

UserPoint-MAG: Abordagem de rede multicamada para o estudo de propagação de contágio no transporte público

**Priscila Santin¹, Fernanda R. Gubert¹, Mauro Fonseca^{1,2},
Anelise Munaretto¹, Thiago H. Silva^{1,2}**

¹Pós Graduação em Engenharia Elétrica e Informática Industrial (CPGEI)
Universidade Tecnológica Federal do Paraná (UTFPR)
Av. Sete de Setembro, 3165 – Curitiba – Brasil

²Pós Graduação em Computação Aplicada (PPGCA)
Universidade Tecnológica Federal do Paraná (UTFPR)
Av. Sete de Setembro, 3165 – Curitiba – Brasil

santin.priscila@gmail.com, fernandagubert@alunos.utfpr.edu.br,

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

Resumo. *Em tempos de pandemia, o transporte público pode ser crucial para a disseminação de vírus, principalmente nas grandes cidades. As vacinas costumam fazer parte das estratégias para reduzir o contágio; no entanto, estas podem ser escassas em cenários pandêmicos. Utilizando dados do sistema de transporte público, este trabalho propõe o uso de redes multicamadas variantes no tempo para identificar os principais locais críticos a serem considerados prioritários em intervenções, como campanhas de vacinação, para ajudar a reduzir o contágio nesse meio de locomoção. Nossa abordagem considera os pontos de ônibus críticos como pontos prioritários de vacinação, indicando que a vacinação nesses locais reduz a propagação da infecção usando menos doses do que uma vacinação aleatória. A abordagem proposta neste estudo não se limita às estratégias de vacinação, sendo também aplicável a outros problemas que compartilham propriedades semelhantes, mesmo em contextos diferentes.*

1. Introdução

Nos últimos anos, fomos surpreendidos pelo aumento de notícias sobre doenças transmitidas pelo ar, como a gripe A (H1N1), a síndrome respiratória aguda grave (SARS) e mais recentemente pela gripe H3N2. Tais doenças acabam impondo sérios encargos financeiros e de saúde pública à nossa sociedade [Zhu et al. 2012]. Atualmente vivemos em meio à pandemia de SARS-CoV-2, que causa a doença COVID-19 e que em janeiro de 2022 já havia afetado mais de 360 milhões de pessoas em todo o mundo, levando à morte mais de 5 milhões de pessoas [World Health Organization 2022]. Nesse sentido, estudar e compreender a transmissão de doenças infecciosas transmitidas pelo ar, principalmente em ambientes com alta densidade de ocupantes, como os sistemas de transporte público [Sun et al. 2014], pode ajudar os governantes a desenvolver formas de minimizar os efeitos dessas transmissões.

O sistema de transporte público pode ser analisado como um sistema de rede complexo [Chodrow et al. 2016, Estrada 2012], sendo uma forma eficiente de modelar e explicar os mecanismos do comportamento coletivo [Newman 2010]. Nessa perspectiva, a ferramenta mais comum para modelar tais redes são os grafos tradicionais, como

aqueles que não levam em conta a dimensão temporal ou outros atributos contextuais. No entanto, essa abordagem pode representar uma simplificação excessiva de uma realidade muito mais complexa, o que pode levar a uma sub-representação do sistema em estudo [Buldyrev et al. 2010, Cardillo et al. 2013].

A modelagem mais sofisticada de sistemas complexos pode ser feita usando vários subsistemas (ou camadas) interdependentes. O conceito de rede multicamadas vai muito além de uma simples intenção de capturar a heterogeneidade dos dados. Portanto, uma camada em uma rede pode variar de acordo com o contexto [Kivelä et al. 2014]. Assim, estudos como disseminação de doenças, navegação e sincronização em redes multicamadas têm atraído significativa atenção [Sahneh et al. 2019, Jacobsen et al. 2018, Lv et al. 2018].

Neste trabalho, utilizamos um caso particular de redes multicamadas em que diferentes vértices, instantes de tempo e aspectos das camadas são os principais recursos a serem representados pelo modelo. Utilizando dados públicos, estabelecemos um modelo de rede multicamadas e avaliamos o impacto da disseminação de doença respiratória e, conseqüentemente, sua prevenção. Identificando pontos críticos de interesse a partir do modelo criado, utilizamos diferentes estratégias de vacinação e medimos seus efeitos com base nos padrões de mobilidade observados na cidade. As principais contribuições deste estudo podem ser resumidas como:

- Criação de um modelo multicamada variante no tempo que integra dados de mobilidade com outros pontos de interesse da cidade independentemente do contexto;
- Estratégia para identificação de pontos críticos de interesse a partir do modelo estabelecido baseado na centralidade de percolação (*Percolation Centrality*) [Piraveenan et al. 2013];
- Estudo e aplicação do modelo proposto na estratégia de vacinação para prevenção de doenças, sendo que neste trabalho foram utilizados dados em grande escala de transporte público.

O restante do estudo está organizado da seguinte forma. A Seção 2 apresenta trabalhos relacionados. A Seção 3 apresenta os procedimentos metodológicos e outras informações relevantes para a compreensão dos resultados. Na Seção 4 são apresentados e discutidos os resultados. Por fim, a Seção 5 apresenta a discussão final e as conclusões.

2. Trabalhos Relacionados

2.1. Redes Multicamadas

A aplicabilidade de modelos multicamadas para extração de conhecimento não é um conceito novo na literatura [Oselio et al. 2014, Belyi et al. 2017]. Um exemplo do uso desses modelos pode ser encontrado na área de transporte. Enquanto no trabalho de Domenico e seus colegas [De Domenico et al. 2013] uma estrutura tensorial é proposta para o estudo de redes multicamadas objetivando auxiliar no desenvolvimento de técnicas de roteamento para sistemas de trânsito multimodais, o trabalho de Kurant e Thiran [Kurant and Thiran 2006] apresenta um modelo em camadas para facilitar a descrição e análise de pequenas redes complexas, permitindo capturar as diferenças fundamentais entre uma carga real e os estimadores de carga comumente usados, explicando por que esses estimadores não conseguem estimar a carga real. Também observa-se o uso de modelos multicamadas na identificação de agentes centrais das redes responsáveis pela

disseminação mais rápida de informações [De Domenico et al. 2015, Zheng et al. 2018], ou ainda na identificação de comunidades e sua diversidade social, interconectando pessoas e lugares [Hristova et al. 2016, Yildirimoglu and Kim 2018]. Alguns modelos de redes multicamadas são propostos para representar a variação no tempo no estudo de mobilidade, como o modelo MAG (Multi-Aspect Graph) [Wehmuth et al. 2016]. Rodrigues e colegas [Rodrigues et al. 2017] propõem uma abordagem para analisar padrões de mobilidade baseados em MAG. Os autores propuseram um algoritmo para análise de mobilidade permitindo a análise da correlação espaço-tempo entre os dados coletados de diferentes fontes de dados, possibilitando ainda o mapeamento de usuários entre camadas. Já o trabalho de Costa e colegas [da Costa et al. 2018] faz o uso do MAG para fornecer uma análise estrutural multicamadas e variante no tempo de uma rede de transporte aéreo brasileira, com o objetivo de entender as estratégias particulares das companhias aéreas para estabelecer e adaptar em um momento de crises suas redes de voos específicas.

2.2. Identificação de Pontos de Interesse

A identificação e recomendação de pontos de interesse (POIs) tem se mostrado cada vez mais um tema de pesquisa relevante, com aplicações no mundo real [Liu and Seah 2015]. O trabalho de Huang [Huang et al. 2018] analisa a relação entre POIs e comunidades de rede de mobilidade humana em vários métodos de detecção de comunidades. No trabalho de Tang [Tang et al. 2020], os autores propõem um framework para descobrir zonas funcionais por meio da análise de estruturas urbanas e comportamentos sociais. Já o trabalho de Thommee e seus colegas [Thomee et al. 2016] propõe um método para localizar automaticamente pontos de interesse representados em fotos tiradas por pessoas ao redor do mundo. Para explorar a análise espacial do uso de bicicletas compartilhadas em Nova Iorque, Bao e colegas [Bao et al. 2018] realizaram um estudo que considerou a diversidade entre diferentes categorias de estações de bicicletas usando dados de cartões inteligentes e dados de POIs, que foram coletados por meio da API do Google Places. E mais recentemente, o trabalho de Chang e seus colegas [Chang et al. 2021] apresentou um modelo de metapopulação suscetível-exposto-infeccioso-removido (SEIR) que integra redes de mobilidade dinâmicas para simular a disseminação do SARS-CoV-2. Utilizando dados de telefonia celular eles mapearam o movimento de milhões de pessoas a pontos de interesse, demonstrando que uma pequena minoria de pontos de interesse é responsável pela grande maioria das infecções e que restringir a ocupação máxima em cada ponto de interesse é mais eficaz do que reduzir uniformemente a mobilidade.

2.3. Estratégias para Redução de Contágio de Doenças no Transporte Público

Os sistemas de transporte público são considerados ambientes de alto risco para a disseminação de infecções devido às condições de confinamento e ventilação limitada nestes meios de transporte [Zhu et al. 2012]. Mo e colegas [Mo et al. 2021] propõe uma rede ponderada de encontros variáveis no tempo para modelar a propagação de doenças infecciosas por meio de sistemas de transporte público usando dados de cartões transporte de Cingapura. O estudo tem como objetivo entender melhor a dinâmica geral espaço-temporal da propagação da pandemia de SARS-CoV-2 no sistema de transporte público e avaliar os efeitos de diversas medidas preventivas. Já Müller et al. [Müller et al. 2020] simulam trajetórias completas de mobilidade individual com base em dados de telefones celulares para calcular os riscos de disseminação do COVID-19 em encontros dinâmicos no trânsito e em edifícios. Goscé e Johansson [Goscé and Johansson 2018] analisam as

viagens de metrô de Londres para verificar a correlação entre o uso dessa modalidade de transporte público e a contaminação por doenças infecciosas transmitidas pelo ar, enquanto que Ventresca e Aleman [Ventresca and Aleman 2013] construíram uma rede ponderada por meio de dados censórios de Ontário, para examinar diferentes aspectos da rede, incluindo a disseminação de uma doença após vacinação hipotética em seis estratégias diferentes. Já Wang et al. [Wang et al. 2020] propõem um método de análise de decisão que visa resolver problemas de resposta a emergências para a prevenção e controle do coronavírus 2019 (COVID-19) no transporte ferroviário urbano.

Ao contrário de todos os esforços anteriores, este artigo explora uma metodologia de rede multicamada variável no tempo para apontar pontos críticos de interesse para vacinação em um cenário com condições análogas à COVID-19 para obter maior eficiência de imunização no transporte público em grandes cidades.

3. Materiais e Métodos

3.1. Curitiba e seu Transporte Público

O município de Curitiba possui, de acordo com o IBGE¹ [Instituto Brasileiro de Geografia e Estatística - IBGE 2010] 1.751.907 habitantes em uma área total de 434.967 km^2 . Administrativamente, Curitiba é dividida em 10 Administrações Regionais, cobrindo seus 75 bairros. Todos esses bairros são atendidos pelo sistema de transporte público de Curitiba. Tal sistema de transporte possui corredores de transporte exclusivos, onde circulam ônibus biarticulados. Esses ônibus biarticulados passam por vários terminais de integração que recebem os ônibus alimentadores da vizinhança, permitindo a integração do sistema. O sistema também possui linhas circulares entre bairros, que permitem a passagem de um bairro para outro, sem ter que percorrer a região central da cidade, além de linhas diretas, que oferecem viagens mais rápidas com menos paradas no itinerário. Essas linhas diretas têm um tipo de ponto de ônibus específico para elas, assim como os terminais de integração. Outra característica importante do transporte público de Curitiba é a tarifa integrada. Pagando apenas uma passagem, o cidadão pode compor sua própria rota, movimentando-se pela cidade. Desde 2002, Curitiba utiliza o sistema de bilhetagem eletrônica para o transporte público. A implementação desse sistema foi necessária para reduzir a circulação de dinheiro em espécie no sistema de transporte, acelerar o embarque e a passagem de usuários pelas catracas, disciplinar e mensurar o uso do sistema de transporte por categorias que gozam de acessos gratuitos, além de reduzir os custos de operação do sistema [Taniguchi and Duarte 2012]. Segundo a URBS² [Urbanização de Curitiba S/A - URBS 2018] em 2018 uma média de 1.365.615 passageiros foram transportados a cada dia útil. Além disso, de acordo com o mesmo relatório, cerca de 60,96% das tarifas foram pagas usando o cartão transporte. O número de cartões transporte ativos em 2018 foi de 1.928.184, dividido em três das quatro categorias de transporte existentes, a saber: Cartão de Usuário, Cartão de Isenção, Cartão de Estudante e Cartão Único (para esta categoria, não há informações nos dados fornecidos pela URBS).

¹IBGE: Instituto Brasileiro de Geografia e Estatística

²URBS (Urbanização de Curitiba S/A): empresa que gerencia o transporte público em Curitiba.

3.2. Pré-processamento dos Dados: Inferindo o Bairro de Residência

O objetivo principal desta etapa foi estabelecer as informações do bairro onde moram os usuários do transporte público. Sabendo que apenas a entrada no sistema é registrada, como resultado desse pré-processamento, mantemos apenas os dados dos quais pudemos inferir a origem com base no histórico de uso de cada cartão transporte. A Figura 1 mostra as etapas de pré-processamento dos dados, desde a obtenção dos dados brutos até a criação do banco de dados utilizado em nosso estudo. Cada uma dessas etapas é detalhada a seguir.

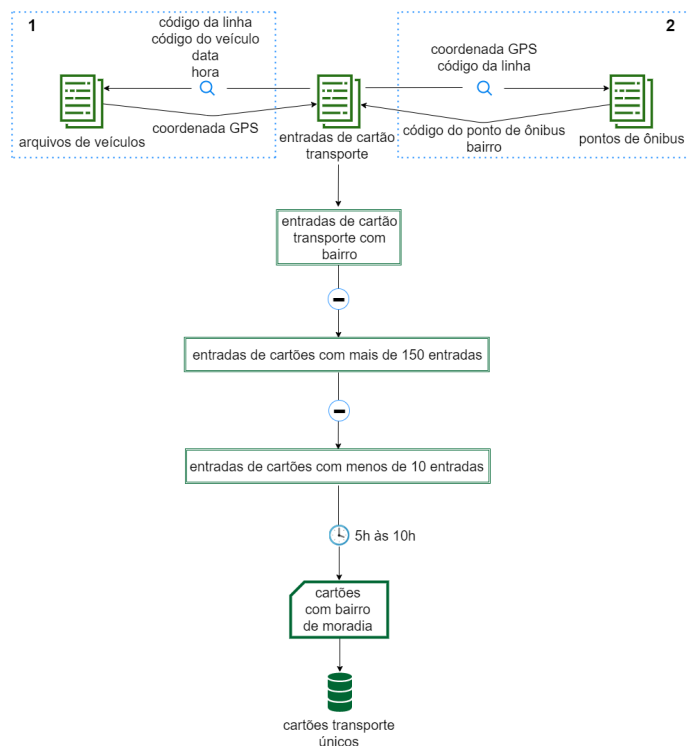


Figura 1. Fluxograma de pré-processamento de dados relacionados ao trânsito

O uso do cartão transporte representa um total de 17.511.710 registros e as informações sobre os ônibus 129.461.200 registros. Para inferir o bairro onde cada cartão transporte foi usado, usamos diferentes abordagens. Quando os usuários entram no sistema de trânsito em estações tubo ou terminais de integração (ponto de ônibus exclusivo para os ônibus biarticulados e algumas linhas especiais), o código da linha de ônibus e o código do veículo vinculado a esta entrada estão diretamente relacionados ao local do ponto de ônibus. Portanto, a identificação do bairro acontece com alta precisão. Quando a entrada do usuário ocorre em pontos de ônibus regulares, procuramos uma correspondência usando o código da linha, código do veículo, data e hora, para obter uma coordenada GPS (*Global Positioning System*) para cada entrada do cartão transporte.

Há casos em que o intervalo de registro da posição GPS é estendido. Em média, em Curitiba, as atualizações dos ônibus, incluindo as coordenadas GPS, são registradas a cada 5 segundos. Com essas informações, infere-se a área aproximada onde o cartão transporte foi utilizado. Assim, mantemos apenas uma atualização por minuto para cada veículo, pois, em um minuto, considera-se que o ônibus não percorreu uma longa distância. Após este filtro, as novas informações sobre os ônibus representam 18.594.495

registros, o que permite que o procedimento de inferência seja realizado de forma mais eficiente sem comprometer consideravelmente a qualidade.

Vale ressaltar aqui que ao combinar os atributos de código da linha, código do veículo, data e hora entre a base dos cartões transporte e os registros do ônibus, observou-se que alguns ônibus apresentavam o código da linha e o código do veículo preenchidos com valor genérico, não sendo possível garantir a identificação da localização GPS da entrada do cartão. Nesses casos, esses registros foram excluídos.

Depois de enriquecer as entradas do cartão transporte com a latitude e a longitude, elas foram associadas com as informações do bairro de um ponto de ônibus específico. Para fazer este procedimento, primeiramente, a latitude e longitude inferidas foram comparadas com o ponto de ônibus mais próximo da linha registrada na entrada do cartão transporte. Este procedimento foi feito usando a estrutura de dados *R-Tree*. O *R-Tree* foi proposto por Antonin Guttman e é amplamente utilizado como método de acesso espacial, permitindo a indexação de informações multidimensionais como coordenadas geográficas [Guttman 1984]. Ao final desta etapa, pode-se inferir o bairro para 5.388.638 entradas de cartões transporte.

Algumas pessoas emprestam seus cartões transporte, com ou sem vantagens financeiras, para outras que trafegam na mesma linha de ônibus, gerando ruído nos registros. Além disso, algumas entradas são feitas com alguns minutos de diferença na mesma linha e no mesmo ônibus nesses casos, gerando muitas entradas para o mesmo cartão transporte. Para minimizar o impacto desses cartões transporte em nossa análise, excluímos cartões transporte com mais de 150 entradas. A lógica por trás deste valor é que o período de estudo é de 69 dias (53 dias úteis e 16 dias de fim de semana) e que cada usuário normalmente usa, em média, pelo menos dois passes por dia [Santin et al. 2020]. Em uma linha de raciocínio semelhante, cartões transportes com menos de dez entradas foram excluídos durante o período de análise. Eles representam um número pequeno, em média menos de uma entrada por semana. Ao desconsiderá-los, também pretendíamos tornar os resultados mais robustos. Após essas exclusões, o número de entradas para identificação do bairro diminuiu para 5.225.573.

Um aumento significativo no número de entradas foi observado na análise de dados em torno de *5am*. Esse crescimento atinge seu pico em torno de *7am*, e foi observado que esse fenômeno praticamente termina às *10am*, conforme mostrado na Figura 2. Devido a este cenário, consideramos as entradas feitas entre *5am* e *10am* para identificar o bairro de origem. Esse período é quando os usuários provavelmente usam o transporte público para se deslocar para uma atividade cotidiana, como trabalhar ou estudar.

De posse das entradas do cartão realizadas entre *5am* e *10am*, agrupamos os dados por cartão transporte e selecionamos o bairro mais frequente como bairro de origem, sendo necessário considerar pelo menos cinco entradas neste bairro. Ao analisar o resultado, pudemos verificar que os limites estabelecidos preservam um número significativo de cartões transporte sem comprometer a análise. Após esse procedimento, obtivemos 92.857 cartões transporte exclusivos contendo o bairro de origem. Apesar das possíveis limitações de nossa abordagem, obtivemos evidências de que representa uma boa aproximação. Construímos um conjunto de dados com entradas de cartões transporte classificadas manualmente, com base em entrevistas com proprietários de cartão transporte que apresentam diferentes padrões de mobilidade urbana e com os quais temos

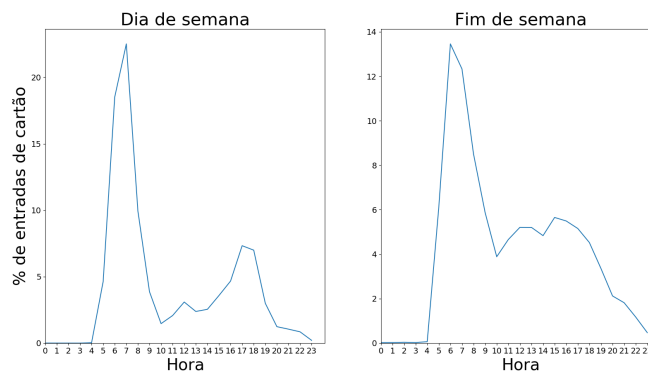


Figura 2. Entradas de cartão transporte por hora

contato frequente, seja por motivos profissionais ou pessoais – total de oito usuários de diferentes domicílios. Desta forma, obtivemos o bairro de origem para esses cartões e verificamos que a nossa abordagem identificou todas as informações corretamente para este conjunto de dados. A construção desse conjunto de dados foi bastante desafiadora por diversos motivos, como a identificação de voluntários e a concordância com a divulgação de informações particulares; no entanto, esta experiência foi vital para nos ajudar a ter uma noção da qualidade da nossa abordagem.

3.3. Modelo e Métricas

3.3.1. Preliminares

Este trabalho é baseado em um modelo multicamada específico, denominado *Multi-Aspect Graph* (MAG), uma estrutura capaz de representar uma rede multicamada variante no tempo. Um modelo MAG é dado por $H = (A, E)$, onde E é o conjunto de arestas e A é a lista de aspectos que compõem o modelo. Cada aspecto $\alpha \in A$ é um conjunto finito e o número de aspectos p representa a ordem do MAG. Cada aresta $e \in E$ é uma tupla com elementos $2 \times p$. Todas as arestas têm a forma $(a_1, \dots, a_p, b_1, \dots, b_p)$, onde a_1, b_1 são elementos do primeiro aspecto de H , a_2, b_2 são elementos do segundo aspecto de H e assim por diante, até a_p, b_p , que são elementos de p -ésimo aspecto de H [Wehmuth et al. 2016]. Para realizar as análises deste trabalho foi criado o modelo MAG específico descrito a seguir.

3.3.2. Modelo *UserPoint-MAG*

O Modelo *UserPoint-MAG* é um grafo multi-aspecto com uma ordem de $p = 4$, ou seja, quatro aspectos. Este modelo tem duas camadas, uma referente aos usuários de cartão transporte e outra referente aos pontos de ônibus, que é um dos aspectos. Os outros aspectos são um identificador (cartão transporte do usuário ou número do ponto de ônibus), um dos 75 bairros de Curitiba (bairro da origem do usuário ou bairro onde está localizado o ponto de ônibus), e por fim a data e hora de registro do cartão transporte. Este modelo apresenta arestas entre as camadas e entre os vértices da camada de parada de ônibus, conforme mostrado na Figura 3. Uma aresta é criada para cada entrada de cartão transporte entre o respectivo usuário e o ponto de ônibus acessado. Caso não seja a primeira entrada

do usuário no dia, também é criada uma aresta entre o último ponto de ônibus acessado e o atual, criando o caminho do usuário durante o dia na camada de ponto de ônibus.

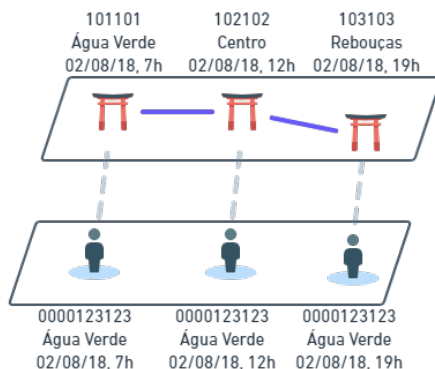


Figura 3. Representação do modelo *UserPoint-MAG*

3.3.3. Centralidade de Percolação

Alguns estudos modelaram com sucesso a propagação de doenças como um exemplo específico de percolação em redes [Sander et al. 2002, Meyers et al. 2006]. Assim, neste trabalho foi utilizada a medida *centralidade de percolação* estabelecida por [Piraveenan et al. 2013], que quantifica a importância relativa, no nosso caso, da propagação de doença para cada vértice na rede com base em sua conectividade topológica, bem como seus estados de percolação. O estado de percolação de um vértice i no tempo t é denotado por x_i^t . Quando não há aspecto temporal, apenas x_i é usado. Se $x_i^t = 0$, isso indica um estado não percolado do vértice. Se $x_i^t = 1$, indica um estado totalmente percolado. Também é possível que um vértice assuma um estado parcialmente percolado, correspondente a $0 < x_i^t < 1$.

A centralidade de percolação de um determinado vértice é definida como a proporção de caminhos percolados que passam por esse vértice. Um caminho percolado significa o caminho mais curto entre um par de vértices, onde o vértice de origem é percolado (em nosso estudo, é o conceito de infectado). O vértice de destino pode ser percolado ($x_i^t = 1$) ou não ($x_i^t = 0$), ou estar em um estado parcialmente percolado. Matematicamente, a centralidade de percolação de um vértice v no tempo t é representada pela Equação 1:

$$PC^t(v) = \frac{1}{(N-2)} \sum_{s \neq v \neq r} \frac{\sigma_{s,r}(v)}{\sigma_{s,r}} \frac{x_s^t}{[\sum x_i^t] - x_v^t} \quad (1)$$

onde $\sigma_{s,r}$ é o número de caminhos mais curtos entre o vértice de origem s e o vértice de destino r , enquanto $\sigma_{s,r}(v)$ é o número de caminhos mais curtos entre o vértice de origem s e o vértice de destino r passando pelo vértice v . Para o processamento desta centralidade, alguns vértices devem ser identificados como infectados *a priori*. O valor crítico de cada vértice na rede é expresso em decimal e está entre 0 e 1. Em nosso modelo, apenas os vértices da camada de usuários de cartão transporte podem ser infectados. Um atributo de vértice chamado “percolação” foi usado para identificar esses vértices, onde 1 indica infectado e 0 não infectado.

4. Experimentos e Resultados

Os resultados descritos a seguir compreendem um período correspondente a sete dias consecutivos, de 02 de outubro a 08 de outubro de 2018, abrangendo tanto os dias úteis como os finais de semana. Três cenários foram estabelecidos para analisar a evolução da infecção em diferentes estratégias de vacinação:

- **Cenário 1:** Vacinação de usuários do sistema de transporte público selecionados de forma aleatória com 0,25 de probabilidade;
- **Cenário 2:** Vacinação dos usuários do sistema de transporte público que acessam os 100 pontos de ônibus mais críticos, com 0,25 de probabilidade;
- **Cenário 3:** Vacinação dos usuários do sistema de transporte público que acessam os 100 pontos de ônibus mais críticos, com 0,5 de probabilidade.

Para essas simulações, foi utilizado o *UserPoint-MAG* em dois bairros da cidade, “Centro” e “Pinheirinho”, que correspondem a regiões estratégicas conforme apresentado em nosso trabalho anterior [Santin et al. 2020]. Cada cenário tem uma infecção inicial, pois a centralidade de percolação é iniciada a partir do vértice previamente infectado, determinando assim os pontos de ônibus críticos. Inicialmente, os usuários que moravam no Centro (pertencente à Regional Matriz) e pegavam ônibus no primeiro dia da semana foram infectados em todos os cenários. A escolha do Centro foi feita por ser a região central da cidade, o bairro mais conectado e de alto tráfego [Santin et al. 2020]. Para fins de comparação, o primeiro e o terceiro cenários também foram analisados sob outra perspectiva de infecção inicial, infectando usuários de cartão transporte que moram no Pinheirinho no primeiro dia da análise. Estudar o Pinheirinho é interessante porque este bairro possui um dos terminais de ônibus mais acessados pela população, gerando alto tráfego. Ainda assim, sua localização é bastante distante da região central da cidade. Portanto, ao todo, foram realizadas cinco variações nos três cenários apresentados; cada variação foi repetida 35 vezes.

Os pontos de ônibus críticos foram determinados aplicando a centralidade de percolação ao modelo apresentado na Seção 3.3.2, mas excluindo o aspecto de data e hora (por subdeterminação). A criticidade refere-se a identificação dos pontos de ônibus mais acessados, independente de data e hora, possibilitando um possível maior contágio entre os usuários do sistema de transporte público. Com essa exclusão, todos os cadastros de usuários de cartão transporte ficam conectados, permitindo que rotas de tráfego mais significativas sejam representadas por arestas mais pesadas.

Para determinar os pontos de ônibus mais críticos nos cenários 2 e 3, onde a infecção inicial é no Centro, todos os usuários que moram no Centro e que usaram transporte público em um determinado dia da semana são infectados. A regra para determinar os pontos de ônibus mais críticos no cenário 3, onde a infecção inicial é no Pinheirinho, é a mesma, mas os usuários que moram no Pinheirinho são infectados. Ao final de cada simulação, os 100 pontos de ônibus com maior valor de percolação são filtrados e exibidos por região na Figura 4. É possível notar uma semelhança no número de pontos de ônibus críticos por regional, diferenciado apenas o bairro onde começou a infecção, que recebe um pouco mais de peso. Após o cenário ser instanciado com usuários infectados no primeiro dia, nos demais dias, novos usuários infectados são contabilizados ao final do dia, e os novos são vacinados no início do dia seguinte. A forma de infecção é a mesma para todos os cenários, em que o usuário é infectado com a probabilidade do valor de

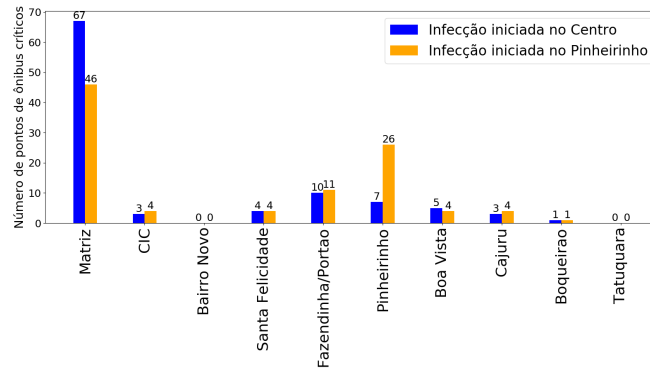
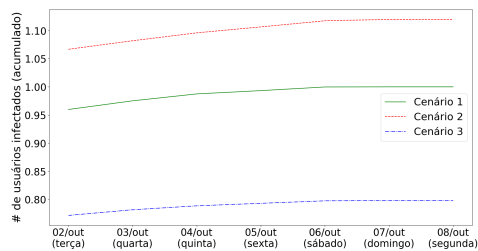
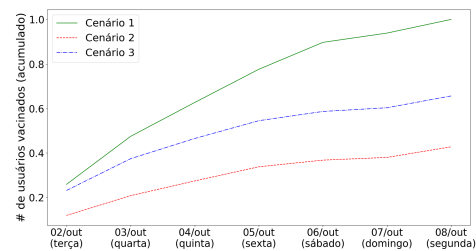


Figura 4. 100 pontos de ônibus mais críticos por regional dos cenários com infecção iniciada no Centro e no Pinheirinho

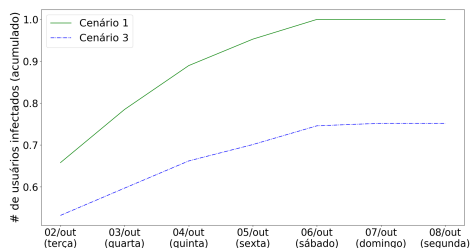
percolação para cada ponto de ônibus acessado no dia. Como a infecção é cumulativa, uma vez infectado, o usuário continua infectado até o final da semana. Um usuário é considerado vacinado quando selecionado com a probabilidade definida em cada cenário. Os usuários vacinados não podem ser infectados e, portanto, não propagam a infecção. Em todos os cenários, foram realizadas 35 repetições para minimizar possíveis erros devido à aleatoriedade na definição de usuários infectados e vacinados.



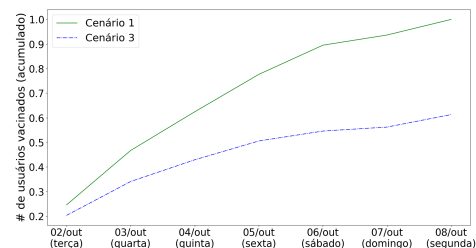
(a) Usuários infectados acumulados - Centro



(b) Usuários vacinados acumulados - Centro



(c) Usuários infectados acumulados - Pinheirinho



(d) Usuários vacinados acumulados - Pinheirinho

Figura 5. Número de usuários infectados e vacinados normalizados pelo cenário de vacinação aleatória de acordo com a infecção inicial

O objetivo de estabelecer diferentes cenários de vacinação é verificar se o uso de uma estratégia pode ser significativamente benéfica em vez de uma vacinação aleatória em termos de crescimento do número de infectados e do número de doses utilizadas. Para uma melhor visualização, calculamos a média do número de infectados/vacinados em cada dia da semana e, em seguida, normalizamos pelo cenário de vacinação aleatória. Para normalizar, dividimos essa média de número de infectados/vacinados de cada dia

pelo número de infectados/vacinados total do cenário de vacinação aleatória. Observando as Figuras 5(a) e 5(b), que mostram os cenários com infecção inicial no Centro, é possível ver que o cenário 3 tem vantagens tanto no número de doses quanto no número de pessoas infectadas em relação à vacinação aleatória. Por outro lado, o segundo cenário mostra que usar uma estratégia com uma quantidade ínfima de doses acaba sendo pior quando comparado a uma vacinação aleatória em termos de disseminação da infecção. As mesmas análises foram realizadas onde a infecção inicial ocorreu no Pinheirinho. Observando as Figuras 5(c) e 5(d), podemos ver que o crescimento do número de infectados é mais gradual, o que é esperado já que o Pinheirinho é um bairro menos conectado comparado ao Centro. No entanto, os resultados são semelhantes. Com a aplicação da estratégia de vacinação nos 100 pontos de ônibus mais críticos, observamos uma diminuição de 30% nas infecções ao final da semana, utilizando menos de 60% das doses do que no cenário de vacinação aleatória.

5. Discussão Final e Conclusões

Neste artigo foi apresentado um estudo sobre estratégias de vacinação para ajudar a reduzir o contágio em pontos de interesse pré-estabelecidos. Foi possível estabelecer pontos críticos de interesse com base na centralidade de percolação usando um modelo de rede multicamada variável no tempo. Apresentamos uma aplicação usando o Multi-Aspect Graph (MAG), denominado *UserPoint-MAG*, focado em cenários de vacinação, onde a análise temporal é de suma importância para determinar o contágio e, conseqüentemente, sua redução. Usando dados de grande escala do sistema de transporte público de Curitiba, estabelecemos um modelo para analisar as estratégias propostas neste trabalho e seu impacto na mobilidade da cidade. Nosso modelo representou uma rede com duas camadas e quatro aspectos. As camadas foram relacionadas aos usuários de cartões transporte e pontos de ônibus da cidade. Os aspectos foram os pontos de ônibus, um identificador, um dos bairros de Curitiba, e a data e hora do cadastro do cartão transporte. Três cenários foram avaliados para uma compreensão mais ampla de nossa abordagem. Concluiu-se que a vacinação em pontos críticos de interesse, identificados por nossa abordagem, especificamente nos 100 pontos de ônibus mais críticos, com 0,5 de probabilidade de vacinar os usuários que passam nesses pontos, apresenta vantagens no número de doses e no número de usuários infectados comparados à vacinação aleatória, utilizando-se menos doses de vacinas e ainda assim reduzindo o número de infectados que utilizam o sistema de transporte público. Ficou claro que as estratégias de vacinação aplicadas em pontos críticos de interesse foram mais eficientes e eficazes do que uma vacinação aleatória. Diferentes fatores devem ser avaliados na contaminação por doenças transmitidas pelo ar, mas o contato entre as pessoas é certamente um dos mais relevantes. A identificação de pontos críticos de interesse pode auxiliar os gestores públicos em campanhas de prevenção de doenças e estratégias de vacinação, otimizando recursos e atingindo uma população maior. Note o potencial que o modelo traz para avaliar outras estratégias. É importante também notar que as aplicações apresentadas neste trabalho estão relacionadas à atual pandemia de SARS-CoV-2. No entanto, o modelo proposto não se limita a isso. Com algumas adaptações, o modelo MAG deste trabalho pode ser aplicado em outros contextos, como otimização em transporte público ou até mesmo com diferentes pontos de interesse. Por exemplo, ao acrescentar uma camada que represente escolas ou universidades, é possível analisar se há acesso adequado e de qualidade a essas instituições que utilizam o transporte público.

Agradecimentos

Este trabalho foi parcialmente apoiado pelo projeto GoodWeb (processo #2018/23011-1 da agência de fomento FAPESP) e pelo CNPq (processo #310998/2020-4).

Referências

- Bao, J., Shi, X., and Zhang, H. (2018). Spatial analysis of bikeshare ridership with smart card and poi data using geographically weighted regression method. *IEEE Access*, 6:76049–76059.
- Belyi, A., Bojic, I., Sobolevsky, S., Sitko, I., Hawelka, B., Rudikova, L., Kurbatski, A., and Ratti, C. (2017). Global multi-layer network of human mobility. *International Journal of Geographical Information Science*, 31(7):1381–1402. PMID: 28553155.
- Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028.
- Cardillo, A., Zanin, M., Gómez-Gardeñes, J., Romance, M., del Amo, A. J. G., and Boccaletti, S. (2013). Modeling the multi-layer nature of the european air transport network: Resilience and passengers re-scheduling under random failures. *The European Physical Journal Special Topics*.
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., and Leskovec, J. (2021). Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87.
- Chodrow, P. S., al Awwad, Z., Jiang, S., and González, M. C. (2016). Demand and congestion in multiplex transportation networks. *PLOS ONE*, 11(9):1–10.
- da Costa, B. M., Bechara, J. V., Wehmuth, K., and Ziviani, A. (2018). A multilayer and time-varying structural analysis of the brazilian air transportation network. In *LADaS@VLDB*, pages 57–64.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S., and Arenas, A. (2013). Mathematical formulation of multilayer networks. *Phys. Rev. X*, 3:041022.
- De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., and Arenas, A. (2015). Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*, 6(1):6868.
- Estrada, E. (2012). *The structure of complex networks: theory and applications*. Oxford University Press.
- Goscé, L. and Johansson, A. (2018). Analysing the link between public transport use and airborne transmission: mobility and contagion in the london underground. *Environmental Health*, 17(1):84.
- Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. *SIGMOD Rec.*, 14(2):47–57.
- Hristova, D., Williams, M. J., Musolesi, M., Panzarasa, P., and Mascolo, C. (2016). Measuring urban social diversity using interconnected geo-social networks. In *Proceedings of the 25th International Conference on World Wide Web, WWW '16*, page 21–30,

- Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.
- Huang, L., Yang, Y., Gao, H., Zhao, X., and Du, Z. (2018). Comparing community detection algorithms in transport networks via points of interest. *IEEE Access*, 6:29729–29738.
- Instituto Brasileiro de Geografia e Estatística - IBGE (2010). Censo 2010. <https://censo2010.ibge.gov.br>. [Online; acessado em 07-Novembro-2019].
- Jacobsen, K. A., Burch, M. G., Tien, J. H., and Rempala, G. A. (2018). The large graph limit of a stochastic epidemic model on a dynamic multilayer network. *Journal of Biological Dynamics*, 12(1):746–788. PMID: 30175687.
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–271.
- Kurant, M. and Thiran, P. (2006). Layered complex networks. *Physical Review Letters*, 96(13).
- Liu, Y. and Seah, H. S. (2015). Points of interest recommendation from gps trajectories. *International Journal of Geographical Information Science*, 29(6):953–979.
- Lv, Q., Qiao, Y., Zhang, Y., Abdesslem, F. B., Lin, W., and Yang, J. (2018). Measuring geospatial properties: Relating online content browsing behaviors to users’ points of interest. *Wireless Personal Communications*, 101(3):1469–1498.
- Meyers, L. A., Newman, M., and Pourbohloul, B. (2006). Predicting epidemics on directed contact networks. *Journal of Theoretical Biology*, 240(3):400–418.
- Mo, B., Feng, K., Shen, Y., Tam, C., Li, D., Yin, Y., and Zhao, J. (2021). Modeling epidemic spreading through public transit using time-varying encounter network. *Transportation Research Part C: Emerging Technologies*, 122:102893.
- Müller, S. A., Balmer, M., Charlton, B., Ewert, R., Neumann, A., Rakow, C., Schlenker, T., and Nagel, K. (2020). Using mobile phone data for epidemiological simulations of lockdowns: government interventions, behavioral changes, and resulting changes of reinfections. *medRxiv*.
- Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.
- Oselio, B., Kulesza, A., and Hero, A. O. (2014). Multi-layer graph analysis for dynamic social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):514–523.
- Piraveenan, M., Prokopenko, M., and Hossain, L. (2013). Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks. *PLOS ONE*, 8(1):1–14.
- Rodrigues, D., Boukerch, A., Silva, T. H., Loureiro, A., and Villas, L. (2017). Sma-framework: Urban data integration framework for mobility analysis in smart cities. In *Proc. of the 20th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Miami, USA.
- Sahneh, F. D., Vajdi, A., Melander, J., and Scoglio, C. M. (2019). Contact adaption during epidemics: A multilayer network formulation approach. *IEEE Transactions on Network Science and Engineering*, 6(1):16–30.

- Sander, L., Warren, C., Sokolov, I., Simon, C., and Koopman, J. (2002). Percolation on heterogeneous networks as a model for epidemics. *Mathematical Biosciences*, 180(1):293–305.
- Santin, P., Gubert, F. R., Fonseca, M., Munaretto, A., and Silva, T. H. (2020). Characterization of public transit mobility patterns of different economic classes. *Sustainability*, 12(22):9603.
- Sun, L., Axhausen, K. W., Lee, D.-H., and Cebrian, M. (2014). Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific Reports*, 4(1):5099.
- Tang, W., Chakeri, A., and Krim, H. (2020). Discovering urban functional zones by latent fusion of users gps data and points of interests.
- Taniguchi, G. and Duarte, F. (2012). Personal smart cards: From transportation to a city smart card—the database integration of public services in Curitiba. In *City Competitiveness and Improving Urban Subsystems: Technologies and Applications*, pages 217–232. IGI Global.
- Thomee, B., Arapakis, I., and Shamma, D. A. (2016). Finding social points of interest from georeferenced and oriented online photographs. *ACM Trans. Multimedia Comput. Commun. Appl.*, 12(2).
- Urbanização de Curitiba S/A - URBS (2018). URBS em números. <https://www.urbs.curitiba.pr.gov.br/institucional/urbs-em-numeros>. [Online; acessado em 07-Novembro-2019].
- Ventresca, M. and Aleman, D. (2013). Evaluation of strategies to mitigate contagion spread using social network characteristics. *Social Networks*, 35(1):75–88.
- Wang, Y., Liang, Y., Sun, H., and Yang, Y. (2020). Emergency response for covid-19 prevention and control in urban rail transit based on case-based reasoning method. *Discrete Dynamics in Nature and Society*, 2020:6689089.
- Wehmuth, K., Fleury, E., and Ziviani, A. (2016). On multiaspect graphs. *Theoretical Computer Science*, 651:50–61.
- World Health Organization (2022). Who coronavirus disease (covid-19) dashboard. <https://covid19.who.int/table>. [Online; accessed in 31-January-2022].
- Yildirimoglu, M. and Kim, J. (2018). Identification of communities in urban mobility networks using multi-layer graphs of network traffic. *Transportation Research Part C: Emerging Technologies*, 89:254–267.
- Zheng, M., Wang, W., Tang, M., Zhou, J., Boccaletti, S., and Liu, Z. (2018). Multiple peaks patterns of epidemic spreading in multi-layer networks. *Chaos, Solitons Fractals*, 107:135–142.
- Zhu, S., Srebric, J., Spengler, J. D., and Demokritou, P. (2012). An advanced numerical model for the assessment of airborne transmission of influenza in bus microenvironments. *Building and Environment*, 47:67 – 75. International Workshop on Ventilation, Comfort, and Health in Transport Vehicles.