

UNIVERSIDADE FEDERAL DO ABC

PROGRAMA DE ESPECIALIZAÇÃO
EM TECNOLOGIA E SISTEMAS DA INFORMAÇÃO

Patrick Natan Gomes dos Santos

**APLICAÇÃO DE REDES COMPLEXAS NO SISTEMA DE
TRANSPORTE COLETIVO URBANO DE ÔNIBUS DA
CIDADE DE SÃO PAULO**

Orientador: Prof. Dr. Mario Alexandre Gazziro

Santo André, SP

2024

PATRICK NATAN GOMES DOS SANTOS

APLICAÇÃO DE REDES COMPLEXAS NO SISTEMA DE TRANSPORTE
COLETIVO URBANO DE ÔNIBUS DA CIDADE DE SÃO PAULO

Trabalho de Conclusão de Curso apresentado ao curso de pós-graduação em TSI - Tecnologia e Sistemas da Informação, da Universidade Federal do ABC, como requisito parcial para obtenção do diploma no curso de Tecnologia.



Assinatura do Orientador

SANTO ANDRÉ

2024

Ficha catalográfica

Santos, Patrick Natan Gomes dos

Aplicação de redes complexas no sistema de transporte coletivo urbano de ônibus da cidade de São Paulo – Santo André, SP: UFABC, 2024

Resumo

O uso de redes complexas na análise de sistemas de transporte urbano está se tornando cada vez mais presente em estudos de mobilidade e planejamento urbano. A teoria de redes complexas oferece uma abordagem poderosa para compreender a estrutura e dinâmica desses e de outros sistemas interconectados. Este trabalho aplica a teoria de redes complexas para analisar o sistema de transporte coletivo de ônibus da cidade de São Paulo, através de dados fornecidos pela SPTrans. O estudo constrói e analisa duas redes distintas: uma baseada na conectividade entre pontos de ônibus e outra na similaridade entre linhas de ônibus. Com o objetivo de entender como métricas de rede complexa podem trazer insights e possíveis melhorias no sistema. A análise da rede de pontos revelou uma estrutura ramificada com alta densidade de pontos no centro e uma boa robustez na rede. A análise da rede de similaridades mostrou uma alta conectividade entre as linhas, contudo, com correlações fracas a moderadas com as métricas de rede complexa. Os resultados obtidos são promissores para estudos futuros e para a busca de melhoria no sistema e na qualidade de vida dos passageiros no sistema de transporte rodoviário público da cidade de São Paulo.

Palavras-chave: Grafos, mobilidade urbana, redes complexas, SPTrans e São Paulo

Abstract

The use of complex networks in the analysis of urban transportation systems is becoming increasingly prevalent in mobility and urban planning studies. Complex network theory offers a powerful approach to understanding the structure and dynamics of these and other interconnected systems. This work applies complex network theory to analyze the public bus transportation system of São Paulo city, using data provided by SPTrans. The study constructs and analyzes two distinct networks: one based on the connectivity between bus stops and another on the similarity between bus lines. The aim is to understand how complex network metrics can bring insights and potential improvements to the system. The analysis of the bus stop network revealed a branched structure with high density of points in the center and good network robustness. The analysis of the similarity network showed high connectivity between lines, however, with weak to moderate correlations with complex network metrics. The results obtained are promising for future studies and for seeking improvements in the system and in the quality of life of passengers in São Paulo's public road transport system.

Keywords: Graphs, urban mobility, complex networks, SPTrans, and São Paulo

SUMÁRIO

1 Introdução	7
1.1 Justificativa	9
2 Objetivo	10
3 Metodologia	11
3.1 Visão Geral	11
3.1 Base de Dados	12
3.1.1 General Transit Feed Specification	12
3.1.2 Total de passageiros	13
3.1.3 Pré-processamento	14
3.2 Construção das redes	14
3.2.1 Cypher	14
3.2.2 Neo4j	15
3.3 Análise da Rede	16
3.3.1 Gephi	16
4 Resultados	17
4.1 Rede de Pontos de ônibus	18
4.1.2 Representação visual	18
4.1.2 Métricas da rede	20
4.1.3 Análise dos nós	21
4.2 Rede de similaridade de linhas	29
4.2.1 Análise da rede	29
4.2.2 Análise dos dados	32
5 Conclusão	37
5.1 Trabalhos futuros	38
Bibliografia	39

1 Introdução

Em um mundo onde tudo é conectado, da ligação entre pessoas até a elos entre conceitos, é de suma importância entendermos a natureza a respeito de cada uma dessas conexões. Para investigarmos ainda mais profundo, utilizamos um modelo matemático que é capaz de representá-las, conhecido como grafo.

Um grafo pode ser definido como uma simples estrutura formada por vértices que se interligam através de nós. Com essa estrutura simples, mas que é poderosa, pode ser utilizada para resolver diversos problemas, como o do menor caminho entre duas cidades ou torneios [1], mostrando a versatilidade desse ramo da matemática.

Uma outra aplicação possível para grafos, são as redes complexas, que de acordo com Barabási [2] é um sistema formado por componentes chamados de nós que tem interações diretos entre eles através de *links* ou pontes. Com essa definição é possível interagir com sistemas complexos [3] e analisá-los, exemplos desses tipos de redes são, estrutura do cérebro humano, rede mundial de computadores, relação em uma rede social, conexões entre aeroportos, cadeias heterotróficas, etc.

A importância de se estudar redes complexas está na observação de que em um sistema complexo qualquer, a soma das partes e o todo não são equivalentes, e alterações em suas conexões ou em seus nós podem afetar consideravelmente o sistema.

Uma grande metrópole pode ser caracterizada como um sistema complexo, seja a conexão de suas ruas, seu sistema de distribuição de água, esgoto, energia e também o seu sistema de transporte, que se conecta ao redor de toda a cidade. Quando as redes complexas são aplicadas em sistemas assim, é possível descrever fatores relacionados ao deslocamento dos indivíduos, melhoria de funções logísticas e até fatores decorrentes de comportamentos coletivos [4].

Com a utilização de redes complexas, é possível aferir diversas métricas, observar propriedades e também executar inúmeras alterações para observar como uma organização de um sistema público de transporte se estabelece.

Um dos grandes aspectos das grandes cidades é a sua mobilidade urbana, que não é algo que deve ser ignorado pelo poder público, pois afeta boa parte da população, e principalmente a que utiliza mais o transporte público. Uma má condição de mobilidade urbana, com o seu serviço de transporte coletivo, pode afetar uma parcela significativa da população a serviços essenciais que as metrópoles oferecem, contribuindo ainda mais para a perpetuação da pobreza urbana e da exclusão social [5], que é algo que demonstra o tamanho da importância de uma mobilidade urbana eficiente.

A utilização de redes complexas em países ao redor do mundo na mobilidade urbana é tema em diversos trabalhos, como a utilização em desenvolvimento de redes de metrô [6], que demonstrou a possibilidade de aumentar a conectividade espacial urbana fornecendo melhores recursos sociais e promoção da qualidade de vida das pessoas na cidade de Porto em Portugal. Outra aplicação, agora no Brasil, foi a análise da rede de transporte de diversas cidades do Paraná [7], que trouxe diversos resultados a respeito da malha rodoviária que podem guiar decisões do planejamento urbano. Já na cidade mais populosa do país, São Paulo, a utilização de redes complexas pode nos fornecer a possibilidade de aperfeiçoar e melhorar a estrutura para melhorar no fluxo de passageiros [8].

1.1 Justificativa

A cidade de São Paulo, de acordo com o IBGE de 2022 [9], consta com uma população de 11 milhões de pessoas, sendo a maior do Brasil. Onde dados da empresa que gerencia o transporte de ônibus da cidade, a SPtrans, informou que cerca de 21% dessa população utiliza os ônibus como transporte público conta com variados problemas, onde boa parte desses problemas são questões de logística [10], que vão desde atrasos de linhas até falta de novas linhas para atender a população.

Em 2023, a prefeitura da cidade de São Paulo aplicou uma pesquisa para conhecer melhor os usuários do transporte público, o resultado foi que a grande maioria dos usuários são trabalhadores [11], dado essa apuração, um sistema que não tem qualidade e tenha ineficiências, que principalmente atinja a dignidade dos trabalhadores, fere o artigo 6 da constituição federal de 1988 e também o Art.22 Código de Defesa do Consumidor, além de, gerar múltiplos prejuízos para a sociedade, desde a qualidade de vida dos usuários [12] até o sistema econômico da metrópole [13]. Por causa disso existe sempre a necessidade da busca de uma melhora do sistema de transporte público..

Com um sistema com cerca de 12 mil ônibus, distribuídos em pelo menos 1300 linhas [14], e com milhares de conexões entre eles, é possível observar o quão complexo é o sistema de transporte urbano da cidade de São Paulo e o quanto que a utilização de redes complexas podem nos ajudar para analisá-lo e principalmente propor melhorias.

2 Objetivo

Este trabalho de caráter exploratório tem como objetivo analisar o sistema de transporte público de ônibus do município de São Paulo, com base na Teoria das Redes Complexas. Para isso, tem-se os seguintes objetivos específicos:

- Fazer o *scraping* de dados que tenham valor para a rede;
- Construir redes para diversos contextos, por exemplo: pontos de ônibus e similaridade entre linhas;
- Utilizar de diversas métricas de Redes Complexas, como centralidade, *Betweenness* etc, para extrairmos informações a respeito da rede;
- Comparar com outras estruturas já existentes;
- Apresentar visualmente a rede complexa com suas métricas a fim de encontrar algum padrão com a estrutura da cidade;
- Efetuar alterações, removendo vértices e arestas e verificando a robustez da rede;
- Analisar correlações das métricas de redes complexas com outros dados.

3 Metodologia

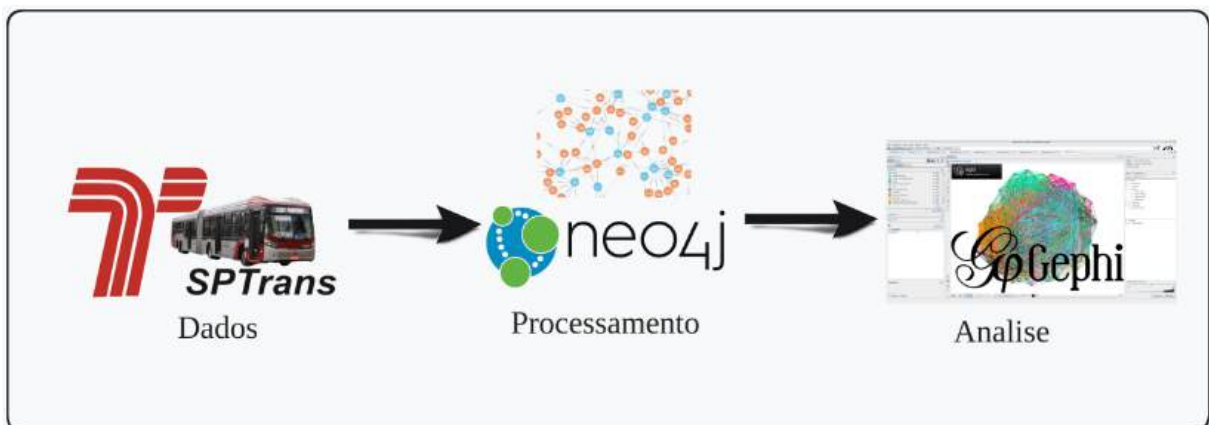
3.1 Visão Geral

Este trabalho tem como foco analisar a rede de mobilidade rodoviária urbana da grande metrópole São Paulo, conforme indica a figura 1. Através do *scraping* de dados por meio de dados abertos fornecidos por Lei de Acesso à Informação da SPTrans; O objetivo em questão é adquirir dados a respeito das linhas de ônibus, pontos de parada e seus trajetos na cidade.

Com os dados coletados, é iniciado o processamento com a ferramenta de banco de dados de grafos Neo4j, onde os dados serão processados através da definição de conexões entre vértices e arestas, definição das propriedades e pela criação do grafo, que é processado e exportado em um formato específico para utilizar na análise.

Na análise é utilizado o software Gephi, onde é possível realizar diversas explorações e experimentações da rede através de variadas métricas e funções do software.

Figura 1: Diagrama do processo para a análise da redes complexa.



Fonte: Elaboração Própria.

3.1 Base de Dados

Uma das possibilidades levantadas para a busca de dados das linhas e pontos de ônibus é a realização do *scraping* de dados diretamente da plataforma da SPTrans, mas embora trouxesse somente o dado necessário e ser bem flexível, poderia ser um processo oneroso por ter diversas limitações, como por exemplo bloqueio do ip e/ou quantidade de requisições por minuto, inviabilizando a coleta contínua de dados. Uma alternativa encontrada para montar a base de dados é registrar-se com uma conta de desenvolvedor na plataforma SPTrans, com isso é possível baixar um arquivo GTFS.

3.1.1 General Transit Feed Specification

O *General Transit Feed Specification* (GTFS) é um formato gerenciado pelo *Google* que possibilita uma padronização para que as agências de trânsito compartilhem seus dados em um formato específico e com isso permitindo que desenvolvedores a utilizem com menos dificuldades. O formato é dividido em até 16 arquivos diferentes no formato .txt em um mesmo arquivo cuja a divisão de dados é por uma vírgula conforme a figura 2 exemplifica.

Figura 2: Exemplo do começo do arquivo trips.txt

```
"route_id","service_id","trip_id","trip_headsign","direction_id","shape_id"  
"1012-10","USD","1012-10-0","Jd. Monte Belo","0","81072"  
"1012-10","USD","1012-10-1","Term. Jd. Britania","1","81073"  
"1012-21","U__","1012-21-0","Jd. Rosinha","0","81195"  
"1015-10","USD","1015-10-0","Chác. Maria Trindade","0","81148"  
"1016-10","USD","1016-10-0","Shop. Center Norte","0","72283"  
"1016-10","USD","1016-10-1","Cem. Do Horto","1","82172"  
"1017-10","USD","1017-10-0","Conexão Vl. Iório","0","72355"  
"1017-10","USD","1017-10-1","Perus","1","72356"
```

Fonte: Elaboração Própria.

Para essa pesquisa foram utilizados os arquivos conforme exemplifica a Tabela 1.

Tabela 1 - Descrição dos arquivos específicos do GTFS utilizados

Arquivo	Conteúdo dos arquivos
stop_times.txt	Horários de partida e chegada para cada linha
routes.txt	Linhas de ônibus
trips.txt	Viagens
stops.txt	Localização dos pontos de ônibus

O arquivo *stop_times.txt* fornece para cada linha de ônibus os IDs dos pontos de parada de forma sequencial. Isso permite que utilizemos esses dados em conjunto com o arquivo *stops.txt* para mapear a localização de cada ponto.

Para nomear cada rota de forma adequada, utilizaremos o arquivo *routes.txt* que contém o nome de cada uma das rotas de acordo com o seu código. A conexão entre as rotas será estabelecida através dos pontos em comum.

O arquivo *trips.txt* descreve como uma linha viaja no seu trajeto de ida e de volta entre os pontos de ônibus definidos no arquivo *routes.txt*, sua utilização é essencial na montagem das relações entre linhas e também auxilia na relação entre pontos.

3.1.2 Total de passageiros

Por meio da Lei de Acesso à Informação, a prefeitura de São Paulo fornece um relatório diário, mensal e anual da quantidade de passageiros transportados em cada uma das linhas de ônibus existentes da SPtrans, esse dado tem a sua importância para dar mais dados para as linhas da rede e permitir sua análise com métricas de rede complexa.

3.1.3 Pré-processamento

Para facilitar a análise de similaridade entre linhas de ônibus, foi desenvolvido um processo de consolidação dos dados do GTFS em um novo arquivo CSV. Este arquivo integra, para cada linha de ônibus, informações como as de viagens realizadas, os pontos de parada de cada viagem, a quantidade total de pontos de parada e o número total de passageiros. A criação deste conjunto de dados constitui uma etapa preliminar para as fases subsequentes de pré-processamento e análise.

3.2 Construção das redes

Para a construção de redes é necessário converter os arquivos pré - processados para um formato de grafos através da utilização da linguagem *Cypher* no aplicativo Neo4j.

3.2.1 Cypher

Da mesma forma que existe a linguagem SQL para bancos de dados relacionais, para banco de dados orientados a grafo como o Neo4j, temos a linguagem *Cypher*, uma linguagem de característica semelhante ao SQL só que com o foco em grafos.

Tabela 2: Alguns dos comandos na linguagem Cypher utilizado

MATCH	Utilizado para procurar padrões, semelhante ao WHERE do SQL
CREATE	Utilizado para criação, tanto de vértices como de arestas
MERGE	Junção dos comandos MATCH e CREATE, se o nó ou vértice já foi criado, não cria novamente
WHERE	Utilizado para filtrar os resultados da consulta baseado em condições específicas

A utilização do *Cypher* neste trabalho será com a função de manipular as redes criadas pela importação dos arquivos de maneira que seja possível filtrar diversos tipos de dados para criar e exportar diversas redes diferentes.

3.2.2 Neo4j

Entre diversas opções de sistemas de gerenciamento de banco de dados especificamente para grafos, existe o Neo4j, de código aberto, embora exista uma versão comercial. Possui uma vasta e ativa comunidade que o mantém, além de possuir processamento de dados nativos em seu núcleo, o que dá a ele uma característica de alta performance de desempenho.

Através do importador automático do Neo4j de um arquivo .csv é possível definir as propriedades de nós e principalmente as suas ligações facilmente através de funções de *importer*.

Também é possível importar

Com o grafo montado no Neo4j agora é possível exportá-lo em um formato CSV compatível com demais ferramentas de análise de redes complexas.

3.3 Análise da Rede

O Neo4j, apesar de oferecer ferramentas em seu ecossistema para a análise de redes complexas, apresenta algumas limitações que o impedem de ser a ferramenta ideal para esse trabalho. O conjunto de métricas e funcionalidades do Neo4j no foco em redes complexas é um pouco mais restrito quando comparado a outras ferramentas especializadas e, por tal motivo, foi escolhido a utilização de ferramentas focadas em análise de redes complexas com as redes modeladas exportadas do Neo4j.

3.3.1 Gephi

O Gephi é um software de código aberto e gratuito utilizado para a realização de análises, integrações e visualizações de redes complexas. Embora tenha sido originalmente projetado para análise de redes sociais, o Gephi se consolidou como uma ferramenta robusta para o estudo de redes complexas em diversas áreas, permitindo a importação das redes construídas na etapa anterior e a utilização de diversas ações com a rede, como:

- Análise da topologia da rede
- Cálculo de métricas de redes complexas
- Simulações de cenários
- Detecção de comunidades
- Análise de sub-redes

4 Resultados

Este capítulo apresenta os resultados obtidos no presente estudo, utilizando uma combinação de tabelas, representações gráficas da rede e diversos gráficos dos dados para facilitar a apresentação e interpretação. A análise foi feita sob duas perspectivas diferentes de modelagem da rede de transporte público:

- Conectividade entre pontos de ônibus: Esta abordagem considera a estrutura física da rede, incorporando a posição geográfica dos pontos de parada.
- Similaridade entre linhas de ônibus: Esta abordagem analisa as conexões entre linhas com base na quantidade de pontos de parada compartilhados.

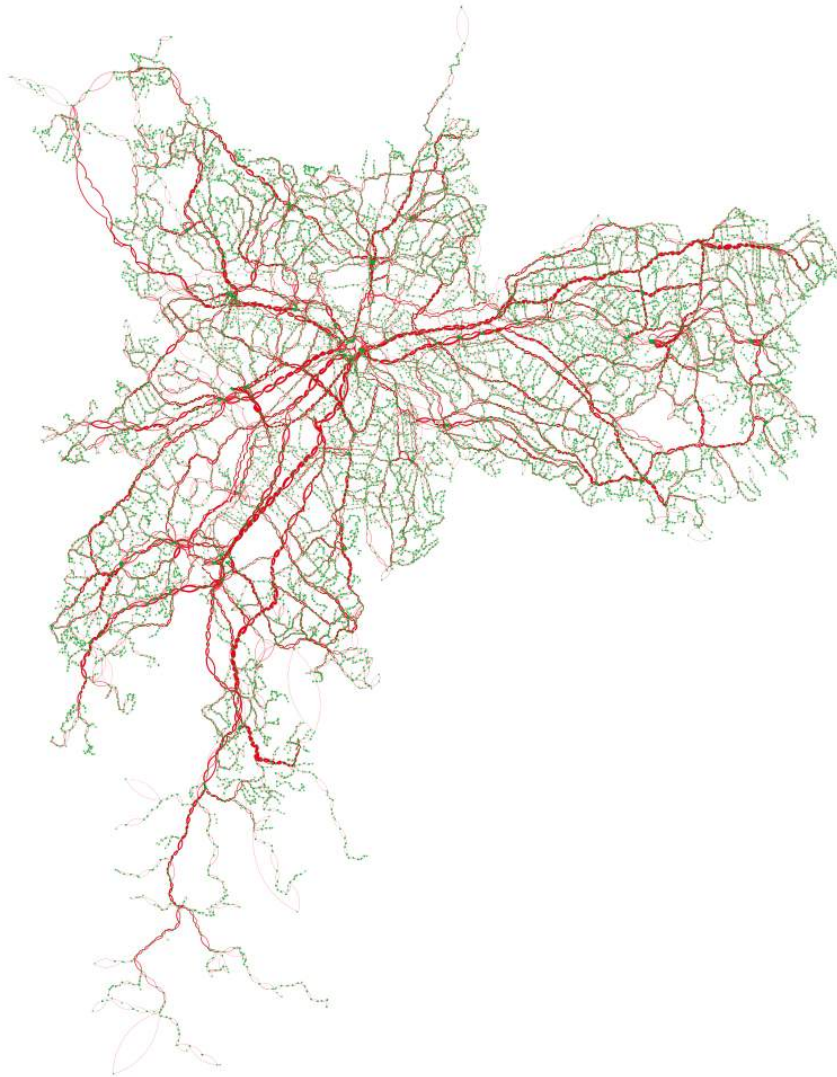
O objetivo desta dupla abordagem é ter uma compreensão abrangente da rede de transporte, permitindo a análise de métricas relevantes e a interpretação do grafo resultante. Busca-se, assim, explicar tanto a estrutura espacial do sistema quanto às relações funcionais entre suas linhas.

4.1 Rede de Pontos de ônibus

4.1.2 Representação visual

Com a cidade de São Paulo possuindo um total de 21469 pontos de ônibus e 28919 conexões entre os pontos no período de maio de 2024, não foram encontrados pontos sem conexão. Ao transformar a rede de pontos utilizando dados de longitude e latitude de cada ponto e mantendo as suas conexões, foi possível construir a representação visual da mesma respeitando os limites geográficos da cidade.

Figura 3: Representação geoespacial da rede de pontos.

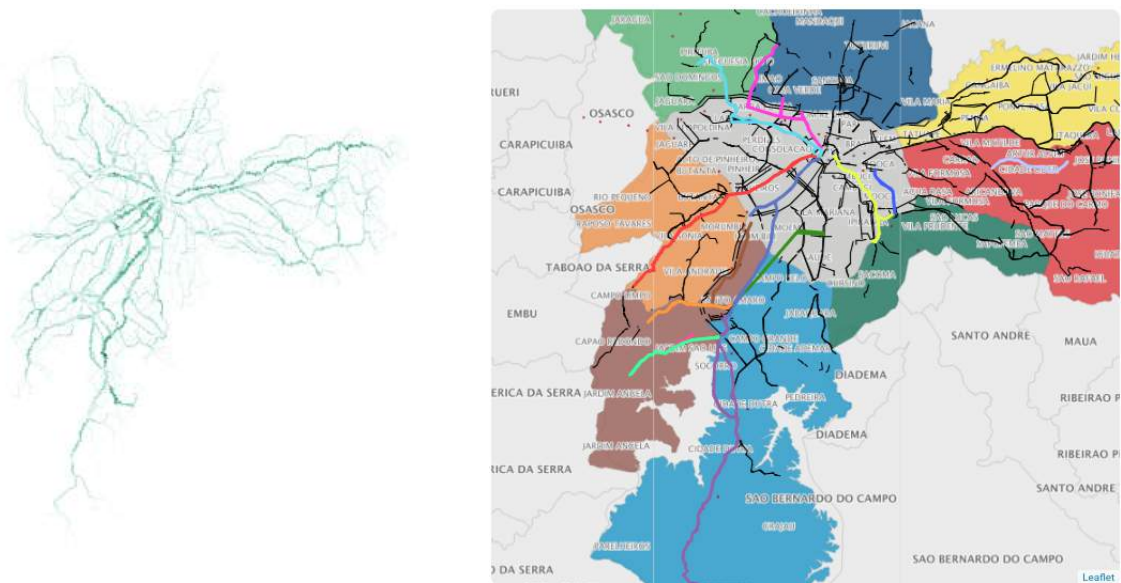


Fonte: Elaboração Própria.

A figura 3 mostra uma representação visual da rede de São Paulo, que se assemelha a uma estrutura ramificada onde os detalhes em verde são os pontos de ônibus e as linhas em vermelho são as suas conexões, é possível observar diversas características da organização da cidade como:

- A região central tem uma alta densidade de pontos, o que é uma característica comum em grandes metrópoles que refletem como o centro é um polo importante ao mesmo tempo que demonstra a hiperconcentração urbana
- Existem diversos pontos da rede com um alto número de conexões, que funcionam como um verdadeiro *hubs* de transporte, esses pontos geralmente estão localizados em locais estratégicos como terminais de ônibus ou perto de estações da malha ferroviária da cidade
- Que ao destacarmos as arestas pelos seus pesos conforme a figura 4, as principais rotas saem do centro e se expandem para todos os hemisférios da cidade com as arestas mais densas o que indica a presença de corredores exclusivos ou faixas exclusivas;

Figura 4: Rede complexa as arestas em comparação da Rede de corredores e linhas exclusivas de ônibus.



Fonte: Elaboração Própria.

4.1.2 Métricas da rede

Realizando a metrificação da rede como um todo foi possível adquirir os seguintes resultados.

Tabela 3: Métricas da rede de pontos de ônibus

Métrica	Valor
Grau médio	1,347
Grau ponderado médio	4,405
Diâmetro da rede	199
Modularidade	0,935
Coefficiente de Agrupamento Médio(direcionado)	0,17

Onde de acordo com a tabela 3, cada métrica significa:

Grau médio: indica que, em média, cada ponto está conectado a aproximadamente 1,347 de outros pontos, o que sugere uma rede relativamente esparsa, com poucas conexões diretas entre os pontos.

Grau ponderado médio: valor que considera o peso das conexões, a diferença do grau ponderado médio para o grau médio indica que existem pontos onde existe maior fluxo de linhas de ônibus por terem conexões mais frequentes.

Diâmetro da rede: representa o caminho mais longo entre quaisquer dois nós da rede, um diâmetro alto indica que a rede é extensa e que pode levar muitas paradas para ir de um extremo ao outro da cidade.

Modularidade: é a força da divisão de uma rede em comunidades, um valor próximo de 1 indica que a estrutura de comunidades é muito forte numa rede, isso indica que existem bairros/conjunto de bairros que são mais densamente conectados do que com o resto da rede.

Coeficiente de Agrupamento Médio (direcionado): métrica que mede como os nós de uma rede tendem a se agrupar, ele indica a tendência dos vizinhos de um nó se conectarem entre si, o valor de 0,17 indica um valor baixo, o que por consequência demonstra que há pouca tendência de agrupamento local na rede. Demonstrando que os pontos de ônibus vizinhos geralmente não estão bem conectados entre si. Na prática isso indica que as conexões são mais direcionadas e menos agrupadas, o que impacta para o usuário mais transferências para alcançar destinos próximos.

4.1.3 Análise dos nós

Na tabela 4 é apresentado o ranking dos pontos de maior número de grau ponderado, o que significa quais são os pontos que têm mais linhas de ônibus passando entre eles, ou seja, quantidade de linhas que entram e que saem de cada nó.

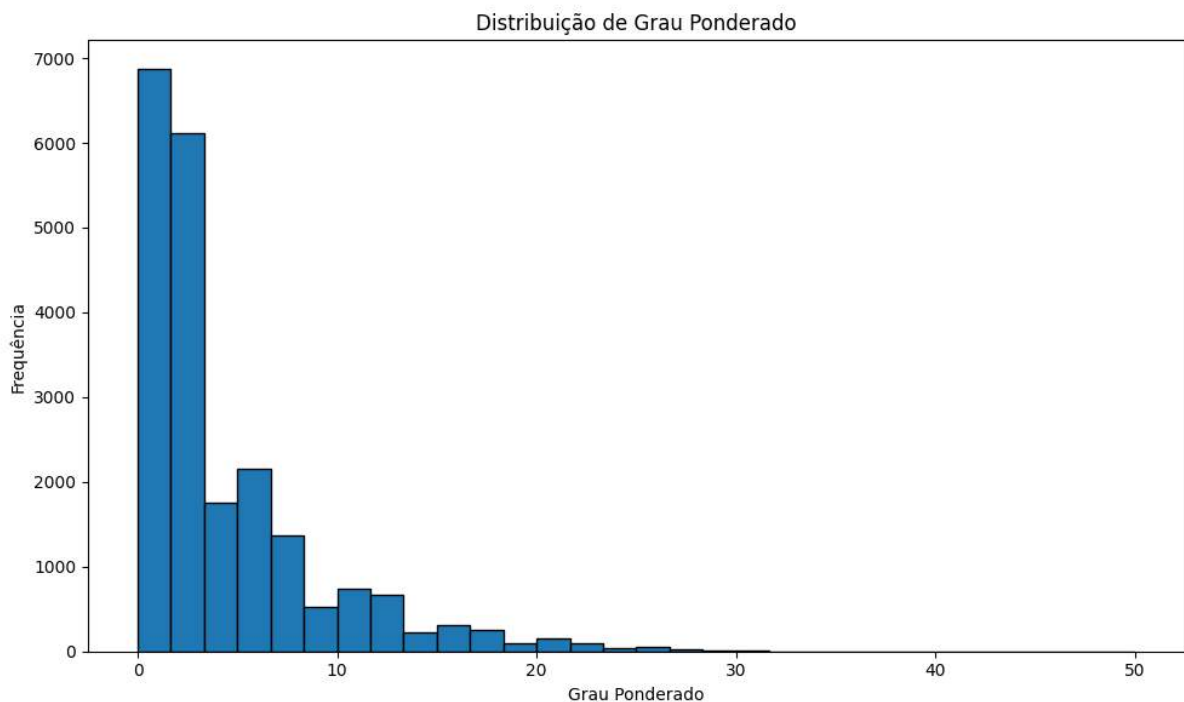
Tabela 4: Ranking de pontos de grau ponderado

Nome do Ponto	Grau Ponderado
R. Rodovalho Júnior, 637	50.0
Av. Pe. José Maria	42.0
Prudente Correia	39.0
Benedito Chaves	39.0
Estr. De Itapecerica, 3646	38.0
Terminal João Dias	37.0
Av. Interlagos, 6422	37.0
Parada 2 - Ferreira De Araújo	36.0
Av. Mal. Tito, 1206	36.0
Jaquirana B/C	34.0

Ao analisar individualmente cada um dos pontos da tabela 4 é possível observar que são pontos próximos de terminais de ônibus ou que estão em áreas de grande fluxo de transporte como avenidas, o bloqueio de qualquer um desses pontos pode causar uma mudança de rota de uma grande quantidade de linhas de uma vez o que pode por consequências afetar as experiências dos usuários e piorar o trânsito local, desse nó de uma maneira muito mais significativa do que em outras linhas com grau ponderado mais baixo.

O gráfico da figura 5 também traz outras informações importantes da rede, demonstrando que a rede é fortemente assimétrica, visto que uma grande concentração de nós está nos graus ponderados baixos, o que indica que em muitos pontos de ônibus existem poucas viagens passando.

Figura 5: Histograma de Distribuição de Grau Ponderado



Fonte: Elaboração Própria

Outra característica a se destacar é que a concentração de diversas linhas passando em um único ponto pode não ser uma boa característica da rede, pois pode gerar consequências negativas como: confusão dos passageiros devido a quantidade de opções, uma superlotação de usuários em um mesmo local, formação de filas extensas de ônibus nas vias e o aumento da probabilidade de congestionamentos na área. Também existem aspectos da qualidade física daquele ponto para os usuários do sistema, como foi observado na figura 5, que apresenta o ponto da R. Rodovalho Júnior, 637. Que conta com a métrica com o valor de 25 nos graus ponderado médio de entrada e de saída, o que indica por volta de 25 linhas passando por esse local. Tal ponto é apresentado como um ponto simples, sem estrutura para os passageiros descobrirem quais linhas estão chegando e nem ao menos bancos para os usuários, o que afeta a qualidade do uso do sistema.

Figura 6: Ponto da R. Rodovalho Júnior, 637.



Fonte: Google Street View.

Ao analisarmos o ranking de centralidade de intermediação que é a métrica que quanto maior o seu valor, melhor é a sua conectividade entre nós da rede, agindo como uma ponte ao longo do caminho mais curto entre dois outros nós da rede, conforme a tabela 5.

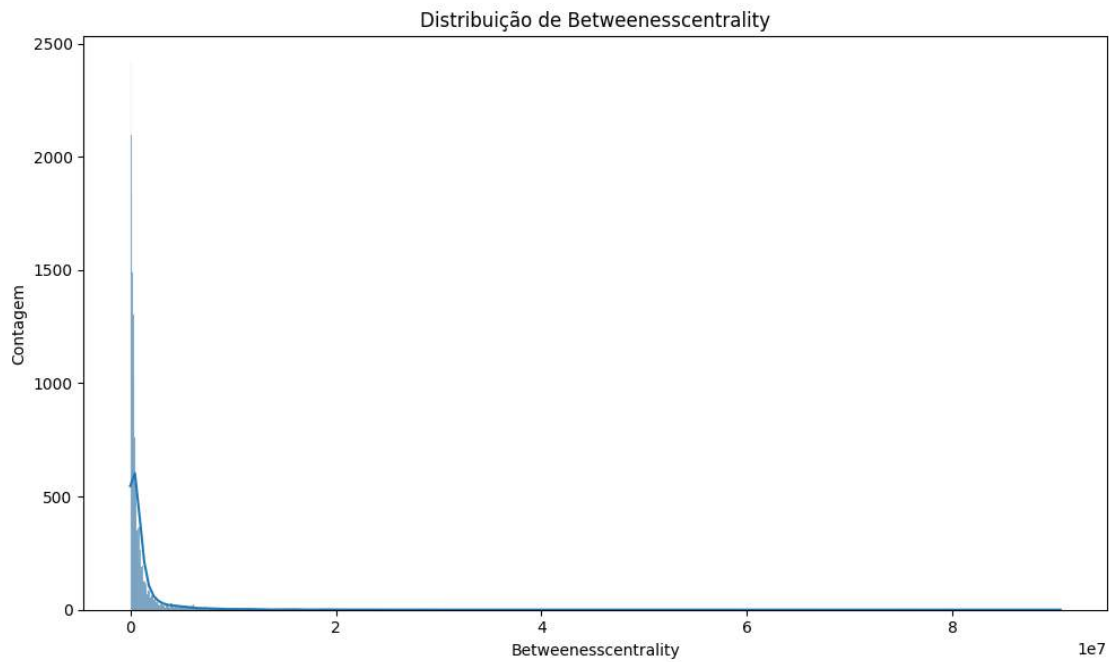
Tabela 5: Ranking do top 5 de maiores centralidade de intermediação

Nome do Ponto	Centralidade de Intermediação	Grau ponderado
Parada Metrô Tatuapé	90497185	21
Terminal Metrô Belém (sul) Pl. D	83294161	21
R. Melo Freire	79563923	24
Av. Cde. De Frontin, 542	77550270	24
Av. Alcântara Machado, 351	76001172	22

Caso ocorra alguma intercorrência nos nós de ranking mais alto de centralidade de intermediação os danos na rede são maiores pois diferente de uma métrica de grau ponderado, a remoção desses nós tem maior poder de danos pois são nós que atuam diretamente com a conectividade global, a remoção deles podem fragmentar a rede afetar caminhos críticos.

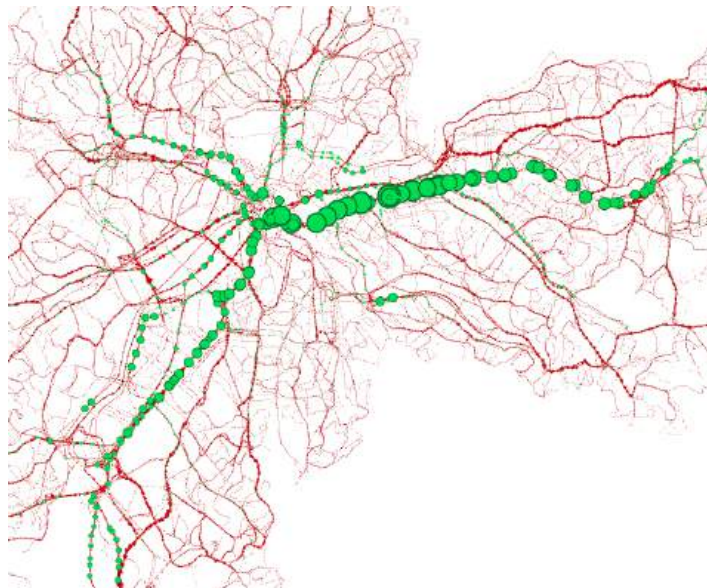
O gráfico da figura 7 mostra a distribuição em formato de cauda pesada da centralidade de intermediação na rede, onde poucos nós são responsáveis pela conectividade global, o que indica que a rede tem vias centrais muito importantes que são apresentadas em destaque por tamanho na figura 8.

Figura 7: Distribuição da medida de Centralidade de Intermediação.



Fonte: Elaboração Própria.

Figura 8: Rede complexa com destaque para vértices com alto grau de intermediação.

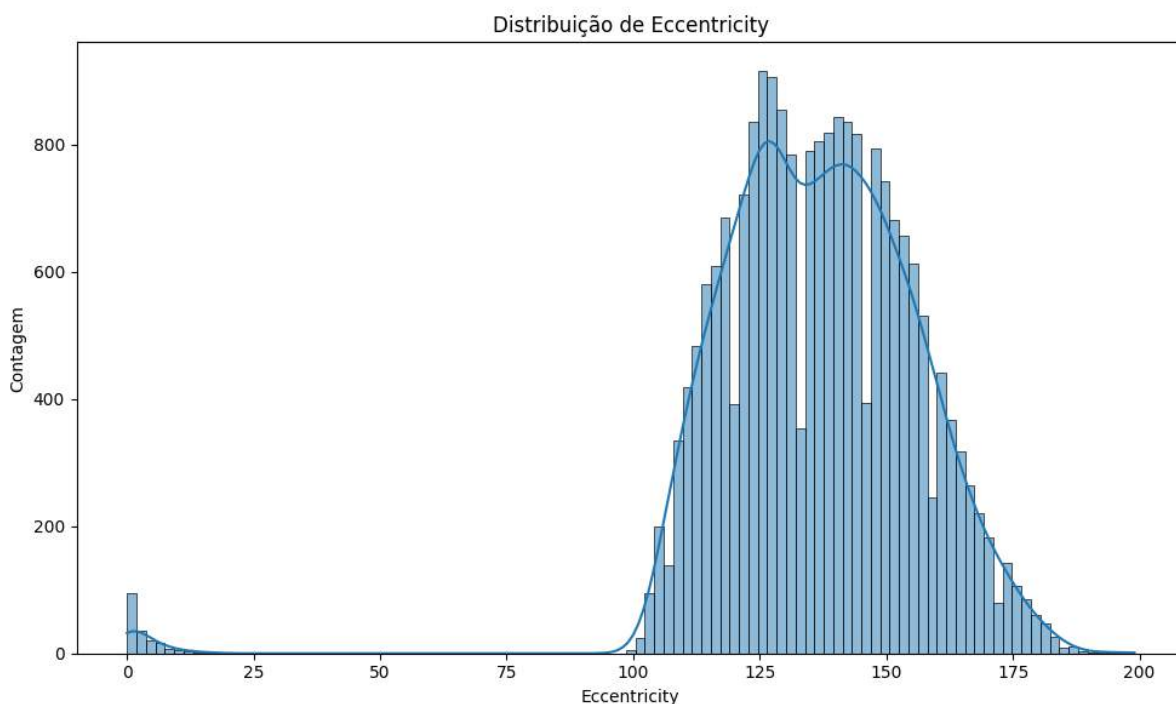


Fonte: Elaboração Própria.

A excentricidade é uma métrica que mede a distância máxima de um nó para outro nó qualquer na rede. No contexto deste trabalho, indica o quão longe um ponto de ônibus está do ponto mais distante dele no sistema. Nós que tem baixa excentricidade estão próximos de todos os outros, enquanto os nós com alta tem pelo menos um ponto muito distante deles na rede.

A distribuição do gráfico figura 9, que é o resultado aplicação da métrica de excentricidade na rede indica que a maioria dos pontos tem como valor entre 100 e 175, com uma distribuição normal e relativamente equilibrada. A falta de muitos pontos com excentricidade baixa ou extremamente alta, indica que a rede tem uma estrutura geral uniforme. No entanto, a presença de alguns pontos com excentricidade muito baixa pode indicar áreas centrais bem conectadas, enquanto os de excentricidade alta indicam regiões fora do centro expandido, conforme indica a figura 10.

Figura 9: Distribuição da medida de excentricidade.



Fonte: Elaboração Própria.

Figura 10: Distribuição da medida de excentricidade nos dois pontos da cidade.

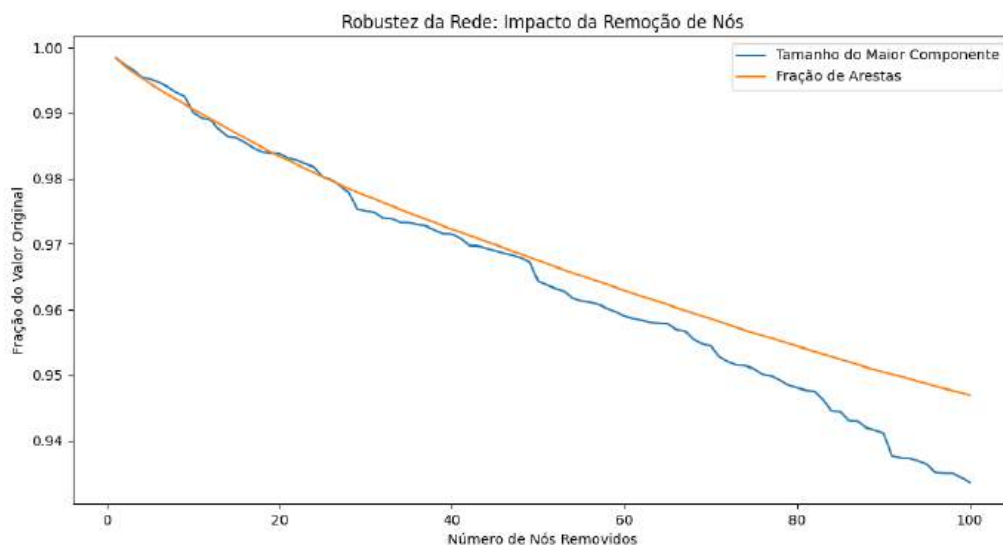


Fonte: Elaboração Própria.

A Robustez, no contexto de redes complexas, refere-se à capacidade de um sistema de manter sua funcionalidade e estrutura geral quando submetido a perturbações. No gráfico da figura 11 é possível observar uma notável resiliência do sistema. Pois mesmo após a remoção dos 100 nós mais importantes, que provavelmente podem representar terminais ou hubs de transferências, o maior componente conectado da rede se manteve com mais de 93% do seu tamanho original, indicando uma estrutura de rede bem distribuída e interconectada, capaz de suportar falhas significativas sem comprometer a funcionalidade global.

A degradação gradual e consistente da rede, sem quedas abruptas na conectividade do maior componente, indica que o sistema não depende excessivamente de poucos pontos críticos. É uma característica crucial para uma metrópole do tamanho de São Paulo, que assegura que o transporte público pode continuar transitando pela cidade e seus pontos mesmo diante de interrupções de áreas importantes. No entanto, isso ignora outros aspectos, como a sobrecarga de outros pontos e a possível piora do trânsito em outras vias.

Figura 11: Robustez da rede.



Fonte: Elaboração Própria.

4.2 Rede de similaridade de linhas

4.2.1 Análise da rede

Com a estratégia de conectar as linhas pelas suas similaridades de pontos de ônibus, a rede complexa das linhas de ônibus contou com um total de 1280 linhas de ônibus e 82514 arestas no período de maio de 2024.

Onde, de acordo com a tabela 6, cerca de 13 linhas são iguais em seus pontos de ônibus, e ,por volta de pelo menos 84% das linhas são no mínimo metade similar a outras linhas entre seus pontos em comum.

Tabela 6: Quantidade de vértices de acordo com a similaridade de arestas

Porcentagem de similaridade	Quantidade de linhas
100%	13
90%	278
75%	601
50%	1081

Ao executar uma query no *Neo4j* para verificar a existência de linhas isoladas, que são linhas que não tem conexão com a maior componente conexa, apenas 2 linhas de ônibus foram encontradas, o que é por volta de 0,001% das linhas. Analisando essas linhas não conectadas para entender o motivo de não estarem no sistema , as linhas encontradas que são a 4055-10 que é uma linha interna e isolada de um bairro e a outra, é a linha 626A-10 que é a linha que representa a travessia pela represa Billings.

A tabela 7 apresenta dados a respeito dessa rede onde:

Tabela 7: Métricas da rede de linhas de ônibus

Métrica	Valor
Grau médio	64,104
Grau médio ponderado	8,055
Diâmetro da rede	6
Modularidade	0,7
Coeficiente de Agrupamento Médio(direcionado)	0,593

Grau médio: Um grau médio alto indica que, em média, cada linha de ônibus está conectada a aproximadamente 64 outras linhas. Isso sugere uma rede muito densa e bem conectada, com muitas opções de conexões diretas entre as linhas de ônibus.

Grau médio ponderado: Apesar do alto grau médio, o grau médio ponderado de 8,055 sugere que, embora existam muitas conexões, elas têm, em média, uma similaridade baixa. Isso pode indicar que, apesar da alta conectividade, as linhas de ônibus não passam com tanta frequência por todas essas conexões.

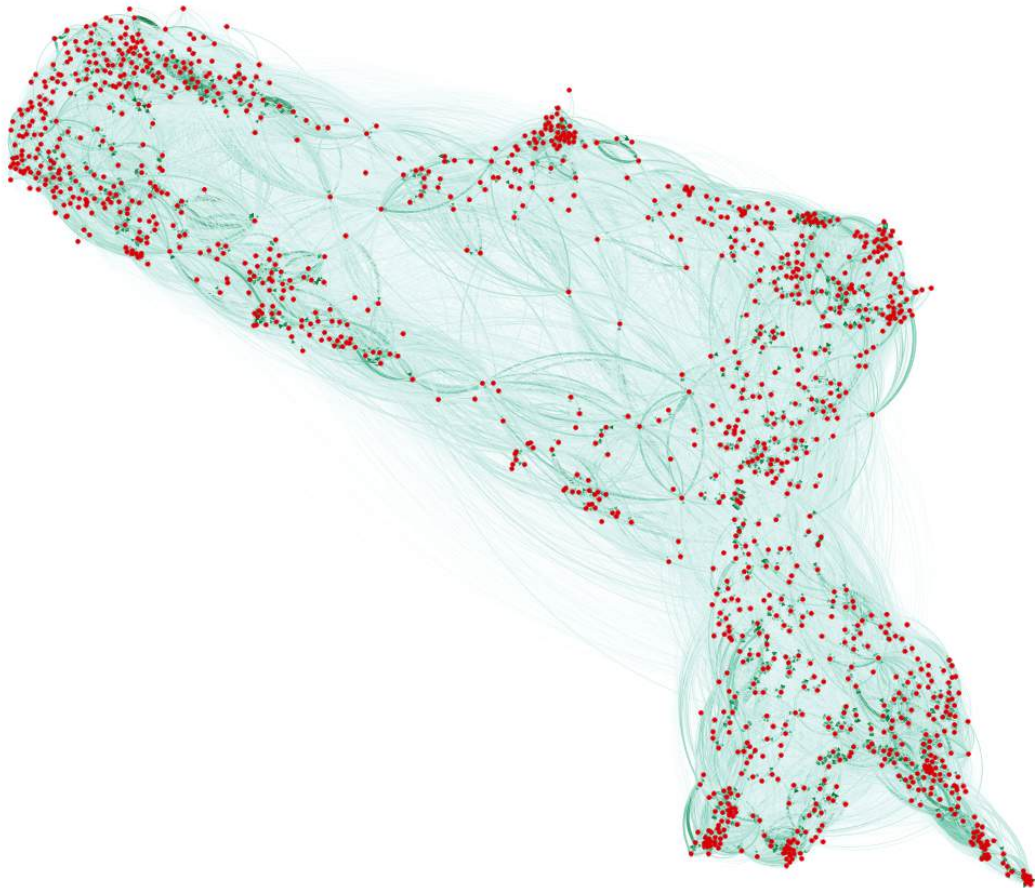
Diâmetro da rede: Um diâmetro de 6 representa o caminho mais longo entre quaisquer duas linhas de ônibus na rede, indicando que a rede de transporte é compacta e que é possível ir de um extremo ao outro da cidade com um número relativamente pequeno de transferências.

Modularidade: Esse valor de modularidade indica uma estrutura de comunidades bem definida na rede de transporte, com grupos de linhas de ônibus mais densamente conectados entre si do que com o resto da rede conforme mostra a figura 9.

Coeficiente de Agrupamento Médio (direcionado): Esse valor moderadamente alto sugere uma tendência significativa das linhas de ônibus vizinhas se conectarem

entre si, facilitando o deslocamento local dentro das comunidades com poucas transferências.

Figura 9: Rede complexa das linhas de ônibus.



Fonte: Elaboração Própria.

Utilizando a *ForceAtlas2* para distribuição visual do grafo na figura 9, embora exista grande quantidade de arestas, é possível observar a formação de comunidades de linhas de ônibus que se conectam fortemente entre si, em destaque as comunidades formadas tem uma similaridade maior visto que a configuração visual utiliza do peso das arestas para definir a sua densidade, a representação visual nesse caso não ajuda muito a entender a rede se existe um padrão.

4.2.2 Análise dos dados

Utilizando dados de abril de 2024, os maiores valores obtidos presentes na tabela 8 servem como referência para as próximas análises deste capítulo.

Tabela 8: Maiores valores de cada coluna

Métrica	Valor	Nome da Linha
Quantidade máxima de passageiros	890509	Jd. Miriam - Itaim Bibi
Quantidade máxima de pontos	216	Ipiranga - Rio Pequeno
Grau ponderado	50.2	Term. Parelheiros - Metrô VI. Mariana
Grau	368	Term. Grajaú - Metrô Brás

Onde grau é quantas conexões aquela linha tem com outras linhas e grau ponderado é essa conexão levando em consideração a quantidade de pontos em comum, a diferença de valores máximos corroboram com os resultado obtidos na etapa anterior.

A tabela 9 é um ranking com o foco na quantidade de passageiros por linha de ônibus da cidade.

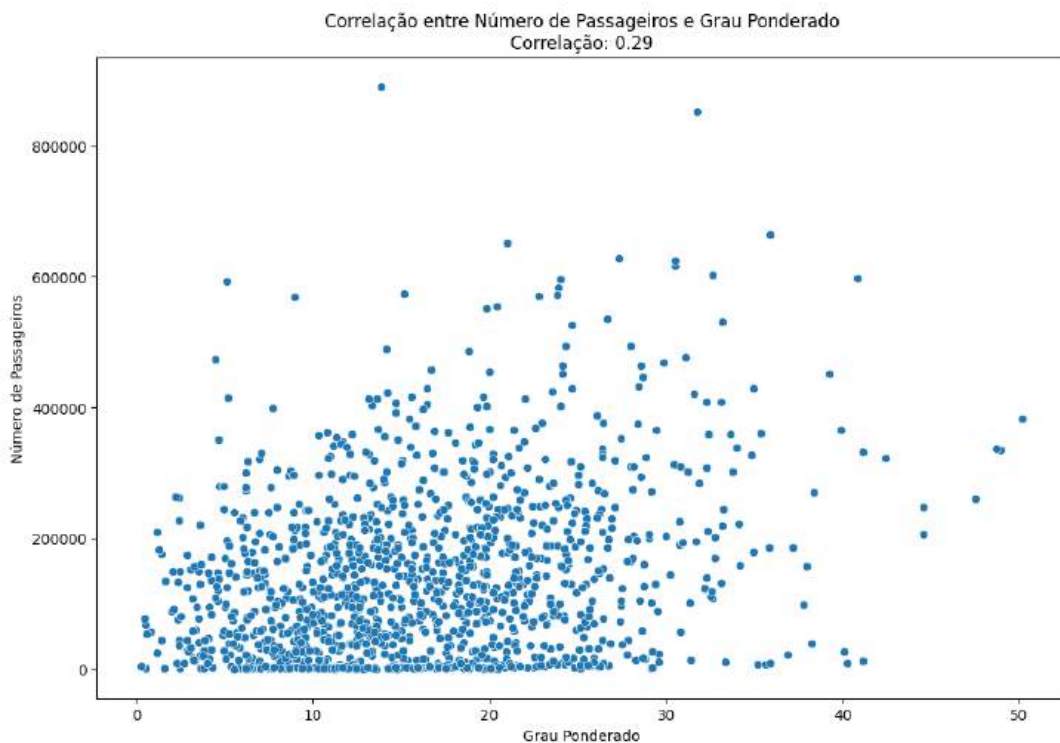
Tabela 9: Top 5 por quantidade de passageiro

Nome da Linha	Quantidade de Passageiro	Quantidade de Pontos	Grau Ponderado	Grau
Jd. Miriam - Itaim Bibi	890509	64	13.9	80
Term. Parelheiros - Term. Sto. Amaro	851603	94	31.7	238
Term. Jd. Ângela - Metrô Sta. Cruz	664024	76	35.9	236
E.t. Itaquera - Term. Pq. D. Pedro li	651478	56	21	158
Term. Capelinha - Lgo. São Francisco	627517	112	27.3	208

Algo a se observar é que algo em comum entre todas essas linhas é que 4 das 5 maiores linhas por total de passageiros têm como origem a zona sul da cidade se conectando o mais próximo à região central.

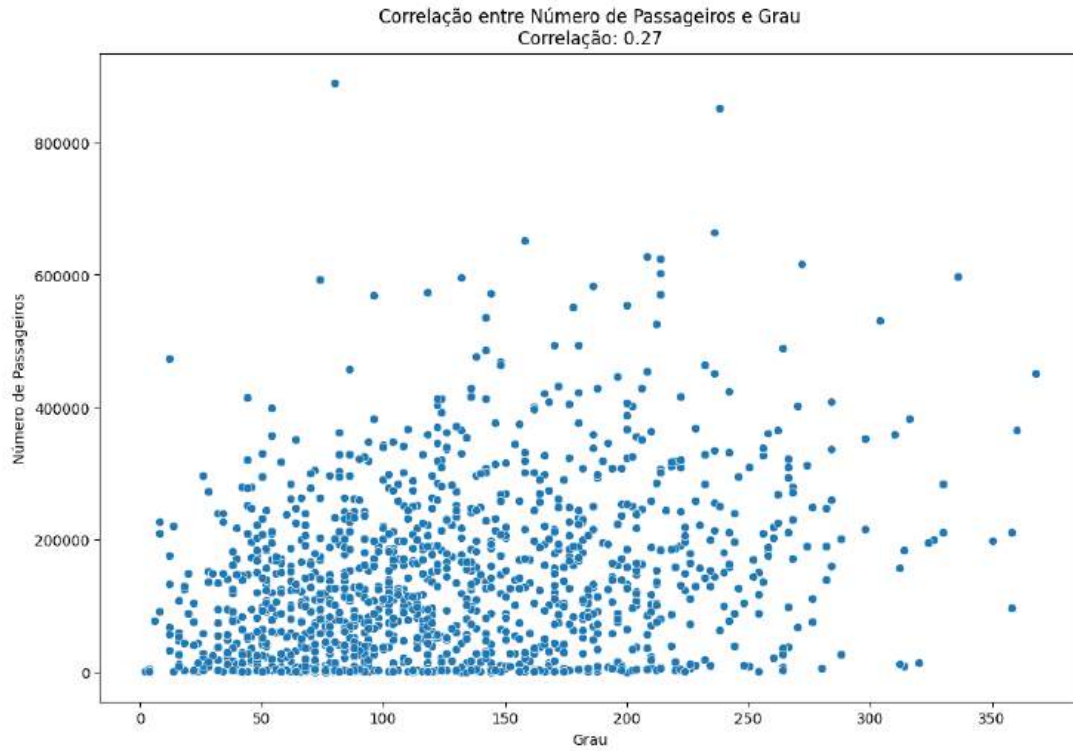
Porém, observa-se que nenhuma das 5 maiores linhas em termos de número de passageiros se aproxima dos valores máximos das métricas analisadas (grau, grau ponderado e total de pontos). Estendendo essa observação para todas as linhas, constata-se que uma das razões para a falta de relações claras entre os dados na tabela 9 é a natureza das correlações. As figuras 10, 11 e 12 mostram correlações positivas, porém fracas a moderadas, entre o número de passageiros e o grau (0.27), grau ponderado (0.29), e total de pontos (0.36), respectivamente. Isso indica que, embora essas métricas tenham alguma influência, não são determinantes substanciais do fluxo de passageiros. A dispersão significativa observada nos gráficos reforça a necessidade de considerar fatores adicionais para explicar adequadamente a complexidade da rede de similaridade entre as linhas de transporte.

Figura 10: Correlação entre número de passageiro e Grau



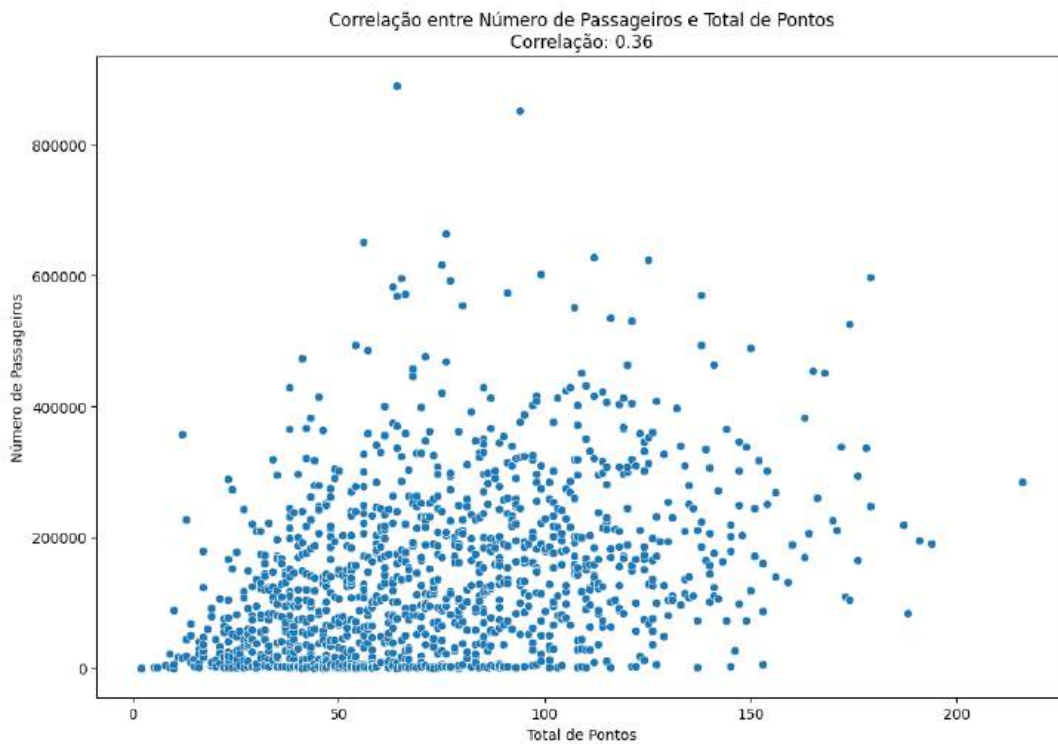
Fonte: Elaboração Própria.

Figura 11: Correlação entre número de passageiro e Grau Ponderado



Fonte: Elaboração Própria.

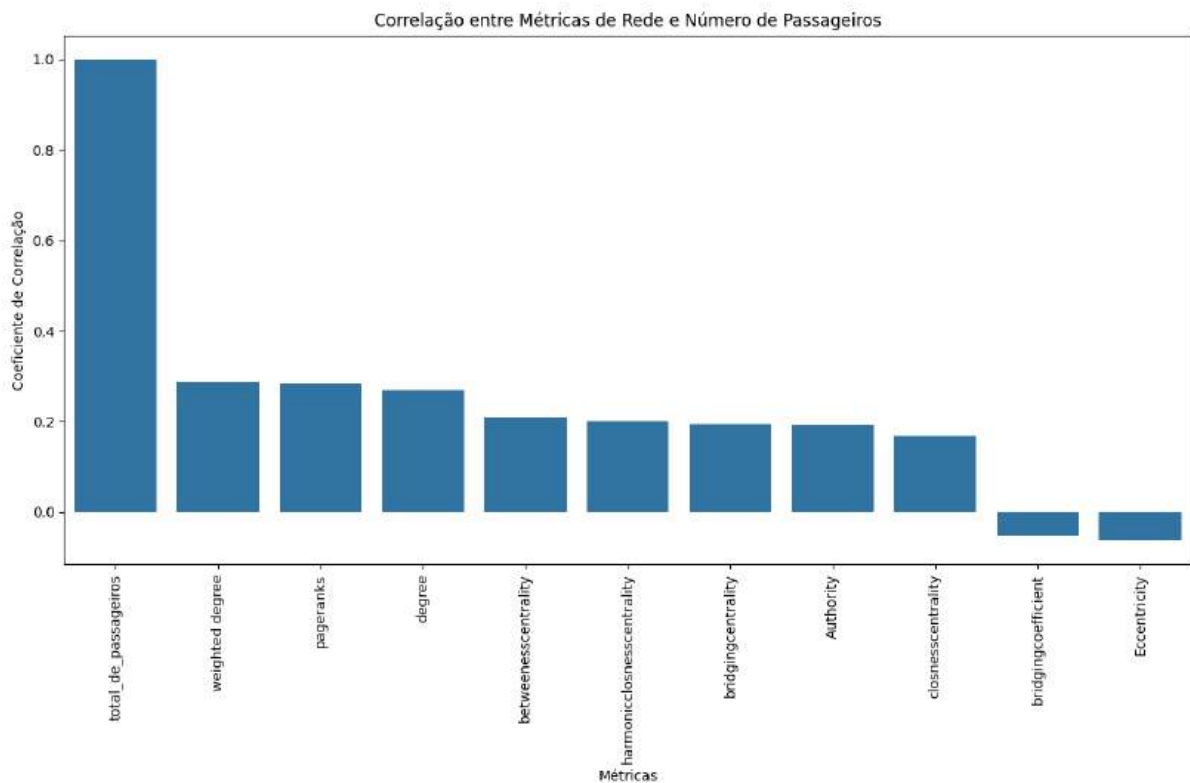
Figura 12: Correlação entre número de passageiro e Total de Pontos



Fonte: Elaboração Própria.

Se estendermos a observação das correlações dos nós com o total de passageiros para outras métricas de rede complexas, como pagerank, centralidade de intermediação (*betweenness centrality*), entre outras, notamos que estas correlações permanecem fracas a moderadas. A figura 13 mostra que, após o grau ponderado com correlação de aproximadamente 0,3, o *pagerank* apresenta uma correlação similar, seguido pelo grau (*degree*). Métricas como centralidade de intermediação, centralidade de proximidade harmônica, e autoridade (*authority*) exibem correlações ainda mais fracas, variando entre 0,2 e 0,1. E por fim, o coeficiente de ponte (*bridging coefficient*) e a excentricidade (*eccentricity*) mostram correlações próximas a zero ou ligeiramente negativas. Essas observações reforçam a ideia de que, embora estas métricas de rede complexa possam oferecer algum insight sobre o fluxo de passageiros, nenhuma delas individualmente apresenta uma forte correlação com o total de passageiros, reforçando que outros fatores não capturados por estas métricas provavelmente desempenham um papel significativo na determinação do número de passageiros em cada linha.

Figura 13: Correlação entre total de passageiros e Métricas de rede Complexa

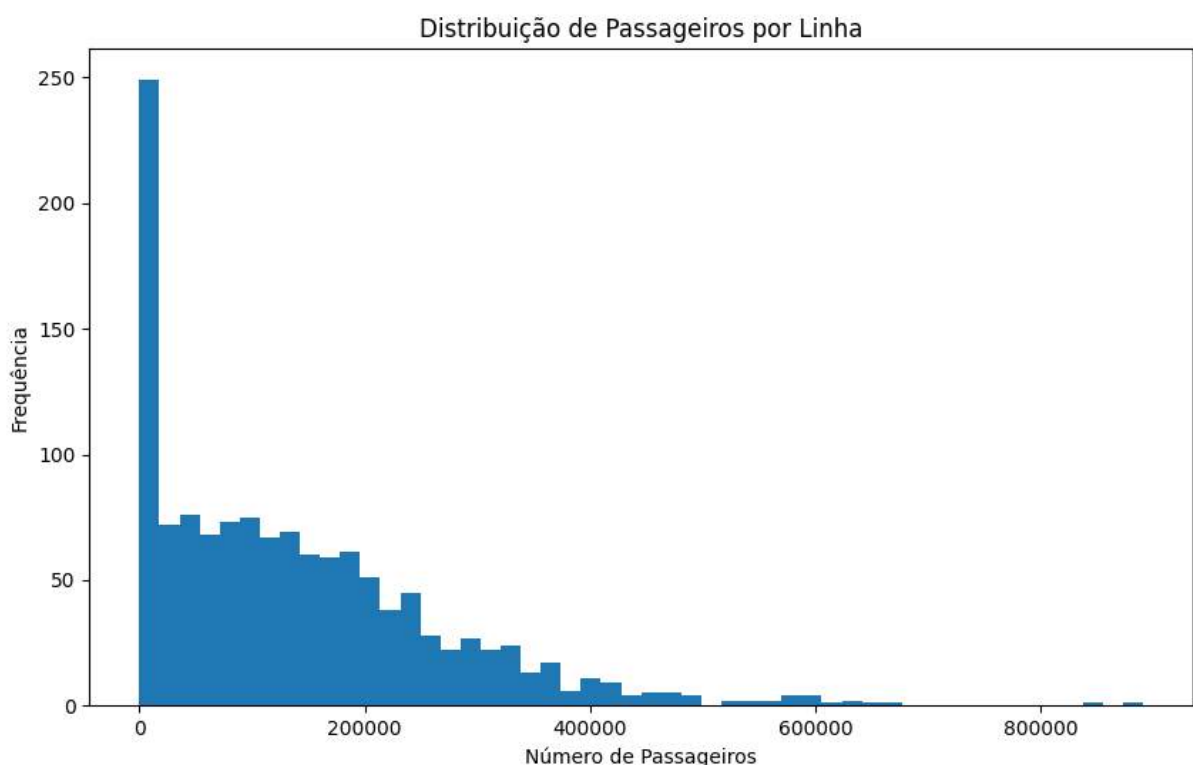


Fonte: Elaboração Própria.

Concluindo as análises relacionados a métrica de total de passageiros, se pegarmos as distribuições do total de passageiros por suas respectivas linhas, conforme a figura 14, observamos um padrão altamente assimétrico e com características de uma possível lei de potência, onde a distribuição demonstra que grande parte das linhas tem poucos passageiros, enquanto poucas linhas são responsáveis por grande parte deles. Tal padrão sugere uma desigualdade desproporcional de uma carga de passageiros na rede de transporte, uma análise mais aprofundada dessa disparidade pode indicar uma necessidade de reestruturação das linhas.

É portanto interessante entender a necessidade dessas linhas com poucos passageiros e até mesmo observar a necessidade de adequá-las para pulverizar os passageiros de linhas com muitos passageiros, a fim de trazer uma melhor qualidade do sistema para os passageiros, com linhas menos lotadas e mais eficientes.

Figura 14: Distribuição de passageiros por linha



Fonte: Elaboração Própria.

5 Conclusão

Conclui-se que as redes complexas são uma ferramenta poderosa e versátil para analisar sistemas interconectados, como demonstrado na análise da rede de pontos de ônibus de São Paulo. Esta abordagem permitiu uma compreensão profunda da estrutura e dinâmica do sistema de transporte, revelando padrões de conectividade e identificando pontos críticos na rede. No entanto, a metodologia utilizada, que considerou pontos individuais em vez de agrupamentos de proximidade de pontos (como terminais), pode ter limitado alguns aspectos da análise. Esta observação ressalta a importância de adaptar a granularidade da modelagem às características específicas do sistema.

No caso da análise da rede de similaridades entre linhas, embora a abordagem de redes complexas não tenha fornecido insights diretos sobre as razões aos padrões de similaridade observados, ela revelou características importantes do sistema. A análise baseada no total de passageiros expôs anomalias significativas na distribuição da carga do sistema de ônibus, indicando uma possível necessidade de reestruturação. A falta de correlações fortes entre as métricas de rede complexa e o número de passageiros sugere que fatores externos à estrutura topológica da rede, como padrões demográficos e estruturais, podem desempenhar papéis cruciais na determinação dos fluxos de passageiros. Esta constatação enfatiza a necessidade de uma abordagem holística que integre dados de transporte com informações socioeconômicas e urbanas para uma compreensão mais completa do sistema. Além disso, a limitação dos dados fornecidos pela SPTrans, particularmente a falta de granularidade nos dados de passageiros, destaca a importância de dados mais detalhados e precisos para análises futuras, que poderiam revelar padrões mais sutis e oferecer insights mais acionáveis para a otimização do sistema de transporte.

5.1 Trabalhos futuros

Para um trabalho futuro pode-se considerar para a análise a inclusão de linhas de ônibus da Empresa Metropolitana de Transportes Urbanos, conhecida pela sigla EMTU, que engloba diversas áreas da região metropolitana de São Paulo além da capital, com ônibus que passam por diversos municípios ao redor da cidade. Também há a possibilidade da adição das linhas de transporte ferroviário (metrô e CPTM), que pode trazer valor para uma análise mais robusta do modal de transporte público existente no município.

Outra alternativa de pesquisa é analisar a rede de maneira temporal, observando principalmente as alterações das linhas de ônibus através do arquivo GTFS durante um período, a fim de entender a evolução da rede conforme a evolução da cidade e outros fatores.

Bibliografia

- [1] BURGUETTI, Renata. **Alguns tipos de grafos e aplicações**. 2022. 81f. Dissertação (mestrado em Matemática) - Universidade Estadual de Maringá, Dep. de Matemática, Programa de Mestrado Profissional em Matemática em Rede Nacional, 2022.
- [2] BARABÁSI, Albert-László et al. **Network science**. Cambridge university press, 2016.
- [3] SOH, Harold et al. (2010). **Weighted complex network analysis of travel routes on the Singapore public transportation system**. Physica A: Statistical Mechanics and its Applications. 389. 5852-5863. 10.1016/j.physa.2010.08.015.
- [4] SHANMUKHAPPA, Tanuja e HO, Ivan Wang-Hei e TSE, Chi e LEUNG, Kin. (2019). **Recent Development in Public Transport Network Analysis From the Complex Network Perspective**. IEEE Circuits and Systems Magazine. 19. 4880-4897. 10.1109/MCAS.2019.2945211.
- [5] GOMIDE, Alexandre de Ávila. **Mobilidade urbana, iniquidade e políticas sociais**. Políticas sociais: acompanhamento e análise, 12, 242-250.
- [6] MOREIRA, Thiago Alves. **O DESENVOLVIMENTO DAS REDES DE METRÔ**. 2023. 73f. Dissertação (mestrado em planejamento e projeto urbano) - Faculdade de Engenharia da Universidade do Porto e Faculdade de Arquitetura Universidade do Porto, Porto, Portugal, Dep. de Engenharia Civil, Programa de Mestrado em Planejamento Urbano, 2023.
- [7] DAFRE, Joana e TATEISHI, Angel Akio (2020). **Redes Complexas aplicadas às Redes de transporte Urbano - Coleta de Dados e Caracterização**. XXIV Seminário de Iniciação Científica e Tecnológica da UTFPR.
- [8] GONÇALVES, Eulaliane e EISENCRAFT, Marcio. (2019). **Centralidade de Intermediação Aplicada a uma Malha de Transportes sobre Trilhos**.
- [9] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA . Censo Brasileiro de 2022. São Paulo: IBGE, 2022.

- [10] DAUER, Letícia, Reclamações de passageiros sobre ônibus em SP aumentam 35%; veja linhas campeãs. **R7**, São Paulo, 08 ago. 2023. Disponível em: <https://noticias.r7.com/sao-paulo/reclamacoes-de-passageiros-sobre-onibus-e-m-sp-aumentam-35-veja-linhas-campeas-07082023>. Acesso em: 02 abr. 2024.
- [11] Pesquisa aponta que trabalhadores são maioria dos passageiros de ônibus em São Paulo. **SECOM - Prefeitura da Cidade de São Paulo**, São Paulo, 27 de out. 2023. Disponível em: <https://www.capital.sp.gov.br/w/noticia/pesquisa-aponta-que-trabalhadores-sao-o-maioria-dos-passageiros-de-onibus-em-sao-paulo>. Acesso em: 02 de abr. de 2024.
- [12] ARAÚJO, M. R. M. de ., OLVEIRA, J. M. de ., JESUS, M. S. de ., SÁ, N. R. de ., SANTOS, P. A. C. dos ., e LIMA, T. C.. (2011). **Transporte público coletivo: discutindo acessibilidade, mobilidade e qualidade de vida**. *Psicologia & Sociedade*, 23(3), 574–582.
- [13] OLIVEIRA, Mariana, 8 dados mostram o impacto da mobilidade urbana nas nossas vidas. **Estadão**, São Paulo, 10 set. 2021. Disponível em: <https://mobilidade.estadao.com.br/mobilidade-para-que/dia-a-dia/8-dados-mos-tram-o-impacto-da-mobilidade-urbana-nas-nossas-vidas/>. Acesso em: 20 abr. 2024.
- [14] SPTRANS. São Paulo Transporte S/A. Página inicial. Disponível em: <https://www.sptrans.com.br/sptrans>. Acesso em: 02 de abr. de 2024.