Criação de Modelo para Simulação de Movimentação de Ônibus a Partir de Dados Reais*

Melissa Wen[1], Thatiane de O. Rosa[1,3], Mariana C. Souza[2], Robson P. Aleixo[1], Camilla Alves[1], Lucas Sá[1], Eduardo Felipe Zambom Santana[1], Fabio Kon[1]

[1]Universidade de São Paulo (USP)

[2]Universidade Federal de Mato Grosso do Sul (UFMS)

[3]Instituto Federal de Educação, Ciência e Tecnologia do Tocantins (IFTO)

{wen,thatiane,lucassa,efzambom,fabio.kon}@ime.usp.br

mariana.caravanti@aluno.ufms.br, robson.aleixo@optimumsolucoes.com,

camilla.almeida.silva@usp.br

Resumo. *A dinâmica socioespacial de uma cidade sofre constantes mudanças ao longo do tempo. Por consequência, a malha viária e o sistema de transporte público precisam de otimizações contínuas para atender às demandas dos cidadãos. Nesse contexto, uma boa alternativa para reduzir custos e impactos na avaliação de soluções é o emprego de simuladores que utilizem modelos consistentes com a realidade. Em vista disso, processamos dados de deslocamento e de planejamento do sistema de ônibus de São Paulo para melhorar o modelo de movimentação de ônibus usado pelo InterSCSimulator, um simulador altamente escalável para cidades inteligentes. Apresentamos um modelo de movimentação baseado em dados reais do serviço de ônibus de São Paulo a fim de tornar o simulador mais eficaz ao recriar cenários de mobilidade urbana.*

Abstract. *The socio-spatial dynamics of a city undergoes constant changes over time. Consequently, the road network and the public transport system need continuous optimization to meet citizen demands. An alternative to reduce costs and impacts on evaluation of solutions is the use of simulators and models consistent with reality. Considering that, we processed vehicle tracking data and bus system planning information of São Paulo to improve the bus movement model used by InterSCSimulator, a highly scalable simulator for smart cities. In this paper, we present a mobility model based on real data from the São Paulo bus service to make the simulator more effective when recreating urban mobility scenarios.*

1. Introdução

Áreas metropolitanas vêm crescendo em todo o mundo e, consequentemente, expandindo em extensão e população [1, 2]. Tal crescimento influencia diretamente as redes de transporte, gerando problemas como congestionamentos, longos tempos de espera para deslocamento e poluição [1, 3]. Para atacar tais problemas é importante investir em soluções

*Este trabalho foi desenvolvido no contexto do INCT da Internet do Futuro para Cidades Inteligentes, apoiado pela FAPESP proc. 2014/50937-1 e CNPq proc. 465446/2014-0.

inteligentes de mobilidade urbana, no planejamento e melhoria do transporte público [4]. Planejar e testar soluções de mobilidade urbana em ambientes reais são atividades complexas, que envolvem altos custos e causam grandes impactos [1, 5]. Nesse contexto, os simuladores apresentam-se como uma boa ferramenta aos gestores de redes de transportes, pois torna possível testar cenários, sem a necessidade da construção de uma infraestrutura física [4]. Entretanto, para que os simuladores atinjam os seus objetivos é fundamental o desenvolvimento de modelos coerentes com a realidade do ambiente estudado [6].

Para preencher as lacunas de escalabilidade e usabilidade dos simuladores de cidades inteligentes, foi criado o InterSCSimulator - um simulador fácil de utilizar e capaz de reproduzir áreas metropolitanas completas a partir de um mapa contendo dezenas de milhares de ruas e milhões de veículos em movimento [4]. Porém, os modelos utilizados por esse simulador ainda não levam em consideração as estocasticidades da realidade do sistema de transporte coletivo. Diante disso, apresentamos neste trabalho um modelo que visa incorporar alguns aspectos estocásticos às simulações de movimentação dos ônibus. Para desenvolvê-lo, utilizamos dados AVL (*Automatic Vehicle Location*) e de planejamento do sistema de transporte público da cidade de São Paulo. Por fim, validamos nosso modelo por meio de análises comparativas entre o comportamento observado a partir de dados reais de movimentação e o comportamento observado a partir de eventos simulados.

Este artigo está organizado da seguinte forma: na Seção 2 apresentamos conceitos de cidades inteligentes e simulação de mobilidade urbana. Na Seção 3 são apresentados alguns trabalhos relacionados. Na Seção 4 a metodologia do estudo é relatada. Na Seção 5 o modelo de movimentação de ônibus é descrita. A Seção 6 apresenta discussões sobre a validação do modelo. Por fim, a Seção 7 aponta as conclusões e trabalhos futuros.

2. Cidades Inteligentes e Simulação de Mobilidade Urbana

O conceito Cidade Inteligente é bastante amplo e possui um conjunto diversificado de termos considerados sinônimos [7, 8, 9]. Diante disso, adotaremos uma definição mais coerente com este estudo, apresentada por Marsal-Llacuna et al. [10]. Segundo os autores, cidades inteligentes são aquelas construídas com o melhor uso de dados, informações e tecnologias de informação (TI) para monitorar e otimizar a infraestrutura existente, fornecer serviços mais eficientes aos cidadãos, aumentar a colaboração entre diferentes atores econômicos e incentivar modelos de negócios inovadores nos setores público e privado. De forma genérica, tal conceito envolve encontrar soluções para lidar com desafios das cidades a partir da utilização de tecnologias da informação e comunicação [2].

Dentre os desafios enfrentados pelas cidades, um dos mais relevantes está relacionado ao domínio de mobilidade [3, 8]. Nesse domínio, problemas como poluição, fluxo de veículos, congestionamento e longos tempos de espera, impactam diretamente a economia da cidade, meio ambiente e a qualidade de vida dos seus cidadãos [3]. Com isso, Neirotti et al. identificam que os estudos sobre mobilidade inteligente possuem três principais objetivos: otimizar a logística e o transporte em áreas urbanas considerando condições de tráfego e consumo de energia, fornecer aos usuários informações dinâmicas e multimodais para obter um fluxo de veículos mais eficiente e, garantir transporte público sustentável por meio de combustíveis ecológicos e sistemas de propulsão inovadores [11].

Diante disso, planejar e testar soluções de mobilidade urbana mostram-se atividades complexas e desafiadoras, devido aos custos (financeiros, de energia e de tempo) e

impactos (sociais, políticos, infraestrutura e meio ambiente) gerados [1, 5]. Como uma estratégia mais viável aos gestores públicos e privados de redes de transportes, é possível simular cenários para testar diferentes soluções envolvendo vários domínios como fluxo de veículos, transporte público e utilização de recursos [4]. No entanto, Ros et al. afirmam que os resultados dessas simulações são de fato significativos apenas quando modelos coerentes com a realidade são processados [6]. De Dios Ortúzar et al. explicam ainda que para projetar modelos que representem contextos futuros é necessário utilizar as variáveis de estudo do presente [5]. Logo, os modelos devem receber, como entrada, dados do cenário atual (ano-base) e possíveis dados do cenário idealizado (dados projetados).

Neste estudo, adotamos o InterSCSimulator, um simulador escalável e de código aberto desenvolvido para o contexto de Cidades Inteligentes. Ele é capaz de simular cenários de trânsito com milhões de agentes, usando um mapa real de uma grande cidade [4] e possui quatro componentes principais, denominados: *Definição do Cenário*, responsável por ler os arquivos de entrada e criar o grafo da cidade e os veículos que serão simulados; *Motor de Simulação* onde são executados os algoritmos e os modelos de simulação; *Visualização do Mapa* que utiliza a saída do Motor de Simulação e o mapa da cidade simulada para criar uma animação visual da movimentação dos veículos no grafo da cidade; e *Visualização de Gráficos* que também utiliza a saída do Motor de Simulação e gera diversos gráficos com análises sobre o trânsito da cidade (Figura 1). Partindo desses quatro componentes, temos que qualquer modelo a ser avaliado com o simulador precisa descrever o mapa da cidade e as viagens a serem simuladas para definição de cenários. A partir disso, o resultado do modelo pode ser visto por meio de animação visual e gráficos.

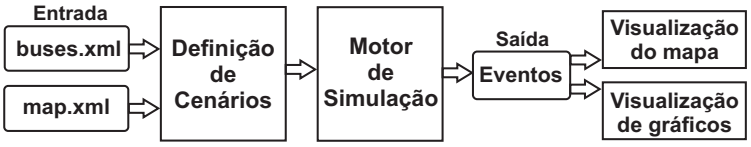

Figura 1. Componentes do InterSCSimulator [4].

Após definir os principais conceitos que sustentam esta pesquisa, a próxima seção explora alguns estudos científicos que analisam aspectos de mobilidade urbana e simulação que vão ao encontro do objetivo aqui proposto.

3. Trabalhos Relacionados

Uma série de estudos analisam diferentes aspectos da mobilidade urbana em cidades inteligentes. Ao considerar o objetivo deste artigo, duas frentes de pesquisa são importantes:

Mobilidade de ônibus nas cidades: estudos que analisam características das viagens dos ônibus de uma cidade com base em dados capturados do GPS (Sistema de Posicionamento Global) dos veículos (denominados AVL) e de outros sensores como smartphones dos usuários. Zhou et al. realizaram um estudo com o objetivo de prever o número de passageiros e o tempo de chegada de ônibus [12]. Para isso, construíram uma base de aprendizagem a partir de um conjunto de modelos de predição, que processava os dados de GPS dos ônibus e de um aplicativo desenvolvido para os usuários do serviço.

Rahman et al. buscaram prever, em tempo real, o horário de chegada dos ônibus [13] a partir da análise das distâncias entre pontos de medição do GPS, denominados "pseudo-horizontes". Para avaliar as variações nos tempos das viagens, consideraram

o comportamento de acordo com o desvio-padrão, coeficiente de variação e obliquidade. Yu et al. visaram prever a ocorrência de situações em que dois ônibus de uma mesma linha chegam a uma mesma parada quase que simultaneamente, fenômeno conhecido como *"bus bunching"* [14]. Para detectar esse fenômeno utilizaram a regressão LS-SVM (*Least Squares Support Vector Machine*), com dados temporais e espaciais dos passageiros.

Simuladores de mobilidade urbana: uma série de simuladores de trânsito consideram diferentes agentes (veículos, pessoas, sensores) e utilizam dados reais para simular cenários de mobilidade urbana. Como exemplo, temos o MATSim e o DTALite. Entretanto nenhum desses simuladores é capaz de reproduzir uma área metropolitana completa contendo um mapa com milhares de ruas e milhões de veículos em movimento [4]. Para preencher essa lacuna, utilizamos neste trabalho o InterSCSimulator - um simulador escalável, fácil de utilizar, paralelizável e distribuido - capaz de reproduzir cenários complexos de Cidades Inteligentes.

4. Metodologia

O InterSCSimulator tem o potencial de agregar múltiplas contribuições aos métodos tradicionais de modelagem em planejamento de transportes. Visando dotar o simulador de maior precisão na descrição da oferta de ônibus, desenvolvemos um modelo de movimentações de ônibus com base em dados GTFS (*General Transit Feed Specification*) de planejamento e dados de AVL do sistema de transporte público da cidade de São Paulo. Nossos estudos foram guiados pelas seguintes questões de pesquisa:

RQ1. *"Quais dados de AVL e de planejamento de transporte público são necessários para desenvolver um modelo que descreva de maneira mais realista a movimentação de ônibus em uma cidade?"*

RQ2. *"Como esses dados devem ser sintetizados para simular adequadamente a movimentação das linhas de ônibus de uma cidade?"*

4.1. Estudo de Caso

São Paulo é uma cidade com grandes dimensões populacionais, espaciais e, consequentemente, grandes desafios de mobilidade urbana. Atualmente, sua frota de ônibus possui cerca de 14,4 mil veículos deslocando 6 milhões de passageiros por dia útil. Em 2017, as suas mais de 2 mil linhas de ônibus atenderam um total de 1.630.604.027 passageiros, segundo a SPTrans, empresa responsável por gerir o sistema de ônibus da cidade.

Para aprimorar as simulações, incorporamos ao modelo utilizado pelo InterSCSimulator informações reais de circulação dos ônibus. Para isso, foram utilizadas duas fontes de dados: GTFS e AVL. Os dados de GTFS refletem o planejamento do serviço, do qual extraímos dados que representam informações estáticas do sistema, tais como: linhas de ônibus, itinerário, localização das paradas de ônibus e o trajeto realizado de uma parada para a outra. Esses dados são abertos e podem ser acessados por meio da página web da SPTrans [1].

Os dados de AVL representam o comportamento real dos ônibus, ou seja, os horários reais de início de circulação, frequências das saídas, assim como velocidade média de deslocamento de cada veículo. Essas informações são obtidas via tecnologia

[1] http://www.sptrans.com.br/desenvolvedores/GTFS.aspx

GPS, e foram fornecidas para este trabalho pela Scipopulis [2], uma *startup* que presta serviços e desenvolve produtos a partir de dados coletados por meio da API do sistema de monitoramento de transporte "Olho Vivo"[3]. Devido à experiência da *startup* em mobilidade urbana, tais dados estavam pré-processados, limpos e relacionados de forma a facilitar o entendimento e extração das informações mais relevantes para construção do modelo. Além disso, alguns erros de geolocalização estavam mitigados, o que favoreceu a precisão do modelo desenvolvido.

Para este trabalho, a *startup* Scipopulis disponibilizou registros diários de monitoramento de uma semana típica (22 a 28 de outubro de 2017), ou seja, 7 dias corridos onde não houveram registros de eventos que impactassem o funcionamento do transporte coletivo. Além desses, foram disponibilizados dados de outros 3 dias considerados atípicos: véspera de feriado prolongado (13 de abril de 2017), clássico de futebol com lotação máxima no estádio Morumbi (24 de setembro de 2017) e intensos protestos pela capital (15 de março de 2017). A combinação dos dados AVL e GTFS trouxe informações da dinâmica de movimentação de 14.139 ônibus e 2.183 linhas no ano de 2017.

4.2. Análise de Dados e Validação do Modelo

Tratando-se da primeira iniciativa de modelagem de dados de AVL e GTFS para utilização no InterSCSimulator, optamos por dotar o modelo de três informações principais: (a) *itinerário de todas as linhas*, representado como arestas entre paradas consecutivas; (b) *velocidades médias das arestas por intervalo de hora*; (c) *o intervalo médio de saída dos ônibus dos terminais por intervalo de hora* para todas as linhas e seu início de circulação;

Assumimos como hipótese que a informação da velocidade média em cada aresta permite que o modelo seja capaz de representar a movimentação de ônibus, incorporando os efeitos da imprevisibilidade e não homogeneidade do mundo real. Tais efeitos implicam em variações nos tempos totais de viagem de uma linha. Também adicionamos as linhas de ônibus representadas no modelo informações referentes à frequência de saída de seus veículos dos pontos terminais. Com isso, busca-se que o modelo reflita o número de veículos da frota em circulação ao longo do dia. Como análise exploratória das fontes de dados, averiguamos a qualidade das informações de circulação com base no registro diário do início de operação dessas linhas, registro de saída do terminal e registro completo do deslocamento dos seus veículos. Também verificamos a compatibilidade entre as frequências planejadas (dados de GTFS) e as observadas (dados de AVL).

A partir da qualidade dos registros, elegemos duas linhas que fazem parte do deslocamento diário da população de baixa renda (residem na periferia e trabalham no centro) para comparar o comportamento real e simulado. O processo de escolha também levou em consideração o estudo da variação dos tempos de viagem como critério de validação do modelo. Dado que, linhas que possuem viagens longas são mais sucetíveis a variações, logo mais adequadas para testar a fidelidade do modelo à realidade, as seguintes rotas foram eleitas: 856R-10-*Socorro/Lapa* e 4311-10-*Term. S. Mateus/Term. Prq. D. Pedro*.

Consideramos a qualidade das informações levantadas acima para validação simplificada do modelo. Assim, para realizar as simulações, elegemos o dia 26 de outubro de 2017, uma quinta-feira na qual verificou-se a ausência de eventos atípicos na cidade que

[2]https://www.scipopulis.com/
[3]http://olhovivo.sptrans.com.br/

pudessem gerar impactos significativos na circulação dos ônibus. Realizamos primeiramente o processamento dos dados AVL para obter as durações reais de todas as viagens com registros completos para cada linha em estudo. Posteriormente, para essas mesmas linhas, verificamos no simulador as durações de todas as viagens iniciadas ao longo das 24 horas do dia. Por fim, comparamos os valores obtidos nesses dois processos anteriores para verificar a similaridade entre os valores de AVL e os resultantes no simulador, aplicando o *Teste t de Student* para validar a eficácia do modelo na representação da realidade.

5. Resultados

A partir dos dados de AVL e GTFS do estudo de caso, construímos um modelo que simula a movimentação diária de 14.139 veículos, operando em 2.183 linhas de ônibus, a fim de representar 8 cenários: 7 dias da semana e 1 dia atípico (véspera do feriado de páscoa). Tais modelos foram desenvolvidos a partir de um conjunto de modificações nos atributos dos arquivos originais de entrada do InterSCSimulator e estão disponíveis na página do projeto InterSCity [4].

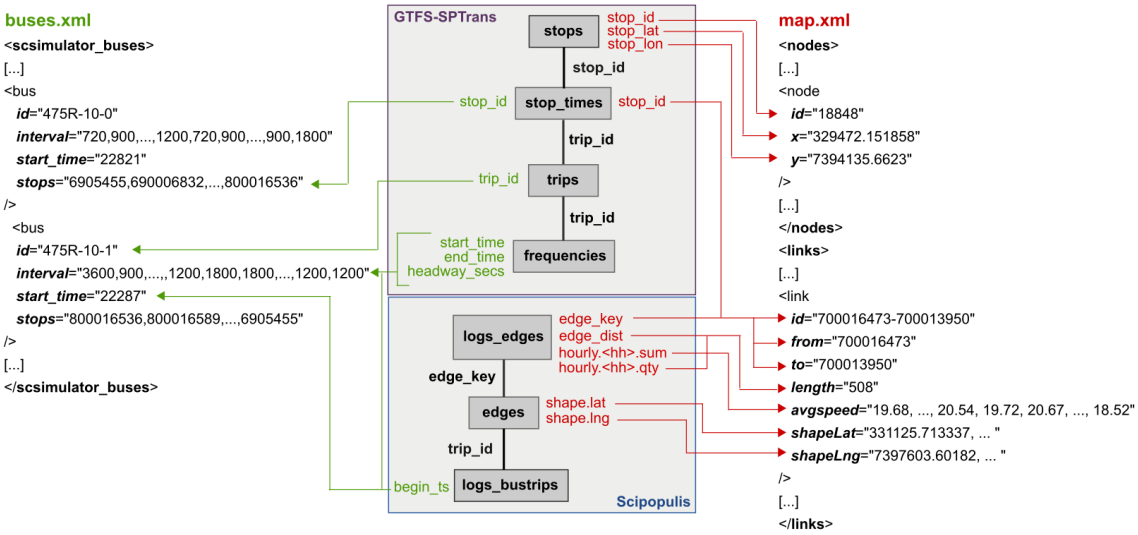

Figura 2. Relacionamento das bases de dados para construção do modelo.

Conforme descrito na Figura 1, o modelo de movimentação dos ônibus é representado por dois arquivos no formato XML (*eXtensible Markup Language*), denominados *map* e *buses*. O arquivo *buses*, descreve as linhas de ônibus do sistema de transporte coletivo *bus* e caracteriza todas as viagens de ônibus que serão realizadas durante a simulação. Já o arquivo *map*, define a malha viária do sistema de ônibus, onde cada parada de ônibus é representada pela marcação *node* e as arestas que ligam duas paradas são representadas por *links*. A Figura 2 apresenta a estrutura desses dois arquivos de entrada e a maneira como seus atributos se relacionam com as informações das fontes de dados.

Considerando o modelo original, modificamos a sua arquitetura para permitir ao simulador representar o comportamento não-homogêneo do serviço de ônibus da cidade ao longo do dia, de forma que seus atributos passam a ter valores variáveis por intervalo de hora. De maneira resumida, os atributos de frequência de saída dos ônibus deixaram de variar apenas em termos de períodos de pico e não-pico, e passaram a variar por intervalo

[4]http://interscity.org/software/interscsimulator/

buses.xml			
buses			
id	Identifica unicamente uma linha		
interval	Lista de 24 posições que armazena a frequência de saída do terminal, em segundos, dos seus ônibus por faixa de hora		
start_time	Horário em que a linha começou a funcionar		
stops	Lista de ID das paradas de ônibus que fazem parte do seu itinerário de viagem		
map.xml			
nodes		**links**	
id	Identificador único da parada de ônibus	id	Identificador único formado pela união do id do *node* origem e *node* destino.
lat	Localização geográfica latitudinal	to	Id da parada de ônibus de destino
lng	Localização geográfica longitudinal	from	Id da parada de ônibus de origem
		length	Distância em metros entre as paradas de origem e destino
		avgspeed	Lista de 24 posições, representando as 24 horas do dia, onde são armazenadas a velocidade média dos ônibus que transitaram no *link*
		shapeLat shapeLng	Coordenadas latitudinais e longitudinais de pontos de referência que estão contidos dentro do espaço que compõe o *link* e determinam a forma do trajeto

Tabela 1. Descrição dos atributos que compõem o modelo.

de hora. Ainda com o processamento dos dados, o deslocamento dos veículos de uma parada de ônibus a outra deixou de simular o melhor caso (velocidade máxima da via e melhor trajeto) e passou a retratar, por faixa de hora, a velocidade média real do percurso real entre duas paradas. A descrição detalhada dos atributos é apresentada na Tabela 1.

Algoritmo 1: Model_generator

Entrada: Dados AVL (edges, logs_bustrips, logs_edges_speed) e dados GTFS(frequencies, stop_times, trips)
Saída: Arquivos buses.xml e maps.xml

```
#Definição do Modelo
modelo : { bus : { id, start_time ← 0
                   interval ← lista[24]
                   stops ← lista  }
           node : { id, lat, lng ← 0 }
           link :  { id, from, to, length ← 0
                     avgspeed ← lista[24]
                     shapeLat, shapeLng ← lista }
         }

Função Principal( )
    dataRef ← <data do modelo_gerado>
    #Gera arquivo buses.xml
    para cada elemento viagem em GTFS.trips faça
        onibus ← novo modelo.bus
        onibus.id ←  viagem.trip_id
        #Obtém horário inicial de saída do ônibus
        primeiraViagem ← encontrePrimeiraViagem(AVL.logs_bustrips, onibus.id, dataRef)
        onibus.start_time ← primeiraViagem.begin_ts
        onibus.interval  ← calculaIntervaloSaidaPorHora(AVL.logs_bustrips, onibus.id, dataRef)
        onibus.stops  ← crieItinerario(GTFS.stops, onibus.id)
        escreva objeto onibus no arquivo buses.xml

    #Gera arquivo Maps.xml
    paradasDeOnibus ← coleteInformacoesParadas(GTFS.stops)
    escreva paradasDeOnibus no arquivo maps.xml
    to, from ← string
    para cada elemento parada em GTFS.stop_times faça
        se parada é o primeiro ponto do itinerário de uma viagem faça
            from ← parada.id
        senão faça
            to ← parada.id
            aresta ← concatene(from, to)
            from ← parada.id
            arestaId ← aresta
            #Obtém velocidade média em cada edge por hora
            arestaVel ← ColeteVelocMediaArestasPorHora(AVL.logs_edges_espeed, arestaId, dataRef)
            #Obtém estrutura das arestas
            arestaEstr ← coleteInfoAresta(AVL.edges, GTFS.stop_times, arestaId, arestaVel, dataRef)
            escreva  arestasEstr no arquivo maps.xml

Função calculaIntervaloSaidaPorHora(AVL.logs_bustrips, onibus.id, dataRef)
    contaOnibus, intervalo ← lista[24] de inteiros
    encontre em AVL.logs_bustrips todos os elementos tal que \
     date = dataRef e trip_id = onibus.id
        #Ponto inicial da viagem
        para todos os elementos viagem em AVL.logs_bustrips onde \
         begin_stop_seq = 0 faça
            horaReferencia ← viagem.begin_ts.hour
            incremente valor de contaOnibus[horaReferencia]
    #Cálculo das frequências de saída para cada hora do dia
    intervalo ← calcFrequenciasPorSegundo(contaOnibus)
    retorne intervalo

Função ColeteVelocMediaArestaPorHora(AVL.logs_edges_espeed, arestaId, dataRef)
    velocMedia  ← lista[24]
    encontre o elemento aresta em AVL.logs_edges_speed tal que \
     date = dataRef e edge_key = idAresta
        idAresta ← aresta.edge_key
        para cada elemento hora em aresta.hourly faça
            #Cálculo de velocidade média por hora
            velocMedia[hora] ← (aresta.edge_dist/(listaVelocPorHora[hora].sum /
             listaVelocPorHora[hora].qty))
    retorne velocMedia
```

Figura 3. Algoritmo utilizado na construção do modelo de simulação.

A Figura 3 apresenta o algoritmo utilizado no processamento das informações das bases de dados para construção do modelo. Além disso, um vídeo ilustrando a simulação (no InterSCSimulator) das linhas adotadas para validação do modelo foi gravado e disponibilizado no YouTube [5].

6. Discussão

De acordo com a estratégia definida na Seção 4.2, foram eleitas duas linhas de ônibus para validação simplificada do modelo. A comparação dos tempos de viagem é ilustrada

[5] www.youtube.com/watch?v=UO4kFBNqhXY&feature=youtu.be

na Figura 4 e refere-se aos horários entre 6:00 e 23:59 do dia 26 de outubro de 2017.

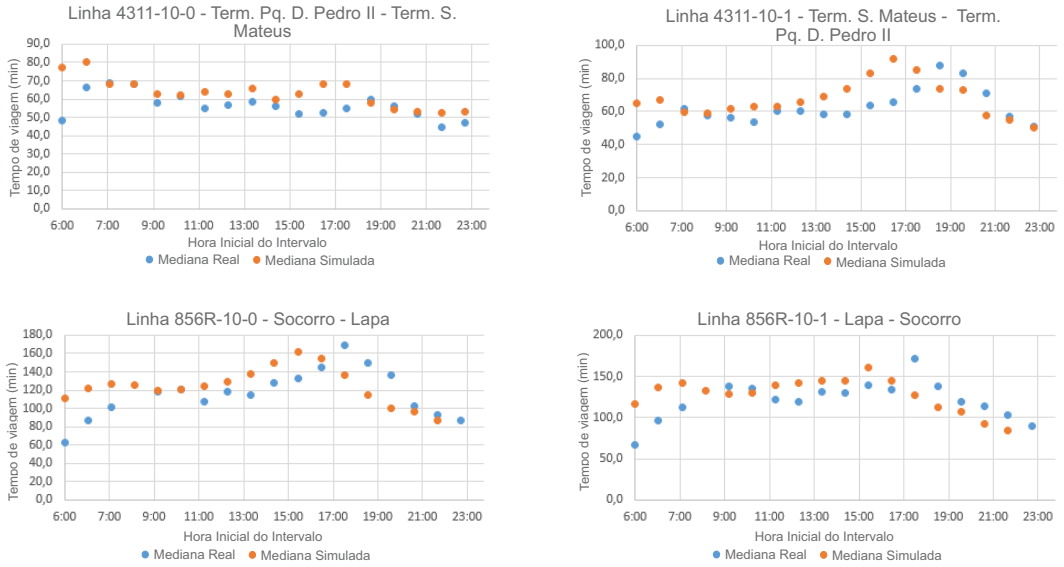

Figura 4. Medianas das durações das viagens reais e simuladas.

Ainda ao analisar os gráficos da Figura 4, podemos verificar que as medianas dos tempos de viagens reais possuem valores maiores em horários de pico, quando comparados aos horários de não-pico. Esse padrão de comportamento também pode ser observado nas medianas dos tempos de viagens simuladas. Sabendo-se que para longas distâncias as velocidades dos ônibus possuem uma distribuição normal [13], também aplicamos o *Teste t de Student* para verificar a similaridade entre as durações das viagens reais e simuladas.

Teste t de Student 856R-10	Real (min)	Simulado (min)
Média	120	127
Número de Observações	57	57
Hipótese nula- Diferença de Médias igual a zero	0	
Valor de t	112	
Graus de Liberdade	-1,53	
P(T<=t) duas-caudas	0,13	
t crítico duas-caudas	1,98	

Teste Z 4311-10	Real (min)	Simulado (min)
Média	60	64
Número de Observações	70	70
Hipótese nula- Diferença de Médias igual a zero	0	
Valor de Z	-1,87	
P(T<=t) duas-caudas	0,06	
t crítico duas-caudas	1,96	

Figura 5. Aplicação de teste de hipótese para comparação dos tempos de viagens das linhas 865R-10 e 4311-10

A Figura 5 apresenta os resultados dos testes de hipótese realizados para as duas linhas em questão. Na primeira linha, 856-10, foram observadas ao todo 57 viagens. Nos dados de AVL, tais viagens obtiveram tempo médio de duração de 120 minutos, contra 127 minutos obtidos nos dados simulados. Uma vez que ambas as amostras apresentaram distríbuições normais com variâncias não diferentes, aplicamos o teste de hipótese *t de student* a fim de validarmos a efetividade do modelo. Podemos observar que o valor de t igual a -1,53 encontra-se dentro do intervalo de confiança, nos garantindo que as médias

de tempo não são diferentes. Já para a segunda linha, 4311-10, foram observadas 70 viagens onde o tempo médio de duração real foi de 60 minutos, contra 64 minutos obtidos pelos dados simulados, como as amostras não apresentaram uma distribuíção normal, para a validação do modelo, aplicamos o teste Z. Nesse caso, notamos que o valor de z, -1,87, encontra-se dentro do intervalo de confiança, nos permitindo concluir que as médias de tempo também não são diferentes. Com esses resultados, garantimos a qualidade de representação do modelo e trazemos maior confiabilidade para as análises.

No estágio atual, o modelo emprega a variável estocástica de velocidade média nas arestas por horário. Apesar de se tratar de uma validação preliminar, podemos perceber que o modelo apresentado permite ao simulador reproduzir as tendências de duração das viagens por faixa de horário, em função das velocidades médias de deslocamento entre paradas consecutivas. Entendemos que melhores resultados podem ser alcançados com a inserção de outros fatores que podem afetar os tempos de viagem, como semaforização e embarque/desembarque de usuários. Por fim, também necessitamos expandir esta análise para as demais linhas do sistema de ônibus simulado.

7. Conclusão

Diversas ferramentas são desenvolvidas para apoiar iniciativas de melhoria de mobilidade urbana. Dentre elas está o InterSCSimulator, um simulador escalável que permite execuções paralelas e distribuídas para representar cenários complexos de cidades inteligentes. Para dotar os cenários gerados pelo InterSCSimulator de aspectos da realidade, desenvolvemos, neste trabalho, um modelo de movimentação de ônibus, utilizando dados reais do sistema de transporte de ônibus da cidade de São Paulo.

A construção da arquitetura deste modelo foi possível a partir da resolução de duas perguntas. Para a primeira, *"Quais dados AVL e de planejamento de transporte público são necessários para desenvolver um modelo que descreva de maneira mais realista a oferta de ônibus de uma cidade?*, utilizamos como estudo de caso a cidade de São Paulo. Avaliamos a integridade dos registros AVL dos seus ônibus quando combinados com os dados GTFS do planejamento de transporte. E assim, definimos quais informações podem ser extraídas das fontes de dados, de modo a incorporar à simulação aspectos reais do sistema de transporte coletivo.

Para responder a segunda pergunta, *"Como esses dados devem ser sintetizados para simular adequadamente a movimentação das linhas de ônibus de uma cidade?"*, apresentamos uma arquitetura para o modelo, onde os seus atributos retratam o comportamento não-homogêneo da circulação dos ônibus, por faixa de horário. Nessa arquitetura, os aspectos estocásticos foram incorporados utilizando a média das velocidades médias dos veículos, por seguimento entre paradas de ônibus. Por fim, também acrescentamos a cada seguimento a descrição real do seu formato.

Novas pesquisas devem ser encaminhadas para incorporar outros aspectos estocásticos de mobilidade urbana ao InterSCSimulator e assim, melhorar a sua acurácia ao prever eventos e retratar a realidade. Como trabalho futuro, em um primeiro momento, é importante realizar uma validação mais consistente do modelo gerado. Além disso, entendemos que é relevante incorporar na modelagem mais elementos que incidam sobre os tempos de viagem de cada linha por intervalo de hora, representando os impactos de outras variáveis além da velocidade média. Também consideramos importante a construção

de modelos que representem o comportamento típico dos ônibus da cidade por dia da semana, compilando dados de diferentes datas de um mesmo dia.

Referências

[1] D. Hall, "Integration opportunities at transit jurisdictional borders," Master's thesis, University of Waterloo, 2013.

[2] A. Caragliu, C. D. Bo, and P. Nijkamp, "Smart cities in europe," *Journal of Urban Technology*, vol. 18, no. 2, pp. 65–82, 2011.

[3] C. Benevolo, R. P. Dameri, and B. D'Auria, "Smart mobility in smart city," in *Empowering Organizations* (T. Torre, A. M. Braccini, and R. Spinelli, eds.), (Cham), pp. 13–28, Springer International Publishing, 2016.

[4] E. F. Z. Santana, N. Lago, F. Kon, and D. S. Milojicic, "Interscsimulator: Large-scale traffic simulation in smart cities using erlang," in *The 18th Workshop on Multi-agent-based Simulation - MABS 2017*, 2017.

[5] J. de Dios Ortúzar and L. G. Willumsen, *Modelling Transport*. John Wiley & Sons, 2011.

[6] F. J. Ros, J. A. Martinez, and P. M. Ruiz, "A survey on modeling and simulation of vehicular networks: Communications, mobility, and tools," *Computer Communications*, vol. 43, pp. 1 – 15, 2014.

[7] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," in *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, dg.o '11, (New York, NY, USA), pp. 282–291, ACM, 2011.

[8] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanović, and E. Meijers, "Ranking of european medium-sized cities," tech. rep., Vienna University of Technology, 10 2007.

[9] E. F. Z. Santana, A. P. Chaves, M. A. Gerosa, F. Kon, and D. S. Milojicic, "Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture," *ACM Comput. Surv.*, vol. 50, pp. 78:1–78:37, Nov. 2017.

[10] M.-L. Marsal-Llacuna, J. Colomer-Llinàs, and J. Meléndez-Frigola, "Lessons in urban monitoring taken from sustainable and livable cities to better address the smart cities initiative," *Technological Forecasting and Social Change*, vol. 90, pp. 611 – 622, 2015.

[11] P. Neirotti, A. D. Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25 – 36, 2014.

[12] C. Zhou, P. Dai, F. Wang, and Z. Zhang, "Predicting the passenger demand on bus services for mobile users," *Pervasive and Mobile Computing*, vol. 25, pp. 48 – 66, 2016.

[13] M. M. Rahman, S. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 453 – 466, 2018.

[14] H. Yu, D. Chen, Z. Wu, X. Ma, and Y. Wang, "Headway-based bus bunching prediction using transit smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 45 – 59, 2016.