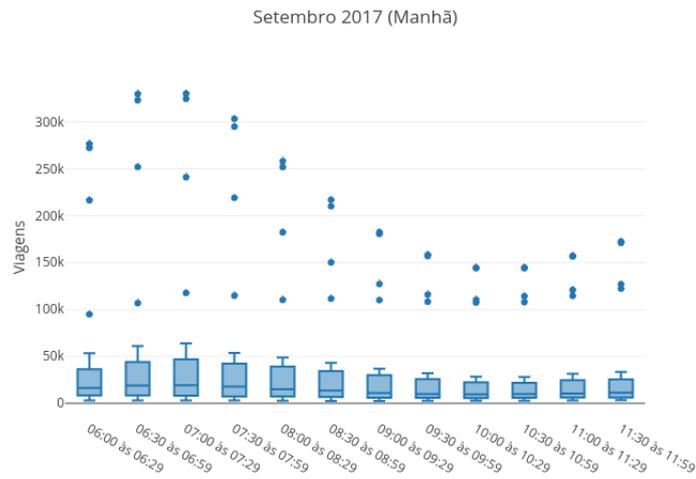
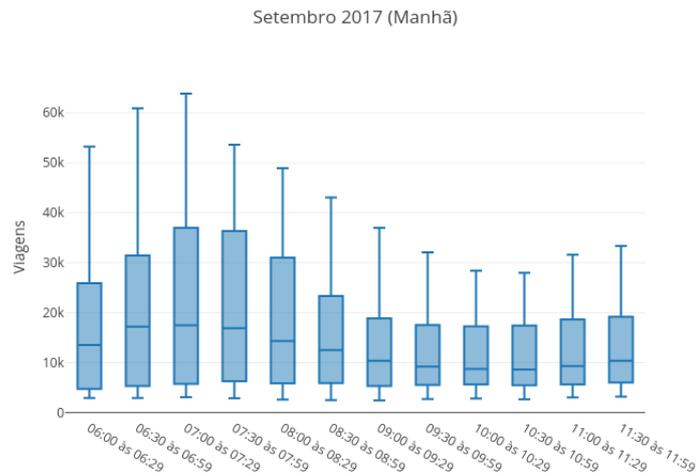


Figura 3 – Boxplot - Manhã (sem outliers).



Determinam-se, portanto, como horários de pico às faixas identificadas pelos maiores intervalos interquartis (IQR), são elas: **Manhã (06:00-08:59)**, **Tarde (16:00-17:59)** e **Noite (18:00-19:29)**.

2.2 Identificação das Principais Linhas

Neste passo, considera-se que as faixas horárias de interesse já foram definidas e assim podemos identificar quais são as linhas de ônibus mais utilizadas. Retomando os *Arquivos Referenciais*, realiza-se contagem de todas as viagens em cada horário de pico. O [Código 4](#) fez este levantamento cujos resultados, para o período da manhã, são apresentados na [Tabela 5](#).

Tabela 5 – Principais Linhas da Manhã

ID	LINHA		Manhã	
			Terminal Principal (TP)	Terminal Secundário (TS)
1877	8700	10	PÇA. RAMOS DE AZEVEDO	TERM. CAMPO LIMPO
478	875C	10	METRÔ SANTA CRUZ	TERMINAL LAPA
1054	3763	10	METRÔ TATUAPÉ	TERM. VL. CARRÃO
1977	607C	10	ITAIM BIBI	JD. MIRIAM
2160	4310	10	TERM. PQ. D. PEDRO II	E.T. ITAQUERA
910	2202	10	CPTM GUAIANAZES	HOSP. ITAIM
2216	209P	10	TERM. PINHEIROS	CACHOEIRINHA
462	9501	10	LGO. DO PAISSANDÚ	TERM. CACHOEIRINHA

Ao fazer o cruzamento da informação, correlacionando a identificação das linhas com seus respectivos nomes, verificamos que os *outliers* são registros virtuais e utilizados em subsistemas internos da SPTrans, não correspondendo a viagens reais.

2.3 Seleção do atributo qualitativo

Considerando o objetivo deste trabalho que consiste na elaboração de um grafo não orientado que represente o fluxo de transporte aplicado a um contexto político, devemos portanto selecionar este contexto. A seleção consiste em obter uma representação gráfica, por meio de uma mapa geodésico, que irá definir contexto à análise.

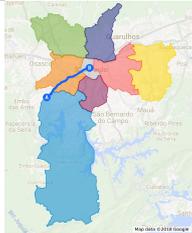
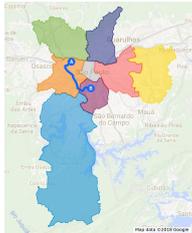
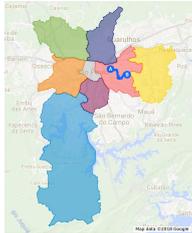
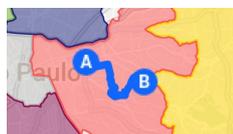
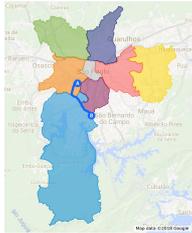
Hoje temos diversos mapas políticos com regiões administrativas da cidade. O mapa do rodízio elaborado da pela CET, assim como o censo demográfico do IBGE são alguns exemplos que poderiam ser aplicados. A SPTrans tem seu próprio mapa de regiões administrativas, [Figura 4](#), neste mapa a cidade é subdividida em 8 regiões além do centro.

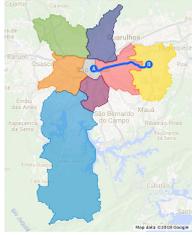
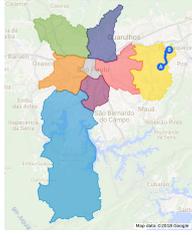
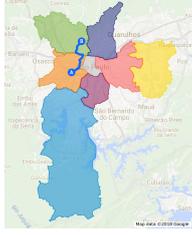
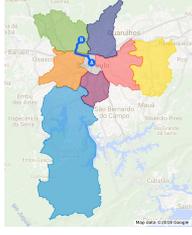
Para fins de exemplificação, foi adotado como base o mapa *Região 8*, por ser elaborado pela prefeitura ^[22], além de possuir poucas subdivisões facilitando sua visualização. O mapa destas regiões é representado na [Figura 5](#).

- Região Leste II: Subprefeituras Cidade Tiradentes, Ermelindo Matarazzo, Guaianases, Itaim Paulista, Itaquera, São Mateus e São Miguel.
- Região Sul I: Subprefeituras Ipiranga, Jabaquara e Vila Mariana.
- Região Sul II: Subprefeituras Campo Limpo, Capela do Socorro, Cidade Ademar, M’Boi Mirim, Parelheiros e Santo Amaro.
- Região Oeste: Subprefeituras Butantã, Lapa e Pinheiros.

2.4 Consolidação de mapas

Nesta seção, o objetivo está em identificar quais são as regiões interligadas pelas principais linhas de transporte coletivo. Isto será feito através de sobreposição de mapas do itinerário das linhas de ônibus e limites político-administrativos da cidade.

Linha	Mapa da cidade	Mapa da região	Rota
8700-10			<ol style="list-style-type: none"> 1. Sul 2 (azul) 2. Oeste (laranja) 3. Centro (cinza)
875C-10			<ol style="list-style-type: none"> 1. Oeste (laranja) 2. Sul 1 (vinho)
3763-10			<ol style="list-style-type: none"> 1. Leste 1 (vermelho) 2. Leste 1 (vermelho)
607C-10			<ol style="list-style-type: none"> 1. Sul 2 (azul) 2. Oeste (laranja)

4310-10			<ol style="list-style-type: none"> 1. Leste 2 (amarelo) 2. Leste 1 (vermelho) 3. Centro (cinza)
2202-10			<ol style="list-style-type: none"> 1. Leste 2 (amarelo) 2. Leste 2 (amarelo)
209P-10			<ol style="list-style-type: none"> 1. Norte 2 (verde) 2. Oeste (laranja)
9501-10			<ol style="list-style-type: none"> 1. Norte 2 (verde) 2. Oeste (laranja) 3. Centro (cinza)

2.5 Elaboração de Grafos

Grafos são representações formadas por *vértices* e *arestas*, sendo bastante simples de construir e muito úteis na análise de malhas de transporte.

Podemos definir um grafo matematicamente como $G(N, A)$ sendo N um conjunto finito de nós (vértices) e A um conjunto finito de arestas que conectam um par de nós. Uma aresta que ligue os vértices $i \in N$ e $j \in N$ é denotado por (i, j) . Se todas as arestas possuírem uma orientação, então o grafo é dito **dirigido (ou orientado)**. Caso as arestas não possuam orientação, o grafo será **não dirigido (ou não orientado)** [23].

Se houver uma aresta e do grafo G que possui o mesmo vértice como extremos, ou seja, (i, i) , então é dito que este grafo possui um **laço** [24].

Uma abordagem para construir os grafos consiste na criação de um vértice para cada região de interesse, assim como uma aresta para cada linha de transporte que as interligue. Uma linha de ônibus pode, eventualmente, interligar dois ou mais vértices.

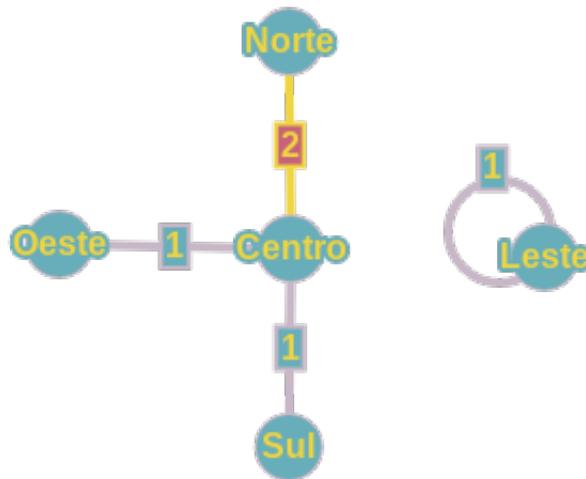
Atribuem-se valores às arestas proporcionais à contagem do número de linhas que as compõem.

Para fins de exemplificação, considere o seguinte cenário:

- Regiões: (N)orte, (S)ul, (L)este, (O)este e (C)entro.
- Existe uma linha de ônibus que interligue (O, C) e (C, N) .
- Existe uma linha de ônibus que interligue (S, C) e (C, N) .
- Existe uma linha de ônibus que seja usada apenas em (L, L) .

No exemplo, existem duas linhas que compõem (C, N) e uma linha pode, eventualmente não estar ligada a outras regiões. É possível verificar a representação deste grafo na Figura 6.

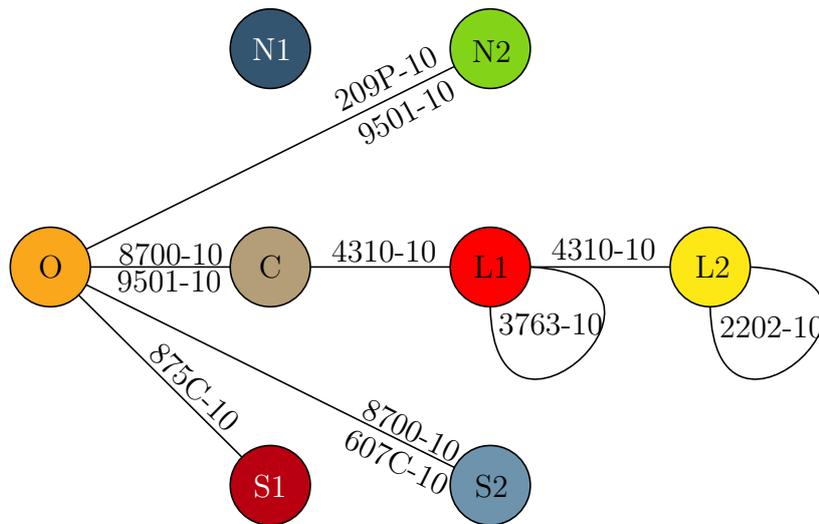
Figura 6 – Grafo Exemplo



3 Resultados

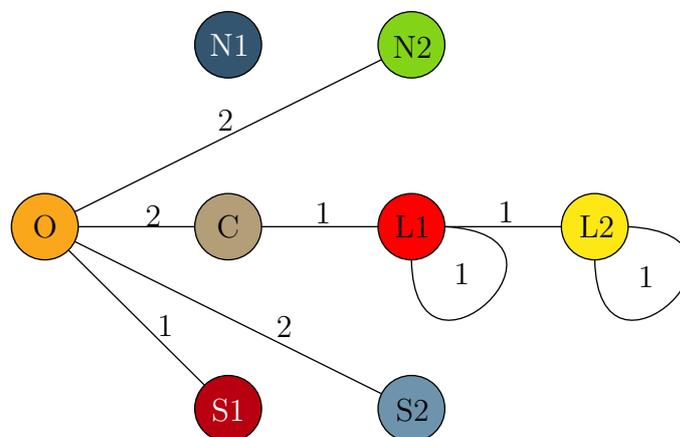
Na [seção 2.4](#), foram mapeados os itinerários das principais linhas de ônibus sobrepostas a um mapa político, permitindo a elaboração do grafo representado na [Figura 7](#). Cada nó deste grafo foi referenciado a uma região administrativa do município, [Figura 5](#).

Figura 7 – Principais Linhas da Manhã



Na [Figura 7](#), observa-se que a indicação por extenso das linhas de ônibus é um fator limitante, quando o modelo for estendido para cenários maiores. Como alternativa, pode-se trabalhar com uma representação simplificada, enumerando o número de linhas de ônibus em cada aresta, [Figura 8](#).

Figura 8 – Principais Linhas da Manhã

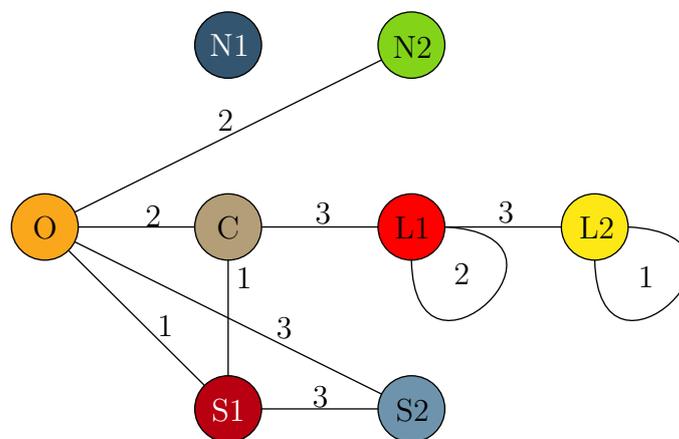


A [Figura 9](#) mostra o resultado ao elaborar o grafo considerando as principais linhas da manhã, tarde e noite.

Uma vez modelados através de grafos, muitas análises podem ser aplicadas para extrair informação útil da malha de transporte. Existem diversos algoritmos para grafos, permitindo trabalhar com problemas de *Caminhos de Custo Mínimo*, *Caminhos de Custo Máximo* e *Fluxo em Redes* [14].

Um exemplo simples de análise para este grafo é atribuir um peso em cada nó correspondente a soma de linhas de ônibus em suas arestas. Podemos inferir que quanto maior o número, maior a demanda por transporte existente neste nó, assim temos: $N1 = 0$; $N2 = 2$; $O = 5$; $C = 6$; $L1 = 8$; $L2 = 4$; $S1 = 5$; $S2 = 6$.

Figura 9 – Principais Linhas M/T/N



Neste modelo, identificamos que a região leste ($L1 + L2$) e sul ($S1 + S2$) são as que concentram os maiores números de linhas de ônibus, com maior quantidade de passageiros. Faltam alternativas de transporte para atender a estas áreas saturadas.

A contribuição deste trabalho é a imediata visualização das regiões e linhas mais demandadas (arestas das figuras) e quais estão adequadas ao fluxo, como exemplo $N1$.

Tal interpretação pode ser confrontada e confirmada ao comparar o mapa do crescimento populacional nos distritos do município, evidenciando a correlação do método proposto com a realidade [25].

Conclusão

Este trabalho conseguiu elaborar um processo para transformar informação bruta de registro de viagens em formato fechado, em informação condensada útil para análise de fluxos, em formato aberto.

Durante a importação e tratamento dos dados brutos, foi verificada uma ausência de padronização de mercado na informação referente às viagens dos passageiros. Cada modal de transporte, bem como suas administradoras usam sistemas internos personalizados, que além de tornar o processo de filtragem de dados igualmente personalizados, dificultam o acesso à informação.

É notável a importância da abertura de dados e suas iniciativas já existentes, porém falta coesão dos dados obtidos nos sistemas vigentes.

Ao propôr um mecanismo para que as redes de transporte sejam analisadas através de grafos, abre-se o formato de acesso à informação. Se este procedimento for integrado aos sistemas internos destas administradoras e disponibilizados à população, será possível elaborar ferramentas de gestão mais robustas, integrando diferentes tipos de modais.

Este trabalho mostrou que o uso de grafos não orientados, não fica restrito apenas à academia, tornando-se instrumento útil para um desenvolvimento planejado e transparente permitindo um acompanhamento simplificado pela população.

O método pode ser flexibilizado para uma rede mais complexa e ampla, com maior número de pontos e linhas. A estratégia pode ser adaptada conforme a escolha de outros mapas políticos e, principalmente, dos acessos aos dados de transporte.

Referências

- 1 IBGE, I. B. de Geografia e E. *Panorama da Cidade de São Paulo*. 2018. Acesso em: 02 abr. 2018. Disponível em: <<https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>>. Citado 2 vezes nas páginas 14 e 17.
- 2 ÂNTICO, C. Deslocamentos pendulares na região metropolitana de São Paulo. *São Paulo em Perspectiva*, Scielo, v. 19, p. 110 – 120, 12 2005. ISSN 0102-8839. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-88392005000400007&nrm=iso>. Citado na página 14.
- 3 MIRANDA, G. *Entenda o transporte da cidade de São Paulo em números*. 2013. Acesso em: 29 abr. 2018. Disponível em: <<http://folha.com/no1351891>>. Citado na página 14.
- 4 GENTILE, R. *Trânsito no horário de pico segue em queda na cidade de São Paulo*. 2018. Acesso em: 29 abr. 2018. Disponível em: <<https://folha.com/fyk5mncz>>. Citado na página 15.
- 5 QUAZ, M. *Paulistanos elegem o Metrô como a melhor prestadora de serviço de transporte*. 2018. Acesso em: 29 abr. 2018. Disponível em: <<http://folha.com/no1966617>>. Citado na página 15.
- 6 SOE, R. M.; MIKHEEVA, O. Combined model of smart cities and electronic payments. In: *2017 Conference for E-Democracy and Open Government (CeDEM)*. [S.l.: s.n.], 2017. p. 194–205. Citado na página 15.
- 7 S/A, S. P. T. *Relatório mensal do número de passageiros transportados*. 2018. Acesso em: 10 jul. 2018. Disponível em: <http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/institucional/sptrans/aceso_a_informacao/index.php?p=245234>. Citado 2 vezes nas páginas 15 e 17.
- 8 GESTÃO, M. do Planejamento Desenvolvimento e. *Plano de Dados Abertos (PDA)*. 2014. Acesso em: 10 jul. 2018. Disponível em: <<http://www.planejamento.gov.br/tema/governo-aberto/plano-de-dados-abertos-pda>>. Citado na página 15.
- 9 INTERNATIONAL, O. K. *Open Data Handbook*. 2018. Acesso em: 02 jun. 2018. Disponível em: <http://opendatahandbook.org/guide/pt_BR/what-is-open-data/>. Citado 2 vezes nas páginas 15 e 17.
- 10 PAULO, P. de S. *Portal da Transparência*. 2018. Acesso em: 10 jul. 2018. Disponível em: <<http://transparencia.prefeitura.sp.gov.br>>. Citado 2 vezes nas páginas 15 e 17.
- 11 SPTRANS, S. P. T. S. A. *A SPTrans*. 2018. Acesso em: 02 abr. 2018. Disponível em: <http://www.sptrans.com.br/a_sptrans/>. Citado na página 15.
- 12 PAULO, P. de S. *Portal da Transparência*. 2018. Acesso em: 02 abr. 2018. Disponível em: <http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/institucional/sptrans/aceso_a_informacao/index.php>. Citado 2 vezes nas páginas 15 e 17.

- 13 Bast, H. et al. Route Planning in Transportation Networks. *ArXiv e-prints*, Provided by the SAO/NASA Astrophysics Data System, 04 2015. Disponível em: <<http://adsabs.harvard.edu/abs/2015arXiv150405140B>>. Citado na página 16.
- 14 FEOFILOFF, P. *Algoritmos para Grafos em C via Sedgewick*. 2018. Acesso em: 02 jun. 2018. Disponível em: <https://www.ime.usp.br/~pf/algoritmos_para_grafos/>. Citado 2 vezes nas páginas 16 e 29.
- 15 SPTRANS, S. P. T. S. A. *Área de Desenvolvedores*. 2018. Acesso em: 02 abr. 2018. Disponível em: <<http://www.sptrans.com.br/desenvolvedores/>>. Citado na página 17.
- 16 PAULO, P. de S. *Mapa digital da Cidade de São Paulo - GeoSampa*. 2018. Acesso em: 02 abr. 2018. Disponível em: <http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx>. Citado na página 17.
- 17 DIE.NET. *sh - command interpreter (shell)*. 2018. Acesso em: 25 Mar. 2018. Disponível em: <<https://linux.die.net/man/1/dash>>. Citado na página 20.
- 18 COLEGAS, J. C. e. *The R Project*. 2018. Acesso em: 25 Mar. 2018. Disponível em: <<https://www.r-project.org/about.html>>. Citado na página 20.
- 19 RUFIBACH, K. *Generate LaTeX Tables of Descriptive Statistics*. 2018. Acesso em: 25 Mar. 2018. Disponível em: <<http://www.kasparrufibach.ch>>. Citado na página 20.
- 20 S/A, S. P. T. *Sistema Informatizados para a gestão do Transporte*. 2018. Acesso em: 02 abr. 2018. Disponível em: <http://www.sptrans.com.br/pdf/biblioteca_tecnica/sistemas_informatizados_para_a_gestao_do_transporte.pdf>. Citado na página 20.
- 21 MAGALHÃES, M. N.; LIMA, A. C. P. de. *Noções de probabilidade e estatística*. [S.l.]: Editora da Universidade de São Paulo São Paulo, 2002. v. 5. Citado na página 21.
- 22 LICENCIAMENTO, S. M. de Urbanismo e. *Região 8 - Divisão do Município de São Paulo em Oito Regiões*. 2018. Acesso em: 02 abr. 2018. Disponível em: <http://www.sptrans.com.br/pdf/biblioteca_tecnica/sistemas_informatizados_para_a_gestao_do_transporte.pdf>. Citado na página 23.
- 23 LARSON, R. C.; ODONI, A. R. *Urban operations research*. [S.l.: s.n.], 1981. Citado na página 26.
- 24 W.WEISSTEIN, E. *Singleton Graph*. 2018. Acesso em: 16 jul. 2018. Disponível em: <<http://mathworld.wolfram.com/SingletonGraph.html>>. Citado na página 26.
- 25 JANNUZZI, P. de M. Cenários futuros e projeções populacionais para pequenas áreas:método e aplicação para distritos paulistanos 2000-2010. *Revista Brasileira de Estudos de População*, scielo, v. 24, p. 109 – 137, 06 2007. ISSN 0102-3098. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-30982007000100008&nrm=iso>. Citado na página 29.

Anexos

ANEXO A – Códigos

```

1 #!/bin/sh
2 # reducao.sh - Filtra os arquivos base fornecidos pela SPTrans,
   limitando-se em fornecer:
3 # Campo 1 - Horário de unilização do bilhete
4 # Campo 4 - Identificação da Linha
5 #
6 # Deste script resultam arquivos nomeados no formato YYYYMMDD.txt
   contendo apenas os campos acima.
7 for arq in *.txt.gz
8 do
9     zcat ${arq} | cut -d "/" -f 1,4 | cut -s -d " " -f 2 > $(basename "$
   {arq}" .gz)
10 done

```

Código 1 – redução.sh

```

1 #!/bin/sh
2 # Este script recebe um 'Datafile', formatado em colunas, separado por |
   (pipes)
3 # Datafile: YYYYMMDD.txt (resultado de reducao.sh)
4 # Como resultado gera-se um arquivo 'top100-YYYYMMDD.txt' contendo:
5
6 # Arquivo referencia para cruzar os ID internos do sistema de bilhetagem
   com
7 # os IDs das linhas comuns, usado pelo OlhoVivo.
8
9 # Para cada arquivo de bilhetagem eletrônica
10 # Cria uma lista com a contagem de passageiros das linhas mais
   frequentadas.
11 # 1.) Ignora-se o cabeçalho.
12 # 2.) Filtra-se pelo id das linhas de onibus.
13 # 3.) Conta-se o total de viagens por linha.
14 # 4.) Filtra-se as 100 mais contabilizadas.
15 # 5.) Sorteia-se 10 linhas dentre o top100.
16 # 6.) Armazena-se o top10, randômico.
17 #
18 # Uso: /bin/sh top10.sh YYYYMM*.txt
19 #
20 echo > faixas-manha.txt
21 echo > faixas-tarde.txt
22 echo > faixas-noite.txt
23 echo > faixas-madrugada.txt
24
25 for arq in "${@}"

```

```

26 do
27     echo "CONT\tID_SYS\t(${arq})"
28     OUTPUTFILE="top10-contagem-${arq}"
29     tail -n +2 ${arq} |\
30     cut -d '|' -f 2 |\
31     sort | uniq -c |\
32     sort -nr | head -n 100 |\
33     sort -R | head -n 10 |\
34     tee "${OUTPUTFILE}"
35
36     /bin/sh contagem-faixas.sh ${OUTPUTFILE} ${arq}
37 done

```

Código 2 – top10.sh

```

1  #!/bin/sh
2
3  IDs_SYS="$(cut -b 9- ${1} | paste -s -d "|")"
4
5  manha="(0[6-9]|1[0-1])"
6  tarde="1[2-7]"
7  noite="(1[8-9]|2[0-3])"
8  madrugada="0[0-5]"
9
10 periodo=${manha}
11 #egrep "${periodo}:...(\|.*){3}\|(${IDs_SYS})\|" ${2} > arq_manha.txt
12 egrep "${periodo}:...(\|.*){3}\|(${IDs_SYS})" ${2} > arq_manha.txt
13
14 cat >> faixas-manha.txt << EOF
15 $(for hh in 06 07 08 09 10 11
16 do
17     echo "${hh}:00 às ${hh}:29\t$(egrep -c "${hh}:[0-2]..." arq_manha.
18     txt)"
19     echo "${hh}:30 às ${hh}:59\t$(egrep -c "${hh}:[3-5]..." arq_manha.
20     txt)"
21 done)
22 EOF
23
24 periodo=${tarde}
25 #egrep "${periodo}:...(\|.*){3}\|(${IDs_SYS})\|" ${2} > arq_tarde.txt
26 egrep "${periodo}:...(\|.*){3}\|(${IDs_SYS})" ${2} > arq_tarde.txt
27 cat >> faixas-tarde.txt << EOF
28 $(for hh in 12 13 14 15 16 17
29 do
30     echo "${hh}:00 às ${hh}:29\t$(egrep -c "${hh}:[0-2]..." arq_tarde.
31     txt)"
32     echo "${hh}:30 às ${hh}:59\t$(egrep -c "${hh}:[3-5]..." arq_tarde.
33     txt)"

```

```

30 done)
31 EOF
32
33 periodo=${noite}
34 #egrep "${periodo}:...:(\|.*){3}\|(${IDs_SYS})\|" ${2} > arq_noite.txt
35 egrep "${periodo}:...:\|(${IDs_SYS})" ${2} > arq_noite.txt
36 cat >> faixas-noite.txt << EOF
37 $(for hh in 18 19 20 21 22 23
38 do
39     echo "${hh}:00 às ${hh}:29\t$(egrep -c "${hh}:[0-2]..." arq_noite.
40     txt)"
41     echo "${hh}:30 às ${hh}:59\t$(egrep -c "${hh}:[3-5]..." arq_noite.
42     txt)"
43 done)
44 EOF
45
46 periodo=${madrugada}
47 #egrep "${periodo}:...:(\|.*){3}\|(${IDs_SYS})\|" ${2} > arq_madrugada.
48 txt
49 egrep "${periodo}:...:\|(${IDs_SYS})" ${2} > arq_madrugada.txt
50 cat >> faixas-madrugada.txt << EOF
51 $(for hh in 00 01 02 03 04 05
52 do
53     echo "${hh}:00 às ${hh}:29\t$(egrep -c "${hh}:[0-2]..."
54     arq_madrugada.txt)"
55     echo "${hh}:30 às ${hh}:59\t$(egrep -c "${hh}:[3-5]..."
56     arq_madrugada.txt)"
57 done)
58 EOF

```

Código 3 – contagem-faixas.sh

```

1 #!/bin/sh
2 # linhas-por-faixa.sh - Retorna as principais linhas para uma
3 # determinada faixa-horaria.
4 # ${1} - Recebe uma expressão regular para a faixa horária de interesse.
5 # ${2} - Recebe o caminho do arquivo para realizar a filtragem.
6 # 0 retorno consiste em 2 colunas, sendo a primeira a contagem e a
7 # segunda a identificação das linhas.
8 #
9 #           ${1}           ${2}
10 # manha: "0[6-8]:...:" arq_manha.txt
11 # tarde: "1[6-7]:...:" arq_tarde.txt
12 # noite: "18:...:|19:[0-2]..." arq_noite.txt
13 egrep "${1}" ${2} | cut -d "/" -f 2 | sort | uniq -c | sort -nr

```

Código 4 – linhas-por-faixa.sh