

Estimação da matriz origem-destino e da distribuição espacial da lotação em um sistema de transporte sobre trilhos a partir de dados de bilhetagem eletrônica

Renato Oliveira Arbex¹, Claudio Barbieri da Cunha²

¹Programa de Pós-Graduação em Engenharia de Transportes, Escola Politécnica, USP, renatoarbex@usp.br

²Programa de Pós-Graduação em Engenharia de Transportes, Escola Politécnica, USP, cbcunha@usp.br

Recebido:

21 de março de 2017

Aceito para publicação:

1 de outubro de 2017

Publicado:

29 de outubro 2017

Editor de área:

Cira Souza Pitombo

Palavras-chaves:

Lotação,
Transporte Público,
Sistema Metroferroviário,
Bilhetagem Eletrônica.

Keywords:

Bus,
Public transportation,
System Metroferroviário,
Electronic Ticketing.

DOI:10.14295/transportes.v25i3.1347

RESUMO

Este artigo apresenta os resultados de uma análise para estimar a lotação por trechos entre estações de um sistema metroferroviário, ao longo do dia, utilizando dados de bilhetagem eletrônica. É proposta uma adaptação da metodologia baseada no encadeamento de embarques no sistema de transportes, que infere o destino dos usuários, uma vez que não há bilhetagem nas estações de destino. Isso permitiu obter, de maneira rápida e eficiente, sem pesquisas de campo custosas, uma estimativa da matriz de origem-destino no sistema de transporte de passageiros sobre trilhos da Região Metropolitana de São Paulo (RMSP). A metodologia proposta também possibilitou obter o padrão espaço-temporal dos níveis de lotação para todas as linhas e sentidos da rede, assim como os volumes de transferências nas estações através da alocação da matriz ao longo de todo o período de operação e não apenas para os horários de pico, como usualmente feito a partir de pesquisas de campo com base em entrevistas com usuários.

ABSTRACT

This paper presents the results of an analysis to estimate station-to-station occupation levels of a metro and urban rail system throughout the day using electronic ticketing data. An adaptation of the trip chaining methodology is proposed, which infers users destination since there is no transaction at the final destination stations. This allowed us to obtain an estimate of the origin-destination matrix quickly and efficiently, without costly field surveys, for the rail passenger transport system of the São Paulo Metropolitan Area (SPMA). The proposed methodology also made it possible to obtain the spatio-temporal pattern of the crowding levels for all lines and directions of the network, as well as the transfer volumes in the stations through the assignment of the matrix throughout the operation period and not only for the peak times, as usually done from field surveys based on user interviews.



1. INTRODUÇÃO

Incremento do uso de transporte individual (carros e motos) para os deslocamentos diários, em detrimento ao transporte coletivo, acarretando congestionamentos crescentes, maiores tempos de viagem e maior poluição são alguns dos desafios atuais da mobilidade nos grandes centros urbanos. Nesse contexto, entender como ocorrem os deslocamentos das pessoas é um dos aspectos importantes a fim de que possam ser propostas novas soluções que contribuam para aprimorar os sistemas de transporte coletivo e a qualidade dos serviços, de forma a atrair mais usuários e diminuir o uso de automóveis, contribuindo assim para melhorar a qualidade de vida nas metrópoles.

Um dos atributos de qualidade dos sistemas de transporte público é o conforto das viagens (Rodrigues e Sorratini, 2008; Ferraz e Torres, 2004). Nesse sentido, o nível de conforto nas viagens urbanas é um dos fatores que influenciam a escolha do modo de transporte. Desta forma, propostas de serviços de transporte urbano devem procurar fornecer um nível adequado de conforto para que este seja atrativo a usuários do transporte individual.

Este artigo propõe uma adaptação de uma metodologia para estimar a distribuição espaço-temporal da lotação nos trechos da rede de um sistema metroferroviário ao longo de intervalos de tempo de um dia através de dados de bilhetagem eletrônica (*smart cards*). Conforme apontam Lin e Hsu (2014), investigar padrões de mobilidade individual é uma área de pesquisa emergente. Diferentemente de outros trabalhos da literatura, no caso analisado há a necessidade de, além de inferir os locais de desembarque, estimar o carregamento nos trechos para calcular a distribuição da lotação. É necessário inferir os locais de desembarque uma vez que os passageiros não validam os seus bilhetes ao deixarem o sistema de transporte metropolitano sobre trilhos. Adicionalmente neste trabalho, para a inferência dos locais de desembarque são utilizados tanto as transações no sistema de trilhos como as transações no sistema de ônibus.

Para encontrar os padrões de lotação espaço-temporais, por sua vez, é necessário realizar a alocação da matriz estimada de viagens entre estações do sistema de trilhos, que é feita através de métodos de caminho mínimo. A etapa de alocação da matriz calculada para estimativa do carregamento é uma etapa adicional proposta neste método, que não aparece nos artigos que procuram apenas estimar as viagens através da bilhetagem eletrônica (Nassir *et al.*, 2011; Munizaga *et al.*, 2012; Alsger *et al.*, 2015). Revelar e investigar os padrões de lotação só é possível com esta etapa adicional.

As análises realizadas e os resultados obtidos possibilitam melhor caracterizar a distribuição da lotação na rede de transporte por trilhos e ao longo das faixas horárias em um dia útil, resultado de difícil obtenção em tais sistemas de transporte, uma vez que é muito difícil identificar onde os usuários desembaram e qual trajeto seguiram. A forma tradicional de fazer isso é através de entrevistas, que apresenta dificuldades decorrentes dos custos de coleta de dados e do tamanho da amostra para a obtenção de resultados representativos do padrão de fluxos; adicionalmente, tais pesquisas não fornecem elementos para uma compreensão do fenômeno mais abrangente, pois são realizadas apenas ou preferencialmente nos horários de pico.

Esta caracterização fornece subsídios aos setores de planejamento de transportes, na medida em que se obtém um panorama da utilização da capacidade ofertada da rede que é muito difícil de se obter por meio de pesquisas em campo nessa modalidade de transporte, em que o usuário não registra a sua saída do sistema, possibilitando inclusive a comparação com modelos estratégicos tradicionais. Além disso, essa caracterização contribui também para fornecer dados sobre a lotação que podem vir a ser aproveitados em sistemas de informação ao usuário, que poderiam ser divulgados em tempo real, através da *internet*, permitindo ao mesmo escolher seu melhor trajeto na rede.

Com relação aos dados de bilhetagem eletrônica, esta tem sido implantada em diversas cidades do Brasil e do mundo com diversos objetivos, dentre eles de melhor estruturação dos sistemas de cobrança. Um benefício direto da implantação desses sistemas automatizados de bilhetagem é uma obtenção de dados contínua a respeito do uso dos serviços de transporte público pelos usuários. Nesse sentido, diversos pesquisadores no mundo estão buscando formas de explorar esse grande volume de dados gerados por estes sistemas para gerar informações relevantes ao planejamento, à operação dos transportes públicos e ao acompanhamento de indicadores de qualidade do sistema.

Novas formas de obtenção de dados que permitam entender como se processam os deslocamentos das pessoas são importantes, pois as metodologias tradicionais, baseadas em pesquisas origem-destino com base domiciliar são muito caras, e quando realizadas, o são com periodicidade elevada (por exemplo, a cada 10 anos na Região Metropolitana de São Paulo), o que as impede de capturar mudanças que ocorrem nesse intervalo, e que poderiam ensejar ações e intervenções mais táticas e operacionais, para assegurar a máxima eficiência dos sistemas de transporte e melhorar o nível de serviço para os usuários.

Por exemplo, a Linha 4 – Amarela do Metrô de São Paulo começou a operar em 2010, após a última pesquisa Origem-Destino, em 2007, alterando significativamente o perfil de utilização do metrô e ocasionando gargalos não previstos em estações de integração que proporcionaram maior conectividade na rede como é o caso das estações Pinheiros e Paulista/Consolação.

A utilização de dados de bilhetagem impõe um desafio grande, pois o volume de dados é muito elevado, e as metodologias tradicionais nem sempre são adequadas para explorar o imenso volume de informações e suas nuances; por exemplo, em São Paulo, a ordem de grandeza diária dos registros é de 28 milhões para os GPS dos ônibus, 12 milhões de bilhetagens realizadas com Bilhete Único (sendo 3 milhões nos trilhos) e 2,5 milhões de caminhos na rede possíveis para ligar as 169 estações entre si.

O presente trabalho está organizado da seguinte forma: a seção 2 examina os trabalhos da literatura relacionados ao uso de dados de bilhetagem para estimativas de lotação em sistema de transporte sobre trilhos. Já a seção 3 detalha a metodologia proposta e os detalhes de como foi feita a sua aplicação ao sistema de transporte sobre trilhos na Região Metropolitana de São Paulo, enquanto os resultados estão na seção 4. Por fim, a seção 5 apresenta as considerações finais.

2. Uso de dados de bilhetagem eletrônica para estimativa de lotação em um sistema sobre trilhos

Pelletier *et al.* (2011) apresentam uma revisão da literatura onde dividem o uso de dados de bilhetagem eletrônica em três grandes níveis: estratégico (planejamento de longo prazo), tático (para ajustes de serviço e desenvolvimento da rede de transportes) e operacional (indicadores de performance e estatísticas de quantidade de passageiros). Tanto para as análises de nível estratégico como para o tático, a estimativa das origens e destinos das transações é essencial. Mais recentemente, Kurauchi e Schomocker (2016) descrevem alguns dos usos recentes em pesquisas de dados de bilhetagem eletrônica, a saber, por exemplo: carregamentos de rotas, variações do uso ao longo do dia e padrões de transferências em estações. A literatura que analisa dados de bilhetagem eletrônica tem crescido nos últimos anos, com destaque para a inferência de matrizes origem-destino de viagens como um dos principais produtos da análise (Nassir *et al.*, 2011; Munizaga *et al.*, 2012; Alsger *et al.*, 2015). Parte dos sistemas de bilhetagem de algumas cidades no mundo, incluindo os implementados nas brasileiras, não exigem uma transação do cartão no momento do desembarque. Assim, para a estimativa dos pares origem-destino, é preciso aplicar métodos de inferência do local de desembarque dos usuários para cada embarque no sistema.

Para a utilização dos dados da bilhetagem eletrônica para estimar a lotação, também é necessário que sejam conhecidos os locais de embarque e desembarque no sistema de metrô e trens dos usuários do sistema. Parte da metodologia aplicada para a elaboração de matrizes origem-destino de viagens se insere nos cálculos de estimativa de lotação nos sistemas. Munizaga *et al.* (2012) utilizam a metodologia descrita como encadeamento de viagens, que, havia sido proposta originalmente por Barry *et al.* (2002), também vem sido aplicada por outros pesquisadores para a estimativa do local de destino. Estas metodologias foram desenvolvidas para resolver o problema de estimar os locais de desembarque para sistemas de transporte onde se registram apenas os embarques.

A metodologia apresentada em Munizaga *et al.* (2012) se baseia nas seguintes hipóteses: i) o destino do embarque corrente é aquele mais próximo do local do embarque seguinte, considerando um limite de distância; ii) o usuário volta ao local do desembarque da viagem anterior para um próximo embarque e iii) na última viagem do dia o usuário retorna ao local de origem da primeira viagem, considerado como próximo ao seu local de residência. Para a aplicação do algoritmo é necessário que os embarques tenham um registro de localização, que em Munizaga *et al.* (2012) foi incorporado aos registros de bilhetagem através do cruzamento com dados de sistemas de GPS dos ônibus, coletados a cada 30 segundos. A metodologia proposta por Munizaga *et al.* (2012) não pode ser aplicada diretamente em viagens onde o usuário não necessita validar o seu bilhete imediatamente ao iniciar a sua viagem, o que ocorre, por exemplo, nas viagens de transporte coletivo por ônibus em cidades brasileiras, onde o usuário pode permanecer um tempo no veículo antes de passar pela catraca. Ao contrário do que ocorre em muitos luga-

res no mundo, sem cobradores, onde a validação é requerida no momento do embarque. Nessas situações, o local inferido de embarque não corresponde ao local real de embarque do usuário. Portanto, cabe salientar que é um processo de estimativa dos locais de transação dos bilhetes. Ressalta-se que nestas metodologias também não está incluído a estimativa de origem real do usuário, já que o trajeto de sua residência ou origem até o ponto de parada de acesso ao sistema é desconhecido. Para elaboração de uma matriz origem e destino de viagens em um sistema de zoneamento tradicional, deve-se espalhar os embarques nas paradas ao redor delas, de forma a representar os trechos de acesso por caminhada. Entretanto, este trabalho foca na matriz origem-destino entre estações de um sistema sobre trilhos, de modo que as desvantagens relatadas não influenciarão os resultados de interesse do método.

Stasko *et al.* (2016) utilizaram dados de bilhetagem para a estimativa de carregamentos no sistema metroferroviário de grande porte da cidade de Nova Iorque. Os autores propuseram um modelo que atingiu três objetivos primários de interesse: i) estimar o carregamento no metrô no nível de viagens de trens específicos; ii) aplicar a escolha de rota com base nos reais movimentos dos trens e iii) ser possível de rodar diariamente. Para a inferência dos destinos, como o sistema nova iorquino é de entrada apenas, os autores aplicaram a metodologia proposta em Barry *et al.* (2009), que por sua vez também utiliza as seguintes hipóteses na determinação dos destinos: que a maioria dos usuários começa sua próxima viagem no destino da viagem anterior ou próximo e que a maioria dos usuários termina sua última viagem do dia no local de início da primeira viagem do dia ou próximo. Essa metodologia obteve 90% de destinos válidos. Entretanto, Barry *et al.* (2009) inferiram a localização dos embarques em ônibus através de quadros de horários e programação, já que não estavam disponíveis dados de GPS na frota de ônibus da cidade. Uma boa precisão na localização das transações nos ônibus é necessária para melhor estimar as estações de destino quando o usuário utilizou o sistema de trilhos na transação anterior.

A lotação nos sistemas de transporte por trilhos afeta usuários na maior parte das grandes redes existentes nas maiores cidades do mundo, especialmente na América Latina e Ásia. Kim *et al.* (2015) também utilizaram dados de bilhetagem eletrônica em Seul, Coréia do Sul, tendo como resultado que os usuários de metrô de fato utilizam outras rotas que não a mais rápida para evitar a lotação em si e os atrasos causados pela lotação. Os impactos da lotação, causados à saúde e bem-estar dos usuários, conforme apresentados em Tirachini *et al.* (2013), vão desde aumento de ansiedade, estresse e sentimentos de exaustão, percepção de risco e insegurança, sentimentos de invasão de privacidade e maior propensão a atrasos em compromissos, causando impacto também na saúde organizacional. Também por isto, é de suma importância um melhor entendimento do padrão espaço-temporal de sua ocorrência nos sistemas de metroferroviários.

Outras alternativas para estudos de mobilidade baseadas em *Big Data* estão surgindo, destacando-se a originária da rede de telefonia móvel no Brasil. Mais especificamente, Feriancic *et al.* (2015) descrevem uma plataforma de *Big Data*, chamada *Smart Steps*, da empresa de telecomunicações Telefônica. De acordo com os autores, o *Smart Steps* é uma plataforma estatística, capaz de inferir a respeito do movimento de multidões dentro de um país a partir da utilização da infraestrutura de telefonia móvel. Os autores ressaltam ainda que a mesma possui cuidados com segurança e sigilo que utiliza dados em massa, encriptados, anônimos, agrupados e extrapolados, preservando a privacidade dos clientes e da população em geral. Entretanto, os custos envolvidos e a não disponibilização dos dados brutos, levando à necessidade de as análises serem feitas exclusivamente pela operadora de telefonia, são barreiras importantes à utilização mais disseminada dessa tecnologia. Outro aspecto importante é quanto à representatividade desses dados, uma vez que se restringem a usuários de uma única operadora de telefonia celular.

Em contraposição, a análise a partir de dados da bilhetagem eletrônica tem a vantagem da ampla utilização do bilhete único, que em São Paulo alcança 95% de uso no sistema de ônibus municipal, conforme estimativa com base nos dados utilizados nesta pesquisa.

3. METODOLOGIA E APLICAÇÃO

A metodologia proposta neste artigo tem por finalidade:

- a) Obtenção da matriz estação de origem-estação de destino de embarques em um sistema metroferroviário urbano de passageiros em que não há bilhetagem na saída do usuário na estação de destino, como ocorre na maioria dos sistemas no mundo;
- b) Estimativa do carregamento e lotação nos trechos da rede ao longo de intervalos discretos do dia;
- c) Cálculo do volume de transferências entre estações do sistema ao longo do dia.

Ao contrário do trabalho de Stasko *et al.* (2016), esta pesquisa incluiu os embarques no sistema de ônibus para auxiliar a inferência da estação de destino. Caso o usuário embarque em uma estação de metrô e o próximo embarque seja em um ônibus, a estação de desembarque estimada é aquela mais próxima do embarque neste veículo. Além disso, esta pesquisa apresenta a espacialização da distribuição da lotação no sistema, que permite assim uma melhor compreensão do uso da capacidade de transportes oferecida.

Ao contrário de Munizaga (2012), que apresentou um método para estimar matrizes de viagens entre zonas de tráfego, neste trabalho, por sua vez, se propõe estimar a matriz de viagens entre pares de estações de um sistema de transporte sobre trilhos com a finalidade de avaliação dos volumes de passageiros no sistema, estimando a distribuição da lotação ao longo de todos os horários do dia. Para tal, é aplicado um método de caminho mínimo para alocar a matriz obtida do processamento, obtendo uma distribuição estimada do carregamento, que é então expandida para obter a distribuição da lotação.

Inicialmente, serão descritos os dados de entrada necessários à aplicação, em seguida serão apresentadas as lógicas dos algoritmos de análise. Por fim, ao final da seção, será apresentada a aplicação que foi realizada no sistema de transporte sobre trilhos da Região Metropolitana de São Paulo. Na subseção seguinte serão descritos os resultados obtidos.

3.1. Dados de Entrada

Os dados de entrada necessários para a metodologia descrita nos próximos parágrafos são os dados de bilhetagem eletrônica, de GPS dos ônibus do sistema e da estrutura da rede de transporte público, organizados nos tópicos a seguir:

- i. Registros de Bilhetagem Eletrônica (*smart cards*) em todo o sistema de transporte público, incluindo estações de metrô, trem e ônibus urbanos;
- ii. Dados de sistema de posicionamento global (GPS) que contém a localização dos veículos que prestam o serviço de ônibus urbano;
- iii. Estrutura da rede de transporte público da região, no formato GTFS (*General Transit Feed System*), que contém informações como a localização das paradas de ônibus e estações, frequências e trajetos das linhas de ônibus, trens e metrô.

A base de dados das linhas de ônibus e da rede metro-ferroviária utilizada está no formato GTFS, que é um formato padrão muito utilizado atualmente em diversas cidades do mundo para a distribuição das informações de seus sistemas de transporte ao *Google Transit* (Google, 2016), assim como a disponibilização das informações ao público em geral e a desenvolvedores interessados em fazer aplicações de transporte. Nessa estrutura estão organizados os dados de transporte em diversos arquivos. Dentre eles, o stops.txt contém a codificação das estações com o campo *stop_id* e sua localização pelas coordenadas de latitude e longitude.

3.2. Processamento dos Dados

A sequência de processamento dos dados está apresentada resumidamente na Figura 1 a seguir. O início do processamento ocorre com a leitura dos arquivos de bilhetagem, GPS e GTFS correspondentes ao dia

de análise em questão. Em seguida, para todas as bilhetagens que ocorreram em estações de metrô e trens, são feitas as associações das catracas utilizadas às estações de embarque correspondentes à localização destas catracas através de uma tabela auxiliar.

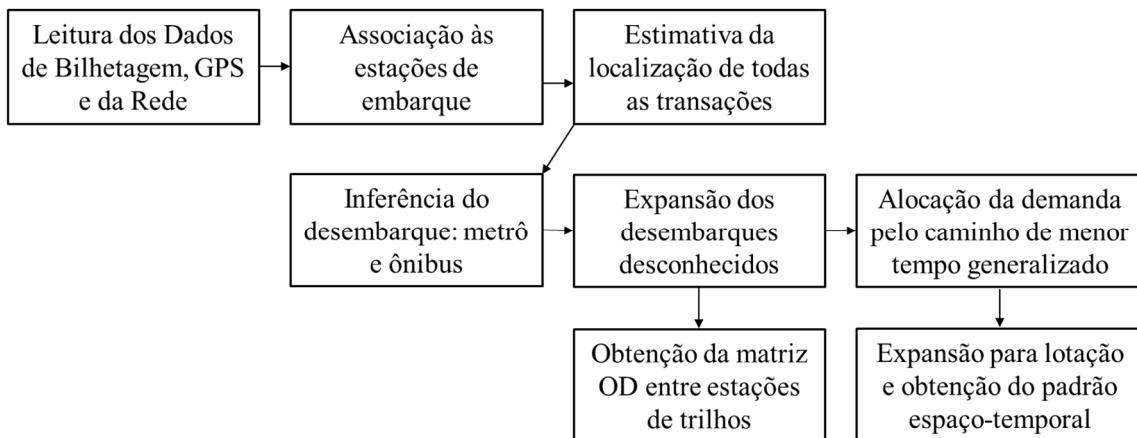


Figura 1: Sequência metodológica de processamento dos dados

Após a associação de todos os embarques no sistema de trilhos às suas respectivas estações, é necessário estimar a localização espacial de todos os embarques no sistema de transporte público ao longo do dia. Para tal, a metodologia adotada para localizar os embarques no sistema de trilhos é diferente dos embarques nos ônibus. Para o sistema de trilhos, o procedimento é através da associação do código da estação à sua localização correspondente no arquivo stops.txt do GTFS, que contém todas as localizações das estações com latitudes e longitudes. Para o sistema de ônibus, entretanto, o processo é menos trivial, já que os veículos estão continuamente se movendo no espaço.

O modo de estimativa da localização dos embarques nos ônibus é um procedimento que percorre todos os veículos da base e analisa as respectivas bilhetagens, feito da seguinte forma: para cada um dos veículos que operaram no sistema naquele dia, são filtradas todas as bilhetagens que ocorreram naquele respectivo veículo. Os horários destas bilhetagens são cruzados com os horários de todos os registros de GPS desde veículo neste dia, obtendo, para cada bilhetagem, o registro de GPS daquele veículo correspondente à menor diferença de tempo entre o horário que ocorreu a bilhetagem e o momento que o veículo emitiu a informação. Esta abordagem de percorrer os veículos, e não as bilhetagens, se mostrou mais eficiente do ponto de vista computacional, já que o número de bilhetagens, em torno de 12 milhões, é muito maior que o número de veículos em operação em um dia, que chega a 15 mil.

Após a estimativa da localização dos embarques, é possível então fazer a estimativa dos códigos de parada em que foram realizados os embarques. A localização dos embarques está atrelada à informação do GPS do veículo, enquanto que se busca agora associar ao ponto de parada onde aquela transação ocorreu. Os códigos de parada são o stop_id do arquivo stops.txt do GTFS. Para as estações na rede de metrô este valor já foi obtido anteriormente, quando do uso da tabela auxiliar com os códigos das catracas utilizadas e suas estações correspondentes. Para os ônibus, o código da estação foi obtido do ponto de parada daquela linha mais próximo do local correspondente ao registro de GPS associado àquela bilhetagem.

Com a inferência do código de parada do embarque, a próxima etapa é a inferência do código da estação ou ponto de parada de desembarque. Esta é a etapa mais importante na medida que a inferência do local de desembarque é a chave para a elaboração da matriz origem-destino entre estações na rede de trilhos. Deve-se destacar que a metodologia aqui proposta não contempla a determinação da matriz origem-destino (OD) de viagens completas na rede, porta-a-porta, envolvendo também o sistema ônibus, uma vez que é necessário diferenciar as viagens e as curtas atividades dos locais de transferência.

Para a inferência do código de parada de desembarque, foram adotadas as hipóteses conforme mencionado na seção 2, onde: i) o destino do embarque atual ocorre nas proximidades do embarque seguinte; ii) os usuários retornam ao destino do embarque anterior para o próximo embarque (parte-se do princípio que não ocorrem viagens intermediárias sem o uso do sistema de transporte público); iii) o destino do último embarque do dia é o local onde ocorreu o primeiro embarque. Desta forma, para cada embarque de cada usuário, o local de desembarque é o ponto de parada daquele serviço que esteja mais próximo espacialmente da localização do embarque seguinte.

A inferência do código de parada de desembarque no sistema de metrô é calculada de forma mais direta que nos serviços de ônibus. No sistema metroferroviário urbano da Região Metropolitana de São Paulo, onde foi aplicada a metodologia, as transferências internas no sistema são gratuitas e não é necessário validar ou registrar o cartão em nenhum momento nas transferências. Desta forma, foi considerado o conjunto de todas as estações do sistema na hora do cálculo da estação de desembarque. Assim, a estação de desembarque em embarques na rede de trilhos é aquela cuja distância à localização do próximo embarque daquele usuário seja mínima. A utilização dos dados de GPS dos ônibus mostra sua importância nesta etapa, pois com estes registros é possível localizar, de forma mais realista e com maior precisão, a localização dos embarques.

Na aplicação aos dados da Região Metropolitana de São Paulo, algumas características dos dados exigiram adaptações e expansões para corrigir uma pequena parte dos registros onde não foi possível a obtenção da estação de destino daquela transação. Na Tabela 1 estão sumarizadas as correções efetuadas levando em consideração que:

- Apenas uma parcela das bilhetagens ocorreu no sistema de trilhos: 24,8%;
- Nem todos os cartões são “rastreáveis”, ou seja, uma parcela dos registros corresponde a cartões de operadores do sistema, como empregados do metrô: correspondem a 4,8% dos registros;
- Para uma parcela de 6,9% das bilhetagens rastreáveis do sistema de trilhos não foi possível inferir o desembarque. A explicação existente para esse percentual é de pessoas que realizaram apenas um único embarque no dia, portanto inferir o destino usando a próxima bilhetagem não foi possível;
- Após a aplicação deste filtro das bilhetagens com desembarque inferido, um percentual de 3,1% dos registros teve como estação de destino a mesma estação onde foi realizado o embarque.

Tabela 1: Descrição dos filtros aplicados aos dados do Bilhete Único de São Paulo

Filtros aplicados	Bilhetagens registradas no dia analisado	Bilhetagens no sistema de trilho	Bilhetagens com cartões rastreáveis	Bilhetagens com desembarque inferido	Bilhetagens com desembarque em outra estação
Total de bilhetagens	12 345 661	3 065 366	2 919 731	2 718 302	2 632 695
Desconsideradas	0	9 280 295	145 635	201 429	85 607
Porcentagem do conjunto anterior	100%	24.8%	95.2%	93.1%	96.9%

Para compensar essa percentagem total de 14,1% dos embarques no sistema de trilhos que não tiveram uma estação de destino válida dentro de sistema, foi aplicada uma inferência utilizando os destinos conhecidos, da seguinte forma: para cada estação de origem e para cada hora do dia, foi armazenada a lista de todas as estações de destino destes respectivos usuários. Assim, foi possível realizar uma amostragem aleatória com reposição desta lista para preencher as estações de destino desconhecidas na mesma proporção de destinos conhecidos considerando estação de embarque e hora do dia. O total de bilhetagens no sistema de trilhos com estações de destino válidas após a expansão foi de 3.065.366 bilhetagens.

A partir dos dados da bilhetagem no metrô e trens expandidos, o primeiro dos objetivos propostos no início da seção 3, tratando da obtenção da matriz origem-destino no sistema de trilhos, pode ser atingido. Como o horário dos dados da bilhetagem tem precisão de segundo, é possível agrupar o dado com qualquer intervalo de interesse. Neste trabalho propõe-se que a demanda seja agrupada em intervalos de 15 minutos e por par estação de origem - estação de destino. Ao todo foram obtidos 599.368 segmentos de demanda com esta divisão. Os resultados das análises são detalhados mais adiante na seção 4.

A fim de alcançar os outros dois objetivos propostos, de estimativa do carregamento nos trechos da rede e do cálculo do volume de transferências estimado, é necessário que seja realizada uma alocação das viagens à rede de trilhos. A alocação distribuirá a demanda entre estações nas linhas da rede, obtendo-se carregamentos nos trechos ao longo do dia. Também será possível obter o volume de transferências entre estações a partir da etapa de alocação.

Para a alocação o método escolhido foi alocar a demanda entre pares de estações utilizando o caminho na rede com o de menor tempo generalizado. O tempo generalizado considera a soma dos tempos nas etapas da viagem com fatores multiplicadores diferentes. As parcelas consideradas na formulação do tempo generalizado são o tempo de acesso ao sistema, o tempo de espera na plataforma, o tempo de viagem e tempos necessários à realização de transferências, conforme Equação 1. Os valores de β_1 , β_2 e β_3 foram escolhidos alinhados com valores relatados em pesquisas sobre valor do tempo em transporte, conforme médias apresentadas em Wardman (2004) para usuários de transporte coletivo urbanos ("urban commuters"). Embora o modelo não tenha sido calibrado para a cidade de São Paulo, os valores são provenientes de extensas pesquisas consolidadas sobre tempo de viagem que melhor retrata a visão do usuário. Para um melhor ajuste poderia ser aplicada uma pesquisa específica no sistema de trilhos da cidade, porém seriam necessárias pesquisas de campo que estão além do escopo do presente trabalho.

$$T_{\text{generalizado}} = \beta_1 * T_{\text{acesso_egresso}} + \beta_2 * T_{\text{espera}} + T_{\text{viagem}} + \beta_3 * T_{\text{transferência}} \quad (1)$$

onde: $T_{\text{generalizado}}$: tempo generalizado [minutos];

$T_{\text{acesso_egresso}}$: tempo de acesso e egresso ao sistema [minutos], adotado 2 minutos;

T_{espera} : tempo de espera pelo serviço na plataforma: metade o intervalo da linha;

T_{viagem} : tempo de viagem nos trechos percorridos [minutos]

$T_{\text{transferência}}$: tempos de transferência [minutos]

β_1 = fator de tempo de acesso e egresso: adotado 1,3;

β_2 = fator de tempo de espera: adotado 2,0;

β_3 = fator de tempo de transferência: adotado 2,0.

A rede metropolitana de transportes sobre trilhos de São Paulo é operada por três empresas: Companhia do Metropolitano de São Paulo (Metrô SP: Linhas 1, 2, 3, 5 e 15), Via Quatro (opera apenas a Linha 4 do transporte metroviário) e Companhia Paulista de Trens Metropolitanos (CPTM: Linhas 7, 8, 9, 10, 11 e 12). A CPTM opera com intervalos maiores entre os trens, enquanto o metrô possui intervalos reduzidos entre as partidas. Ao todo são 169 estações, ou seja, nós da rede, e 564 links, sendo 319 links entre estações e 245 links de transferência.

O tempo de viagem utilizado foi baseado no dado disponibilizado no site de elaboração de trajetos para o usuário do sistema do Metrô SP (Metrô SP, 2017). Foram feitas pesquisas *online* de trajetos em todas as linhas do sistema, de modo a se obter os tempos de viagem entre pares de estações adjacentes e elaborar uma base de dados com os tempos de viagem dos trechos. O tempo de viagem disponibilizado no site é único para todo o dia, assim, as rotas de menor custo generalizado são as mesmas para cada par durante todo o dia. Esta é uma simplificação necessária devido à falta de dados específicos de tempos de viagem por hora. De toda forma, o sistema de trilhos possui uma regularidade nos tempos de viagem maior que o sistema de ônibus, portanto com tempos de viagem mais regulares ao longo o dia. Para os tempos de transferência, foram feitas pesquisas em campo para obtenção dos valores nas estações: Sé, República, Barra Funda, Luz, Brás, Santo Amaro, Ana Rosa, Paraíso, Pinheiros e Tamanduateí.

3.2.1. Algoritmo de Caminho Mínimo Utilizado

A metodologia de cálculo do caminho utilizado para cada um dos pares origem-destino entre estações nos trilhos segue o conceito de *breadth-first search*, e procura elencar todas as possibilidades de combinações e caminhos entre todas as estações da rede. No total foram encontrados 2.505.485 caminhos na rede, que servem os 28.392 pares (estação de origem – estação de destino), uma vez que a rede é composta de 169 estações com 168 possibilidades de destino. Desta forma estruturou-se uma listagem com todos os caminhos possíveis na rede, e o caminho de menor tempo generalizado pode ser calculado através da inserção posterior dos tempos de viagem, acesso, egresso, tempos de espera e transferência em todas as rotas e escolhendo a de menor tempo generalizado para cada par.

A enumeração completa de todos os caminhos na rede foi realizada uma vez que a rede de transporte sobre trilhos da RMSP não apresenta um volume de ligações que crie uma grande dificuldade de processamento computacional. Assim, foi possível calcular todos os caminhos para então escolher o caminho com o menor tempo generalizado para cada par. Também poderiam ter sido usados algoritmos tradicionais de busca de caminhos em rede, como o algoritmo de Dijkstra (1959), ou outros mais recentes. Em um trabalho recente, Stasko *et al.* (2016) aplicaram o algoritmo de Dijkstra para o cálculo dos menores caminhos na rede de trilhos de Nova Iorque, utilizando uma biblioteca específica chamada *Lemon* na linguagem C++. Esta análise, utilizando um algoritmo eficiente de busca de menor caminho, é necessária já que a rede de trilhos de Nova Iorque (468 estações) é consideravelmente maior que a de São Paulo (169 estações).

A validação deste algoritmo de escolha de rotas é um desafio uma vez que seriam necessários dados adicionais não disponíveis, como volume de transferências por catraca nas estações, pesquisa com usuários sobre qual a rota que foi escolhida no sistema e um levantamento da percepção das impedâncias das transferências nas diferentes estações do sistema.

3.2.2. Estimativa do Carregamento, Lotação e Volume de Transferências

Por fim, com a listagem dos caminhos de menor tempo generalizado calculada para cada par, o carregamento é feito na rede da seguinte forma: para cada par OD entre estações, e a cada 15 minutos ao longo do dia, a demanda é alocada no caminho de menor tempo generalizado (considerado único durante o dia), carregando todos os trechos ao longo do caminho. Uma matriz de carregamentos com o trecho nas linhas e o período do dia nas colunas é preenchida.

Neste trabalho, foi tomado o cuidado de, conforme a viagem ocorre, o tempo é atualizado com os tempos de viagem e transferências. Assim, os trechos subsequentes vão ser carregados no bloco de tempo em que o usuário estaria passando. Desta forma, para viagens mais longas, os carregamentos nos trechos mais adiante no trajeto ficam carregados no horário adequado.

Para a estimativa do volume de transferências nas conexões do sistema ao longo do dia, durante a etapa de alocação são contados todos os movimentos de transferências realizados, incluindo a estação da transferência, linha de origem e linha de destino e período do dia que ocorreu, nos intervalos de 15 minutos. De forma a representar melhor o tempo de viagem nos intervalos de 15 minutos, foram considerados, para cada viagem em cada par origem destino, os respectivos tempos de viagem de forma a assegurar a consideração dos intervalos de tempos corretos. Ou seja, se uma viagem começou às 6h00 da manhã, e durou 1h, ela vai carregar os respectivos trechos nos intervalos de 15 minutos correspondentes, considerando os tempos necessários para percorrer cada trecho. Esta consideração é importante pois as viagens com origens mais distantes tendem a começar em um intervalo horário mais cedo, e as mesmas vão carregar os trechos da região central em horários mais próximos do início da atividade correspondente a esta viagem, como o trabalho. Procedendo dessa forma, o carregamento refletirá mais corretamente a distribuição da demanda espacial e temporalmente.

Após a finalização da matriz de carregamento, o cálculo da lotação pode ser feito através da capacidade do serviço naquele intervalo de tempo. Como a capacidade dos trens em metros quadrados por hora não é um dado trivial de obter, e varia ao longo do sistema e ao longo do dia, foram utilizados valores

máximos de lotação por linha divulgados pelas empresas operadoras para a imprensa (Barbosa, 2016). Desta forma, considerando que o trecho de carregamento máximo naquela linha e período do dia correspondia ao valor da lotação divulgado, todo o padrão espaço-temporal da lotação foi obtido para todas as linhas do sistema por proporção. Apenas a operadora Via Quatro não havia divulgado a lotação máxima, portanto para a Linha 4 do Metrô foi estimado o máximo em 5,0 passageiros/m².

4. RESULTADOS

A metodologia descrita anteriormente na seção 3 foi aplicada aos dados de bilhetagem eletrônica do Bilhete Único na Região Metropolitana de São Paulo. Ressalta-se que além do Bilhete Único existem outras formas de pagamento para acesso ao sistema de trilhos, como o bilhete BOM e o bilhete *Edmonson*, com pouco uso. Entretanto, os autores não tiveram acesso a estes bancos de dados. O sistema de metrô e trens de São Paulo possui 169 estações, onde ocorreram no dia de análise, 6 de abril de 2015, um total de 3.065.366 embarques utilizando Bilhete Único. Esta demanda foi agrupada em intervalos de 15 minutos para cada par origem destino (OD), totalizando 638.371 combinações. Nesta seção serão apresentados os resultados sobre a matriz origem-destino entre estações, a estimativa da lotação nos trechos da rede e o volume de transferências estimado entre estações.

4.1. Matriz Origem-Destino entre Estações

Dos 28.561 pares possíveis (entre 169 estações para as outras 168), 22.472 (78,7%) tiveram ao menos 1 usuário percorrendo cada par OD em algum intervalo do dia. A Tabela 1 mostra os cinco pares OD com maior demanda pelos dados do Bilhete Único, e a participação destes pares no volume total da rede. É possível observar que a soma da demanda dos 5 maiores pares em volume de viagens ao longo do dia soma apenas 1,02% do total das 3.056.366 bilhetagens. Para se chegar a 10% da demanda, são necessários 92 pares. Os 21.044 pares com menor demanda geram 50% do total. Isso demonstra uma alta dispersão da demanda na matriz de viagens entre estações.

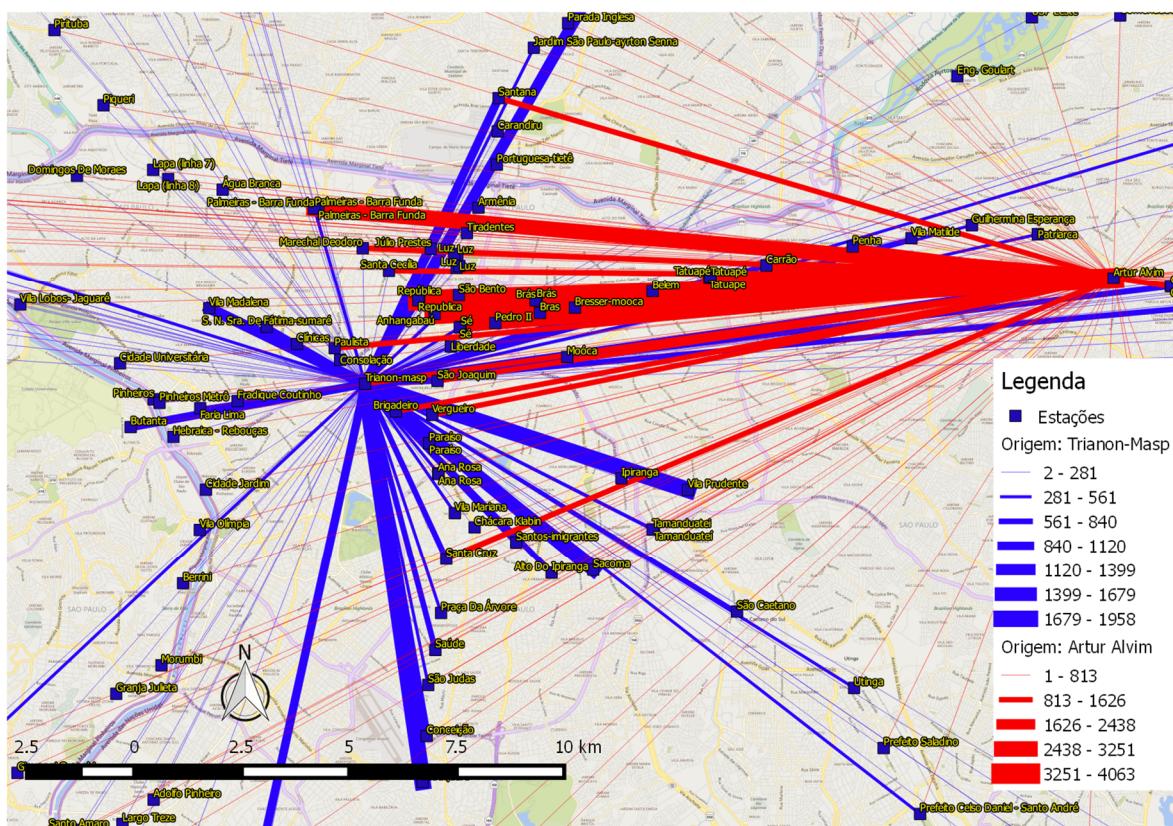


Figura 2: Volume de Viagens do dia de análise partindo da Estação Trianon-Masp e Artur Alvim.

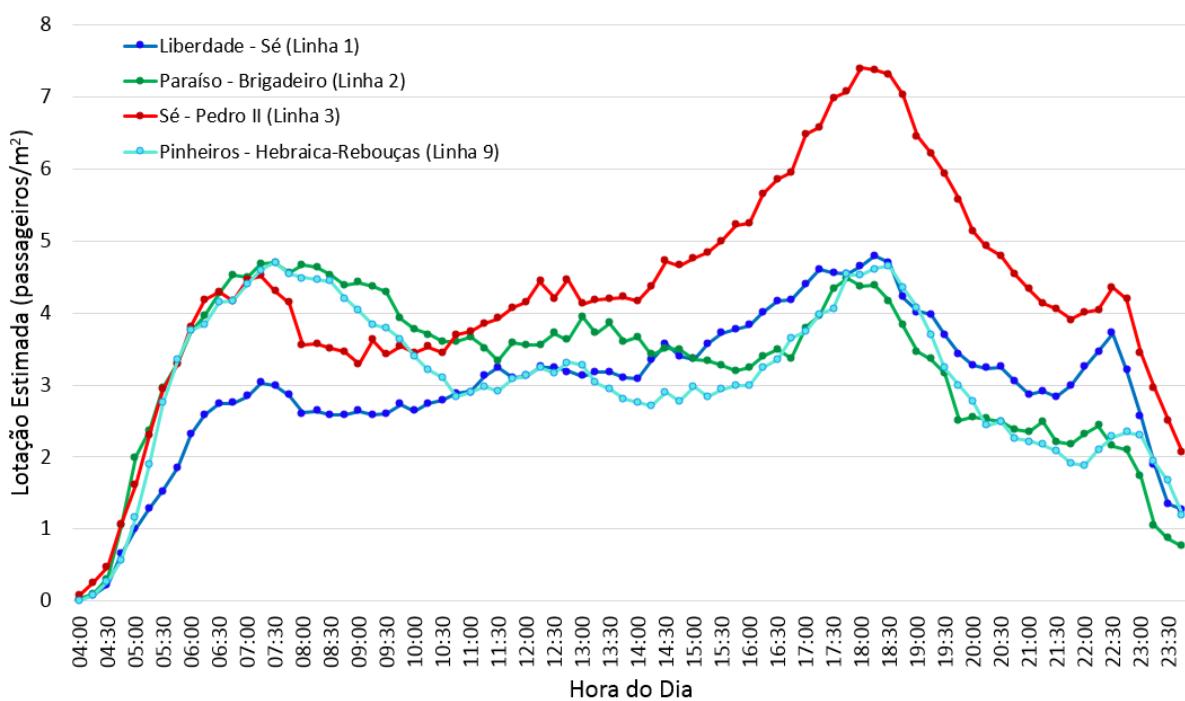
Tabela 1. Maiores Volumes de Viagens entre estações na rede de SP.

Código de Origem (Stop_id)	Código de Destino (Stop_id)	Origem	Destino	Total De Viagens no Dia	Percentual do Total
18944	18895	Tatuapé	Guaianazes	8 149	0.266%
1010054	18895	Brás	Guaianazes	6 535	0.213%
18895	18944	Guaianazes	Tatuapé	5 540	0.181%
18890	18867	Itaquera	Anhangabaú	5 466	0.178%
18867	18888	Anhangabaú	Artur Alvim	5 422	0.177%

A Figura 2 ilustra uma parcela da matriz obtida, com o volume de viagens com origens na estação Trianon-Masp (Metrô Linha 2 – Verde) e na estação Artur Alvim (Metrô Linha 3 – Vermelha), e destinos para todas as outras 168 estações do sistema. A estação Trianon-Masp é uma das estações localizadas na Avenida Paulista, área com alta concentração de atividades e a estação Artur Alvim se encontra em uma área residencial de menor renda na zona leste. É interessante ressaltar o contraste existente, já que os principais destinos da Trianon-Masp são algumas estações localizadas nos extremos das Linhas 1 e 2, como Jabaquara e Vila Prudente, enquanto que os principais destinos dos embarques na estação da zona leste é a área do centro da cidade e em polaridades da zona leste.

4.2. Estimativa da Lotação nos trechos da Rede

Com base nos dados processados, segundo a metodologia descrita na seção 3, foi obtida a estimativa da lotação nos trechos da rede de transporte sobre trilhos de São Paulo, em intervalos discretizados em períodos de 15 minutos. Para ilustrar a distribuição espacial da lotação, o mapa que representa a lotação durante o intervalo 6h45-7h00 está apresentado na seção 4.3 juntamente com o volume de transferências estimado para as estações.

**Figura 3:** Variação da Lotação ao Longo do Dia para as Linhas 1, 2, 3 do Metrô e Linha 9 da CPTM.

O gráfico apresentado na Figura 3 mostra a evolução, ao longo do dia, da lotação estimada para quatro trechos diferentes da rede. Para fins de comparação no gráfico, foram selecionados os trechos das Linhas 1, 2 e 3 do Metrô e da Linha 9 da CPTM que apresentaram maior lotação. É possível observar na Figura 3 que os padrões de variação temporal da lotação ao longo do dia são distintos para esses quatro trechos.

selecionados. É notório o pico acentuado no período de 22h30-22h45min no trecho Liberdade – Sé (Linha 1), correspondente ao período de saída de estudantes do período noturno nas universidades e faculdades localizadas nas proximidades das estações desse trecho. É importante destacar que esse resultado aparece em função da maior discretização do tempo, em intervalos menores, permitido pela coleta contínua dos dados do bilhete único.

4.3. Volumes de Transferência Estimados entre Estações

Considerando a alocação através do caminho de menor tempo generalizado, conforme detalhado na seção 3, foi possível obter o total de transferências realizadas por cartões do tipo Bilhete Único em cada uma das 16 estações do sistema onde é possível a transferência entre linhas. Na Tabela 3 estão apresentadas as dez estações com maiores volumes de transferência ao longo do dia, com base na alocação pelo caminho de menor tempo generalizado: observa-se que as estações Luz, Sé e Pinheiros são, nessa ordem, as com maior demanda de transferência. A estação da Luz contempla 2 linhas de trens (Linhas 7 e 11) e 2 linhas de metrô (Linhas 1 e 4), sendo possíveis diversas combinações de trocas. A Sé interliga as Linhas 1 e 3 do Metrô, as duas primeiras linhas de metrô construídas em São Paulo. Certamente essa situação era diferente antes da inauguração da Linha 4 do Metrô, em 2010, o que evidencia a importância da metodologia proposta, pois possibilita estimar as alterações decorrentes de ampliações na rede e alterações na oferta de transporte, sem depender de pesquisas de base domiciliar, como discutido anteriormente.

Na Figura 4 está representado o mapa com a lotação calculada para o intervalo das 6h45-7h00. Esse intervalo foi escolhido por fazer parte do período de tempo onde a média da lotação em todos os 320 trechos da rede é a mais alta no período da manhã (2,17 passageiros/m²), que vai de 6h45 até 7h30. No período da tarde, a lotação média é ainda mais alta, com média de 2,73 passageiros/m² entre 18h00-18h30. Na Figura 4 é possível observar o trecho Sé - Pedro II, o mais lotado durante esse período. Além disso, o volume de transferências mostra as estações Luz, Sé, Pinheiros, Barra Funda e Tatuapé com maiores volumes de transferências também neste período de tempo da manhã.

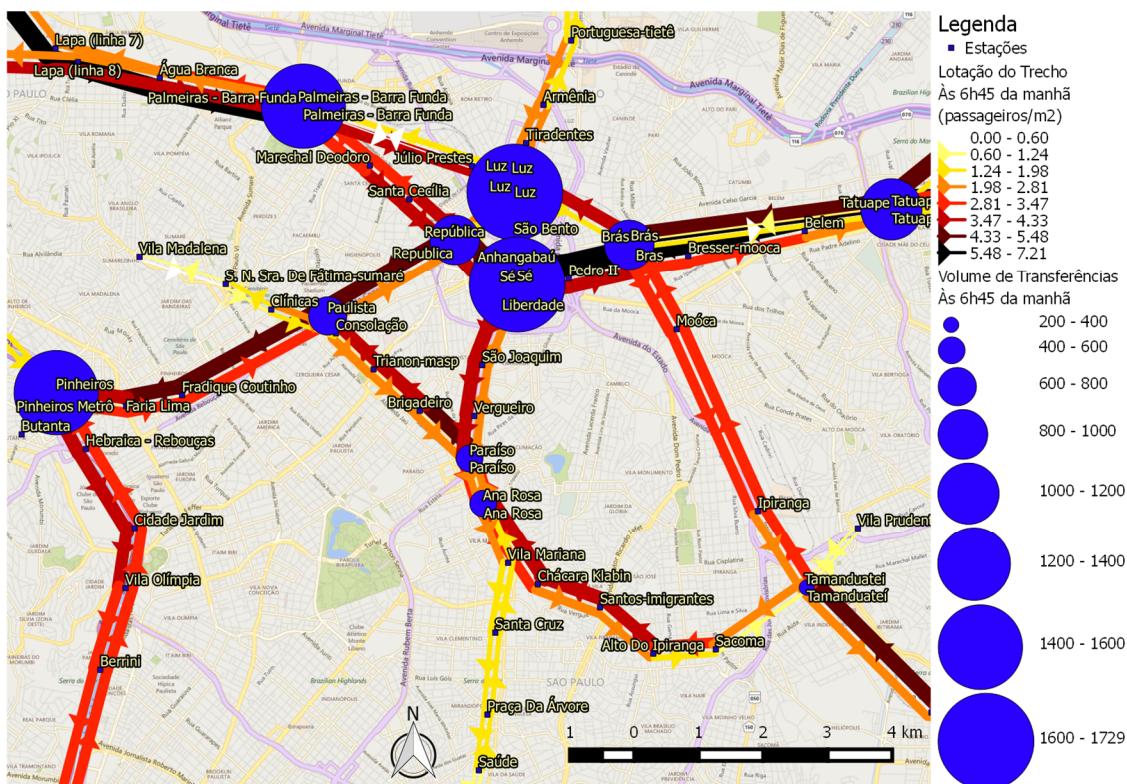


Figura 4: Mapa da Lotação e Volume de Transferências estimados às 6h45 da manhã do dia analisado.

Tabela 3: As dez estações com maiores volumes de transferência ao longo do dia

Estação	Total de Transferências no Dia inferidos pelos Dados de Bilhetagem
Luz	103374
Sé	101618
Pinheiros	93564
Palmeiras - Barra Funda	83431
Tatuapé	63959
República	61692
Paulista	49191
Brás	40785
Santo Amaro	36666
Ana Rosa	35630

CONSIDERAÇÕES FINAIS

Neste trabalho foi proposta uma adaptação de uma abordagem que permite utilizar dados de bilhetagem eletrônica para a obtenção da matriz origem-destino entre estações de um sistema metroferroviário urbano de passageiros para o caso em que os passageiros não validam seus bilhetes no destino, assim como para a estimativa da lotação nos trechos da rede e o volume de transferências entre estações. Foi possível obter, devido à maior riqueza dos dados, uma distribuição espacial e temporal desses atributos em intervalos 15 minutos ao longo de todo o dia de operação e para cada trecho da rede. Uma análise similar não seria possível utilizando métodos tradicionais de pesquisa, como por coleta de dados em campo.

Os resultados evidenciaram as especificidades do sistema de transporte sobre trilhos de São Paulo, confirmado o grande volume de viagens e consequentemente maiores lotações em trechos da rede na zona leste. A estação de maior volume de transferências, com as conexões possibilitadas pela Linha 4 do Metrô é a estação da Luz.

Concluindo, a metodologia apresentada possibilita o acompanhamento periódico e permanente da lotação e carregamentos do sistema de trilhos a baixo custo, quando comparada a métodos tradicionais de pesquisa de campo baseados em entrevistas com usuários. Esta metodologia permite inclusive avaliar a distribuição do volume de transferências realizadas pelos usuários que entraram no sistema de trilhos através do Bilhete Único ao longo dos dias. Assim, é possível avaliar a utilização do sistema de transportes em quaisquer dias do ano, avaliando eventos atípicos, como greves, reflexos de falhas na operação e eventos diversos com grande impacto na demanda, auxiliando no planejamento da operação dos serviços.

Este trabalho utiliza o grande volume de informações hoje disponíveis através de sistemas de bilhetagem eletrônica e GPS dos veículos, incluindo transações tanto no sistema de trilhos como no sistema de ônibus para obter uma matriz estimada de viagens entre estações do sistema de trilhos das transações do Bilhete Único de São Paulo. A matriz é então alocada através do método *Dijkstra* de caminho mínimo, de forma a obter a distribuição do carregamento ao longo do dia, que é então utilizado para estimar a distribuição da lotação e do aproveitamento da oferta ao longo do dia no sistema.

Os resultados indicam ainda que a metodologia utilizada pode ser um valioso recurso para validação e complementação de resultados em pesquisas mais abrangentes como as tradicionais pesquisas origem e destino tradicionais, que têm por finalidade, através de entrevistas domiciliares, investigar o padrão de viagens que as pessoas fazem diariamente numa região, incluindo os motivos e os modos de transporte dessas viagens.

As possibilidades de extensão deste trabalho são diversas. Os dados de bilhetagem eletrônica e de dados de GPS de veículos, assim como outros dados gerados passivamente pelos sistemas de transporte público, permitem uma avaliação técnica e um acompanhamento de implementação de mudanças muito mais preciso e contínuo. Estas informações possibilitam um mapeamento diário de carregamentos e da

qualidade oferecida dos serviços à população, de forma a mais rapidamente ajustar a oferta à demanda observada. Em suma, espera-se também que este artigo estimule a inovação e a pesquisa nesta área de análise de grandes volumes de dados de transporte no Brasil.

Agradecimentos

Os autores agradecem à CAPES pelo financiamento da pesquisa, à SPTrans pelos dados analisados, à CPTM e à Via Quatro pelos esclarecimentos necessários para melhor compreensão dos dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alsger, A. A., M. Mesbah, L. L. Ferreira e H. Safi (2015) Use of Smart Card Fare Data to Estimate Public Transport Origin-Destination Matrix. *Transportation Research Record: Journal of the Transportation Research Board*, (2535), 88-96. DOI: 10.3141/2535-10
- Barry, J., R. Newhouser, A. Rahbee e S. Sayeda (2002) Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record: Journal of the Transportation Research Board*, (1817), 183-187. DOI: 10.3141/1817-24
- Barry, J., R. Freimer e H. Slavin (2009) Use of entry-only automatic fare collection data to estimate linked transit trips in New York City. *Transportation Research Record: Journal of the Transportation Research Board*, (2112), 53-61. DOI: 10.3141/2112-07
- Barbosa, B. (2016) Alívio na lotação de Metrô e CPTM deve ficar para 2017. UOL Notícias Cotidiano. Disponível em: <<http://noticias.uol.com.br/cotidiano/ultimas-noticias/2016/01/11/alivio-na-lotacao-de-metro-e-cptm-deve-ficar-para-2017.htm>>. Acesso em: 22 de junho de 2016.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1), 269-271. DOI: 10.1007/BF01386390
- Feriancic, G., F. R. Celeiro e L. N. B. Silva (2015) Planejamento da Mobilidade com Big Data de Telefonia Móvel. 20º Congresso Brasileiro de Transporte e Trânsito, Santos.
- Ferraz, A. C. P. e I.G.E. Torres. (2004) *Transporte público urbano*. RiMa Editora.
- Google Transit (2016) Google Transit, documento de referência para desenvolvedores. Disponível em: <<https://developers.google.com/transit/gtfs/reference>>. Acesso em: 1 de julho de 2016.
- Lin, M. e Hsu, W-J. (2014). Mining GPS data for mobility patterns: A survey. *Pervasive and Mobile Computing*, 12, p.1-16. DOI: 10.1016/j.pmcj.2013.06.005.
- Kim, K. M., S. P. Hong, S. J. Ko e D. Kim (2015) Does crowding affect the path choice of metro passengers?. *Transportation Research Part A: Policy and Practice*, 77, 292-304. DOI: 10.1016/j.tra.2015.04.023
- Metrô SP (2017). Site oficial do Metrô de São Paulo. <http://www.metro.sp.gov.br/>. Acesso em 13 de março de 2017.
- Munizaga, M. A. e C. Palma (2012) Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18. DOI: 10.1016/j.trc.2012.01.007
- Nassir, N., A. Khani, S. Lee, H. Noh e M. Hickman (2011) Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system. *Transportation Research Record: Journal of the Transportation Research Board*, (2263), 140-150. DOI: 10.3141/2263-16
- Ortúzar, J. e L. G. Willumsen (1994) *Modelling transport*. New Jersey: Wiley. DOI: 10.1002/9781119993308
- Pelletier, M. P., M. Trépanier e C. Morency (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568. DOI: 10.1016/j.trc.2010.12.003
- Rodrigues, M. A e Serratini, J. A. A qualidade no transporte coletivo urbano. In: XXII Congresso de Pesquisa e Ensino em Transportes, 2008, Fortaleza, CE. Panorama Nacional da Pesquisa em Transportes 2008. XXII ANPET. Rio de Janeiro, p. 1081-1092.
- Stasko, T., B. Levine e A. Reddy (2016) A Time-Expanded Network Model of Train-level Subway Ridership Flows Using Actual Train Movement Data at New York City Transit. In *Transportation Research Board 95th Annual Meeting* (No. 16-4090). DOI: 10.3141/2540-11
- Tirachini, A., D. A. Hensher e J.M. Rose (2013) Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation research part A: policy and practice*, 53, 36-52. DOI: 10.1016/j.tra.2013.06.005
- Wardman, M. (2004) Public transport values of time. *Transport policy*, 11(4), 363-377. DOI: 10.1016/j.tranpol.2004.05.001