

Integração de Dados Públicos de Saúde e Transporte: Caracterização para Modelagem Multicamadas

Priscila L. L. Santin¹, Anelise Munaretto¹, Mauro S. P. Fonseca¹, Thiago H. Silva²

¹Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial (CPGEI)

²Programa de Pós-Graduação em Computação Aplicada

Universidade Tecnológica Federal do Paraná (UTFPR)
80.230-901 – Curitiba – PR – Brasil

{priscilasantin, anelise, maurofonseca, thiagoh}@utfpr.edu.br

Abstract. *The abundance of data from different contexts has been driving a revolution in the digital data industry. The information from these data can help public managers to provide better assistance to their citizens in different areas such as health care and urban mobility. Problems in the quality of public transport affect people's ability to travel, making it difficult for citizens to access basic health services provided by governments. This paper presents the characterization of the public transport and health system of Curitiba aiming at the employability of a MultiAspect Graph (MAG) data structure for the integration of these systems.*

Resumo. *A abundância de dados de diferentes contextos vem impulsionando uma revolução no setor de dados digitais. As informações provenientes desses dados podem auxiliar gestores públicos a prover melhor assistência aos seus cidadãos em diferentes áreas, como no cuidado à saúde e a mobilidade urbana. Problemas na qualidade do transporte público afetam a capacidade de deslocamento das pessoas, dificultando a acessibilidade dos cidadãos aos serviços básicos de saúde disponibilizados pelos governos. Este trabalho apresenta a caracterização do sistema público de transporte e de saúde de Curitiba visando a empregabilidade de uma estrutura de dados MultiAspect Graph (MAG) para a integração desses sistemas.*

1. Introdução

A quantidade de dados gerados, coletados e compartilhados por administrações públicas, empresas, entidades sem fins lucrativos e até mesmo pela comunidade científica vem aumentando consideravelmente nos últimos anos [Agarwal e Dhar 2014]. A análise dessas informações digitais e a inteligência para usá-las têm proporcionado uma significativa diferença nos negócios e na forma de viver das pessoas [Labrinidis 2012][Mervis 2012][Nature Editorial 2008]. Alguns trabalhos propõem maneiras pelas quais os governos podem utilizar-se de dados estratégicos para prover melhor atendimento aos seus cidadãos e lidar com outros desafios, tais como custos com cuidados de saúde, criação de emprego, desastres naturais e mobilidade urbana [Manyika et. al. 2011] [Pena 2018].

O debate sobre a mobilidade urbana no Brasil tem sido um tópico bastante discutido entre administradores públicos de grandes cidades. A mobilidade urbana refere-

se à facilidade de deslocamento da população e de bens no espaço geográfico das cidades, em função das complexas atividades nelas desenvolvidas, constituindo um componente da qualidade de vida aspirada por seus habitantes [Araújo et. al. 2011]. Segundo Pena (2018), a principal causa dos problemas de mobilidade urbana no Brasil relaciona-se ao aumento do uso de transportes individuais em detrimento da utilização de transportes coletivos, embora esses últimos também encontrem dificuldades com a superlotação. Neste contexto, problemas de infraestrutura e de qualidade do transporte comprometem a mobilidade e a capacidade de deslocamento das pessoas, dificultando a acessibilidade, ou seja, a facilidade em distância, tempo e custo em alcançar fisicamente os destinos desejados [Araújo et. al. 2011], tais como as unidades de saúde públicas.

Atualmente menos de um quarto da população brasileira possui plano de saúde privado [Oliveira 2017], o que significa que a maioria da população faz uso da rede pública de atenção à saúde. Dessa maneira, entender como acontece o deslocamento da população das grandes cidades até as unidades de saúde poderia contribuir com propostas que facilitariam o atendimento a esses cidadãos no que diz respeito à saúde, evitando, inclusive, que os mesmos utilizem o sistema de transporte público fazendo deslocamento desnecessário. Assim, este trabalho analisa o uso do transporte público e das unidades de saúde de Curitiba por meio dos registros dos usuários desses dois domínios. Para o transporte público foram analisados os registros dos cartões transporte (bilhetagem eletrônica) e o deslocamento da frota do sistema. A junção dessas informações foi realizada com o objetivo de aprimorar o entendimento dos dados estudados. No domínio da saúde pública, foram analisados os registros de atendimentos médicos das unidades de saúde da cidade.

Uma caracterização foi realizada nas bases de dados consideradas neste estudo de forma a entender melhor as questões referentes à utilização dos sistemas em questão, tais como o uso por períodos e faixas etárias, além da dinâmica de deslocamento por meio do sistema de transporte público de Curitiba. Entende-se que é fundamental verificar as propriedades desses sistemas para uma adequada integração de dados. Diante da caracterização realizada, o trabalho propõe a empregabilidade de uma estrutura de dados *MultiAspect Graph* (MAG) para analisar a dinâmica dos sistemas. Acredita-se que essa modelagem é fundamental para auxiliar na proposição de melhorias na mobilidade e no atendimento à saúde da população de grandes cidades, além de possibilitar a geração de dados sintéticos mais próximos da realidade dos sistemas estudados, possibilitando a simulação de usos em diversos cenários.

O restante do trabalho está organizado da seguinte forma. A seção 2 descreve alguns dos principais trabalhos relacionados, enquanto na seção 3 são apresentados os dados utilizados no trabalho e como foi realizado o tratamento desses dados. A seção 4 apresenta a análise e caracterização desses dados. A seção 5 apresenta a empregabilidade do MAG no modelo multicamadas incluindo a modelização do sistema nas bases estudadas. Por fim, a seção 6 apresenta as considerações finais.

2. Trabalhos Relacionados

A caracterização de dados se mostra relevante tanto para se obter um melhor entendimento dos sistemas analisados, bem como para aprimorar a qualidade do atendimento aos seus usuários. Nessa direção, o trabalho de Braz et. al. (2018) utiliza as bases do transporte público de Curitiba para estimar a ineficiência nos tempos de viagem

realizadas no sistema de ônibus da cidade. Já o trabalho de Ahmad (2019) apresenta um estudo direcionado à população idosa do Paquistão de modo a compreender as características de mobilidade existentes dos idosos, suas necessidades e fatores limitantes. O trabalho de Zhang et. al. (2018) propõem um quadro analítico para comparar os padrões de mobilidade urbana extraídos de duas fontes de dados, os cartões de transporte do sistema de transporte público de Singapura e as trajetórias dos táxis dessa cidade. Resumidamente, o estudo demonstra a importância de análise comparativa de dados para pesquisa urbana e de transporte.

Com a caracterização de dados é possível implementar de forma mais adequada modelos que representem os sistemas estudados. Nesse contexto, os modelos multicamadas têm sido empregados em diferentes estudos com esse objetivo. Kurant e Thiran (2006) apresentam em seu trabalho um modelo em camadas para facilitar a descrição e a análise de pequenas redes complexas, que são partes de sistemas maiores, onde uma série de topologias coexistentes interage e depende um do outro. Já Hristova et. al. (2016) apresentam uma rede interconectada de pessoas e lugares. Foram definidas quatro métricas pelos autores que concluem que a análise feita por eles espelha a relação entre a prosperidade das pessoas e dos lugares. O trabalho de Rodrigues et. al. (2017) propõem uma ferramenta chamada *SMAFramework* para analisar padrões de mobilidade baseados em MAG. Os autores utilizaram-se de duas fontes de dados diferentes e propuseram um algoritmo para análise de mobilidade.

Este presente trabalho se diferencia dos anteriores em diversos aspectos. Foi feita uma caracterização de dois importantes sistemas públicos de uma grande cidade brasileira, o que pode ser útil, por exemplo, para a geração de dados sintéticos mais realistas. Além disso, a caracterização permite encontrar de maneira mais apropriada a relação entre os dados analisados. Por exemplo, a dimensão temporal se mostra essencial para o problema estudado, indicando que a construção do modelo multi-aspecto MAG temporal é mais adequada. Esse também é um ponto de diferenciação de trabalhos relacionados anteriores, pois nenhum leva esse aspecto em consideração no modelo multicamadas.

3. Descrição dos Conjuntos de Dados

Nesta seção descrevemos as bases de dados estudadas, sua disponibilização e processamentos dos dados realizados para a busca do conhecimento.

3.1. Bases de Dados Referentes ao Transporte Público

O controle do sistema de transporte público de Curitiba é feito por uma empresa de economia mista, a URBS (Urbanização de Curitiba). É a URBS quem disponibiliza as bases de dados referentes ao transporte público utilizadas neste estudo.

3.1.1. Base do Cartão Transporte

A base de dados dos usuários do cartão transporte de Curitiba é uma base de dados disponibilizada exclusivamente para convênios com universidades e outras instituições conveniadas à URBS. Por meio desta base de dados são disponibilizados diariamente os

registros dos cartões transporte utilizados no dia anterior em arquivos *.json*¹. As informações contidas nessa base de dados estão distribuídas em 7 campos: Código da Linha, Nome da Linha, Código do Veículo, Número do Cartão, Data de Utilização, Data de Nascimento e Sexo.

3.1.2. Base do Transporte Público de Curitiba

Disponibilizada pela URBS para o público em geral, trata-se de uma base com diversos arquivos de informações, tais como as linhas de ônibus existentes e os pontos de ônibus que as linhas atendem, com GPS e a tabela horária dos veículos. Há dados estáticos, atualizados uma vez ao dia, e outros atualizados praticamente em tempo real. Todas essas informações são disponibilizadas por meio de arquivos *.json*. Neste estudo foram utilizados dois arquivos referentes ao transporte público de Curitiba: (I) *Veículos* – arquivo com a localização de todos os veículos que circularam por Curitiba, com informações do Código do Veículo, Latitude, Longitude, Data de Utilização e Código da Linha; (II) *Pontos* – relação de todos os pontos de ônibus existentes em Curitiba com informações como Nome do Ponto, Número do Ponto, Latitude, Longitude, Sequência, Grupo, Sentido, Tipo, Código da Linha.

3.1.3. Informações das Bases Referentes ao Transporte Público

A Tabela 1 apresenta um resumo das informações contidas nas bases de dados do transporte público estudadas neste trabalho.

Tabela 1 – Resumo das informações das bases do transporte público

<i>Descrição</i>	<i>Valor</i>
Número de leituras de cartões	22.703.308
Quantidade de cartões distintos	390.383
Quantidade de linhas operantes	407
Quantidade de pontos de ônibus	6.770

3.1.4. Aprimoramento dos Dados do Transporte Público

Alguns pré-processamentos foram realizados para que uma melhor análise do uso do transporte público fosse possível.

Como principal base de dados de estudos foi utilizada a base de dados dos cartões transporte, porém um aprimoramento foi feito a essa base acrescentando as coordenadas latitude e longitude, por meio do cruzamento das informações das bases apresentadas nessa seção. Além disso, com as informações de latitude e longitude inseridas foi possível identificar os bairros de registro do cartão transporte. Calculou-se, por meio da fórmula de *Haversin*² o ponto mais próximo da linha utilizada com base nas informações geoespaciais. Com as informações da latitude e longitude do ponto mais próximo, buscou-se por meio da função *geocode* do *Google Maps API*³ o bairro de localização daquele ponto.

¹ *JavaScript Object Notation*. Formato compacto, de padrão aberto independente, de troca de dados simples e rápida entre sistemas, que utiliza texto legível a humanos, no formato atributo-valor.

² $haversin\left(\frac{d}{R}\right) = haversin(\Delta\phi) + \cos(\phi_1) \cos(\phi_2) haversin(\Delta\lambda)$

³ <https://cloud.google.com/maps-platform/?hl=pt-BR>

3.2. Base de Dados da Saúde

Para atendimento médico, Curitiba possui atualmente 6 centros de especialidades, 112 unidades municipais de saúde (UMS), 9 unidades de pronto atendimento (UPA) e um hospital de atendimento exclusivo para idosos.

Obedecendo a Lei da Transparência⁴, a Prefeitura Municipal de Curitiba (PMC) disponibiliza na *Internet* a base de dados referente ao Sistema E-Saúde Médicos que contém o registro dos atendimentos médicos nas unidades municipais de saúde. Essa base é disponibilizada mensalmente e contempla o registro dos atendimentos dos três meses anteriores à sua publicação em formato *.csv*⁵.

3.2.1. Campos de Preenchimento das Bases

Cada registro da base possui 38 campos de preenchimento: Data do Atendimento, Data de Nascimento, Sexo, Código do Tipo de Unidade, Tipo de Unidade, Código da Unidade, Descrição da Unidade, Código do Procedimento, Descrição do Procedimento, Código do CBO, Descrição do CBO, Código do CID, Descrição do CID, Solicitação de Exames, Qtde Prescrita Farmácia Curitibana, Qtde Dispensada Farmácia Curitibana, Qtde de Medicamento Não Padronizado, Encaminhamento para Atendimento Especialista, Área de Atuação, Desencadeou Internamento, Data do Internamento, Estabelecimento Solicitante, Estabelecimento Destino, CID do Internamento, Tratamento no Domicílio, Abastecimento, Energia Elétrica, Tipo de Habitação, Destino Lixo, Fezes/Urina, Cômodos, Em Caso de Doença, Grupo Comunitário, Meio de Comunicação, Meio de Transporte, Município, Bairro, Nacionalidade.

Analisando as informações contidas nesses registros observa-se que alguns desses campos são de preenchimento obrigatório, já outros não. É possível observar também a duplicidade de informações. Pode-se citar como exemplo os campos ‘Código do Tipo de Unidade’ e ‘Tipo de Unidade’. Esses campos estão correlacionados, ou seja, há apenas um código para cada tipo de unidade.

3.2.2. Informações da Base

Na Tabela 2 é apresentado em resumo das informações obtidas na base da saúde estudada.

Tabela 2 – Resumo das informações da base da saúde

<i>Descrição</i>	<i>Valor</i>
Número de registros	960.062
Quantidade provável de pacientes distintos	355.697

Uma dificuldade encontrada ao trabalhar com a base de dados do Sistema E-Saúde que contém o registro dos atendimentos médicos nas unidades municipais de saúde é a falta de identificação do paciente. Desde 1º de março de 2012, o cadastro e uso do Cartão Nacional de Saúde (CNS) passou a ser obrigatório para usuários do Sistema Único de Saúde (SUS) segundo portaria do Ministério da Saúde⁶. Cada usuário do SUS possui um

⁴ Lei Nº 12.527, de 18 de novembro de 2011.

⁵ *Comma Separated Values*. Formato de arquivo texto que usa a vírgula para separar os dados de uma planilha permitindo a exportação a um programa editor de planilhas ou ainda utilizar alguma linguagem de programação para ler o arquivo e executar operações.

⁶ Portaria Nº 940, de 28 de abril de 2011.

número único de cartão, e a disponibilização desse número já contribuiria para a identificação precisa do número de pacientes distintos atendidos, assim como ocorre no trabalho de Jasmim [Jasmim et. al. 2017], ao invés do número de atendimentos.

Até o momento os dados disponibilizados pela Secretaria Municipal de Saúde de Curitiba não trazem a informação do número único do paciente. Assim, a técnica utilizada nesse trabalho combina informações referentes aos pacientes das unidades de saúde de Curitiba por inferência e calcula um valor único por meio de uma função *Hash*. Lembrando que uma função *hash* mapeia dados grandes e de tamanho variável em dados pequenos de tamanho fixo.

3.2.3. Processamento da Base da Saúde

Uma pré-análise dos registros da base considerada neste trabalho apontou que alguns desses registros são repetidos. Essas repetições às vezes são no mesmo instante de tempo, em outras ocorrem com diferença de pequenos segundos de tempo, porém apresentam o mesmo conteúdo em todos os demais campos preenchidos. Esses registros repetidos foram retirados da base, e representavam menos de 0,2% do total de registros da base. A Tabela 2 já traz no campo ‘Número de registros’ a quantidade sem as repetições.

4. Caracterização das Bases de Dados

Nesta seção identificamos as principais características do uso do transporte público e das unidades de saúde de Curitiba por meio de seus usuários.

4.1. Caracterização do Uso do Transporte Público

De posse dos dados do cartão transporte e realizados os processamentos informados na seção 3.1.4 foi possível obter uma visão da utilização do transporte público de Curitiba ao longo dos 3 meses analisados.

Observou-se que o mês de agosto/2018 foi o que apresentou o maior número de registros (325.728). Já os meses de setembro/18 e outubro/18 apresentaram uma pequena queda nos registros (294.756 e 312.965, respectivamente). Esse comportamento pode ser justificado se observarmos que esses dois últimos meses contemplam dias de feriados, reduzindo o uso do transporte público.

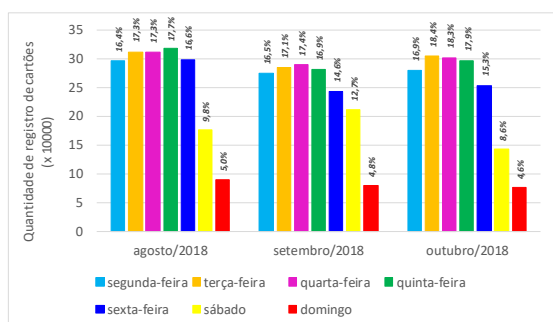


Figura 1 – Utilização de cartões transporte no período de análise

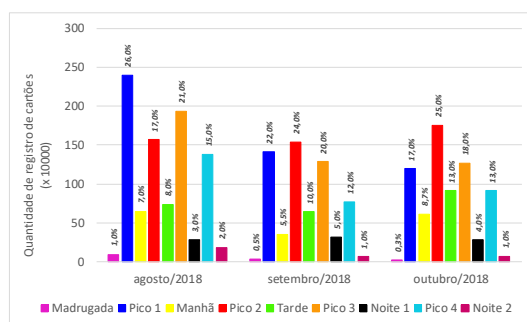


Figura 2 – Utilização de cartões transporte por período

Ainda com relação à utilização do transporte público por meio dos cartões transporte, observa-se na Figura 1 o comportamento do usuário ao longo dos dias da semana. Pode-se verificar que os dias de maior utilização são os dias úteis da semana.

Observa-se também uma queda maior nas sextas-feiras do mês de setembro/18 e outubro/18, coerente com os dias de feriado desses meses (Independência do Brasil, Dia de Nossa Senhora da Luz dos Pinhais – padroeira de Curitiba, Nossa Senhora Aparecida – padroeira do Brasil e Dia dos Professores).

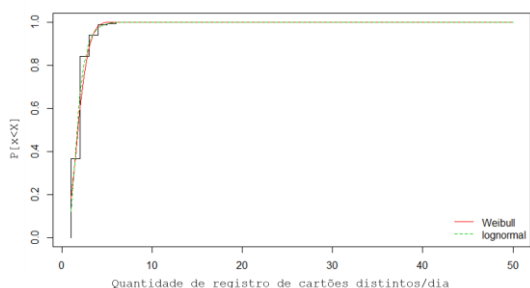


Figura 3 – Registro de cartões por dia

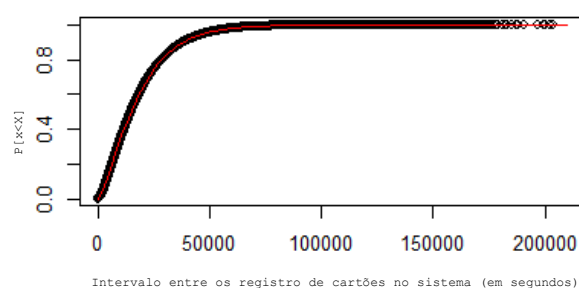


Figura 4 – Intervalo entre os registros de cartões

A Figura 3 mostra a CDF (*Cumulative Distribution Function*) da quantidade de registros de cartões distintos para um mesmo dia. Percebe-se que 80% dos usuários do cartão transporte se registram até 2 vezes por dia no sistema. Observa-se também que há o registro de até 50 vezes no sistema por um mesmo cartão, um padrão inesperado para a utilização do transporte público. A distribuição *Lognormal*⁷ ($\mu = 0.525$ e $\sigma = 1.492$) ou a distribuição *Weibull*⁸ ($\alpha = 2.143$ e $\delta = 2.122$) melhor representam os dados reais analisados. Já a Figura 4 mostra a CDF do intervalo entre os registros sucessivos de um mesmo cartão. Observa-se que cerca de 75% dos usuários de cartão transporte faz o uso do sistema com um intervalo de 8h. Esse valor corrobora com os dados levantados estatisticamente, já que a grande maioria dos usuários do cartão transporte são os adultos, que na média trabalham 8h por dia. O intervalo médio entre os registros de um mesmo cartão é melhor representado por uma distribuição *Gamma* ($\alpha = 1.53$ e $\beta = 8.32e-05$).

4.1.1. Horários de Uso do Transporte por Período

A URBS organiza o transporte público de Curitiba conforme o horário. Atualmente existem 9 períodos considerados pela gestão dos transportes: Madrugada, Pico 1, Manhã, Pico 2, Tarde, Pico 3, Noite 1, Pico 4 e Noite 2.

A Figura 2 apresenta o número de registros de cartões por período de utilização. Observa-se que mesmo havendo variância quanto ao número de registros, os horários de maior utilização do sistema de transporte público de Curitiba obedecem à distribuição de períodos estabelecidos pela URBS, o que vem ao encontro também do aumento da frota nestes períodos, com a operação de contingência, que pode ser observada na base de dados do cartão transporte.

Vale salientar aqui que não se trata de leitura de cartões distintos, uma vez que um mesmo cartão pode ser registrado repetidas vezes num pequeno intervalo de tempo.

$$^7 f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2}\right)$$

$$^8 f(x) = \frac{\delta}{\alpha^\delta} x^{\delta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\delta\right]$$

Mesmo o cartão transporte sendo de uso pessoal e intrasferível⁹, observa-se na Figura 4 que há repetidos registros com intervalo de tempo de poucos segundos para um mesmo cartão.

4.1.2. Uso do Transporte por Sexo e Faixa Etária

Cerca de 60% do uso do transporte público de Curitiba é feito por pessoas do sexo feminino, conforme pode ser observado na Figura 5. Analisando os dados do DETRAN-PR (Departamento de Trânsito do Paraná) [DETRAN 2016], 65% de pessoas habilitadas para dirigir são do sexo masculino no levantamento de 2016, o que talvez justifique o maior número de mulheres utilizando o transporte público.

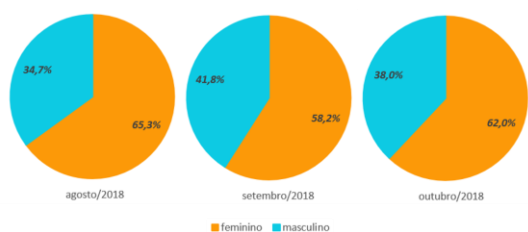


Figura 5 - Utilização de cartões transporte por sexo

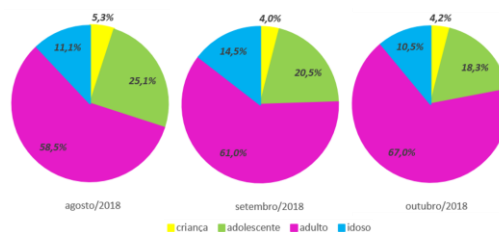


Figura 6 - Utilização de cartões transporte por faixa etária

Para uma interpretação ainda melhor da utilização do transporte público, para esse estudo, a população de Curitiba foi dividida nas 4 faixas etárias, baseadas no Estatuto da Criança e do Adolescente [ECA2018] e no Estatuto do Idoso [EI2018]: Crianças, Adolescentes, Adultos e Idosos. Na Figura 6 observa-se o número de cartões distintos por faixa etária. As crianças representam o menor número de cartões. Cabe ressaltar que até os 6 anos de idade elas possuem gratuidade no uso do transporte público, e costumam não passar pelas catracas dos veículos. Mesmo assim, há na base de dados dos cartões transporte registros de crianças com idade inferior aos 6 anos. Tais registros se justificam por essas crianças possuem algum tipo de benefício para o uso do transporte público, como gratuidade por necessidades especiais, devendo então já possuírem o cartão. As crianças são seguidas pelos Idosos na utilização de cartões. Neste caso, mesmo os Idosos maiores de 65 anos, que por lei¹⁰ possuem gratuidade no uso do transporte, devem possuir o ‘Cartão Isento’ para se utilizarem do sistema. Ao verificar a utilização do sistema de transporte público observou-se que cerca de 1% dos registros não incluíam a data de nascimento do usuário. Esses registros foram desconsiderados nessa caracterização pois a data de nascimento é uma das chaves de identificação do usuário do sistema.

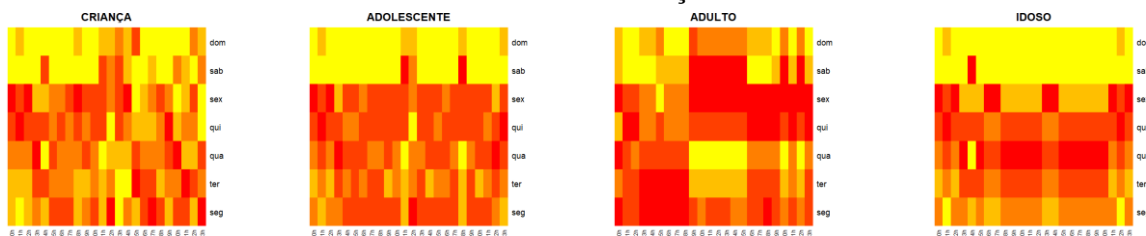


Figura 7 - Utilização do transporte por dia, horário do dia e faixa etária

⁹ <https://www.urbs.curitiba.pr.gov.br/utilidades/cartao-transporte>

¹⁰ Artigo 230 da Constituição Federal Brasileira.

A Figura 7 apresenta um resumo do uso do transporte público por faixa etária, baseado nos dias da semana e horários dos dias. O gradiente das cores varia do amarelo (menor valor) ao vermelho (maior valor). Observa-se que nos finais de semana há uma redução no uso do transporte público por todas as faixas etárias, com exceção dos adultos, que fazem um maior uso entre o período compreendido entre 9h e 15h. Ao apresentar a utilização do sistema separado por horário e faixa etária o gradiente de cores utilizado foi o mesmo para todas as faixas etárias. Para análise da Figura 7 deve-se observar o percentual de pessoas em cada faixa etária para que não haja erro de interpretação. Se observarmos que dentro do total de usuários do sistema de transporte apenas 5% dele representa as crianças, ou seja, em torno de 20.000 crianças, o maior gradiente das crianças representa, percentualmente, um número bem pequeno se comparado às demais faixas etárias, e pode ser justificado como um aumento de demanda pontual para aquele período. Da mesma forma acontece com os adultos, uma pequena variação da utilização aparenta visualmente uma grande queda no gradiente.

4.2. Caracterização do Acesso à Serviços de Saúde

Os mesmos 3 meses analisados na base do transporte público foram analisados para o sistema de saúde de Curitiba. Observou-se que o mês de agosto/2018 foi o que apresentou o maior número de registros (345.242). Já os meses de setembro/18 e outubro/18 apresentaram uma pequena queda nos registros (308.135 e 306.685, respectivamente).

4.2.1. Horários de Uso do Sistema de Saúde

Por convenção, as horas de um dia são divididas em quatro períodos: Madrugada, Manhã, Tarde e Noite.

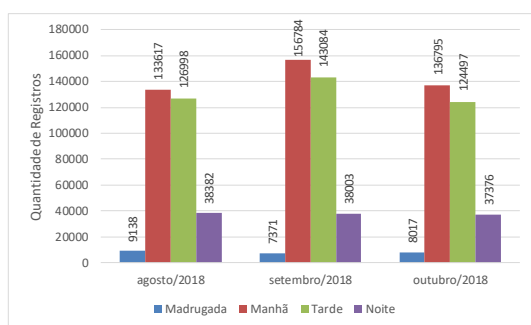


Figura 8 – Registro de Atendimentos por período

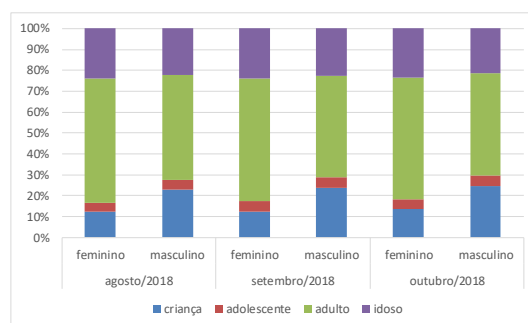


Figura 9 – Atendimentos por faixa etária e por sexo

A Figura 8 apresenta o número de registros de atendimentos por período de utilização do sistema. Observa-se que os horários de maior registro de atendimentos ocorrem no período da manhã e da tarde. Isso pode ser explicado, porque são nesses horários que a maioria das Unidades Municipais de Saúde estão em seu período de funcionamento. Nos períodos da madrugada e da noite somente algumas unidades, e as UPAs, é que se encontram abertas ao público.

4.2.2. Uso do Sistema de Saúde por Sexo e Faixa Etária

Cerca de 60% dos registros de atendimento de saúde de Curitiba são feitos por pessoas do sexo feminino, conforme pode ser observado na Figura 10. Esse comportamento também foi verificado em estudo do IBGE (Instituto Brasileiro de Geografia e Estatística)

[IBGE16], onde mostra que foi observado esse percentual maior de procura por atendimento entre as mulheres.

Na Figura 11 observa-se número de registros por faixa etária. Os adolescentes representam o menor número de atendimento. O maior número de registros de atendimentos é para os adultos, seguidos dos idosos e por fim das crianças.

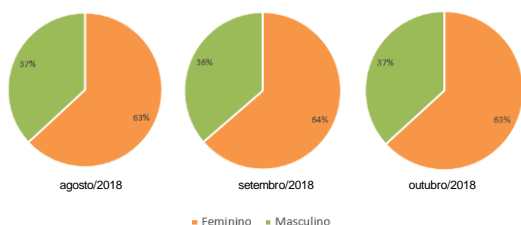


Figura 10 – Utilização do sistema de saúde por sexo

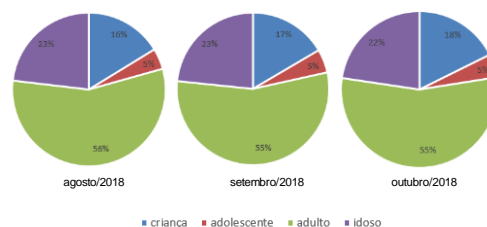


Figura 11 - Atendimentos por faixa etária

Ao fundir as informações dos registros por sexo e faixa etária (Figura 9), observa-se que as crianças apresentam um comportamento inverso ao número de atendimento por sexo. Neste caso, há um maior registro de atendimento para crianças do sexo masculino em todos os meses observados. Segundo dados do IBGE [IBGE19], anualmente há um número maior de nascimento de crianças do sexo masculino. É também na infância que há uma maior dependência dos pais, responsáveis por proporcionar o atendimento à saúde de seus filhos.

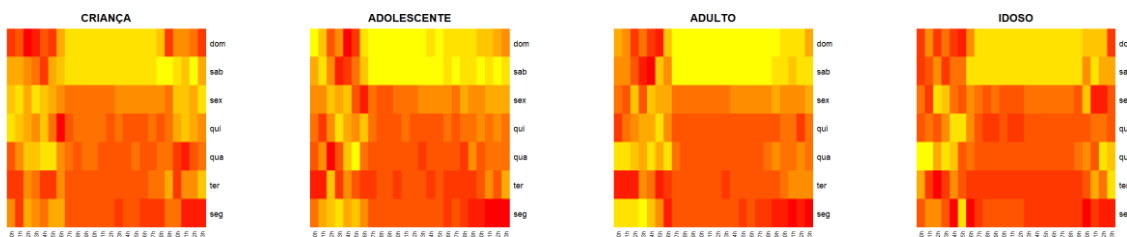


Figura 12 - Utilização do transporte por dia, horário do dia e faixa etária

A Figura 12 apresenta um resumo do registro de atendimentos por faixa etária, baseado nos dias da semana e horários dos dias. Observa-se um comportamento padrão entre as faixas etárias quanto a busca de atendimento à sua saúde, sendo que nos finais de semana há uma redução no número de registros para todas as faixas etárias, ficando os dias úteis e períodos matutinos e vespertinos com maior utilização do sistema.

5. Utilizando o *MultiAspect Graph* para a Criação de um Modelo Multicamadas

Este trabalho também busca a integração de diferentes bases de dados para a criação de um modelo multicamada. A caracterização é importante para a escolha dos aspectos do *MultiAspect Graph*, representando melhor a dinâmica dos sistemas.

5.1. MultiAspect Graph

O *MultiAspect Graph* (MAG) é uma generalização de grafos capaz de representar redes variantes no tempo, redes multicamadas e redes simultaneamente multicamadas e variantes no tempo. Há diferentes descrições do MAG na literatura, porém os modelos

propostos por Kivelä et. al. (2014) e Wehmuth et. al. (2016) são mais bem aceitos pela comunidade acadêmica.

Uma rede multicamada apresenta um conjunto de vértices V igualmente a um grafo, porém acrescentam-se outras camadas. É permitido que cada vértice pertença a qualquer camada, e as arestas podem combinar qualquer conexão entre vértices ou camadas. Além dessa propriedade, o MAG permite a representação de diferentes características dos dados por meio de aspectos. O aspecto pode ser um conjunto de vértices, camadas, instantes de tempo ou qualquer outro recurso independente. Em uma rede multicamada vários aspectos podem ser incluídos. Desta forma, não é possível restringir a rede em um único conjunto de camadas. Por exemplo, em uma rede onde o primeiro aspecto é o tipo de interação e o segundo aspecto é o tempo, é necessário ter um conjunto de camadas para tipos de interação e um segundo conjunto de camadas para o tempo.

Como se pretende construir um modelo multicamadas, as camadas do modelo são formadas por informações extraídas das bases estudadas nesse trabalho e apresentadas na seção 3. Por se tratar de um modelo heterogêneo, cada camada pode ser formada por dados descrevendo aspectos diferentes, tais como localização geográfica, usuário, dados de especialidade ou ainda um híbrido entre esses aspectos. Esse modelo utiliza-se de localização geográfica, identificação do usuário e análise no tempo para criar a relação entre as camadas.

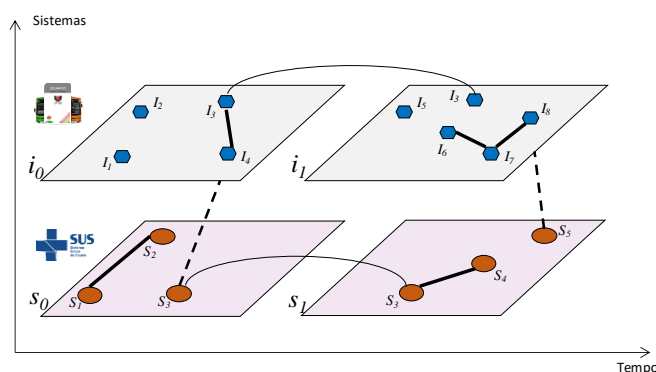


Figura 13 – Representação de um MAG para as bases de estudo

O caso de uso apresentado na Figura 13 traz a representação de um MAG baseado nos dados desse estudo. Neste mapeamento são apresentados três tipos de relacionamentos em dois aspectos (sistemas e tempo). A linha contínua fina representa a movimentação de um vértice no aspecto tempo (no caso de uso, um paciente do sistema de saúde e um usuário do sistema de transporte). A linha tracejada representa o vértice em uma camada diferente. Aqui é importante ressaltar que os dados trabalhados neste estudo não possuem essa informação explicitamente. A conexão para esse aspecto deve ser criada utilizando-se de métodos que avaliem característica similares, localização GPS e compatibilidade de horário entre os sistemas. Por fim, neste modelo, a linha contínua grossa representa a relação dentro da mesma base de dados. Acreditamos que essas três representações no MAG são importantes para o objetivo deste estudo.

5.2. Descrição Formal do Modelo

Para o modelo, os dados em uma camada devem conter os seguintes atributos: t : intervalo de tempo em que os dados foram criados ou observados; l : localização GPS onde os dados

foram gerados ou observados, sendo possível também trabalhar com um raio r a partir da localização informada para minimizar erros; e : informações especiais; u : identificador do usuário que gerou os dados.

A Figura 14 apresenta o mapeamento para a aplicação do MAG. A cidade de Curitiba fornece os dados públicos de uso dos sistemas de saúde e transporte. Esses dados são coletados e armazenados num Banco de Dados de Entrada. Os registros armazenados no banco de dados contêm as informações das bases descritas na seção 3. Todos os registros são organizados temporalmente, e após o pré-processamento apresentado na seção 3.1.4, os dados do cartão transporte são enriquecidos com as informações de latitude e longitude provenientes das informações dos ônibus, formando um novo registro.

Após organização, processamento e caracterização dos registros, esses são armazenados no Banco de Dados do MAG para criar a estrutura MAG apresentada na Figura 13. Os vértices dessa estrutura MAG são elementos formados conforme a seguinte descrição. Seja $U = \{u_1, u_2, u_3, \dots, u_n\}$ um conjunto de usuários de um sistema. No mapeamento proposto neste trabalho trata-se do sistema de transporte e do sistema de saúde, assim, seja $O = \{o_1, o_2, o_3, \dots, o_n\}$ um conjunto de tipo de transporte, neste caso o ônibus, e $C = \{c_1, c_2, c_3, \dots, c_n\}$ um conjunto que represente informações de deslocamento, provenientes das informações do cartão transporte. Cada usuário $u_i \in U$ utiliza-se de um tipo de transporte $o_j \in O$ compartilhando informações de deslocamento $c_k \in C$. Cada n -ésima informação de deslocamento compartilhada é representada pela tripla $I_n^{c_k} = \langle t, u, o \rangle$, onde t se refere ao registro de data e hora quando o usuário $u_i \in U$ do sistema de transporte utiliza-se do ônibus $o_j \in O$. De maneira análoga, o m -ésimo usuário de um sistema de saúde é representado pela tripla $S_m^{u_i} = \langle t, l, e_x \rangle$, onde t se refere ao registro de data e hora em que esse usuário $u_i \in U$ foi registrado em uma das unidades de saúde l , apresentando alguma informação especial de atendimento e_x . Observa-se que as triplas $I_n^{c_k}, S_m^{u_i}$ são os vértices de cada uma das camadas i e s , representadas na Figura 13.

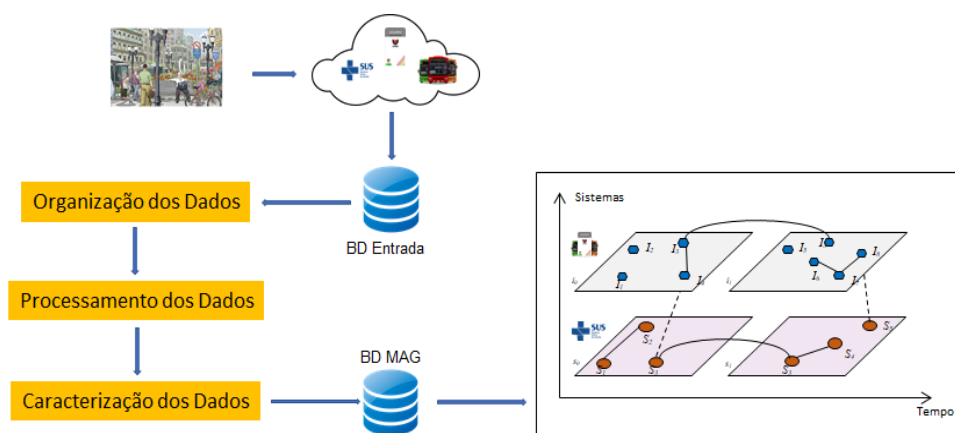


Figura 14 – Mapeamento para a Criação da Estrutura MAG

A rede multicamadas formada contém 2 aspectos que contém uma sequência de conjuntos de camadas elementares. Camada elementar representa um elemento do conjunto de camadas, ou seja, há um único conjunto de camadas elementares L_a para cada aspecto a ($L_{s_0}, L_{s_1}, L_{i_0}$ e L_{i_1}). O conjunto de todas as camadas no grafo é o produto cartesiano de todas as camadas elementares $L_{s_0} \times L_{s_1} \times L_{i_0} \times L_{i_1}$. Qualquer um dos vértices $V = \{I_n^{c_k}, S_m^{u_i}\}$ pode estar em qualquer camada, assim, o conjunto de vértices que

estão presentes no grafo é representado por $V_M \subseteq V \times L_{s_0} \times L_{s_1} \times L_{i_0} \times L_{i_1}$. As arestas do grafo devem conectar vértices diferentes em qualquer camada ou aspecto. Desta forma, o conjunto de aresta do modelo é representado por $E_M \subseteq V_M \times V_M$. Com essas informações, o modelo possui a forma $M = (V_M, E_M, V, L)$.

Essa representação no MAG é um passo chave para entender como acontece o deslocamento da população das grandes cidades até as unidades de saúde. Isso pode contribuir com propostas que podem facilitar o atendimento a esses cidadãos no que diz respeito à saúde, desde a implantação de especialidade médica nas unidades, evitando o deslocamento desnecessário, ou até mesmo a criação de uma nova rota de ônibus na busca de um atendimento especializado.

6. Considerações Finais

Esse trabalho estuda a utilização dos sistemas de transporte público e de saúde da cidade de Curitiba. Uma caracterização desses sistemas foi realizada durante um período de 3 meses, entre agosto e outubro de 2018. Esse estudo inicia apresentando os principais resultados acerca da caracterização dos dados dessas duas grandes e representativas bases de estudos da cidade de Curitiba. Finalmente, esse estudo também apresenta como contribuição uma formalização do MAG para implementação do modelo pretendido. Como trabalhos futuros, pretende-se utilizar o modelo proposto para auxiliar na resposta de perguntas complexas relacionadas com a relação entre o sistema de transporte público e de saúde, como propagação de doenças na cidade entre outras.

7. Referências

- Agarwal, R.; Dhar, V.. “Editorial – Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research”. *Information Systems Research*, Volume 25, Issue 3, pg. 443–448. September 2014.
- Ahmad, Z.; Batool, Z.; Starkey, P.. “Understanding mobility characteristics and needs of older persons in urban Pakistan with respect to use of public transport and self-driving”. *Journal of Transport Geography*. Volume 74, January 2019, Pages 181-190.
- Anuário Estatístico 2016 – DETRAN-PR. Disponível em <<https://bit.ly/2Wpsbdd>>. Acessado em 28 de novembro de 2018.
- Araújo, M. R. M. et. al. “Transporte público coletivo: discutindo acessibilidade, mobilidade e qualidade de vida”. *Psicologia & Sociedade*, Volume 23, pg. 574-582. 2011.
- Braz, T. et. al. “Estimating Inefficiency in Bus Trip Choices From a User Perspective With Schedule, Positioning, and Ticketing Data”. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 11, November 2018.
- Estatuto da Criança e do Adolescente. Disponível em <<https://bit.ly/1NqqzW6>>. Acessado em 14 de dezembro de 2018.
- Estatuto do Idoso. Disponível em <<https://bit.ly/2CFDyqa>>. Acessado em 14 de dezembro de 2018.

- Hristova, D. et. al. “Measuring Urban Social Diversity Using Interconnected Geo-Social Networks”. Proceedings of the 25th International Conference on World Wide Web, pg. 21-30. April 16.
- Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional de Saúde. Disponível em <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv97329.pdf>>. Acessado em 24 de janeiro de 2019.
- Instituto Brasileiro de Geografia e Estatística. Sistema IBGE de Recuperação Automática. Disponível em <<https://sidra.ibge.gov.br/tabela/2679>>. Acessado em 04 de fevereiro de 2019.
- Jasmim, L. O.; Ziviani, A. et. al. Caracterização de Atendimentos em uma Rede de Atenção à Saúde. 17º Workshop de Informática Médica. Julho 2017.
- Kivelä, M. et. al. “Multilayer networks”. Journal of Complex Networks, Volume 2, Issue 3, pg. 203–271. September 2014.
- Kurant, M.; Thiran, P.. “Layered Complex Networks”. Physical Review Letters, Volume 96. April 2006.
- Labrinidis, A.; Jagadish, H. V.. “Challenges and Opportunities with Big Data”. Proceedings of the Very Large Data Bases Endowment, Volume 5, Issue 12, pg. 2032-2033. August 2012.
- Manyika, J. et. al. “Big Data: The next frontier for innovation, competition, and productivity”. Available in <<https://goo.gl/2jX7W5>>. Acessado em 17 de janeiro de 2018.
- Mervis, J.. U.S. Science Policy: Agencies Rally to Tackle Big Data. Science, Volume 336, Issue 6077, pg. 22. April 2012.
- Nature Editorial. Community cleverness required. Nature, Volume 455, Issue 7209, pg. 1. September 2008.
- Oliveira, N.. Gastos com saúde crescem mesmo em meio à crise e atingem 9,1% do PIB. Agência Brasil. Disponível em <<https://goo.gl/4ijGnF>>. Acessado em 03 de abril de 2018.
- Pena, R. F. A. Mobilidade urbana no Brasil. Brasil Escola. Disponível em <<https://goo.gl/n3CkrS>>. Acessado em 03 de abril de 2018.
- Rodrigues, D. O. et. al. SMAFramework: Urban Data Integration Framework for Mobility Analysis in Smart Cities. 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems. November, 2017.
- Wehmuth, K.; Fleury, E.; Ziviani, A.. On MultiAspect graphs. Theoretical Computer Science, Volume 651, pg. 50–61. 2016.
- Zhanga, X.; Xub, Y.; Tuc, W.; Rattid, C.. Do different datasets tell the same story about urban mobility — A comparative study of public transit and taxi usage. Journal of Transport Geography 70 (2018).