

Análise da Qualidade de Serviço do Sistema Público de Ônibus Urbanos através de Monitoramento por GPS

Tiago R. Borzino¹, Fernando D. M. Silva¹, Aline C. Viana²,
Luís Henrique M. K. Costa¹

¹COPPE – Universidade Federal do Rio de Janeiro (UFRJ) – Brasil

²INRIA, França

{borzino, fernandodias, luish}@gta.ufrj.br, aline.viana@inria.fr

Resumo. A introdução de dispositivos IoT equipados com GPS nos ônibus urbanos possibilita a coleta de dados para monitorar a frota e avaliar a qualidade do serviço prestado. Este trabalho analisa dados de GPS dos ônibus do Rio de Janeiro, incluindo trajetórias de viagens, linha e identificador do veículo. A partir dessas informações, são estimadas métricas de desempenho de cada linha. Em particular, são investigadas as métricas de intervalo de passagem, número de agrupamentos de ônibus, tempo para percorrer uma fração da trajetória e entropia como indicadores de qualidade de serviço. Os resultados demonstram a regularidade ou imprevisibilidade das diferentes linhas analisadas. Por fim, investiga-se a correlação entre as diferentes métricas a fim de mostrar se a piora ou melhora em um critério de desempenho reflete em algum dos outros.

1. Introdução

A conexão de veículos através de dispositivos de Internet das coisas (*Internet of Things* – IoT), aliada a um servidor centralizado que agrega dados desses veículos, permite uma visão abrangente de todos os veículos conectados, assim como de diversos aspectos relacionados à mobilidade no ambiente urbano. Em uma cidade conectada, o transporte público é um dos meios de transporte que têm suas informações coletadas, permitindo o monitoramento da frota existente. Usando esses dados coletados pelos veículos, é possível analisar a qualidade do serviço oferecido por diferentes linhas e empresas de ônibus urbanos, auxiliando em decisões de planejamentos urbanos e de novos trajetos de ônibus. Os ônibus urbanos do Rio de Janeiro são equipados com sensores capazes de captar suas informações de GPS, data e hora, identificador de ônibus e a trajetória que está sendo percorrida. Essas informações são transmitidas para um servidor usando rede celular e disponibilizadas pela prefeitura carioca pelo site DataRio [Prefeitura do Rio de Janeiro, 2025].

Em um contexto em que se deseja monitorar a frota de ônibus urbanos, é possível estimar diferentes métricas que permitem avaliar a qualidade do serviço prestado. Algumas das métricas permitem analisar características de cada linha. O intervalo de passagem mede o tempo necessário para dois ônibus de uma mesma linha passarem pelo mesmo local. O agrupamento de ônibus conta o número de ônibus muito próximos de outros que realizam o mesmo trajeto. O tempo para percorrer um trecho da linha computa o tempo que um mesmo ônibus levou para transitar entre dois locais do percurso. Por fim, a entropia mede a regularidade com que as diferentes viagens de ônibus são feitas.

Além disso, o aparelho de localização dos ônibus, como qualquer aparelho de GPS, está sujeito a imprecisões nas coordenadas transmitidas. Dessa forma, algumas das

posições enviadas podem estar localizadas no sentido contrário da rua ou ainda dentro de prédios ou rios. Um algoritmo de correção de trajetória é necessário para mitigar esse problema e eliminar dados incongruentes.

Este trabalho propõe uma análise de diferentes métricas para qualidade de serviço dos ônibus. Essa análise consiste na correção na trajetória dos ônibus, baseando-se na linha por eles informada, e em calcular diferentes métricas de qualidade de serviço e realizar uma comparação entre elas. Primeiramente realiza-se a correção de cada viagem de ônibus analisada e, após isso, calcula-se as diferentes métricas de qualidade de serviço. Este trabalho utiliza as métricas de intervalo de passagem, agrupamento de ônibus, tempo para percorrer uma fração da trajetória e entropia para avaliar a qualidade do serviço dos ônibus.

Este trabalho está assim organizado: A Seção 2 discute trabalhos relacionados. A Seção 3 apresenta o pré-processamento aplicado sobre os dados. A Seção 4 descreve as métricas utilizadas para a avaliação da qualidade. A Seção 5 descreve a abordagem utilizada para calcular as métricas. A Seção 6 apresenta os resultados e observações utilizando as métricas de qualidade do serviço. A Seção 7 conclui o artigo.

2. Trabalhos Relacionados

A mobilidade urbana é um tópico relevante na maioria das cidades atualmente, com estudos abordando diferentes temas. O trabalho de Li *et al.* [Li et al., 2022] estudou como o tráfego varia ao longo do dia em uma estrada dos Países Baixos. Um estudo feito sobre o sistema de ônibus de Los Angeles [Nguyen et al., 2023] analisou diferentes aspectos referentes à qualidade do serviço, como agrupamento de ônibus e regularidade no horário em que um transporte coletivo chega em alguma de suas paradas.

Outros trabalhos da literatura também analisam a ocorrência de agrupamentos de ônibus como indício da qualidade de serviço. Du *et al.* [Du e Dublanche, 2018] pesquisaram a influência que o dia da semana tem na ocorrência de agrupamentos. Já Xuhe [He, 2015] e Wang *et al.* [Wang et al., 2021] analisaram o problema sob a óptica de propor novas estratégias para lidar com agrupamentos, uma vez que os ônibus saem de seu ponto inicial com uma diferença de tempo entre dois consecutivos e, caso dois ou mais desses coletivos estejam muito próximos existe uma indicação de que ocorreu alguma situação não planejada que afetou diretamente o bom funcionamento daquela linha. Por fim, Zhou *et al.* [Zhou et al., 2022] verificou como disponibilizar informações em tempo real da lotação dos coletivos impactaria na escolha dos passageiros por um veículo e seu efeito no agrupamento.

Para estimar a incerteza na movimentação de pessoas, diferentes trabalhos utilizaram a entropia. O trabalho de Song *et al.* [Song et al., 2010] utiliza a entropia como forma de estimar a quantidade de incerteza que existe na rotina de locomoção de um indivíduo. O uso de informações de contexto e a remoção de próximos lugares impossíveis de se alcançar em uma amostragem foram estudadas para diminuir as incertezas por Teixeira *et al.*, Cuttone *et al.* e Huang *et al.* [Teixeira et al., 2019, Cuttone et al., 2018, Huang et al., 2024]. O estudo feito por Teixeira *et al.* [Teixeira et al., 2021] separou os locais visitados em rotina (os que são sempre visitados) e novidade (visitados pela primeira vez), verificando o impacto de cada uma na entropia. Ikanovic *et al.* [Ikanovic e Mollgaard, 2017] variaram as resoluções espacial e temporal do cálculo

da entropia e verificaram seu impacto na incerteza da posição. A entropia também foi utilizada em outros contextos para prever incertezas, como prever a demanda de táxi em uma cidade chinesa [Zhao et al., 2016] e prever o tráfego celular de internet [Huang et al., 2024].

3. Monitoramento dos Ônibus e Processamento dos Dados de Trajetórias

Nesta seção, são descritos os conjuntos de dados utilizados neste trabalho, coletados a partir do sistema de monitoramento por GPS da prefeitura, e o pré-processamento aplicado sobre eles.

3.1. Conjuntos de Dados

Os dados sobre os ônibus urbanos do Rio de Janeiro são extraídos de dois diferentes conjuntos de dados. O primeiro conjunto é composto por entradas de dados GPS brutos, já o segundo contém informações sobre cada uma das viagens realizadas.

O primeiro conjunto contém dados de GPS dos ônibus de 2021 até atualmente. A Tabela 1 mostra um exemplo das entradas desse conjunto. Cada registro contém informações sobre a posição do ônibus (latitude e longitude), hora e dia, rota e identificador do veículo. Neste trabalho utiliza-se o período de dezembro de 2022 a janeiro de 2023. O segundo conjunto, exemplificado na Tabela 2, consiste em vários registros de viagens completas das várias linhas que operam na cidade do Rio de Janeiro. Esse conjunto contém informações sobre a hora de início e de fim da viagem, rota percorrida, sentido e identificador do veículo que realizou a viagem.

Tabela 1: Extrato do conjunto de dados de GPS.

Estampa de tempo	ID do Veículo	Rota	Latitude	Longitude
2022-12-15 13:59:39	B32502	779	-22.85807	-43.38753
2022-12-15 13:59:40	B32602	779	-22.83073	-43.39779
2022-12-15 13:59:43	B32813	779	-22.85642	-43.38489
2022-12-15 13:59:43	B32822	779	-22.87675	-43.33789
2022-12-15 13:59:43	B32523	779	-22.80759	-43.36720

Tabela 2: Extrato do conjunto de dados de viagens.

ID Veículo	Rota	Sentido	Horário de partida	Horário de chegada
C50113	805	I	2022-12-10 10:31:29	2022-12-10 10:59:29
B58098	497	I	2022-12-10 10:31:30	2022-12-10 11:24:30
A29191	457	I	2022-12-10 10:31:42	2022-12-10 11:53:42
B27150	SP265	I	2022-12-10 10:32:03	2022-12-10 11:37:03
D86074	864	I	2022-12-10 10:32:07	2022-12-10 10:51:37

3.2. Seleção de Trajetórias

Para o cálculo das métricas, cada viagem em cada rota é considerada de forma independente. Para isso, cruzam-se as informações dos conjuntos de dados de GPS e de viagens. Além disso, por filtrar por viagens realizadas, eliminam-se casos em que o ônibus possa transmitir dados estando parado ou ainda em que envie dados errados ou insuficientes.

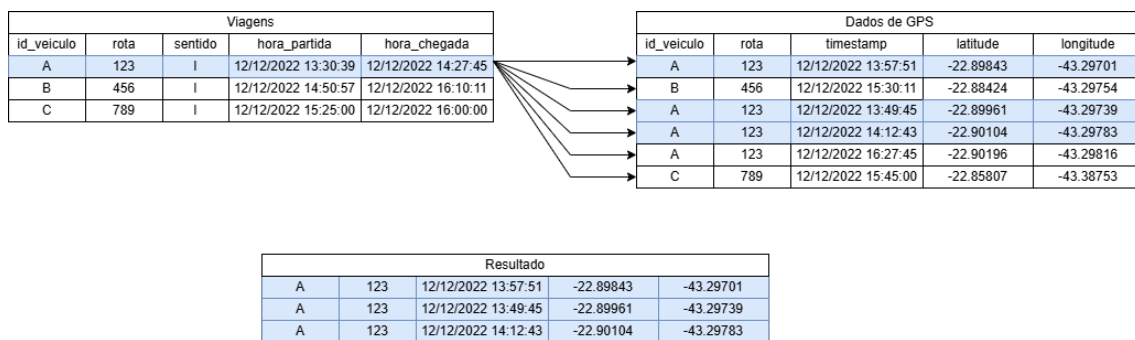


Figura 1: Exemplo de combinação dos *datasets*.

Na Figura 1 o processo é ilustrado. A viagem referente à primeira entrada da tabela de viagens é escolhida. É feita então uma busca na tabela de dados de GPS dos ônibus, procurando registros que contenham identificador de veículo, linha e horário de coleta dos dados, selecionando apenas esses registros. As entradas escolhidas são então separadas, formando o conjunto de dados referentes a uma viagem.

3.3. Correção da Trajetória

As posições coletadas dos ônibus podem ser imprecisas devido ao próprio instrumento de GPS. Como a trajetória de cada linha é conhecida, pode-se aplicar um algoritmo de correção nas coordenadas com base na linha informada. Primeiramente, as informações enviadas pelos coletivos são separadas por linha, em seguida, combinando o conjunto de dados que contém as informações dos ônibus com as informações das viagens, é feita uma separação de cada uma das viagens que foram realizadas no período analisado.

Posteriormente, cada coordenada de uma viagem é comparada com a trajetória ideal da linha, disponibilizada no DataRio, e o valor da posição real será substituído pelo valor de posição ideal que tenha menor distância para a real. Caso essa distância seja maior que um limiar, o dado real que originou essa distância será descartado.

4. Métricas para Qualidade de Serviço de Ônibus

A metodologia utilizada para avaliar a qualidade do serviço dos ônibus envolve a correção dos dados de GPS das trajetórias, antes de transformá-los em sequências de símbolos. A partir destas sequências são calculadas as diferentes métricas para avaliação do serviço.

4.1. Intervalo de Passagem

O intervalo de passagem é uma métrica calculada a partir da diferença de tempo que dois ônibus da mesma linha consecutivos levam para passar no mesmo local. Do ponto de vista de um passageiro, essa métrica avalia o tempo que ele provavelmente esperará caso opte por deixar um coletivo passar e escolha o seguinte.

Para calcular o intervalo de passagem, é escolhido um ponto na trajetória da linha e definido um raio em torno desse ponto. Caso a distância do ônibus para a parada seja menor que esse raio, será considerado que aquele ônibus passou por aquela parada. Para verificar se algum ônibus passou pelo local escolhido, primeiro a correção de trajetória é executada e, em seguida, as coordenadas das diferentes viagens são comparadas com a posição escolhida. Se a distância de algum dado para o ponto escolhido for menor que

o limiar definido, o horário da primeira ocorrência será gravado. A fim de evitar que um mesmo ônibus não possa ser detectado em diferentes horários do dia, para cada coletivo registrado é atribuída uma variável de controle com a informação do horário da detecção e, caso haja uma nova detecção em um instante diferente, ela também será registrada. Repetindo o processo para todas as viagens e organizando as ocorrências temporalmente, o intervalo entre passagens de ônibus da mesma linha é calculado.

4.2. Agrupamento de Ônibus

A distância entre ônibus que operam realizando a mesma trajetória no mesmo sentido, ou agrupamento de ônibus, é outra métrica que pode ser utilizada para avaliar a qualidade do serviço oferecido [Nguyen et al., 2023]. Um maior número de ônibus da mesma linha que trafegam agrupados normalmente é um sintoma da irregularidade do serviço.

Para calcular o número de ônibus próximos de outros, a distância de cada um para todos os demais da mesma linha é calculada durante uma janela de tempo. Caso essa distância seja inferior a um valor em algum momento da janela, os ônibus são considerados agrupados e parte de um mesmo grupo. Dois ônibus podem fazer parte de um mesmo grupo mesmo estando mais distantes que o valor determinado, para isso, basta que ambos distem menos que o limite para um mesmo terceiro ônibus. Como pode existir mais de um grupo formado, o número de agrupamento em uma janela será a soma do número de ônibus que compõem cada grupo. Como alguns ônibus podem ficar parados em seus pontos inicial e final, os dados enviados nessas regiões são descartados.

4.3. Intervalo de Tempo para Percorrer uma Fração da Trajetória

As linhas de ônibus possuem cronogramas que devem seguir, definindo o horário que devem chegar em seu ponto final. Essa informação é de conhecimento dos motoristas que, para cumprir com o horário tabelado, podem percorrer o final da viagem de forma acelerada ou lenta. Como esta medida analisa apenas parte da trajetória, ela seria capaz de capturar as reais condições em que a viagem foi feita.

Com o conhecimento do trajeto que o ônibus deve percorrer, é possível dividi-lo em partes igualmente espaçadas. A partir do fracionamento da rota, uma das partes é escolhida e o tempo que cada ônibus levou para percorre-la é computado. Para descobrir o tempo necessário para um coletivo atravessar o espaço escolhido, os símbolos que compõem a trajetória são gerados. Após isso, de acordo com a parte escolhida, os símbolos que estão localizados no início e no fim da parte são destacados. Um exemplo de fração de trajetória de uma linha, a 606, está na Figura 2, onde a fração que tem seu tempo de deslocamento calculado é aquela que se encontra entre os dois marcadores presentes na figura.

4.4. Cálculo da Entropia

A entropia H de uma variável aleatória X pode ser interpretada como uma medida de incerteza existente nessa variável aleatória. De acordo com Shannon [Shannon, 1948], a entropia pode ser calculada por

$$H(X) = -E[\ln(p(x))] = -\sum_x p(x) \ln(p(x)), \quad (1)$$

onde $p(x)$ é a probabilidade de cada saída da variável aleatória.



Figura 2: Exemplo de divisão de uma das linhas analisadas.

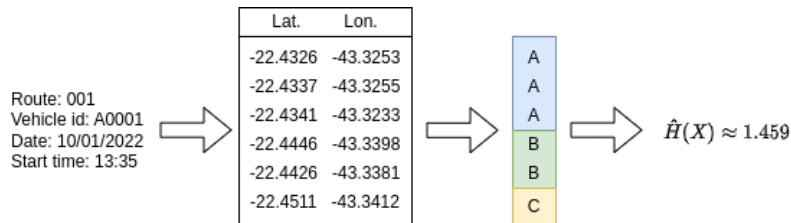


Figura 3: Exemplo de cálculo da entropia.

No contexto de ônibus urbanos, a variável aleatória representa o conjunto de possibilidades de localizações que um ônibus pode assumir ao percorrer uma trajetória. Para estimar a entropia, em um primeiro momento a sequência de posições que forma uma viagem de ônibus é transformada em símbolos, como será descrito na Seção 4.4.1. A seguir a entropia da viagem é estimada através da Equação 1.

4.4.1. Conversão de Coordenadas para Símbolos

O cálculo da entropia é feito sobre uma sequência de símbolos. Uma vez que a entropia será computada usando pares de latitude e longitude dos ônibus, que são dois valores diferentes, é necessário que haja uma conversão de coordenadas bidimensionais para símbolos unidimensionais que sejam capazes de representar a posição original do ônibus em apenas uma dimensão. Para isso, a latitude e a longitude são multiplicadas por um fator, seus valores são truncados, e o resultado das operações para a latitude e longitude é concatenado resultando em um único número. Assim, cada viagem é representada por uma sequência de símbolos que representam as posições coletadas enquanto o ônibus percorria sua trajetória. E, após isso, sua entropia é calculada. Esse processo é ilustrado na Figura 3.

O processo que envolve a multiplicação e o truncamento define a área que cada símbolo irá representar. Ele é feito de forma a escolher o número de casas decimais nas coordenadas GPS que serão usadas, com o truncamento sendo feito para eliminar as casas decimais restantes após a multiplicação. O uso de um maior número de casas

decimais para a formação dos símbolos representa um símbolo que compreende uma área menor, uma vez que seria observada uma maior precisão na latitude e longitude. Da mesma forma, o uso de um menor número de casas decimais leva a um aumento na área representada por cada símbolo, indicando uma menor certeza. Dessa forma, caso o símbolo represente uma grande área, ele irá se repetir várias vezes devido à natureza da velocidade de deslocamento do ônibus, tornando a localização real dentro da célula pouco precisa. Em contrapartida, caso o símbolo represente uma área pequena, a célula em que o ônibus está pode ser trocada devido às imprecisões inerentes ao GPS ou ainda símbolos desperdiçados por serem incompatíveis com a velocidade do ônibus.

5. Metodologia

Esta seção descreve primeiramente duas escolhas que impactam o número de dados a serem analisados na estimativa da qualidade de serviço: o período de tempo coberto e o tamanho de célula espacial, que define o número de símbolos diferentes existentes. Em seguida, as escolhas de parâmetros para as métricas utilizadas são definidas.

5.1. Período de Tempo Analisado

O período escolhido para análise das linhas de ônibus foi de 10 de dezembro de 2022 até 9 de janeiro de 2023, totalizando 31 dias. Porém, optou-se por não se usar todos os dados de todos os dias, pois há uma variação natural entre dias da semana e horários do dia que poderia atrapalhar a análise. Assim, consideram-se apenas os dias úteis, uma vez que algumas linhas do Rio de Janeiro não operam nos finais de semana, podem ter suas trajetórias alteradas, número de ônibus reduzido, além do fato de finais de semana apresentarem uma dinâmica de tráfego diferente. Pelo mesmo motivo, foi aplicado um filtro para selecionar apenas dados emitidos entre 6 horas da manhã e 10 horas da noite, por serem horários com maior circulação de veículos do que a madrugada.

Foram selecionados dois grupos de linhas que operam na cidade do Rio de Janeiro. O primeiro grupo compreende as linhas que possuem aproximadamente 15 km de extensão e o segundo grupo, linhas que possuem aproximadamente 30 km de extensão.

5.2. Escolha do Tamanho de Célula

Para a escolha do tamanho de célula para a criação dos símbolos, foi feito um experimento verificando a distribuição dos símbolos presentes em diferentes realizações de uma mesma linha. Após contabilizar o total de ocorrências de cada símbolo nas diferentes viagens realizadas, foi feita uma divisão pelo número total de viagens analisadas. Com isso, a média de ocorrências de cada símbolo por viagem foi obtida.

A escolha do tamanho de célula foi feita verificando-se qual o tamanho mais adequado para a maioria das linhas escolhidas. Para escolher o tamanho, diferentes valores foram testados nas linhas e o selecionado foi aquele que apresentava melhor equilíbrio entre células que apareciam em toda viagem, células que não apresentassem um grande número de repetições e poucas células que não apareciam em todas as viagens.

Foram testados os tamanhos de célula na Tabela 3. Analisando a Tabela 3, é possível ver que, conforme a célula diminui, o número médio de símbolos únicos por viagem aumenta de forma conjunta à entropia. Isso ocorre pois a quantidade de repetições de símbolo diminui gradativamente conforme um símbolo representa um menor espaço.

Tabela 3: Exemplo de tamanhos de célula testados para a linha 871.

Célula	Entropia média	Número médio de símbolos únicos	Número médio total de símbolos
100	6.214	76.214	81.0
200	5.936	65.357	81.0
250	5.715	57.643	81.0
300	5.520	50.214	81.0
333	5.380	45.571	81.0

Para a medida de 100 metros, existiam muitas células que não eram usadas em toda viagem. Para a medida de 333 metros, a maioria das células era utilizada mais de uma vez por viagem. O valor escolhido para a célula foi de 250 metros, pois apresenta um bom equilíbrio entre o número de células que aparece em toda viagem, e poucas células abaixo de uma ocorrência ou muito acima de uma.

5.3. Cálculo das Métricas

Após o pré-processamento dos dados descrito na Seção 3, as métricas de qualidade de serviço são calculadas. Para cada uma delas, são feitas escolhas de parâmetros que influenciam nas medidas e em suas análises. A entropia das linhas de ônibus escolhidas será analisada por viagem feita por cada ônibus que informou estar realizando uma trajetória. Para o cálculo do intervalo de passagem, por sua vez, foi feita a escolha da penúltima posição de cada linha como local de referência e definido um raio de 500 metros em torno dessa localização. Caso algum dado fosse enviado dentro da região, suas informações eram utilizadas. Para o cálculo do número de ônibus agrupados, a cada janela de tempo de três minutos foi verificado o número total de ônibus agrupados em raios de 200 metros. A análise do tempo necessário para percorrer uma fração da trajetória foi feita separando a trajetória do ônibus em três pedaços de igual tamanho. Para a determinação da medida, foi calculado o tempo que o ônibus demora para ir de um terço até dois terços do percurso.

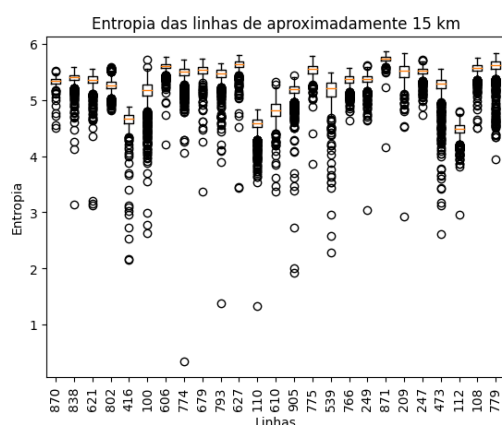
6. Resultados

Esta seção apresenta os resultados observados com as quatro métricas de qualidade do serviço analisadas: entropia, intervalo de passagem dos ônibus, quantidade de agrupamentos e tempo necessário para percorrer uma fração da linha.

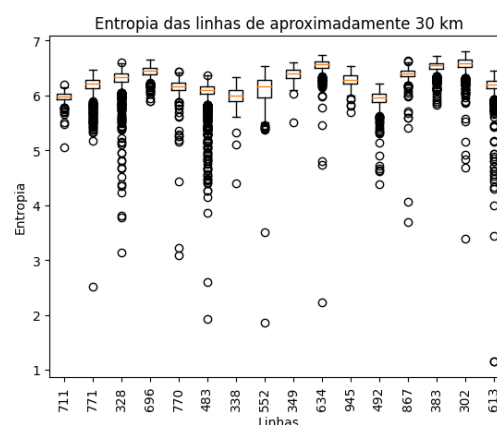
6.1. Resultado das Métricas Individuais

A Figura 4 expõe o resultado da entropia para as diferentes linhas de ônibus analisadas. Nela é possível notar que as linhas de 15 quilômetros possuem entropia em torno de cinco e, as linhas de 30 quilômetros, em torno de seis. Além disso, existem linhas em que a entropia está mais concentrada, como as linhas 416, 871 e 696, e outras em que existe uma variação maior na entropia, como as linhas 610 e 552.

Embora as linhas nos dois gráficos da Figura 4 apresentem tamanhos semelhantes, existem algumas que possuem seus valores de entropia menores que as demais. Isso pode ocorrer pois as linhas trafegam em regiões com menor volume de trânsito ou os motoristas da linha em questão podem andar mais rápido que os das demais linhas, fazendo com que as viagens sejam completadas mais rapidamente, gerando menos símbolos.



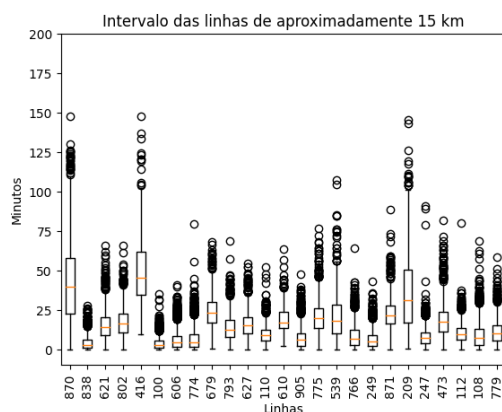
(a) Linhas de 15 km.



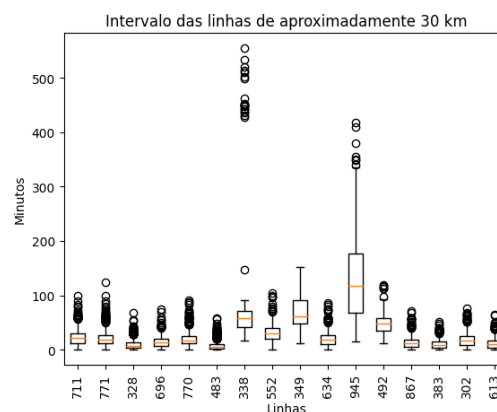
(b) Linhas de 30 km.

Figura 4: Entropia das linhas analisadas.

A Figura 5 mostra o resultado da métrica intervalo de passagem. As diferentes linhas analisadas mostram diferentes distribuições para esta medida, como é possível observar na Figura 5. Algumas linhas mostram uma maior regularidade no intervalo de passagem, como as linhas 247, 838 e 328, por exemplo. Outras demonstram maior irregularidade, caso das linhas 416, 209 e 945. Linhas que apresentam maiores intervalos de passagem são logicamente candidatas a receber mais veículos, pois as longas esperas desencorajam seu uso.



(a) Linhas de 15 km.



(b) Linhas de 30 km.

Figura 5: Intervalo de passagem das linhas analisadas.

A Figura 6 mostra os resultados de números de agrupamentos. Embora na maioria das linhas a situação normal tenha sido não encontrar ônibus agrupados, em outras o que mais se observou foi a existência de dois ou mais ônibus realizando suas viagens de forma conjunta.

As linhas em que ter ônibus agrupados é a situação normal indicam que poderiam ser utilizados menos ônibus naquela trajetória, pois é comum que dois ou mais trafeguem de forma conjunta. É importante ressaltar que o número de ônibus circulando na linha impacta diretamente o número de agrupamentos, uma vez que se existirem duas linhas de mesmo tamanho mas com números diferentes de ônibus circulando, aquela que tiver uma

maior quantidade tenderá a ter mais agrupamentos.

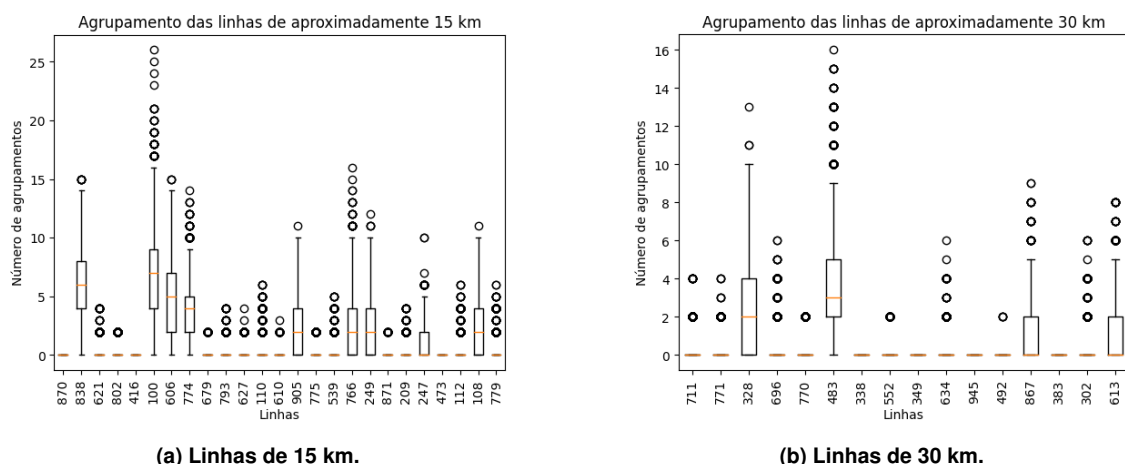


Figura 6: Número de agrupamentos de ônibus observados nas linhas analisadas.

A Figura 7 exibe o resultado da métrica de tempo necessário para percorrer de um terço até dois terços da trajetória. Os tempos encontrados para as diferentes linhas investigadas apresentam uma maior variação em seus valores mais comuns, indo de 10 minutos até 30 minutos para as linhas com 15 km (Figura 7a) e de 15 minutos até 35 minutos para as linhas com 30 km (Figura 7b). Essa variação reflete as condições de trânsito dinâmicas ao longo do dia, com a intensidade dos engarrafamentos variando e, também, os diferentes limites de velocidade e características geográficas dos locais por onde o percurso da linha passa.

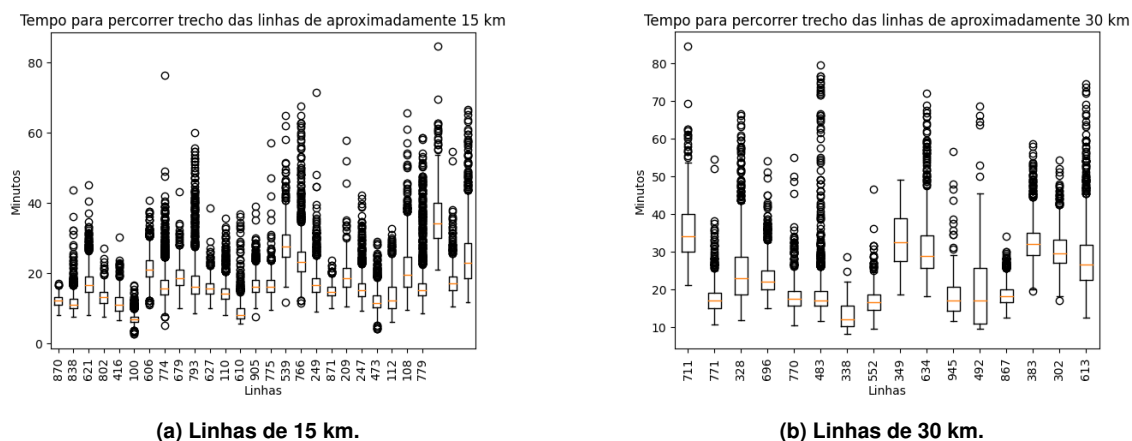


Figura 7: Tempo para percorrer uma parte das linhas analisadas.

6.2. Correlação entre as Métricas

A primeira comparação feita é entre a entropia e o intervalo de passagem. A Figura 8a exibe uma grande variação tanto na entropia quanto no intervalo de tempo dos ônibus. Nela, o intervalo da maioria dos ônibus mostra-se inferior a 50 minutos porém, a variação da entropia para esse mesmo intervalo começa em valores próximos a 4,5 indo até 6,5, indicando uma maior variação na entropia sem que o intervalo de passagem acompanhe.

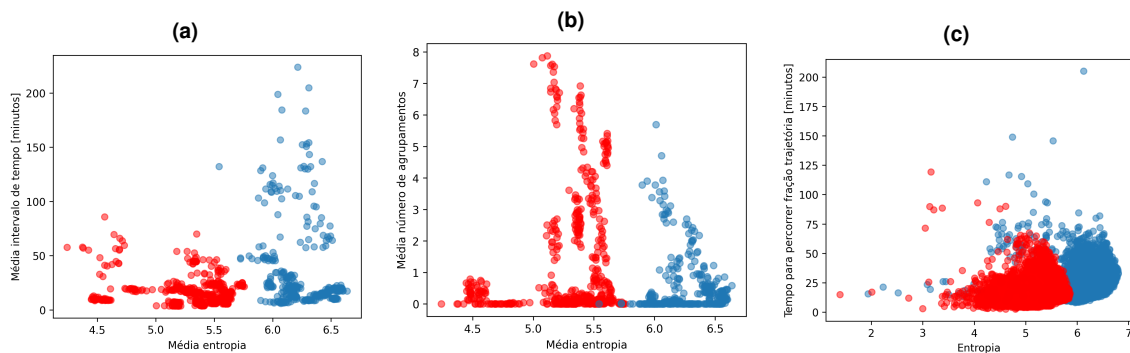


Figura 8: Comparação entre entropia e intervalo de tempo, agrupamento e tempo para percorrer fração da trajetória. Pontos em vermelho correspondem a linhas de 15 km e azuis linhas de 30 km.

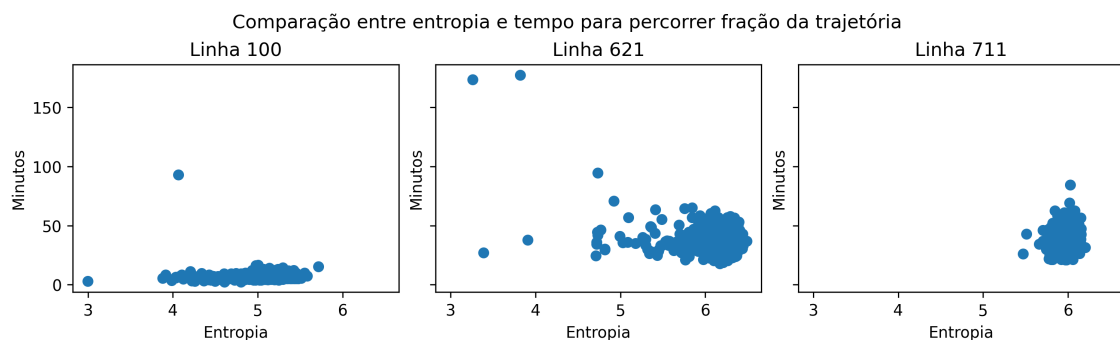


Figura 9: Diferentes comportamentos da entropia e tempo para percorrer fração da trajetória.

A comparação entre a entropia e o número de agrupamentos é feita observando a média dos dois valores para as diferentes linhas no período analisado (Figura 8b). Nela, enquanto a maioria do número de agrupamentos continua em zero, a entropia varia entre 4,5 e 6,5. Embora existam valores maiores de agrupamentos, não existe uma variação na entropia que aconteça de forma conjunta aos valores mais elevados de agrupamento.

A Figura 8c mostra uma comparação entre a entropia de todas as viagens de todas as linhas e o tempo necessário para o mesmo ônibus percorrer o trecho investigado. Analisando o gráfico, é possível perceber que, para os dois grupos de linhas estudados, existe uma grande dispersão na entropia, mesmo que a maioria dos tempos esteja entre 25 e 50 minutos.

A Figura 9 mostra as variações que ocorrem em diferentes linhas. Enquanto para a linha 100 o intervalo se mantém relativamente constante, para a linha 711 a entropia mostra-se mais regular em detrimento de uma maior variação temporal. A linha 621, por sua vez, exibe variação na entropia e no tempo.

Na Figura 10 está a comparação entre intervalo de passagem, agrupamento e tempo para percorrer uma fração da trajetória. Analisando a Figura 10a, existe uma indicação de relação entre as médias do intervalo de tempo e do agrupamento. Isso ocorre pois, com uma maior frequência de passagem de ônibus em uma linha, menor será a

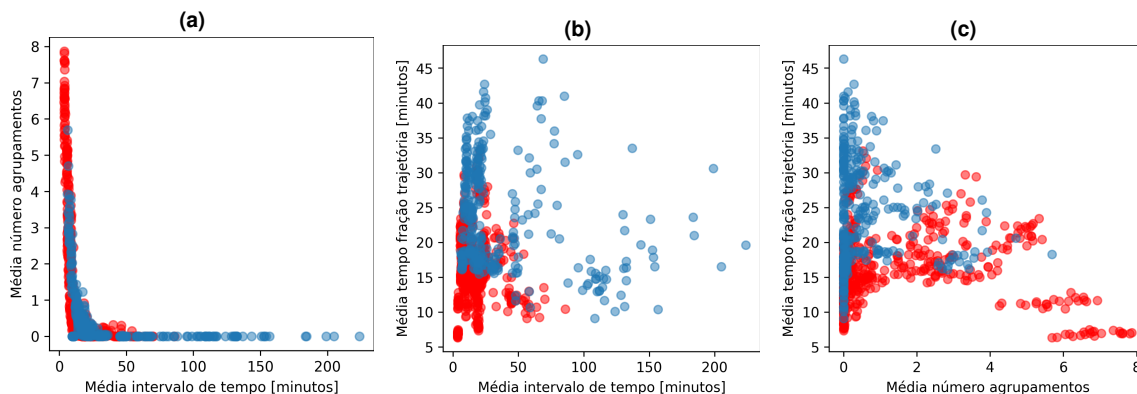


Figura 10: Comparação entre agrupamento, intervalo de tempo e tempo para percorrer fração da trajetória. Pontos em vermelho correspondem a linhas de 15 km e azuis linhas de 30 km.

distância de um ônibus para o próximo e para o anterior. Essa é a situação ideal para que ocorra a formação de grupos de ônibus percorrendo seus trajetos de forma conjunta.

Já a Figura 10b e a Figura 10c comparam o tempo médio para percorrer uma fração da linha com a média do intervalo de tempo e com a média de agrupamentos, respectivamente. Esperava-se encontrar uma relação entre a média do intervalo de tempo e a média do tempo necessário para percorrer uma fração da linha, pois caso um ônibus demore mais para realizar seu percurso que o normal, o tempo de chegada na parada de ônibus seria afetado. Entretanto, observou-se que as medidas variam de forma independente, mostrando um ponto que merece maior investigação. Uma possível explicação é que quando é feita a média do intervalo de passagem do ônibus para o anterior e para o próximo, o comportamento anômalo seria compensado pela média entre os dois valores, um maior e outro mais baixo.

Como o agrupamento e o intervalo de passagem mostram-se relacionados, um comportamento semelhante à comparação entre intervalo de passagem e tempo necessário para percorrer uma fração da linha foi observado na comparação entre agrupamentos e tempo necessário para percorrer uma fração da linha. Ônibus que realizam sua trajetória de forma mais lenta que os demais tendem a formar grupos com aqueles que percorrem de forma mais rápida. Entretanto, quando a velocidade reduzida é causada por engarrafamentos, ambos os veículos utilizam velocidades semelhantes.

6.3. Discussão

A entropia dos dois grupos de linhas de ônibus mostra-se regular de acordo com seu comprimento, sendo possível diferenciar aquelas que têm 15 km das que possuem 30 km pelo valor típico da entropia. Da mesma forma, caso alguma viagem dessas linhas apresente um valor destoante dos demais, existe uma indicação de que algo atípico ocorreu, seja falha no envio de dados, seja alguma condição do percurso. O elevado número de linhas que apresentam ônibus agrupados, seja aquelas em que a ocorrência é o comum, seja aquelas em que são exceção, indica que o fenômeno do agrupamento é normal no Rio de Janeiro. Entretanto, é importante destacar que ele ainda assim é indesejado, pois prejudica a regularidade na frequência de passagem dos ônibus.

Os resultados das comparações entre a entropia e as demais métricas indicam que não existe correlação aparente entre as medidas mais típicas de qualidade de serviço e a entropia. A comparação entre o agrupamento e o intervalo de tempo, por sua vez, revelou que linhas que têm altos intervalos entre ônibus consecutivos tendem a não formar grupos de ônibus, pois a distância entre os coletivos é elevada. A análise simultânea de tempo para percorrer uma fração do percurso e intervalo de passagem e agrupamento revelou que as duas últimas medidas variam de forma independente da primeira.

7. Conclusão

Este trabalho verificou a qualidade do serviço de linhas de ônibus urbanos. Para isso, diferentes métricas de análise de qualidade foram utilizadas e comparadas entre si. Com a intenção de evitar dados incoerentes, um algoritmo de correção de trajetória baseado na rota real que o ônibus deveria seguir foi aplicado para, em seguida, as métricas serem calculadas.

Ao analisar as métricas de forma individual, foi possível separar os grupos de linhas de tamanhos diferentes utilizando a entropia das viagens. O intervalo de passagem, por sua vez, apresentou uma variação de alguns minutos na maioria das linhas. Já a ocorrência de agrupamentos mostrou-se um cenário comum na maioria das linhas do Rio de Janeiro. Por fim, o intervalo de tempo para percorrer uma fração da trajetória foi capaz de capturar as variações no tempo de deslocamento nas diferentes linhas.

Comparando a entropia com as demais medidas, este trabalho concluiu que não existe uma relação clara entre a entropia e as demais medidas de qualidade de serviço de ônibus mais comumente utilizadas. Quando analisando o intervalo de tempo em comparação com o número de agrupamentos, observou-se que as medidas variam de forma conjunta. A comparação entre o tempo para percorrer uma fração da trajetória e intervalo de passagem e agrupamento, por sua vez, mostrou comportamento semelhante nos dois casos mas sem que uma correlação entre as medidas aparecesse de forma evidente.

Os próximos passos dessa pesquisa visam o teste de novas métricas de qualidade de serviço e de novas representações para o cálculo de entropia, buscando novas formas de avaliar o serviço. Por fim, planeja-se utilizar as métricas para a detecção online de anomalias no serviço dos ônibus.

Referências

- Cuttone, A., Lehmann, S. e González, M. C. (2018). Understanding predictability and exploration in human mobility. *EPJ Data Science*, 7(1):2.
- Du, B. e Dublanche, P.-A. (2018). Bus bunching identification using smart card data. Em *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, p. 1087–1092.
- He, S.-X. (2015). An anti-bunching strategy to improve bus schedule and headway reliability by making use of the available accurate information. *Computers & Industrial Engineering*, 85:17–32.
- Huang, Z., Xu, S., Wang, M., Wu, H., Xu, Y. e Jin, Y. (2024). Human mobility prediction with causal and spatial-constrained multi-task network. *EPJ Data Science*, 13(1):22.

- Ikanovic, E. L. e Mollgaard, A. (2017). An alternative approach to the limits of predictability in human mobility. *EPJ Data Science*, 6(1):12.
- Li, G., Knoop, V. L. e van Lint, H. (2022). Estimate the limit of predictability in short-term traffic forecasting: An entropy-based approach. *Transportation Research Part C: Emerging Technologies*, 138:103607.
- Nguyen, K., Yang, J., Lin, Y., Lin, J., Chiang, Y.-Y. e Shahabi, C. (2023). Los Angeles Metro Bus Data Analysis Using GPS Trajectory and Schedule Data. Tech Report CA18-3124, University of Southern California, Califórnia, USA.
- Prefeitura do Rio de Janeiro (2025). Data.rio - portal de dados abertos da cidade do rio de janeiro. Acesso: 14-Mar-2025.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.
- Song, C., Qu, Z., Blumm, N. e Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, 327(5968):1018–1021.
- Teixeira, D. d. C., Almeida, J. M. e Viana, A. C. (2021). On estimating the predictability of human mobility: the role of routine. *EPJ Data Science*, 10(1):49.
- Teixeira, D. d. C., Viana, A. C., Alvim, M. S. e Almeida, J. M. (2019). Deciphering predictability limits in human mobility. Em *SIGSPATIAL 2019 - 7th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, United States.
- Wang, P., Chen, X., Zheng, Y., Cheng, L., Wang, Y. e Lei, D. (2021). Providing real-time bus crowding information for passengers: A novel policy to promote high-frequency transit performance. *Transportation Research Part A: Policy and Practice*, 148:316–329.
- Zhao, K., Khryashchev, D., Freire, J., Silva, C. e Vo, H. (2016). Predicting taxi demand at high spatial resolution: Approaching the limit of predictability. Em *2016 IEEE International Conference on Big Data (Big Data)*, p. 833–842.
- Zhou, C., Tian, Q. e Wang, D. Z. (2022). A novel control strategy in mitigating bus bunching: Utilizing real-time information. *Transport Policy*, 123:1–13.