

# Como Melhorar a Representação de Traces: Um Estudo de Caso Usando Dados do Transporte Público

Hernani de B. Vieira Jr<sup>1</sup>, Stefany Gaspar<sup>1</sup>, Felipe D. Cunha<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação  
Pontifícia Universidade Católica de Minas Gerais  
R. Dom José Gaspar, 500 – 30535-901 – Belo Horizonte – MG – Brazil.

hernani.vieira@sga.pucminas.br, stefanygaspar@gmail.com,

felipe@pucminas.br

**Abstract.** *Simulation is the most commonly used approach to evaluate protocols and algorithms for vehicular networks (VANETs). Simulation tools typically use mobility traces to recreate the network topology based on the existing contacts between the mobile nodes. However, in terms of spatial and temporal granularity, the quality of the traces is an essential factor that directly affects the network topology. Therefore, the reliability of the results depends strongly on the accurate representation of the real network topology by the mobility model. We show that a real bus mobility trace has gaps, leading to results that do not represent reality. In this work, we use a solution to fill these gaps, leading to more refined features, which lead to more reliable simulation results. Besides, we also present the evaluation results that compare the communication metrics of the original and calibrated traces. The results reveal that the gaps induce network topologies that are different from reality, reducing the quality of the evaluation results. Although mobility data for vehicle networks are constructive, they must undergo a quality improvement process.*

**Resumo.** *Simulação é uma abordagem mais comumente adotada para avaliar protocolos e algoritmos para redes veiculares. Geralmente, as ferramentas de simulação utilizam traces de mobilidade para recriar a topologia de rede e avaliar o desempenho das soluções. No entanto, a qualidade dos traces em termos de granularidade espacial e temporal é um fator importante, que pode afetar diretamente na confiabilidade dos resultados. Neste trabalho, é apresentado a caracterização de dados em um trace que representa a mobilidade real de ônibus na cidade do Rio de Janeiro. Por apresentar lacunas e falhas de leitura, os dados passaram por um processo de preenchimento das lacunas, levando a traces mais refinados, e que geram resultados de simulação mais confiáveis. Além disso, todas as métricas computadas foram comparadas entre os dados originais e calibrados. Resultados revelam que as lacunas realmente induzem topologias de rede diferentes da realidade, diminuindo a qualidade dos resultados da avaliação.*

## 1. Introdução

Com o crescimento das grandes cidades, diversos problemas de mobilidade urbana têm se apresentado. Dentre eles, pode-se citar a grande concentração de pessoas no mesmo lugar e a falta de planejamento da exploração do solo, dificultando a movimentação diária das pessoas que as habitam essas regiões. Aliado a isso, as vias públicas também apresentam diariamente problemas estruturais, acidentes e até mesmo as mudanças climáticas podem dificultar a mobilidade dos veículos no dia a dia, gerando atrasos e muitos engarrafamentos [Cunha et al. 2016a, Al-Sultan et al. 2014].

O entendimento da evolução do tráfego de veículos desempenha um papel fundamental na criação de soluções para redes veiculares inteligentes. Neste tipo de rede, a comunicação entre os veículos depende da proximidade entre eles, e nesta direção entender como os veículos transitam pelas cidades se torna um ponto importante no projeto de soluções. Diferentes tipos de veículos possuem características diferentes de mobilidade, e estudá-la se torna uma questão importante a ser investigada [Celes et al. 2019].

O estudo da mobilidade dos veículos em uma região pode ser feito por meio do uso de base de dados conhecidos como *traces*. Esses trazem consigo informações relevantes que ajudam na compreensão e criação de cenários para simulação mais realistas. No entanto, esses *traces* podem conter falhas de leituras e baixa amostragem que podem levar a conclusões inconsistentes e, conseqüentemente, afetam o desempenho das soluções propostas [Celes et al. 2017].

Nesse sentido, neste trabalho será apresentado um estudo de uma base de dados de mobilidade de ônibus da cidade do Rio de Janeiro [Dias and Costa 2018], considerando seus aspectos espaciais e temporais. Para isso, esta base foi analisada e, com o uso de técnicas de calibragem, lacunas de tempo e espaço foram preenchidas com o objetivo de aumentar a granularidade da base de dados, o que a torna mais rica e com maior representatividade.

Neste estudo também foi feita uma comparação entre os dados reais e os dados calibrados avaliando métricas importantes no que diz respeito a conectividade de uma rede veicular como: tempo de contato, tempo entre contatos, a capacidade de transmissão, e números de encontros computados ao longo dos dias [Cunha et al. 2014]. O objetivo dessas métricas, é investigar o comportamento dos ônibus ao longo dos dias e durante o mês inteiro. Esta análise pode ajudar a compreender melhor as características do transporte coletivo, e como a conectividade desse tipo de transporte pode ser explorada para transmissão de dados além de computar a capacidade de transmissão possível de alcançar. Os resultados mostram que o uso de dados de mobilidade para redes de veículos é extremamente relevante, sendo possível sua exploração para a caracterização da conectividade da rede.

O restante do artigo está organizado da seguinte forma: na Seção 2 serão apresentados aos trabalhos relacionados; na Seção 3 será descrita a metodologia utilizada para a análise, os conjuntos de dados, os processos utilizados, e métricas avaliadas. Os resultados serão apresentados para as métricas de conectividade e de comunicação na Seção 4. Por fim, na Seção 5, será apresentado as observações finais e trabalhos futuros.

## 2. Trabalhos Relacionados

Os trabalhos relacionados estão divididos em três grupos. O primeiro grupo destaca a importância do entendimento da mobilidade dos ônibus para as redes veiculares, o segundo cobre os trabalhos que investigam os *traces* reais de mobilidade, detectando lacunas e utilizando métodos de calibragem para que apresentem mais confiabilidade em seus resultados para criação de aplicações, e o terceiro grupo que analisa e compara o resultado de diversas métricas de contato e comunicação em *traces* de mobilidade.

Como as redes veiculares são altamente impactadas pela mobilidade dos nós, um dos principais desafios é identificar estratégias que forneçam a comunicação entre esses nós. Nesse sentido, alguns estudos propuseram arquiteturas de rede veicular nas quais os ônibus desempenham um papel fundamental para garantir a disponibilidade de comunicação e sensoriamento [Wong et al. 2004, Luo et al. 2010, Jiang and Du 2015].

A motivação para o uso de ônibus como nós especiais refere-se ao fato de que eles possuem um conjunto de recursos de mobilidade que diferem dos outros padrões diários de mobilidade de carros particulares ou táxis. Para mostrar isso, os autores em [Celes et al. 2019] mostram alguns recursos da mobilidade de ônibus em um dia de trabalho típico em Dublin, na Irlanda. Em termos de dimensão espacial, a mobilidade de ônibus é restrita a rotas predefinidas realizadas ao longo do dia. E é possível observar que as linhas de ônibus cobrem todas as regiões da cidade. Além disso, em geral, as linhas de ônibus tendem a conectar as regiões mais periféricas à região central, seguindo caminhos previsíveis. Também mostram que para cada linha de ônibus existe um tempo médio de viagem esperado, e que essas viagens começam em horários programados, outra característica interessante da mobilidade de ônibus.

Com a dificuldade de se obter *traces* reais, a simulação acaba saindo como a maneira menos custosa para avaliar protocolos e algoritmos para redes veiculares. Para as ferramentas de simulação criarem topologias de rede factíveis, elas usam como base *traces* de mobilidade reais com base nos contatos existentes entre os nós móveis. No entanto, a qualidade desses *traces*, em termos de granularidade espacial e temporal, pode ser insuficiente devido a vários fatores, desde custos do armazenamento até falha nos dispositivos que causam baixa taxa de amostragem, e acarretam a perda de confiabilidade nos resultados. O que o trabalho [Celes et al. 2017] propõe é uma metodologia para deixar esses *traces* mais próximos da realidade, aumentando a granularidade espaço temporal, preenchendo possíveis lacunas. Utilizando o *trace* real como base, é apresentada uma forma de aumentar a quantidade de pontos, mantendo o padrão e sendo condizente com a realidade, os resultados mostram como a calibração exibe números melhores nas métricas de contato do que os *traces* reais.

Alguns trabalhos focaram em investigar a mobilidade em termos de contato e comunicação entre veículos, como é o caso em [Cunha et al. 2016b], os autores analisaram *traces* reais de três cidades diferentes, Roma, São Francisco e Xangai para poder comparar com sua versão calibrada. O objetivo era mostrar que as lacunas existentes afetam as métricas principais, como a duração do contato, o tempo de entre contatos e a capacidade da rede. Quanto aos contatos entre os veículos, a duração do contato mede por quanto tempo dois veículos conseguem se comunicar. O tempo de entre contatos avalia o tempo decorrido entre dois encontros sucessivos de dois veículos. E, a capacidade da rede demonstra como a capacidade da rede é afetada pelas lacunas existentes nos *traces* originais.

Devido a necessidade de se estudar os ônibus, e com a disponibilidade desse *trace* do Rio disponível para análise, decidimos calibrá-lo e disponibilizar para a comunidade científica essa análise. Diversos trabalhos na literatura apresentam análises temporais de *traces* de veículos. Entretanto, poucos exploram aspectos espaciais destas bases de dados. Assim, neste trabalho pretende-se aproveitar a localização espacial dos encontros entre os veículos de forma a extrair padrões e características que venham ajudar no projeto de novos protocolos e serviços para as redes veiculares.

### 3. Metodologia

Esta seção apresenta a metodologia utilizada para calibrar e analisar o *trace* de ônibus do Rio de Janeiro de uma maneira espaço temporal, para mostrar a relação de contato e a comunicação entre os veículos ao longo do dia. Nesta análise, foi usado o *trace* do Rio de Janeiro, Brasil, modelado como um grafo temporal.

Após uma limpeza de ruídos dos dados, o *trace* é passado pelo processo de calibração, onde tem como objetivo preencher espaços faltantes dada a baixa amostragem. E então foi feita uma análise comparativa dos resultados das métricas com o *trace* original e o gerado pela calibração. O diagrama da Figura 1 apresenta as etapas do processo executado e nas seções seguintes essas etapas serão descritas com mais detalhes.

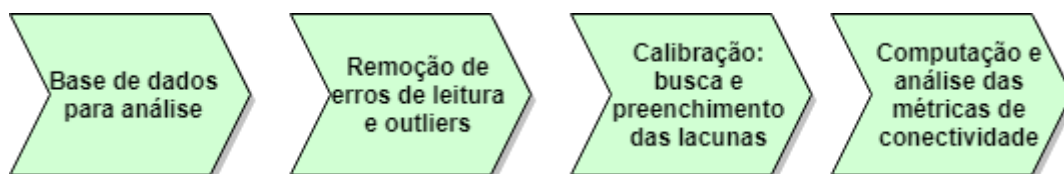


Figura 1. Etapas da análise efetuada na base de dados.

#### 3.1. Base de dados

O *trace* original contém registros da mobilidade dos ônibus. Os dados foram coletados da plataforma de dados abertos da prefeitura do Rio de Janeiro. São coordenadas GPS de cerca de 12.456 ônibus, coletados ao longo de 30 dias. As coordenadas foram coletadas a cada minuto de 1<sup>o</sup> a 30 de Outubro de 2014 [Dias and Costa 2018]. A Tabela 1 apresenta os dados das linhas de ônibus para o período analisado.

Dados do mês de Outubro/2014	
Quantidade de ônibus	12.456
Quantidade de viagens	1.832.461
Quilômetros rodados	72 milhões
Passageiros	117 milhões

Tabela 1. Dados das linhas de ônibus do mês de Outubro/2014

Cada ônibus é equipado com um aparelho de GPS, e envia informações sobre sua localidade (identificador, *timestamp*, latitude e longitude) periodicamente para um servidor central. Os *traces* reais de mobilidade de veículos apresentam problemas quanto a qualidade na granularidade espacial e temporal. Um dos motivos para isso é a baixa taxa

de amostragem devido às grandes distâncias, tempo entre dois veículos e até mesmo a qualidade dos sensores.

Quando se trata de Redes Veiculares, as análises podem ter resultados que não condizem com a realidade, causados pelos *traces* de mobilidade com baixa fidelidade que geram impacto na topologia da rede. Para preencher essas lacunas espaciais e temporais, os autores em [Celes et al. 2017] adotaram uma solução baseada em algoritmos de agrupamento. Eles usaram esses algoritmos para descobrir e preencher as lacunas que existiam entre os veículos, e assim obter uma versão com uma maior granularidade.

A diferença entre os *traces* calibrados e originais é que no original, os dados foram coletados a cada minuto, enquanto no calibrado existem mais coordenadas de GPS para preencher as lacunas. O trabalho apresentado em [Cunha et al. 2016b] comparou *traces* original e o calibrado em termos de contatos entre veículos. Eles analisaram métricas importantes de rede, como duração de contato, tempo de inter-contato e capacidade de rede, em Roma, São Francisco e Xangai. Os resultados mostraram que as lacunas afetam essas métricas.

### 3.2. Pré-processamento

Para que o processo de calibragem seja efetivo, é necessário que se tenha um conjunto de pontos do *trace* que sejam condizentes com a realidade dos veículos, pois são esses pontos que serão adicionados no conjunto de dados e preencherão os *gaps* espaço-temporais. E um conjunto bruto de dados recém coletados do servidor contém ruídos que podem atrapalhar essa amostragem.

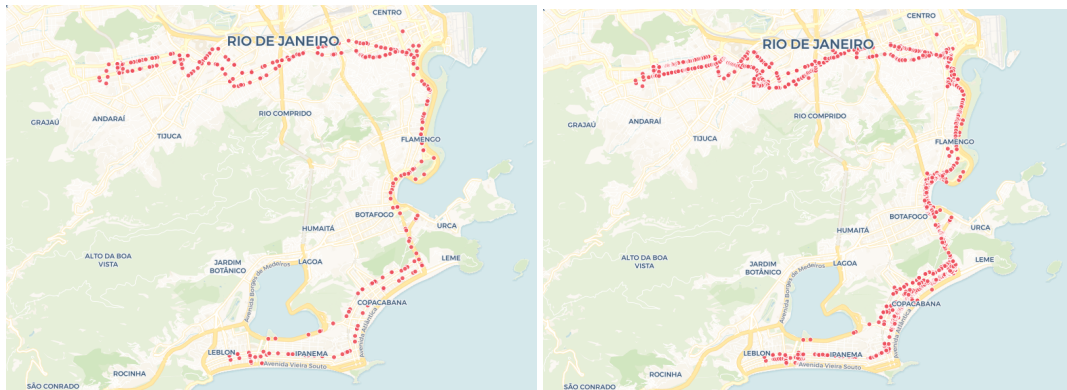
A primeira etapa foi de remoção de pontos que não estivessem dentro da cidade do Rio de Janeiro, após isso para otimizar o processamento, foram escolhidos 5 veículos aleatoriamente de cada linha para que o tempo de geração dos resultados fosse diminuído e ainda assim a base mantivesse sua cobertura espacial.

Ainda foram removidos pontos que estavam fora das ruas e avenidas, como pontos em garagens dos ônibus, como o intuito da pesquisa são os encontros feitos nas rotas cotidianas dos veículos esses pontos significavam *outliers*. Após essa limpeza passamos para a fase de calibração.

### 3.3. Calibragem

A integridade da topologia da rede é um fator chave para um bom desempenho nas VANETs. Os contatos entre veículos que ocorreram na realidade, mas não foram computados devido a lacunas nas trajetórias, afetam a avaliação das métricas de algoritmos e protocolos, uma vez que a transmissão de dados depende desses contatos. Uma trajetória é definida como uma sequência de pontos espaço-temporais  $T = (p_1, \dots, p_n)$ , onde  $p_i = (x, y, t)$ , para  $i = 1 \dots n$ , e  $x, y$  são coordenadas espaciais,  $t$  é um *timestamp*, e  $p_i.t < p_{i+1}.t$ . E a definição formal para as lacunas, ou intervalos sem amostragem dos dados é dada uma trajetória  $T$  de um veículo e um limite  $\theta$ , uma lacuna existe quando a distância  $\Delta s$  entre dois pontos consecutivos de  $T$  é maior que  $\theta$ , isto é,  $\Delta s = d(p_i, p_{i+1}) > \theta$ . A Figura 2 apresenta dois cenários comparativos de trajetória, o antes de depois do processo de calibragem.

Para preencher esses espaços será utilizado o método da calibração proposto por [Celes et al. 2017], que é dividido em duas etapas. A primeira etapa extrai um sistema



(a) Antes da calibração

(b) Depois da calibração

**Figura 2. Exemplo de rota de um ônibus com espaçamento entre os pontos (a) e com espaços preenchidos depois da calibração (b).**

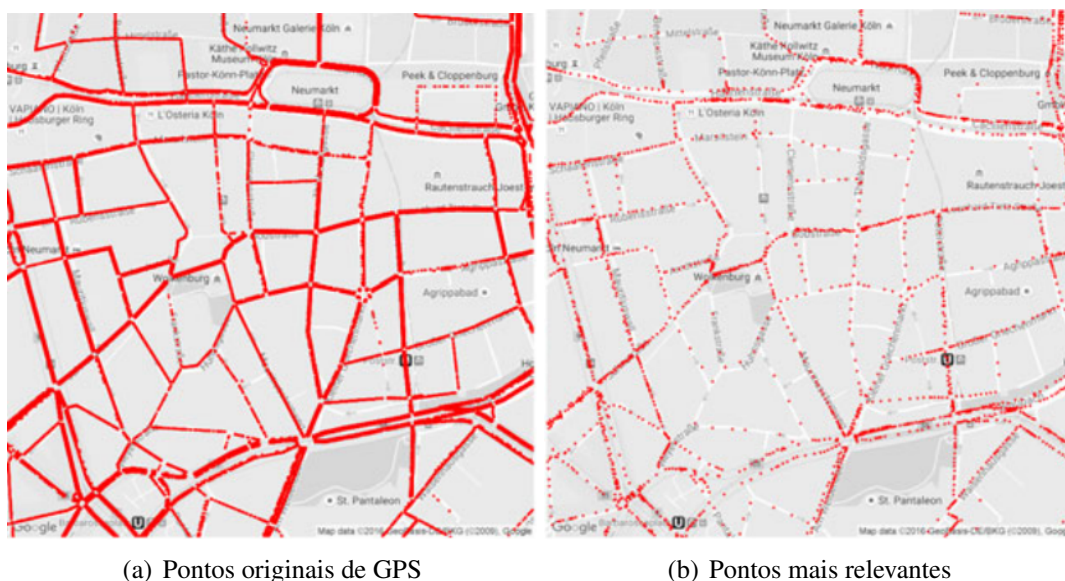
de referência do conjunto de dados de trajetórias dos veículos. O segundo estágio aplica um método de calibração, usando um subconjunto de pontos relevantes do previamente construído sistema de referência.

O sistema de referência consiste em um conjunto de pontos resultantes de um processo de agrupamento que usa trajetórias históricas. Cada ponto, chamado centroide, representa um grupo de pontos de GPS próximos um do outro, que representam as localidades dos veículos no *trace*. Assumindo que essas localizações representam trajetórias reais, é razoável supor que cada centroide é um local potencial para um novo ponto em uma trajetória. Em outras palavras, é muito provável que um centroide represente um ponto correto em uma rua, ou avenida pela qual os veículos trafegam. Neste trabalho também foi adotado o método de agrupamento *k-means* para particionar os dados em *k* grupos, de acordo com a densidade dos pontos GPS, e então, obtemos o ponto centroide de cada grupo para formar o sistema de referência.

Na segunda etapa, realizamos a calibração, quando existe uma lacuna na trajetória *T*, obtemos o sistema de referência daquela determinada região e, em seguida, selecionamos os pontos centroides entre os pontos inicial e final da lacuna. O método de calibração recebe os seguintes parâmetros como entrada: *T*, um conjunto de *n* pontos consecutivos com informações espaço-temporais; o sistema de referência obtido anteriormente; a distância mínima limite para se considerar a existência de uma lacuna espacial; e, por último, o tempo mínimo limite para considerar um intervalo temporal entre duas coordenadas consecutivas. Como resultado, é obtida uma nova trajetória *T'* com os pontos originais de *T* e um conjunto de pontos calibrados adicionados para preencher as lacunas existentes em *T*.

Nessa análise foi considerado que existia uma lacuna, se a distância entre dois pontos de uma trajetória, era maior que 200 metros, e intervalo de tempo entre esses pontos era menor que 70 segundos. Considerando a velocidade média dos ônibus como 13,7 km/h ou 3,5 m/s [Tatto 2015], e com amostragem média de 65 segundos, os veículos se locomovem aproximadamente 230 metros a cada envio de localização. O algoritmo passa a identificar as lacunas quando a distância entre um ponto e outro for maior que 200m e o tempo entre eles menor que 70 segundos, sendo assim adiciona um ponto entre eles, se

existir algum nessa localização no sistema de referência.



**Figura 3. Pontos originais de GPS vs pontos mais relevantes [Celes et al. 2017].**

### 3.4. Grafos Temporais

A base de dados foi modelada da seguinte forma: um grafo temporal  $G(t) = (V, E)$ . Onde  $G$  representa um grafo não direcionado em um tempo  $t$ , e  $V$  um conjunto de veículos  $V_i$ , e  $E$  o conjunto de arestas  $E_{ij}$ . A aresta  $E_{ij}$  só existe durante o tempo  $t$ , se  $V_i$  e  $V_j$  são dois veículos diferentes [Domingos Da Cunha et al. 2014].

Com o intuito de extrair as métricas de conectividade e comunicação, analisamos cada dia separadamente, agrupando as arestas a cada 15 minutos e criando grafos separados. Foi escolhido 15 minutos para poder ter uma amostra significativa de dados. Um valor menor, poderia gerar uma amostra muito pequena de dados com um grafo muito esparsos, sem conectividade. E um maior valor, a amostra pode ser muito grande, com densidade próxima de um grafo temporal.

Já para as métricas dos tempos de contato e entre contatos, foi escolhido um dia aleatoriamente, e dele foram gerados grafos temporais a cada segundo, seguindo a mesma regra para criação das arestas. E deles foram extraídos o exato momento em que dois ônibus se encontraram, ou seja, o *timestamp* em que estavam a menos de 100 metros distância um do outro.

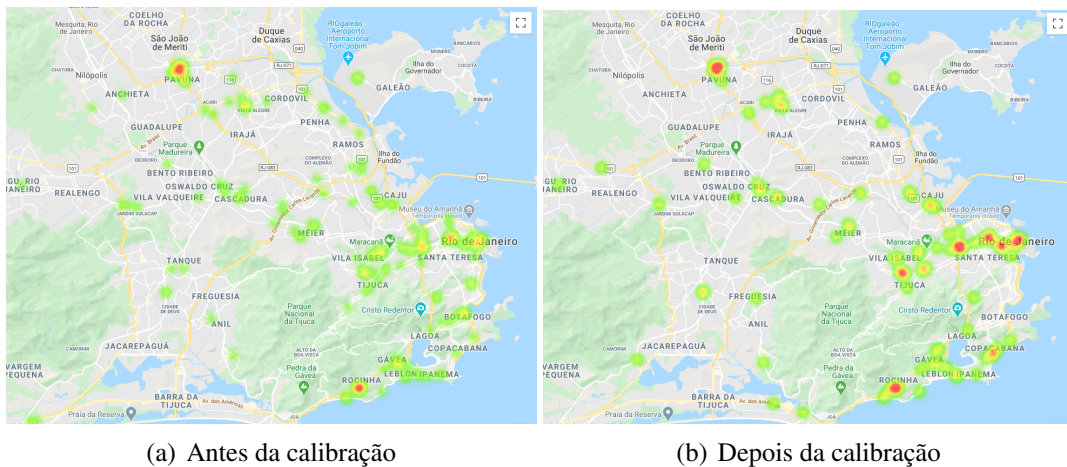
Irá sempre existir uma aresta entre dois veículos caso eles consigam estabelecer uma comunicação entre si, levando em conta que ambos os valores de  $t$  compartilham a mesma hora e minuto. Adotando o padrão de protocolo IEEE 802.11p, a comunicação ocorre se os dois carros estiverem no máximo a 100 metros um do outro.

### 3.5. Métricas

As métricas têm como base a análise espacial temporal de contatos entre os veículos: quantidade encontros ao longo do dia, duração dos encontros, intervalo entre os encontros e capacidade da rede.

## Encontros ao longo do dia:

A primeira métrica avaliada é a quantidade de encontros entre os ônibus ao longo do dia. Ou seja, quantas vezes dois ônibus diferentes conseguiram se comunicar, em um dia específico. Quanto mais encontros acontecerem entre os veículos, melhor é para a rede. Isso significa que as informações compartilhadas entre os ônibus na rede fluem facilmente, e em situações em que todos os veículos precisam ser informados, a informação chega aproximadamente ao mesmo tempo para todos. Pode-se observar na Figura 4 um comparativo dos encontros antes e depois do processo de calibração.



**Figura 4. Mapa de calor de encontros repetidos na mesma localização.**

## Duração do encontro:

A duração do encontro mede quanto tempo dois veículos estão em contato, e consequentemente, aptos a trocar informações. Esta métrica é importante pois revela como aplicações podem explorar o contato entre veículos para oferecer serviços. Quanto maior o tempo de contato, mais dados podem ser transmitidos. Um contato existe quando a distância Euclidiana  $d(v,u)$ , entre dois veículos  $v$  e  $u$ , é menor em relação ao *range* de comunicação deles. Neste trabalho, o alcance de transmissão dos veículos é de aproximadamente 100m.

## Tempo entre contatos:

O tempo entre contatos também é um fator importante para os protocolos de encaminhamento de pacotes em termos de atraso de ponta-a-ponta. Esta métrica refere-se ao tempo decorrido entre dois encontros sucessivos de dois veículos. Ou seja, é a janela de tempo que começa no final de um contato entre dois veículos e termina no início do próximo contato entre os mesmos veículos. Quanto menor o tempo entre contatos, mais eficaz é a aplicação na transmissão dos dados, uma vez que mais oportunidades existirão para transferi-los.



## Capacidade da rede:

O objetivo dessa métrica é evidenciar como a capacidade da rede é afetada pelas lacunas existentes no *trace* original. Partindo disso, vão ser determinados quantos *frames* podem ser transmitidos durante as horas do dia. Para isso, é feito um cálculo para encontrar qual é a quantidade de *frames* transmitidos através do contato. Considerando um *frame* de 4.095 Bytes, e a menor taxa de transmissão do protocolo 802.11p sendo 3Mbps, o total de dados necessários para transmitir um *frame* é:  $(4.095 \times 8) + (34 \times 8) + (5 \times 8) + 6 = 33.078$  bits, que podem ser representados em  $\frac{33.078}{24} = 1.379$  símbolos. Dado que cada símbolo requer  $8\mu s$  e mais  $40\mu s$  para o *preamble*, o tempo total de transmissão de um *frame* é de  $11.072\mu s$ , ou aproximadamente  $0.011s$ . Isto quer dizer que é possível transmitir  $\frac{T}{0.011}$  *frames* durante um contato de tempo T.

## 4. Resultados

Nesta seção, são apresentados os resultados, eles estão agrupados de acordo com as métricas definidas previamente, para as métricas de conectividade e de comunicação. A base contém 30 dias, mas alguns desses dias não possuem dados suficientes para se analisar as 24 horas, foi escolhida então uma semana que possuíam mais dias com dados completos, e esta semana foi dos dias 13 a 19 de outubro de 2014.

### 4.1. Quantidade de encontros ao longo do dia:

**Tabela 2. Porcentagem de dados vs quantidade de encontros real.**

Dia	50%	90%
Domingo	3	7
Segunda	11	25
Terça	3	6
Quarta	14	32
Quinta	15	35
Sexta	4	11
Sábado	5	10

**Tabela 3. Porcentagem de dados vs quantidade de encontros calibrado.**

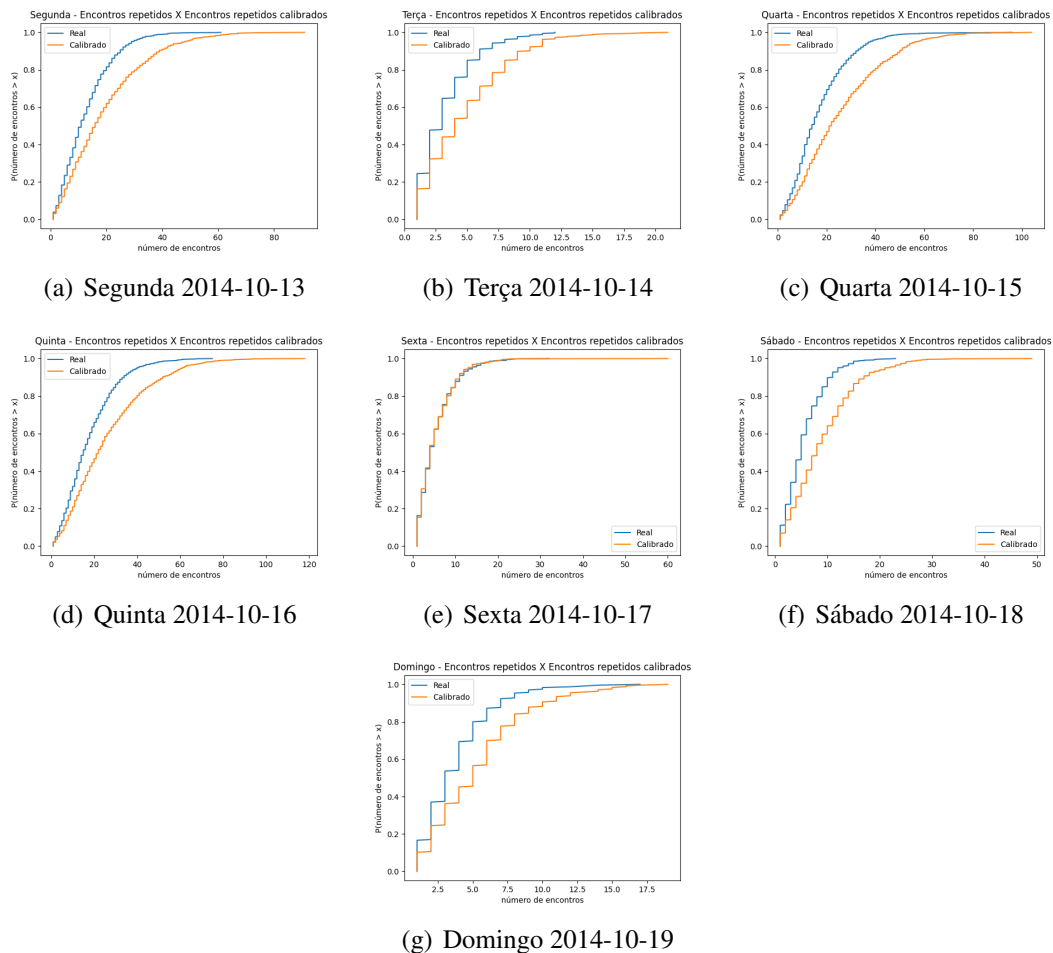
Dia	50%	90%
Domingo	5	9
Segunda	16	39
Terça	4	9
Quarta	22	50
Quinta	22	52
Sexta	4	11
Sábado	8	17

Durante esta análise dos resultados da métrica de encontros, percebe-se que para cada um dos dias, não houve grandes mudanças, isto é, todas as Segundas obtiveram um mesmo padrão, os dias do fim de semana também e assim por diante. Por esse motivo, o foco se manteve em apenas uma semana do conjunto de dados, ao invés de todo o mês. A Figura 5 apresenta o gráfico da função de distribuição cumulativa (CDF) para cada um dos dias da semana, gerado a partir dos dados reais e dos calibrados para a comparação.

É perceptível também a evolução na quantidade de encontros ao longo do dia, a análise apresenta a quantidade com 50% e em seguida 90% dos dados da base processados, e vemos que ao longo do processamento de cada dia o número cresce, demonstrando a maior conectividade dos veículos, representado nas tabelas 2 e 3.

Comparado os encontros do *trace* real (linha azul) e *trace* calibrado (linha laranja), pode-se observar que para todos os dias da semana a quantidade de encontros do calibrado

é maior. Isso significa que o *trace* calibrado preenche espaços do *trace* real fazendo com que haja mais encontros como o esperado, ou seja, encontros não computados pelos dados reais são possíveis de serem capturados após o processo de preenchimento. Além disso, nota-se que a base calibrada apresenta um total de 50% a mais de encontros do que comparando com a base real.



**Figura 5. Total de encontros por dia da semana.**

#### 4.2. Duração de contato:

A Figura 6 apresenta dois histogramas para a duração de contato do *trace* de ônibus do Rio de Janeiro. Um histograma com a comparação do *trace* real *versus* calibrado em um dia útil (quarta-feira), e no outro em um final de semana. Pode-se perceber que o *trace* calibrado apresenta bem mais contatos entre os veículos do que o *trace* real. O histograma também mostra que os contatos no *trace* apresenta duração muito maior ao comparado com o *trace* real. Esse comportamento se deve por contatos ausentes no *trace* real acarretam uma menor duração de contato.

Comparando o dia da semana com o final de semana, pode-se observar também que devido ao fluxo maior de tráfego, maior quantidade de veículos em circulação e por consequência uma maior probabilidade para trânsitos lento e engarrafamento os dias da

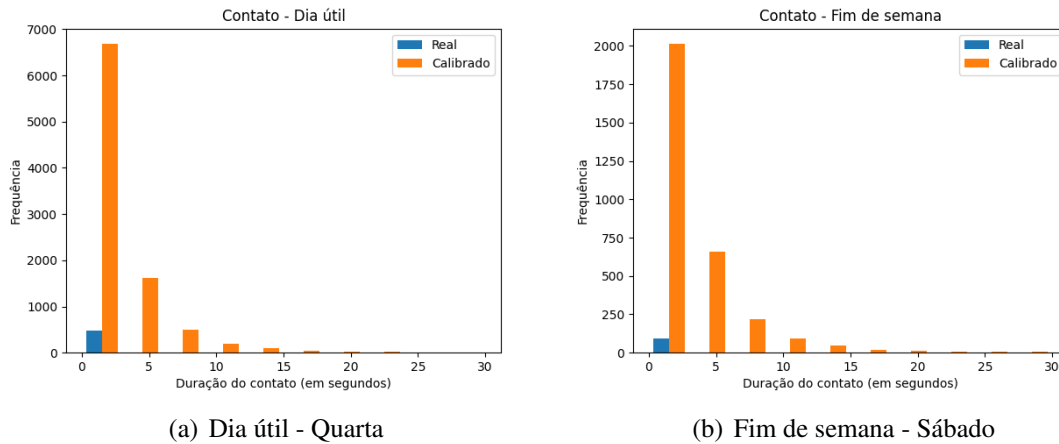


Figura 6. Duração dos encontros dos *traces* Real vs Calibrado.

semana apresentam duração maior de contato e uma frequência maior de contatos com mais duração (Figura 6-(a)).

### 4.3. Tempo entre contatos

O tempo entre contatos é uma métrica importante que pode retratar bem a conectividade da rede e guiar o desenvolvimento de protocolos capazes de lidar com essas desconexões. A Figura 7 ilustra duas Funções de Distribuição Acumulada Complementar (CCDF) do tempo entre contatos do *trace* real e calibrado, sendo na primeira apresentando o gráfico de um dia útil, e na segunda de um dia do fim de semana. Pode-se observar que o *trace* calibrado possui uma redução expressiva deste tempo comparado ao *trace* real. Dados faltantes provocam desconexões e o processo de calibração consegue contornar esse problema.

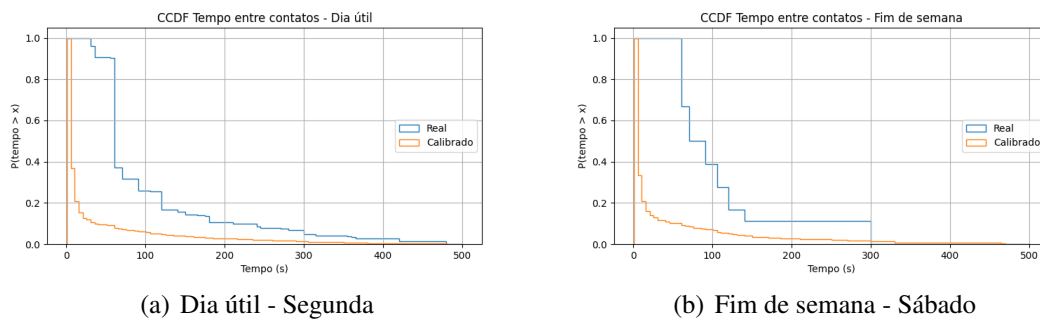


Figura 7. Tempo entre contatos dos *traces* Real vs Calibrado.

### 4.4. Capacidade da rede

A Figura 8 apresenta a capacidade da rede, em termos de quantidade de *frames* que podem ser transmitidos, para cada hora do dia. Com base nessa figura, é possível observar que o *trace* calibrado apresenta maior capacidade que o real, em todas as horas do dia. Durante alguns períodos, o aumento da capacidade do *trace* calibrado chega a ser até 46 vezes maior às 10h da manhã nos dias úteis. Os resultados revelam que o *trace* calibrado aponta para uma maior capacidade na rede quando o quesito a ser avaliado é a quantidade de *frames* que podem ser transmitidos usando o protocolo IEEE 802.11p.

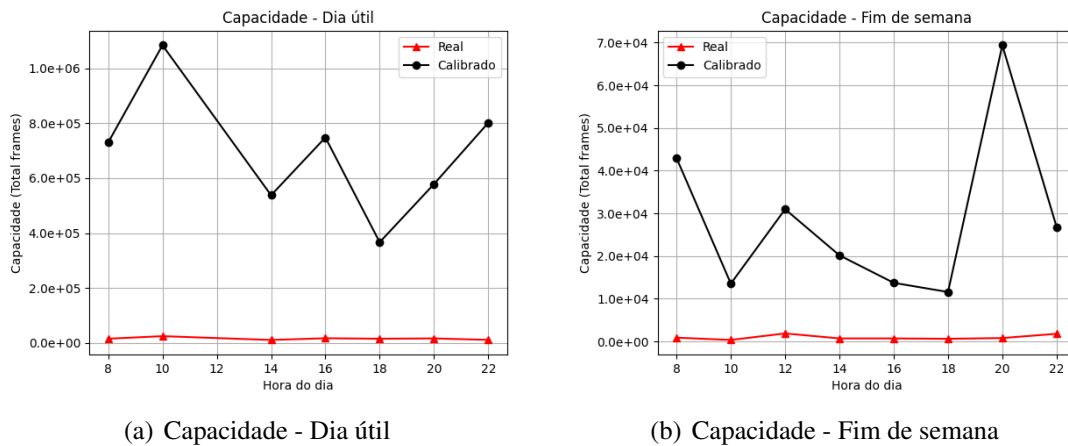


Figura 8. Capacidade de transmissão por hora do dia.

## 5. Conclusão e Trabalhos Futuros

Neste trabalho, foi aplicado um método de calibração em um *trace* de ônibus e feita uma análise espaço temporal da conectividade e de comunicação dos ônibus em toda a cidade do Rio de Janeiro-BRA. Para o desenvolvimento utilizou-se grafos temporais, *traces* de ônibus, como base de dados, e quatro métricas: encontros repetidos ao longo do dia, duração dos contatos, tempo entre os contatos e a capacidade da rede. Sendo que cada dia da semana foi analisado separadamente para uma possível comparação.

Os resultados para cada dia repetido da semana foram bem próximos um do outro, então escolhemos focar em apenas uma semana com mais dados, foi possível a identificação de padrões em dias úteis e fins de semana. O estudo comprova que os *traces* calibrados representam melhor a realidade e, e que o processo de calibragem deve ser aplicado nos *traces* que apresentam baixa amostragem para serem utilizados nas simulações.

Como trabalhos futuros, pretende-se testar outras formas de calibração, com outros métodos de agrupamento dos pontos, analisar outros *traces* de diferentes tipos de topologias. Além de analisar outras métricas específicas dos ônibus como número de encontros de diferentes linhas, e encontro por horário que podem auxiliar os sistemas de transporte a criarem aplicações.

## Referências

- Al-Sultan, S., Al-Doori, M. M., Al-Bayatti, A. H., and Zedan, H. (2014). A comprehensive survey on vehicular ad hoc network. *Journal of network and computer applications*, 37:380–392.
- Celes, C., Boukerche, A., and Loureiro, A. A. F. (2019). Towards understanding of bus mobility for intelligent vehicular networks using real-world data. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6.
- Celes, C., Silva, F. A., Boukerche, A., Andrade, R. M. d. C., and Loureiro, A. A. F. (2017). Improving vanet simulation with calibrated vehicular mobility traces. *IEEE Transactions on Mobile Computing*, 16(12):3376–3389.

- Cunha, F., Villas, L., Boukerche, A., Maia, G., Viana, A., Mini, R. A., and Loureiro, A. A. (2016a). Data communication in vanets: Protocols, applications and challenges. *Ad Hoc Networks*, 44:90–103.
- Cunha, F. D., Silva, F. A., Celes, C., Maia, G., Ruiz, L. B., Andrade, R. M. C., Mini, R. A. F., Boukerche, A., and Loureiro, A. A. F. (2016b). Communication analysis of real vehicular calibrated traces. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6.
- Cunha, F. D., Vianna, A. C., Mini, R. A., and Loureiro, A. A. (2014). Is it possible to find social properties in vehicular networks? In *2014 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–6. IEEE.
- Dias, D. and Costa, L. H. M. K. (2018). CRAWDAD dataset coppe-ufrj/riobuses (v. 2018-03-19). Downloaded from <https://crawdad.org/coppe-ufrj/RioBuses/20180319>.
- Domingos Da Cunha, F., Viana, A. C., Rodrigues, T., Mini, R. A. F., and Loureiro, A. A. F. (2014). Extração de Propriedades Sociais em Redes Veiculares. In *XXXII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Florianópolis, Brazil. SBC.
- Jiang, X. and Du, D. H. (2015). Bus-vanet: A bus vehicular network integrated with traffic infrastructure. *IEEE Intelligent Transportation Systems Magazine*, 7(2):47–57.
- Luo, J., Gu, X., Zhao, T., and Yan, W. (2010). Mi-vanet: A new mobile infrastructure based vanet architecture for urban environment. In *2010 IEEE 72nd Vehicular Technology Conference - Fall*, pages 1–5.
- Tatto, J. A. (2015). Faixas exclusivas aumentaram a velocidade média dos ônibus - usp. <https://www5.usp.br/96824/faixas-exclusivas-aumentaram-a-velocidade-media-dos-onibus/>. Acessado em 05/03/2020.
- Wong, K.-J., Lee, B.-S., Seet, B.-C., Liu, G., and Zhu, L. (2004). Busnet: Model and usage of regular traffic patterns in mobile ad hoc networks for inter-vehicular communications.