

## Visualização de Dados de Origem-Destino - Foco em Unidades de Saúde e Educação

Fernando X. De Souza<sup>1</sup>, Paulo R. Bauer<sup>1</sup>, Keiko Fonseca<sup>1</sup>, Tatiana Gadda<sup>1</sup>,  
Rita Berardi<sup>1</sup>, Nádía P. Kozievitch<sup>1</sup>

<sup>1</sup>Universidade Tecnológica Federal do Paraná (UTFPR)  
Avenida Sete de Setembro, 3165  
Departamento Acadêmico de Informática – DAINF – Curitiba – Brasil

{fernando.1994,paulobauer}@alunos.utfpr.edu.br,  
{keiko,tatianagadda,ritaberardi,nadiap}@utfpr.edu.br

**Abstract.** *Understanding the dynamics of displacement of people in the public transport system in order to improve this service is a constant challenge in large urban centers. This article presents a visualization tool for origin-destination data (using heatmap, clustering and flux map), with focus on health and education units (citizens attractions and traffic generators). A preliminary evaluation using data from Curitiba showed positive results from the users point of view.*

**Resumo.** *Compreender a dinâmica de deslocamento das pessoas no sistema de transporte público visando melhorar este serviço é um desafio constante em grandes centros urbanos. Este artigo apresenta uma ferramenta de visualização para dados de origem-destino (usando mapa de calor, clusterização e mapa de fluxo), com foco em unidades de saúde e educação (atratoras de cidadãos e geradoras de tráfego). Uma avaliação preliminar da aplicação usando dados de Curitiba mostrou resultados positivos do ponto de vista de usuários.*

### 1. Introdução

O crescimento das cidades traz desafios relacionados à mobilidade urbana, e técnicas que envolvem processos automatizados, onde é possível produzir, consumir e distribuir grandes quantidades de informação em tempo real [Lemos 2013] vêm sendo utilizadas para propor novas soluções. Nesse cenário, a administração pública de cidades como Paris<sup>1</sup>, Nova Iorque<sup>2</sup> e Moscou<sup>3</sup>, disponibilizam seus dados seguindo a tendência de dados abertos. Na mesma tendência, a cidade de Curitiba disponibiliza seus dados através de várias fontes (Instituto de Pesquisa e Planejamento Urbano (IPPUC)<sup>4</sup>, Urbanização de Curitiba (URBS)<sup>5</sup> e dados abertos da Prefeitura<sup>6</sup>). Em Curitiba, diariamente, uma média de 1229 veículos transportam aproximadamente 1.365.615 passageiros (sendo 60% deles usando cartões inteligentes) por 251 rotas diferentes e 24 terminais de ônibus, resultando, em média, 14.166 viagens de ônibus em 300.373 km por dia, em 6500 pontos de ônibus.

<sup>1</sup>[opendata.paris.fr](http://opendata.paris.fr) - Acessado em Mar.2020.

<sup>2</sup>[opendata.cityofnewyork.us](http://opendata.cityofnewyork.us) - Acessado em Mar. 2020.

<sup>3</sup>[data.gov.ru](http://data.gov.ru) - Acessado em Mar. 2020.

<sup>4</sup>[ippuc.org.br](http://ippuc.org.br) - Acessado em Mar.2020.

<sup>5</sup>[urbs.curitiba.pr.gov.br/](http://urbs.curitiba.pr.gov.br/) - Acessado em Mar.2020.

<sup>6</sup>[www.curitiba.pr.gov.br/dadosabertos/](http://www.curitiba.pr.gov.br/dadosabertos/) - Acessado em Mar.2020.

Do ponto de vista de dados e geoprocessamento, o sistema de transporte público possui componentes geográficos e temporais, além de dados sobre sua dinâmica, como por exemplo, viagens. Uma viagem de ônibus (na qual um usuário vai de uma origem a um destino e volta a uma origem) pode ser caracterizada por datas e coordenadas de partidas e chegadas (dados de destino de origem) associados a uma linha de ônibus (rota predeterminada). Uma viagem também pode ter informações como nome e código de linha, código do veículo, ID do cartão inteligente, sexo e data de nascimento do passageiro e latitude e longitude do ônibus (por exemplo, a cada cinco minutos).

Relatórios de origem-destino (OD) como [IPPUC 2017] já indicaram a importância de pontos de interesse, como entidades de saúde e educação (que concentram uma grande quantidade de população), para entender a dinâmica destes dados. Entretanto, não há nenhuma ferramenta desenvolvida para os dados de Curitiba que aborde as variáveis citadas (espaciais/temporais) permitindo compreender a dinâmica do transporte na cidade. A partir das demandas identificadas em sua pesquisa, [Parcianello 2020] desenvolveu uma interface para auxiliar na análise do deslocamento dos passageiros. Este artigo tem como objetivo expandir as opções de análises do trabalho desenvolvido por [Parcianello 2020], por meio de uma extensão de sua ferramenta, incluindo dados da localização de Pontos de Interesse (POI) como escolas e hospitais (equipamentos públicos e privados que concentram população) junto à adaptações da interface para consulta e apresentação desses dados.

## 2. Trabalhos Relacionados

O transporte público é uma das áreas mais críticas sob a perspectiva da cidade. Os desafios de mobilidade já ganharam atenção da comunidade científica no Brasil <sup>7</sup>. Nesta direção, diferentes serviços telemáticos de transporte já foram propostos em [Diniz Jr. 2017] (como localização de rotas sem ônibus, média de lotação em ônibus, alerta para diferentes rotas e alertas de velocidade), usando os mesmos dados propostos neste artigo. Já dados do Rio de Janeiro foram usados em [Cruz et al. 2018] para a identificação e a classificação de anomalias através de um processo que envolve integração de *open data* e mineração de dados via uso do algoritmo *Apriori*. A pesquisa de [Spadon et al. 2018] também compara dados abertos da rede viária, da demografia, da extensão territorial e de outros indicadores urbanos de 645 cidades do estado de São Paulo. Este último estudo também aplica conceitos de redes complexas e algoritmos de clusterização para classificar tais cidades por semelhança sob diferentes perspectivas.

Por outro lado, as imagens no geoprocessamento são importantes mas no transporte o movimento é crucial (exemplos em [Andrienko and Andrienko 2013]). Como já mencionado em [Ferreira et al. 2013]: grande parte do trabalho é baseada em dados de trajetória. Já em OD, os dados multivariados têm apenas o início e o posições finais juntamente com atributos associados ao movimento e ao tempo gasto. No mesmo artigo, a visualização de dados de táxi é estudada e uma solução é proposta para não especialistas, com diferentes abordagens de visualização. Ao se tratar de apresentação de dados, vários projetos podem ser mencionados, como DataViz<sup>8</sup>, VisualComplexity<sup>9</sup> ou

<sup>7</sup><http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/802-grandesdesafiosdacomputao/no-brasil>- Acessado em Jun. 2020.

<sup>8</sup>[datavizproject.com](http://datavizproject.com) - Acessado em Jun. 2020.

<sup>9</sup>[www.visualcomplexity.com](http://www.visualcomplexity.com) - Acessado em Jun. 2020.

CityGeographics<sup>10</sup>. Por outro lado, dados abertos já estão sendo disponibilizados com soluções interativas, como Manhattan Population Explorer<sup>11</sup> ou Transit Accident Dashboard from IPPUC<sup>12</sup>. Algoritmos de clusterização ou gráficos (como marcadores, mapa térmico) também podem ser usados na visualização, como em [Vila 2016], onde é proposta uma solução web-mobile que permite a visualização espaço-temporal de dados georreferenciados.

### 3. O Protótipo

Os dados de OD utilizados no protótipo são os dados mencionados em [Parcianello 2020] (e detalhes mais específicos podem ser verificados em [Diniz Jr. 2017]), e os dados de saúde (hospitais, laboratórios, CAPS, unidade de saúde) são do portal de dados abertos de Curitiba. Para o armazenamento dos dados, foi utilizado PostgreSQL<sup>13</sup> e Postgis<sup>14</sup>. Os pontos de interesse escolhidos foram equipamentos urbanos da cidade: transporte (pontos de ônibus e terminais), educação (escolas de educação básica, de jovens e adultos, especial, infantil, profissional técnica, superior, ensino médio).

A ferramenta em [Parcianello 2020] é dividida em duas partes: busca e visualização (usando clusterização e mapa de calor) e o painel com gráficos (total de embarques e desembarques por faixa etária, total de embarques por sexo, embarques por sexo ao longo do período, total de embarques por idade, e embarques por faixa etária ao longo do período). Este trabalho adicionou na visualização o mapa de fluxo, técnica de visualização que ilustra as viagens realizadas através de setas no mapa ligando os pontos de origem e destino das viagens, além da adição de dados de localização de instituições de saúde e educação como Pontos de Interesse.

A Figura 1 mostra um diagrama de como as tecnologias foram combinadas para o desenvolvimento do protótipo. Na arquitetura original desenvolvida por [Parcianello 2020] foram utilizadas as bibliotecas Bootstrap, Leaflet, JQuery e Open Street Map para o lado do cliente, e neste trabalho foi adicionado a biblioteca Leaflet.Migration.Js para utilização de mapas de fluxo. Para o lado do servidor foram utilizadas tecnologias Apache e PHP, além da base de dados.

Na prática, a interface foi modificada nos seguintes itens: a Figura 2-A mostra a nova opção para selecionar a categoria do tipo de Ponto de Interesse a ser pesquisado, na Figura 2-B é possível selecionar uma subcategoria para afinar os resultados, na Figura 2-C foi adicionada uma nova opção para utilizar a técnica de mapa de fluxo na exibição dos resultados da pesquisa, e por fim na Figura 2-D encontra-se os botões para realizar a pesquisa, configurar e retirar os resultados da última pesquisa do mapa.

Como exemplo, considere as viagens realizadas no dia 01 de Outubro de 2017, entre as 0:00 até às 23:59, tendo como ponto de destino as instituições "UTFPR - Sede Central" e "UFPR - Setor de Ciências Sociais Aplicadas", ilustrada na Figura 3. As setas verdes ilustram quais as que tiveram como destino a UTFPR, e as setas azuis quais foram

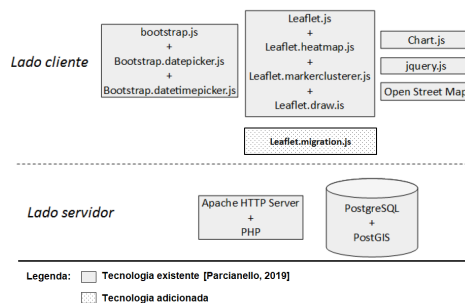
<sup>10</sup>[citygeographics.org](http://citygeographics.org) - Acessado em Jun. 2020.

<sup>11</sup>[manpopex.us](http://manpopex.us) - Acessado em Jun. 2020.

<sup>12</sup>[ippuc.org.br/mapasinterativos/AcidentesDeTransito/dashboard.html](http://ippuc.org.br/mapasinterativos/AcidentesDeTransito/dashboard.html) - Acessado em Jun. 2020.

<sup>13</sup><https://www.postgresql.org/> - Acessado em Jun 2020.

<sup>14</sup><https://postgis.net/> - Acessado em Ago 2020.



**Figura 1. Tecnologias utilizadas no protótipo.**

destinadas à UFPR.

A Figura 4-esquerda apresenta as instituições de ensino superior que obtiveram em suas imediações (400 metros) a maior concentração de usuários como destino, entre os dias 23 a 29 de Outubro de 2017. A Figura 4-direita indica a quantidade de usuários que tiveram como destino as imediações de cada Instituição. Note que a instituição Faculdades Integradas Camões possui em suas imediações a maior quantidade de usuários.

Com o intuito de avaliar de maneira preliminar a aplicação, foi desenvolvido um teste contendo 3 tarefas, aplicado a um grupo de cinco pessoas no período entre os dias 11/11/2019 a 20/11/2019. O perfil dos entrevistados é de estudantes da instituição UTFPR, graduandos e mestrands de computação. Os testes foram realizados separadamente, sob observação direta, com os entrevistados tendo acesso à aplicação, às perguntas e a um mapa da cidade de Curitiba. As tarefas escolhidas foram: (1) Descobrir a origem das pessoas que chegaram de ônibus nas redondezas da UTFPR-Centro, no dia 04 de Outubro de 2017; (2) Descobrir para que bairro mais foram e a porcentagem de pessoas do sexo feminino que utilizaram ônibus na região do Terminal Guadalupe na semana de 8 a 14 de Outubro de 2017 e (3) Descobrir quantas pessoas utilizaram ônibus no Centro com destino à todas as instituições de saúde Especializadas em Saúde Mental, do dia 01 ao dia 31 Outubro de 2017. Como resultados, todos os entrevistados conseguiram resolver as 3 tarefas, apresentando respostas compatíveis com o que era esperado de cada tarefa. Um vídeo da aplicação pode ser encontrado em <sup>15</sup>.

**Lições aprendidas.** Vários fatores devem ser considerados para projetar e impactar os protótipos OD: 1) os filtros e sua classificação (quais são relevantes para o usuário, quão fáceis são seu uso e implementação, e quais resultados são mais bem compreendidos); 2) o usuário final e quais restrições e conhecimentos de tecnologia estão presentes (percepções de como a experiência do usuário com as tecnologias aplicadas impacta na compreensão dos resultados); 3) a visualização geral dos dados e quais estatísticas básicas impactam o usuário (em que visão geral o usuário final está interessado); 4) variações de movimento e região (intra-região, regional, variações menores (como K-Means)); 5) a análise da arquitetura, dados, estrutura de banco de dados e consultas a fim de melhorar a otimização (técnicas como otimização de consulta, índices, partições de mesa); 6) teste e integração de diferentes bibliotecas (às vezes não disponíveis gratuitamente, ou difíceis de integrar, ou indisponíveis para diferentes sistemas operacionais).

<sup>15</sup><https://youtu.be/aj7-TMSBXO4> - Acessado em Jun.2020.

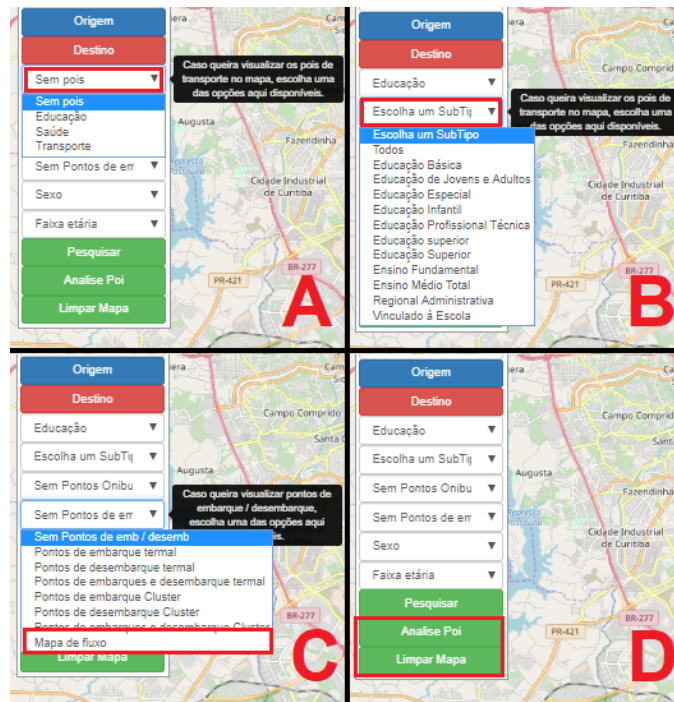


Figura 2. Novidades apresentadas na interface.

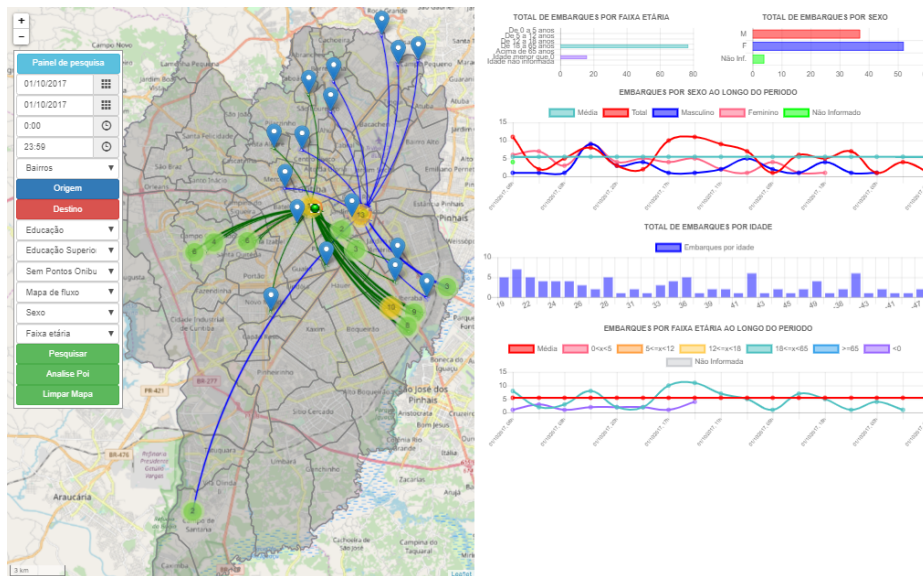
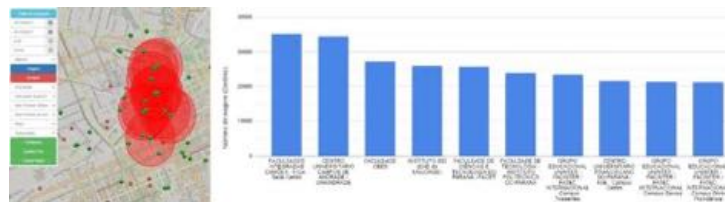


Figura 3. Resultado da consulta de origem-destino.

#### 4. Conclusão

Este artigo apresentou uma aplicação de visualização de dados de Origem-Destino, com um painel de busca e visualização (mapa de calor, clusterização e mapa de fluxo) e um painel de análise (categorização dos dados em gênero, horário e idade). O objetivo é auxiliar no estudo dos padrões de deslocamento dos usuários do transporte coletivo, entre



**Figura 4. Instituições de Ensino Superior que obtiveram em suas imediações a maior quantidade de usuários como destino (esquerda), e a quantidade respectiva de cada uma (direita).**

equipamentos urbanos de transporte, saúde e de educação. Uma avaliação preliminar obteve uma aceitação positiva, onde todos os usuários amostrados conseguiram concluir as tarefas propostas. Como trabalhos futuros, podemos listar a inclusão de outros filtros de pesquisa, as rotas e horários de maior fluxo de pessoas, aplicar técnicas de clusterização ao mapa de fluxo e inserir uma interface que permita realizar consultas diretamente em SQL.

## Referências

- Andrienko, N. and Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24.
- Cruz, A., Ferreira, J., Carvalho, D., Mendes, E., Pacitti, E., Coutinho, R., Porto, F., and Ogasawara, E. (2018). Detecção de anomalias frequentes no transporte rodoviário urbano. In *SBBB: Brazilian Symposium on Databases*, pages 271–276. SBC.
- Diniz Jr., P. C. (2017). Serviços telemáticos em uma rede de transporte público baseados em veículos conectados e dados abertos. Master’s thesis, Universidade Tecnológica Federal do Paraná.
- Ferreira, N., Poco, J., Vo, H. T., Freire, J., and Silva, C. T. (2013). Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158.
- IPPUC (2017). Consolidação de dados de oferta, demanda, sistema viário e zoneamento - relatório 5 - pesquisa de origem-destino domiciliar. disponível em [http://www.ippuc.org.br/visualizar.php?doc=http://admsite2013.ippuc.org.br/arquivos/documentos/D536/D536\\_002\\_BR.pdf](http://www.ippuc.org.br/visualizar.php?doc=http://admsite2013.ippuc.org.br/arquivos/documentos/D536/D536_002_BR.pdf). Technical report, IPPUC, Curitiba.
- Lemos, A. (2013). Cidades inteligentes. *GV-executivo*, 12(2):46–49.
- Parcianello, Y. (2020). Análise de origem-destino do uso do sistema de transporte coletivo de Curitiba sob o ponto de vista de regions of interest. Master’s thesis, Universidade Tecnológica Federal do Paraná.
- Spadon, G., Scabora, L. C., Nesso-Jr, M. R., Traina Junior, C., and Rodrigues Junior, J. F. (2018). Caracterização topológica de redes viárias por meio da análise de vetores de características e técnicas de agrupamento. In *SBBB’18*, pages 157–168. SBC.
- Vila, J. J. F. R. (2016). Clusterização e visualização espaço-temporal de dados georreferenciados adaptando o algoritmo marker clusterer: um caso de uso em Curitiba. Master’s thesis, Universidade Tecnológica Federal do Paraná, Campus Curitiba.