

Universidade de São Paulo  
Escola de Artes, Ciências e Humanidades

SANDRO FERREIRA SOUSA

**Estudo das propriedades e robustez da rede de transporte  
público de São Paulo**

São Paulo

2016



SANDRO FERREIRA SOUSA

**Estudo das propriedades e robustez da rede de transporte público de São Paulo**

Versão corrigida

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção de título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos.

Área de concentração: Sistemas Complexos

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 08 de Junho de 2016. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Camilo Rodrigues Neto

São Paulo

2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

### CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

Sousa, Sandro Ferreira

Estudo das propriedades e robustez da rede de transporte público de São Paulo / Sandro Ferreira Sousa ; orientador, Camilo Rodrigues Neto. – São Paulo, 2016.

87 p. : il

Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Modelagem de Sistemas Complexos, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo. Versão corrigida.

1. Transporte urbano - São Paulo (SP) - Simulação computacional. 2. Transporte público - São Paulo (SP). 3. Redes complexas. 4. Sistemas dinâmicos. I. Rodrigues Neto, Camilo, orient. II. Título

CDD 22.ed.– 388.4011098161

Dissertação de autoria de Sandro Ferreira Sousa, sob o título “**Estudo das propriedades e robustez da rede de transporte público de São Paulo**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos, na área de concentração Sistemas Complexos, aprovada em 08 de Junho de 2016 pela comissão julgadora constituída pelos doutores:

**Prof. Dr. Camilo Rodrigues Neto**

Presidente

Instituição: EACH/USP

**Prof. Dr. Fernando Fagundes Ferreira**

Instituição: EACH/USP

**Prof.<sup>a</sup> Dr.<sup>a</sup> Carolina Moutinho Duque de Pinho**

Instituição: CECS/UFABC

**Prof. Dr. Fabiano Lemes Ribeiro**

Instituição: DFI/UFLA

São Paulo

2016



# Agradecimentos

Agradeço profundamente ao meu orientador, Prof. Dr. Camilo Neto, ao Prof. Dr. Fernando Fagundes e ao colega Rolf Simões pelas inúmeras contribuições. Estendo meus agradecimentos aos outros professores e amigos que encontrei pelo caminho. Agradeço, ainda, à CAPES pelo fomento e apoio à produção dessa pesquisa.





# Resumo

SOUSA, Sandro. **Estudo das propriedades e robustez da rede de transporte público de São Paulo**. 2016. 87 p. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2016. Versão original.

Sistemas Complexos são característicos por possuir uma rede interna representando o relacionamento estrutural entre seus elementos e uma forma natural de interpretar essa interação é através de um grafo. Neste trabalho, o sistema de transporte público urbano de São Paulo é reinterpretado de forma acoplada (ônibus e metrô juntos) como uma rede complexa, abstraindo detalhes operacionais e focando na conectividade. Pelo grafo empiricamente gerado, é feita uma caracterização estatística nas métricas de redes complexas, onde diferentes valores de raio de distância são usados para agrupar pontos e estações próximas que antes se apresentavam desconectados. Esse agrupamento pode ser interpretado como uma ferramenta de política pública, representando a disposição do usuário em se locomover ao ponto mais próximo para acessar o transporte. O processo mostrou que aumentar essa disposição gera grande redução na distância e número de passos entre ônibus, trens e linhas de metrô para atingir todos os destinos da rede. É utilizado um modelo exploratório que testa a robustez da rede aleatoriamente, deterministicamente e probabilisticamente tendo como alvo pontos e linhas. De acordo com os raios de agrupamento, definido como disposição, diferentes valores de fragmentação foram obtidos diante dos ataques simulados. Esses resultados suportam duas principais características observadas na literatura de redes deste tipo: possuem um elevado grau de robustez à falhas aleatórias, mas são vulneráveis a ataques tendo como alvo nós ou links importantes.

**Palavras-chaves:** Sistemas Complexos. Redes Complexas. Ciência de Redes. Rede de Transporte Público. Robustez. Vulnerabilidade à ataques.



# Abstract

SOUSA, Sandro. **Study of properties and robustness of the public transport network of São Paulo**. 2016. 87 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2016. Version original.

Complex systems are characteristic by having an internal network representing the structural relationship between its elements and a natural way to interpret this interaction is through a graph. In this work, the urban public transport system of São Paulo is reinterpreted as a coupled (bus and subway) complex network, bypassing operational details and focusing on connectivity. Using the empirically generated graph, a statistical characterisation is made by network metrics where different radius values are used to group nearby stops and stations that were disconnected before. That can be interpreted as a public policy tool, representing the user's willingness to get around the nearest point to access transportation. This process has shown that increasing this willingness generates great reduction in the distance and in the number of jumps between buses, trains and subways lines to achieve all the network destinations. An exploratory model is used to test the robustness of the network by randomly, deterministically and preferentially targeting the stops and service lines. According to the grouping radius, aka willingness, different fragmentation values were obtained under attack simulations. These findings support two main characteristics observed in such networks literature: they have a high degree of robustness to random failures, but are vulnerable to targeted attacks.

**Key-words:** Complex Systems. Complex Networks. Network Science. Public transport network. Robustness. Attack vulnerability.



# Lista de ilustrações

Figura 1 – Emissões de $CO_2$ em São Paulo . . . . .	26
Figura 2 – Modelo computacional como uma função . . . . .	31
Figura 3 – Diagrama do fluxo do processo de agrupamento e simulação de ataques . . . . .	33
Figura 4 – Pontes de Königsberg e modelo proposto por Euler . . . . .	35
Figura 5 – Grafo dirigido e com múltiplos <i>links</i> . . . . .	36
Figura 6 – Grafo regular circular conectado . . . . .	38
Figura 7 – Grafo aleatório Erdős e Rényi . . . . .	38
Figura 8 – Grafo pequeno-mundo Watts e Strogatz . . . . .	39
Figura 9 – Processo de reconexão em um Grafo pequeno-mundo . . . . .	39
Figura 10 – Grafo livre de escala Barabasi Albert . . . . .	41
Figura 11 – Modelo de parte da rede de metrô como um grafo . . . . .	46
Figura 12 – Arquivo <i>stop_times</i> no padrão GTFS . . . . .	48
Figura 13 – Ciclo de uma viagem não fechado . . . . .	48
Figura 14 – Algoritmo de agrupamento por raio . . . . .	49
Figura 15 – Representação no espaço L e P . . . . .	51
Figura 16 – Modelo simbólico dos cenários de ataque no espaço L . . . . .	54
Figura 17 – Resultados empíricos: Mapa com componentes desconectados . . . . .	55
Figura 18 – Resultados empíricos: Comportamento de nós e <i>links</i> em função de $\rho$ . . . . .	56
Figura 19 – Resultados empíricos: Grau médio e mediana do nó em função de $\rho$ . . . . .	57
Figura 20 – Resultados empíricos: Clusters da rede em função de $\rho$ . . . . .	58
Figura 21 – Resultados empíricos: Caminhos da rede em função de $\rho$ . . . . .	59
Figura 22 – Resultados empíricos: Histograma dos caminhos médios . . . . .	60
Figura 23 – Resultados empíricos: Eficiência das redes em função de $\rho$ . . . . .	61
Figura 24 – Resultados empíricos: Distribuição de grau dos nós em escala logarítmica . . . . .	62
Figura 25 – Resultados empíricos: Grafo da rede de São Paulo em $\rho = 65$ . . . . .	66
Figura 26 – Resultados ataques: Impacto nos <i>links</i> ao remover nós . . . . .	67
Figura 27 – Resultados ataques: Impacto no grau máximo ao remover nós . . . . .	68
Figura 28 – Resultados ataques: Impacto nos componentes ao remover nós . . . . .	69
Figura 29 – Resultados ataques: Impacto no caminho médio ao remover nós . . . . .	70

Figura 30 – Resultados ataques: Impacto no diâmetro ao remover nós . . . . .	71
Figura 31 – Resultados ataques: Nós necessários para dividir a rede em partes . . . . .	72
Figura 32 – Resultados ataques: Impacto no grau máximo ao remover <i>links</i> . . . . .	74
Figura 33 – Resultados ataques: Impacto nos componentes ao remover <i>links</i> . . . . .	75
Figura 34 – Resultados ataques: Impacto no caminho médio ao remover <i>links</i> . . . . .	76
Figura 35 – Resultados ataques: Impacto no diâmetro ao remover <i>links</i> . . . . .	77
Figura 36 – Resultados ataques: Grau máximo nos cenários nó e <i>link</i> nas estratégias alvo determinístico e probabilístico na rede $\rho = 65$ . . . . .	77
Figura 37 – Resultados ataques: Componentes, caminho médio e diâmetro nos cená- rios nó e <i>link</i> na estratégia alvo determinístico na rede $\rho = 65$ . . . . .	78
Figura 38 – Resultados ataques: Componentes, caminho médio e diâmetro nos cená- rios nó e <i>link</i> na estratégia alvo probabilístico na rede $\rho = 65$ . . . . .	79

# Lista de tabelas

Tabela 1 – Tabela com os resultados obtidos no <i>fit</i> da distribuição de grau dos nós nas 5 redes $\rho$ analisadas. . . . .	63
Tabela 2 – Tabela com os resultados do teste de pequeno-mundo para as 5 redes $\rho$ comparadas com redes aleatórias do tipo Erdős e Rény. . . . .	64
Tabela 3 – Tabela com <i>ranking</i> dos 10 pontos de maior grau e sua descrição/localidade correspondente na rede de $\rho = 65$ . . . . .	65
Tabela 4 – Tabela com a quantidade de nós necessários para dividir a rede em pelo menos dois componentes no cenário de pontos de articulação. . . . .	73





# Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
CET	Companhia de Engenharia de Tráfego
CETESB	Companhia Ambiental do Estado de São Paulo
CMTC	Companhia Municipal de Transportes Coletivos
CPTM	Companhia Paulistana de Trens Metropolitanos
CTA	Centrais de tráfego em área
DSV	Departamento de Operação do Sistema Viário
DTP	Departamento de Transportes Públicos
EMTU	Empresa Metropolitana de Transportes Urbanos de São Paulo
FGV	Fundação Getúlio Vargas
GTFS	General Transit Feed Specification
PTN	<i>Public Transport Network</i>
RMSP	Região Metropolitana de São Paulo
SMT	Secretaria Municipal de Transportes
SPTrans	São Paulo Transportes S.A.



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
<b>2</b>	<b>TRANSPORTE PÚBLICO E COMPLEXIDADE</b>	<b>23</b>
2.1	Problema do trânsito em São Paulo	25
2.2	O transporte público de São Paulo	27
2.3	Modelagem de Sistemas Complexos	29
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>33</b>
<b>3.1</b>	<b>Análise empírica</b>	<b>34</b>
3.1.1	Redes Complexas	34
3.1.2	Propriedades de redes complexas	37
3.1.3	Métricas de redes complexas	41
3.1.4	Sistema de transporte como rede complexa	46
3.1.5	GTFS	47
3.1.6	Algoritmo de agrupamento dos pontos	49
3.1.7	Abordagem de espaço	50
<b>3.2</b>	<b>Modelo exploratório de simulação de ataques</b>	<b>51</b>
3.2.1	Cenários de ataques	52
<b>4</b>	<b>RESULTADOS</b>	<b>55</b>
4.1	Análise empírica da rede de transporte de São Paulo	55
4.2	Simulação de ataque à rede de transporte de São Paulo	66
<b>5</b>	<b>CONCLUSÕES</b>	<b>81</b>
	<b>REFERÊNCIAS</b>	<b>83</b>



# 1 Introdução

Redes estão em todo lugar, como indivíduos somos parte de vários tipos diferentes de redes sociais, podendo representar relações pessoais, comerciais, profissionais, acadêmicas e assim por diante. Como sistema biológico, possuímos diversas redes de reações bioquímicas ocorrendo simultaneamente em nosso corpo, do cérebro ao sistema imunológico (BOCCALETTI et al., 2006).

A análise de redes sociais no campo das Ciências Sociais figuram como uma das áreas pioneiras na obtenção de grandes avanços no uso de Redes Complexas. Entretanto, a Ciência de Redes atraiu grande interesse de estudos a partir dos trabalhos de Watts e Strogatz (1998), com a introdução do modelo de rede denominado pequeno-mundo (*small-world*), que tentava explicar propriedades em redes encontradas na natureza, que não eram aleatórias como muitos modelos matemáticos existentes. Posteriormente, Barabási e Albert (1999) apresentaram o modelo livre de escala (*scale-free*), levando em consideração as propriedades de crescimento e preferência de conexão observadas em redes reais. Conseqüentemente, diferentes tipos de redes foram examinadas, passando por áreas desde a sociologia (NEWMAN; WATTS; STROGATZ, 2002; KOSSINETZ; WATTS, 2006) até redes ecológicas (DUNNE; WILLIAMS; MARTINEZ, 2002; DOROGOVTSSEV; MENDES, 2013).

Também foi dada atenção às redes de infraestrutura, através da exploração das características estatísticas de aeroportos (LI; CAI, 2004; BAGLER, 2008) e propostas de modelagem dos mesmos (GUIMERA; AMARAL, 2004). De mesmo modo, redes de transporte público também foram estudadas, tendo suas propriedades estatísticas investigadas através da Teoria de Grafos (LATORA; MARCHIORI, 2002; SIENKIEWICZ; HOŁYST, 2005; MA et al., 2011; FERBER; HOLOVATCH; PALCHYKOV, 2005; FERBER et al., 2007; FERBER et al., 2009). Numa análise mais específica, foi dedicado esforço no entendimento das propriedades de robustez dessas redes (FERBER; HOLOVATCH; HOLOVATCH, 2007; FERBER et al., 2012; DERRIBLE; KENNEDY, 2010), resultando em propostas de topologias que diminuíaam suas vulnerabilidades diante de falhas.

Mais recentemente, estudos foram conduzidos com o intuito de avaliar os mecanismos

de crescimento de redes de transporte (QING et al., 2013; LENG; ZHAO; XIONG, 2014), bem como suas vulnerabilidades diante de ataques e, mais especificamente, seu comportamento em situações de falha em cascata (HUANG et al., 2015; YANG; HUANG; GUAN, 2014). Da mesma forma, Mouronte e Benito (2015) realizaram uma análise estrutural da rede de transporte de Madrid e a relação de seu fluxo com a população, propondo um modelo gravitacional capaz de descrever este comportamento.

Entretanto, a análise do transporte público urbano pela ótica de Ciência de Redes é relativamente limitada na literatura local e contemplando sistemas do Brasil. A maioria dos estudos são direcionados aos aspectos operacionais (STRAMBI; BILT, 1998) e governamentais (ROLNIK; KLINTOWITZ, 2011), apresentando os sistemas de ônibus e metrô de forma desacoplada (ARBEX; CUNHA, 2015). Esta abordagem faz com que propriedades importantes sejam perdidas, visto que formam um sistema integrado e a falha em uma dessas sub-redes impacta diretamente todo o sistema. Apesar da rede de São Paulo ter sido abordada anteriormente, não é comum estudos focados na investigação de sua robustez e modelagem como rede complexa.

Essa ausência de estudos em nível macro se torna preocupante diante do constante aumento da migração de pessoas das regiões rurais para urbanas. No mundo, em torno de 54% da população vivia em áreas urbanas em 2014, com projeção de chegar à 66% até 2050, entretanto, no Brasil temos 85% da população residindo em áreas urbanas com concentração próxima de 21 milhões de habitantes na Região Metropolitana de São Paulo (RMSP) (UNITED, 2014). Essa população precisa se locomover pela cidade para alcançar postos de trabalho e espaços de lazer. Megacidades como São Paulo não comportam esses deslocamentos de forma individual via veículos automotores por meras limitações físicas, requerendo um sistema de transporte coletivo que suporte o imenso fluxo de viagens.

Para dimensionar este fluxo, são transportados mais de 9,5 milhões de passageiros por dia útil<sup>1</sup>, que precisam se locomover na quinta<sup>2</sup> maior aglomeração urbana do mundo de forma rápida e eficaz, o que não é observado na prática. O impacto de uma parada no sistema de metrô ou ônibus pode ser imenso, visto que quase metade da população da região metropolitana utiliza o transporte público diariamente para se locomover. A cidade já apresenta um alto índice de 0,42 veículos por habitante, quase o dobro da média

---

<sup>1</sup> Fonte: <[http://www.sptrans.com.br/a\\_sptrans/](http://www.sptrans.com.br/a_sptrans/)>

<sup>2</sup> Fonte: United Nations (2014)

nacional (0,22) (CINTRA, 2013), ter mais usuários optando pelo transporte individual pode acabar colapsando o sistema viário e impedi-los de chegar ao seu destino, aumentando os congestionamentos já presentes.

Aliado à isso, Cintra (2013) dimensionou o problema do trânsito de São Paulo em valores monetários, revelando que “as despesas geradas pelos congestionamentos são crescentes ao longo do tempo passando de R\$ 7 bi em 2002 para R\$ 10 bi em 2012. Os custos do tempo perdido saltaram de R\$ 10,3 bi para R\$ 30,2 bi no mesmo período”. Dessa forma, o transporte público coletivo revela-se como um dos pilares para sustentar os deslocamentos urbanos, pensar em soluções que tragam mais eficiência e confiabilidade ao mesmo mostra-se essencial para acompanhar a dinâmica da cidade.

Diante do exposto, este trabalho tem como propósito investigar o sistema de transporte público urbano do município de São Paulo de forma acoplada (ônibus e metrô juntos), através de uma análise descritiva de suas propriedades estatísticas nas métricas de redes complexas. São usados dados empíricos para criação da rede, onde diferentes valores de raio de distância são usados para agrupar pontos e estações próximas. Esse agrupamento mostra-se útil como uma medida de política pública, podendo representar a disposição do usuário em se locomover ao ponto mais próximo para acessar o transporte. Em segundo momento, é formulado um modelo exploratório que testa a robustez da rede, com parâmetros e variáveis definidos pelos resultados empíricos. São realizados simulações de ataque direcionados às estações e linhas, com o intuito de observar a capacidade do sistema em continuar conectando destinos diante de rupturas.

Na introdução foi apresentado um breve resumo da abordagem de sistemas de transporte com o método de redes complexas. Na sessão seguinte é feita uma revisão mais detalhada sobre o estudo de robustez nessas redes, apresentando seu papel nas cidades, impactos e complexidade. Na sessão 3 é apresentada a formalização matemática, métodos e dados usados para representação do sistema de transporte como uma rede complexa. Segue-se então com os resultados na sessão 4, sendo divididos pela análise empírica, onde é usado o algoritmo que agrupa pontos desconectados. Este processo mostra o quanto amplia-se a disponibilidade de destinos e linhas à medida que aumenta-se o raio de distância. Na mesma sessão, são apresentados os resultados das simulações de ataque à rede para as diferentes estratégias alvo ou aleatório, seguido pela conclusão na sessão 5.





## 2 Transporte público e Complexidade

Constantemente redes de transporte são reportadas como alvo de atentados, causando impacto em grande escala aos seus usuários e à cidade. Estabilidade em redes é um problema de fator essencial, principalmente em redes de infraestrutura crítica, onde conhecer os elementos mais importantes (ou vulneráveis) são cruciais para que se estabeleça sistemas mais eficientes e robustos. Nesse contexto, robustez é entendida como a capacidade da rede em absorver perturbações aos seus nós (pontos de parada) ou *links* (linhas de serviço) e continuar operando nas mesmas condições encontradas numa situação normal, isto é, continuar conectando usuários aos seus destinos (RODRÍGUEZ-NÚÑEZ; GARCÍA-PALOMARES, 2014).

Robustez pode ser facilmente confundida com resiliência. Entretanto, resiliência está mais associado à adaptação, à capacidade de se auto recuperar de perturbações em larga escala, de retornar ao seu estado original ou se ajustar à um novo estado (FOLKE, 2006). Enquanto robustez se refere à absorção de rupturas e continuar operando mesmo com essas rupturas ainda presentes. Este conceito é aqui utilizado para se referir ao sistema de transporte de São Paulo, já que o mesmo não apresenta a capacidade de se auto recuperar, requerendo intervenção externa diante de falhas.

Nos últimos anos, vários sistemas de transporte tem sido investigados usando conceitos de Física Estatística e Redes Complexas. Muito pelo fato de essas *Public Transport Network* (PTN) desempenharem um papel crucial nas regiões urbanas da nossa sociedade moderna. Mais especificamente, Ferber et al. (2005, 2007, 2009) analisaram diferentes aspectos de PTNs em trabalhos distintos, obtendo o comportamento livre de escala para as redes de Berlin, Düsseldorf e Paris em uma análise de suas propriedades estatísticas. Em um trabalho posterior, os mesmos autores compararam estatisticamente 14 PTNs de grandes metrópoles pelo mundo, considerando suas informações geográficas em diferentes representações espaciais, sendo proposto um modelo de crescimento baseado em regras simples que suportasse as propriedades estatísticas obtidas em um trabalho seguinte.

Na mesma linha, Sienkiewicz e Holyst (2005) realizaram uma análise estatística de

22 PTNs na Polônia, obtendo distribuição da conectividade dos pontos seguindo lei de potência e função exponencial em diferentes abordagens de espaço<sup>1</sup>, mas ambas topologias apresentaram o comportamento pequeno-mundo. [Latora e Marchiori \(2002\)](#) constataram o mesmo fenômeno anteriormente para a PTN de Boston, adicionalmente propondo uma medida de eficiência que permite lidar com redes de transporte. Para as cidades de Beijing e Chengdu na China, [Ma et al. \(2011\)](#) obteve o mesmo tipo de lei de potência na distribuição da conectividade dos pontos, acrescentando ao estudo um peso para os *links* com base no fluxo de passageiros entre estações. Num estudo ainda maior utilizando o conceito de espaço  $P^2$ , [Qing et al. \(2013\)](#) analisou empiricamente 330 cidades na China, confirmando o efeito pequeno-mundo para todas elas, mas em contraste foi obtida uma distribuição da conectividade dos pontos como função exponencial, revelando um mecanismo similar ao aleatório de evolução da rede.

Apesar disso, poucos trabalhos tiveram como finalidade investigar a robustez dessas redes diante de ataques. [Ferber, Holovatch e Holovatch \(2007\)](#) analisaram os efeitos provocados nas propriedades de 14 PTNs pelo mundo em diferentes cenários de ataque, reportando que parecem robustas sob ataques aleatórios, mas vulneráveis quando o nó desempenha um papel importante na rede, como um *hub*. Considerando 33 diferentes sistemas de metrô pelo mundo, [Derrible e Kennedy \(2010\)](#) realizaram uma análise da robustez utilizando uma abordagem de espaço própria similar à representação  $P$ , propondo métricas de impacto na experiência do passageiro. Obtiveram como resultado o característico comportamento pequeno-mundo e maior robustez em redes mais agrupadas. Como recomendação, foi proposta a ampliação de redundâncias, criando novas estações de transferência para redes pequenas e ampliando estas nas regiões periféricas ao centro em redes grandes.

Para as PTNs de Londres e Paris, [Ferber et al. \(2012\)](#) realizou uma análise quantitativa comparando qual topologia apresenta maior robustez no espaço  $L^3$ , observando como a remoção de linhas ou pontos afeta sua funcionalidade geral. A lei de potência na distribuição da conectividade dos pontos foi observada para ambas as redes, mas Paris mostrou-se mais robusta que Londres. Por fim, foi proposto como medida adicional investigar a formação de *cluster* pelo processo de percolação e a representação do sistema

---

<sup>1</sup> Consultar a sessão de métodos para mais detalhes.

<sup>2</sup> Representação focando nas transferências entre as linhas e sistemas.

<sup>3</sup> Representação focando nos pontos e paradas em sequência como ocorrem no sistema.

no espaço P, permitindo investigar o efeito em cascata nas rotas ao serem removidos seus pontos.

O fenômeno pequeno-mundo discutido tão frequentemente em redes de transporte é interessante por apresentar uma topologia com proporções que remetem à robustez e eficiência do sistema, ou seja, um agrupamento dos pontos relativamente alto (redundâncias) e um caminho médio baixo para conectá-los, característico pela presença de atalhos que atingem maiores distâncias em menos passos. A propriedade livre de escala é característica pela presença de nós altamente conectados (*hubs*), podendo apresentar dezenas, centenas ou milhares de *links*. Nesse sentido, a rede parece não apresentar escala. Entretanto, esses *hubs* quando removidos causam grande impacto na rede como um todo, tornando-a vulnerável. A identificação desses elementos é essencial para que medidas de proteção desses *hubs* possam ser tomadas.

Diante das várias abordagens e resultados obtidos fica evidente que sistemas de transporte não possuem relações de causa e efeito que seguem um resultado trivial, os resultados variam drasticamente de acordo com a abordagem usada. São compostos por diversos elementos em diferentes escalas que estão interagindo constantemente, seja diretamente com seus subsistemas ou com os usuários. Enxergar a PTN de São Paulo como um sistema complexo é observar como esses elementos interagem, como são dependentes e que perturbações podem ocasionar impactos no sistema como um todo.

## 2.1 Problema do trânsito em São Paulo

O município de São Paulo é composto por um núcleo central rodeado por diversos outros centros residenciais e comerciais. Seu sistema viário tem uma lógica semelhante, caracterizado pela presença de grandes artérias radiais que dão fluidez ao trânsito, na expectativa de que haja maior velocidade de circulação nos deslocamentos (CINTRA, 2013). O que não é observado na prática, já que quanto mais atrativa é a oferta viária, maior a demanda sobre ela, anulando rapidamente o efeito de fluidez dessas grandes avenidas.

Sendo composta por 39 municípios, grande parte da população da RMSP se locomove sentido ao centro do Município de São Paulo, motivados pelo trabalho, estudos ou em busca de atividades de lazer. Um dos fatores causadores do trânsito na cidade deve-se

à grande concentração de postos de trabalho na região central (ÂNTICO, 2005), aliado à falta de equilíbrio na ocupação do solo urbano (PIETRANTONIO; STRAMBI; GUALDA, 1996), contribuindo diretamente nos padrões de deslocamento e surgimento de aglomerações em determinadas regiões. Entretanto, políticas de uso do solo e descentralização de regiões econômicas exigem planejamento de longo prazo e variáveis que podem estar aquém da gestão pública, o processo é lento e os resultados demorados. Por outro lado, atuando no transporte público isoladamente, é possível propor ações de forma mais rápida visto que as linhas de ônibus são mais flexíveis, com alguns impactos quase em tempo real, como por exemplo incluir novas linhas de ônibus para atender demandas específicas (eventos de grande porte) ou alterar trajetos em interdições temporárias nas vias.

Implementar um sistema de transporte público eficiente está longe de ser relevante apenas para o transporte em si. Veículos nas ruas emitem grande quantidade de poluentes na atmosfera, sua redução pode ser associada ao aumento do transporte em massa, que é mais eficiente no sentido de transportar mais pessoas com menos recursos. Existem diferentes aspectos de ganhos nesse contexto, tanto para o usuário final que pode ter melhor qualidade de vida e conforto, quanto para a sociedade como um todo que tem um meio de transporte menos poluente em relação ao uso do automóvel individual. A atratividade e adoção dessa medida tem relação com a qualidade do sistema. Para que os usuários estejam dispostos a abandonar o veículo pessoal é preciso oferecer uma alternativa de serviço que atenda suas necessidades no custo e confiabilidade desejados.

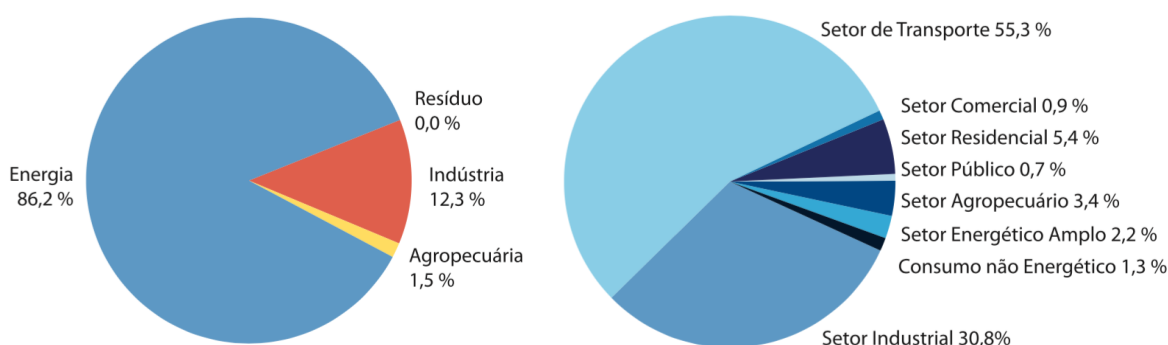


Figura 1 – Emissões de  $CO_2$  por setor e subsetor de energia em 2008 no estado de São Paulo. (Fonte: (CETESB, 2011))

Essa adoção é fator essencial na melhora dos índices de saúde e redução dos problemas respiratórios gerados pela poluição. A opção por um transporte público menos poluente mostra-se como um caminho importante para reverter os altos índices de emissão

de gases na cidade. Em um primeiro levantamento de emissões de gases de efeito estufa do estado de São Paulo, a Companhia Ambiental do Estado de São Paulo (CETESB), mostra que 86% das emissões de  $CO_2$  estão diretamente ligadas ao setor de energia, onde 55% correspondem ao setor de transporte, como pode ser observado na figura acima (CETESB, 2011).

Na mesma linha, em 2013 o Instituto Saúde e Sustentabilidade avaliou o impacto da poluição atmosférica no estado de São Paulo, apontando que os gastos públicos e privados (saúde suplementar) de internações por doenças cardiovasculares, pulmonares e câncer de pulmão atribuíveis à poluição foram respectivamente de R\$ 76 milhões e R\$ 170 milhões em 2011. O mesmo estudo aponta outro fato alarmante, no Estado morrem 17.000 pessoas precocemente e ocorrem 68.500 internações relacionadas à poluição. Excluindo os fatores econômicos, a grande quantidade de vidas perdidas devido à estas complicações é motivo suficiente para redefinir prioridades nos meios de transporte.

## 2.2 O transporte público de São Paulo

A história dos transportes coletivos em São Paulo tem registro a partir de 1865 (SPTRANS, 2006), com a regulamentação dos serviços no ano seguinte e início da operação dos bondes com tração animal em 1872. A primeira linha de bonde elétrico foi instalada em 1890 e por muito tempo permaneceu sob operação de investidores privados. A partir de 1939 começa-se os estudos em transportes coletivos, com o objetivo de criar um programa para a administração pública assumir o controle da operação. Em 1946 foi então fundada a Companhia Municipal de Transportes Coletivos (CMTC), que passou a operar a rede e fazer a gestão da frota, incorporando os ônibus particulares e o trólebus em 1958. O sistema de bondes é então desativado em 1968 e sete anos depois tem início a operação do metrô para atender o crescimento populacional da cidade, que começou a ter sua integração ao sistema de ônibus em 1983. Com a municipalização em 1991, foi aberta a primeira licitação para operação por terceiros.

Enfrentando dificuldades financeiras e baixa produtividade operacional, a CMTC foi privatizada e nasceu então a SPTrans em 1995, que começou a implementar programas de corredores e terminais de integração. Em 2001 inicia-se o processo de bilhetagem eletrônica, com a implementação do novo sistema de operação por áreas em 2003. O Bilhete Único

é oficializado no ano seguinte, oferecendo aos usuários um sistema que permite o uso do transporte via ônibus por um período de tempo pagando apenas uma tarifa.

Desde então, o transporte público coletivo no município de São Paulo tem sua operação, gestão e planejamento realizados por uma estrutura organizacional<sup>4</sup> constituída de três principais instituições:

- Secretaria Municipal de Transportes (SMT) - Criada em 1967 pela Lei 7.065 de 30 de outubro que tem por finalidade estudar o planejamento, integração, supervisão, fiscalização e controle dos transportes coletivos, táxis, veículos de carga e outros.
- Companhia de Engenharia de Tráfego (CET) - Empresa de economia mista com capital majoritário da Prefeitura do Município de São Paulo que opera sob supervisão do Departamento de Operação do Sistema Viário (DSV). Criada em 1976 com o intuito de desenvolver novas tecnologia de centrais de tráfego em área (CTA), registradores eletrônicos de infração, desenvolvimento de novas placas de sinalização, operação 24 horas do sistema viário, estudos de planejamento e projeto, implantação e manutenção da sinalização e estudos estratégicos.
- São Paulo Transporte S.A. (SPTrans) - Criada em março de 1995 como sociedade de economia mista sob supervisão do Departamento de Transportes Públicos (DTP), tem por finalidade a gestão do sistema de transporte da cidade de São Paulo, operando todo o sistema de transporte realizado por ônibus.

Nesse sistema, o município se utiliza dos modais terrestre, rodoviária e ferroviário para deslocamento dos passageiros. O modal rodoviário é operado por empresas privadas sob gestão da SPTrans, tendo 16 consórcios responsáveis pela operação de 15 mil veículos em mais de 1.300 linhas. Existe ainda o sistema metropolitano de ônibus, que sob responsabilidade da EMTU<sup>5</sup>, conecta o município às demais cidades da região. Sob gestão do Governo Estadual, o modal ferroviário tem sua operação pela Companhia do Metropolitano de São Paulo (Metrô), possuindo 5 linhas em funcionamento, 69,4 quilômetros de rede, 61 estações e 150 trens. Em conjunto com o Metrô, a Companhia Paulista de Trens Metropolitanos (CPTM) atende 22 municípios da RMSP com suas 92 estações operacionais, 6 linhas, e um total de 260,8 km de extensão.

<sup>4</sup> Fonte: Secretaria Municipal de Transportes (SMT).

<sup>5</sup> Empresa Metropolitana de Transportes Urbanos de São Paulo, gestão do Governo Estadual.

Apesar dessa infraestrutura desempenhar um papel fundamental na cidade, é preciso sempre ter em mente que sua demanda é derivada e não primária, é gerada à partir das necessidades individuais das pessoas, que resultam em atividades, interações e conseqüentemente necessidades de deslocamentos. Sua natureza depende das oportunidades disponíveis em determinadas regiões, que podem estar materializadas em forma de edificações e instituições. Entretanto, quando o acesso ao transporte passa a ser uma barreira, essa demanda tem sua importância ampliada, visto que é preferível residir perto das estações de metrô ou local de trabalho quando o deslocamento é muito custoso<sup>6</sup>.

Desse modo, percebe-se limitações na estimativa de demanda por transportes sem uma previsão adequada do crescimento da cidade e uso do solo. Mas é extremamente desafiador integrar modelos de uso do solo e demandas de transporte devido à suas escalas. São tantas variáveis existentes que torna-se impraticável propor um modelo que incorpore fatores econômicos, sociais e de crescimento da cidade junto com acesso ao transporte e padrões de deslocamento. Dessa forma, torna-se mais viável uma análise do sistema de transporte isoladamente, que mesmo não considerando os aspectos demográficos da cidade, possibilita a extração de informações importantes quanto à sua topologia e robustez.

Assim, a simulação computacional torna-se essencial nessas observações, visto que os modelos computacionais possibilitam “voltar” no tempo e alterar parâmetros que foram responsáveis por um determinado comportamento. Com essas ferramentas, vários cenários podem ser testados em um ambiente dinâmico e interativo por um custo baixíssimo se comparado aos estudos tradicionais *in loco*.

## 2.3 Modelagem de Sistemas Complexos

Reduccionismo tem sido a abordagem utilizada pela ciência nos últimos 4 séculos, ou seja, dividir o problema em partes e analisá-las individualmente da melhor forma possível, tendendo à começar pela parte mais fácil e gradualmente ascender à mais difícil (MITCHELL, 2009). É possível perceber o reflexo desse comportamento na grande quantidade de linhas de pesquisa extremamente específicas e especializadas. O problema dessa abordagem é que a soma das partes de um sistema não necessariamente representa o todo, existem características emergentes que só podem ser observadas ao analisar o

---

<sup>6</sup> Custo podendo ser caracterizado como tempo, conforto, financeiro, etc.

resultado do comportamento coletivo e como as interações dessas partes se dão. Essa necessidade de entendimento do nível macroscópico – entre outras linhas de estudo – fez surgir a Ciência dos Sistemas Complexos.

Como comparação, sistemas simples tem comportamento tipicamente previsível, as relações de causa e efeito podem ser conhecidas, bem como pequenas mudanças na causa implicam pequenas mudanças no efeito. Sistemas complexos são frequentemente imprevisíveis, tem causalidade dada pela sequência de fatos, existe um processo de retroalimentação (*feedback*), pequenas mudanças na causa implicam efeitos dramáticos e padrões podem emergir à partir das interações locais entre os elementos sem que haja um comando externo ou central (auto-organização) (ÉRDI, 2008).

Segundo Newman (2011), um sistema complexo é “um sistema composto de muitas partes que interagem, frequentemente chamadas de agentes, exibindo um comportamento coletivo que não segue trivialmente à partir dos comportamentos das partes individuais”. Ou como entende Boccara (2010), um sistema é dito como complexo se existe uma dinâmica emergente global como resultado da ação de suas partes ao invés de ser imposto por um controle central. Como exemplos, temos a população de formigas que respondem à diferentes tarefas no ninho através de sinais bioquímicos sem que haja intervenção da rainha; o cérebro que gera pensamentos, consciência, percepção, sentimentos e outras importantes atividades através da interação e troca de sinais elétricos e químicos dos neurônios; a internet como um sistema social auto-organizado, com pessoas trocando informações e serviços sendo criados e extintos de acordo com o uso e também um sistema de transporte, com vários subsistemas trocando informações e interagindo constantemente com usuários que não se comportam de uma forma previsível.

Para entender esses sistemas, modelos surgem como uma alternativa às ferramentas tradicionais. São um dos principais instrumentos da ciência moderna, com dezenas de variações criadas para investigar distintos fenômenos na natureza, tendo o papel de possibilitar uma representação matemática simplificada de um sistema de interesse (FRIGG; HARTMANN, 2012). Um dos principais aspectos na definição de um modelo está na seleção das variáveis e características do sistema a serem exploradas, faz sentido manter apenas o que é relevante e desempenha um papel essencial na interpretação do fenômeno que se quer observar.



Entretanto, existem diferenças entre modelos e a simulação tradicional. Um modelo eficiente requer o mínimo de detalhes possíveis, é orientado às teorias gerais do fenômeno explorado, possibilitando que conceitos e observações sejam portados para outras áreas da ciência. Na Simulação isolada, quanto maior a quantidade de detalhes melhor o resultado, mas a mesma acaba se tornando muito específica, não explora teorias gerais do fenômeno analisado e tem utilidade reduzida ao escopo analisado (BOCCARA, 2010).

Apesar do modelo ser uma redução natural de um problema, é diferente da prática reducionista pois o sistema não é dividido em partes, mas sim reduzido às suas propriedades essenciais para que a dinâmica da interação dos indivíduos possa ser observada e teorias gerais possam ser construídas. Mantém-se a estrutura ao mínimo necessário sem desfigurar o sistema. É possível executar esses modelos diversas vezes, testar hipóteses para diferentes valores de variáveis e a partir disso, comparar os resultados com o observado empiricamente. De fato, um modelo tem validade e utilidade para o problema ao qual se destina quando seu comportamento é semelhante ao fenômeno natural estudado.

Um modelo computacional é uma abstração do fenômeno real implementado em um programa de computador, geralmente associado à algum modelo matemático que o suporte. Como em modelos estatísticos, recebe *inputs* de dados, que são processados com base nas regras pré definidas e retorna *outputs* com os resultados da execução. Frequentemente, os *inputs* estão associados aos atributos necessários para tornar o modelo equivalente a algum contexto específico e os *outputs* são os comportamentos do modelo ao longo do tempo. Com isso, seu funcionamento pode ser interpretado de forma simplista como uma função (figura 2).

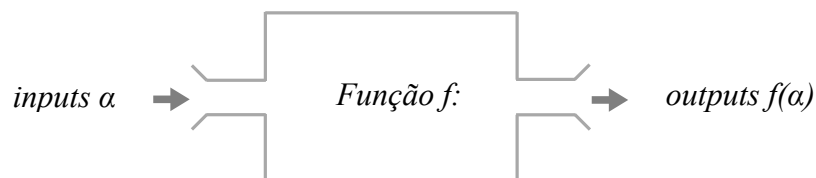


Figura 2 – Exemplo do funcionamento de um modelo computacional abstrato representado como uma função.

Apesar de conduzir naturalmente ao pensamento de gerar previsões, um modelo possibilita explicar os fatores de um fenômeno e não necessariamente prevê-los, como por exemplo um subsistema de transporte desconectado da rede pode ser explicado pelo rompimento ou ausência de linhas, mas não previsto pelo modelo visto que diversos

fatores externos ao sistema podem ter provocado o rompimento. Seu uso tem maior abrangência, visa esclarecer a dinâmica central de um sistema, sugerir analogias, propor novas questões, visualizar cenários, confrontar teorias diante de resultados, dentre outras aplicações (EPSTEIN, 2008).

Um modelo capaz de testar diferentes cenários, que comporte a aplicação de intervenções locais e a avaliação da adesão dessas soluções mostra-se uma ferramenta de extrema utilidade para a gestão pública. O investimento em transporte de massa envolve altos valores monetários e diversas barreiras de implementação, que vão desde restrições ambientais à aceitação dos usuários. O custo do erro na implementação de transporte não permite experimentação ou prototipação reduzida, é preciso sempre levar em consideração o estado atual do sistema com as variáveis que representam os padrões de deslocamento. A modelagem possibilita o planejamento de transporte público de forma dinâmica e não estática, obtendo como resultado um processo ao invés de um plano, mais adequado às constantes mudanças da cidade.

### 3 Material e Métodos

Na sessão de análise empírica são apresentados os métodos e dados utilizados para construir-se o grafo que representa o sistema de transporte público de São Paulo, abordando a formalização matemática de redes, dados coletados, processo de agrupamento de pontos de ônibus bem como abordagem de espaço utilizada. Na sessão seguinte apresenta-se os métodos utilizados no modelo exploratório de simulação de ataques, com os respectivos cenários e estratégias aplicadas e como os mesmos são executados.

Uma visão geral desses processos pode ser consultada no diagrama da figura 3 contendo o fluxo do agrupamento de pontos e o uso de seus resultados na execução do modelo de simulação de ataques. As etapas destes processos e respectivos métodos serão abordados com mais detalhes nas sessões seguintes.

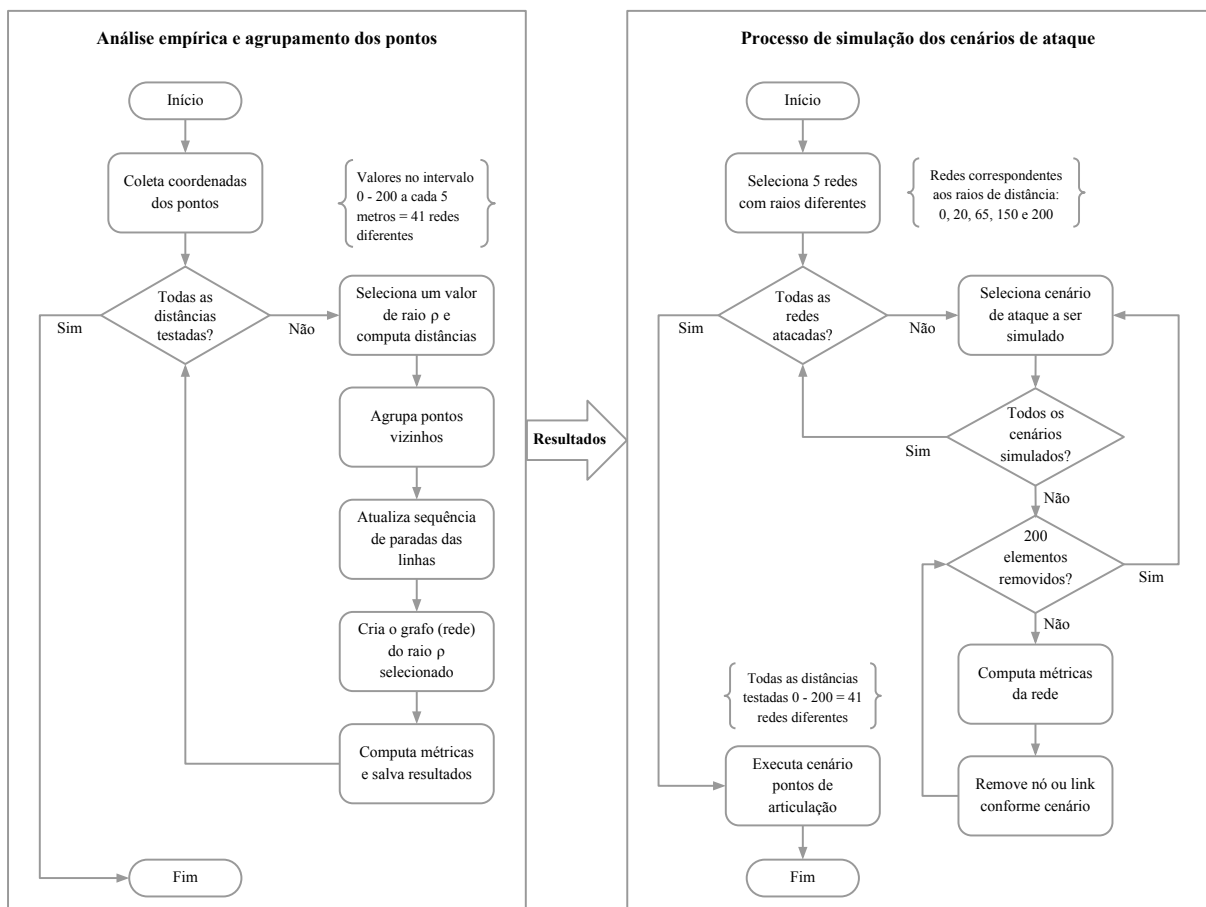


Figura 3 – Diagrama representando o fluxo do processo de agrupamento dos pontos onde os resultados são utilizados no processo seguinte de simulação de ataques.

## 3.1 Análise empírica

A investigação do sistema de transporte público urbano do município de São Paulo é feita de forma acoplada, ou seja, linhas de ônibus, metrô e trem são interligadas, mas sem nenhuma distinção entre elas. Para isso, são utilizados dados públicos de trânsito, disponíveis para quem possa interessar, bastando apenas que seja feito um cadastro no site de desenvolvedor da SPTrans. Entretanto, os dados apresentam lacunas que necessitam de ajustes para apresentar maior fidelidade ao sistema real. O problema mais grave encontrado refere-se à subsistemas totalmente desconectados da rede, requerendo um reagrupamento dos mesmos através da criação de um *link* com o ponto mais próximo. Os resultados obtidos neste processo norteiam o modelo exploratório que testa a robustez da rede diante de ataques.

Outro ponto importante refere-se a abrangência da RMSP, deve-se levar em consideração que as linhas da EMTU não estão contemplados na análise, bem como dos municípios vizinhos de São Paulo. Isso ocorre pelo fato da gestão dos sistemas de transporte serem municipais e descentralizadas. A maioria das prefeituras da RMSP não disponibiliza os dados de transporte no formato GTFS, possuindo apenas (quando possuem) portais locais e padrões diversos de armazenamento. Não está no escopo deste trabalho coletar, compatibilizar e agrupar esses dados, entendendo-se aqui como RMSP somente a área abrangida pelo sistema ferroviário e as linhas de ônibus que se expandem além do município de São Paulo.

### 3.1.1 Redes Complexas

Os primeiros estudos de redes tem sido domínio do campo da matemática discreta conhecida como Teoria de Grafos, onde o matemático Suíço Leonhard Euler foi responsável por apresentar a primeira prova verdadeira de um grafo em 1736 ([NEWMAN, 2003b](#)). Seu experimento consistia em provar se era possível atravessar as 7 pontes que conectavam as duas grandes ilhas da cidade de Königsberg ao continente, sem passar pelas pontes mais de uma vez e retornar ao ponto de partida. Transformando as pontes em linhas e as partes de solo em pontos, Euler provou matematicamente que não existia caminho possível com a restrição imposta, criando o primeiro grafo que se tem conhecimento.

Apesar da resposta parecer trivial nos dias de hoje, não era tão óbvio na época

propor uma prova matemática fazendo uso de grafos (Figura 4). Barabási (2016) esclarece que existem diferenças de nomenclatura da Ciência de Redes, o sistema representado tem o nome de rede, seus pontos são denominados nós e suas linhas *links*, diferente da Teoria de Grafos onde a rede recebe o nome de grafo, os pontos são vértices e as linhas são arestas. Será utilizada essa mesma convenção ao longo deste trabalho para referir-se à rede real ou à um grafo.

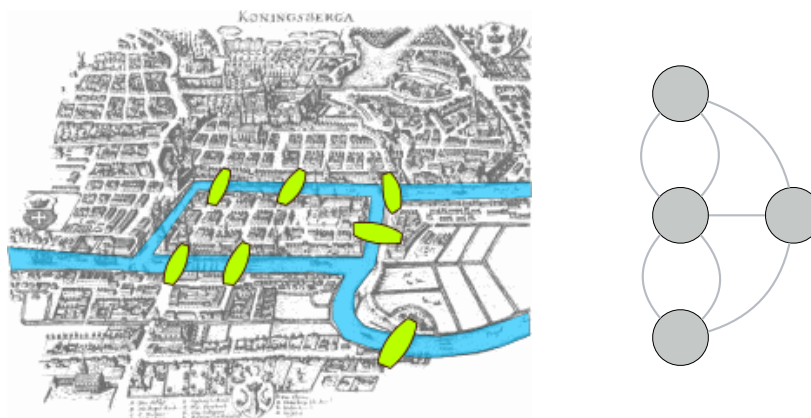


Figura 4 – Pontes da cidade de Königsberg e à direita modelo proposto por Euler, transformando terrenos em pontos e pontes em *links*. Fonte: <[https://commons.wikimedia.org/wiki/File:Konigsberg\\_bridges.png](https://commons.wikimedia.org/wiki/File:Konigsberg_bridges.png)>

Desde então, Teoria de Grafos tem sido usada como abordagem para observar diversos fenômenos do cotidiano. A análise de redes sociais no campo das Ciências Sociais figuram como uma das áreas pioneiras na obtenção de grandes avanços, como o famoso estudo de Milgram (1967) que media a quantidade de laços de amizade necessárias para que uma carta chegasse de um remetente A à um destinatário B sem que esses se conhecessem. Na ocasião, Milgram apresentou um valor médio de 6 laços necessários, resultado obtido posteriormente em estudos mais recentes com redes de colaboração científica (NEWMAN, 2001). Essas evidências de que temos em média 6 graus de separação de qualquer pessoa no mundo cunharam o termo conhecido como pequeno-mundo (*small-world*).

Formalmente, um grafo é definido por Newman, Barabasi e Watts (2006) como um objeto matemático constituído de pontos (vértices) e linhas (arestas), fazendo abstração dos detalhes físicos e focando na conectividade. De fato, um grafo pode existir sem que haja um sistema real na natureza equivalente, como grafos aleatórios que são uma linha de estudo em Teoria de Grafos. Já a rede é a representação de um sistema existente através de um grafo, geralmente se refere ao sistema real, natural ou construído pelo homem que

se quer representar, é uma ciência essencialmente baseada em dados e empirismo. Para [Érdi \(2008\)](#) e [Boccaletti et al. \(2006\)](#) é a técnica matemática mais importante e natural para representar um relacionamento estrutural entre elementos de um sistema social ou da natureza, quando estes estão contidos em uma rede.

Em termos simples, uma rede é uma coleção de nós conectados por *links*, onde os nós correspondem aos indivíduos e os *links* às conexões entre eles ([MITCHELL, 2009](#)). Como exemplo temos paradas e estações de ônibus como nós e suas rotas de serviço como *links* em uma rede de transporte público; neurônios e suas sinapses no cérebro; computadores e o cabeamento físico na internet, pessoas e suas relações em redes sociais; dentre diversos outros exemplos encontrados na epidemiologia, gestão e economia.

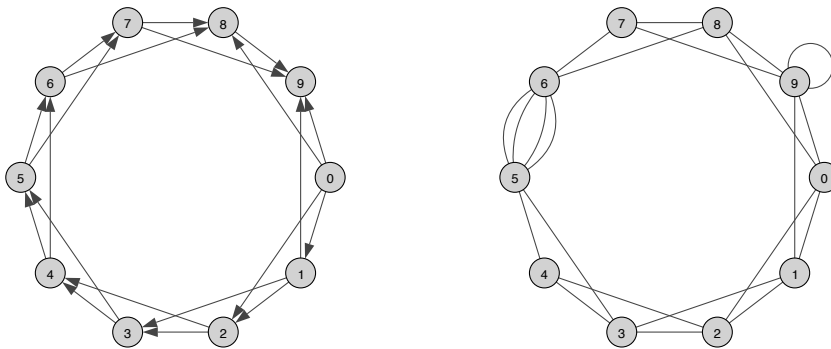


Figura 5 – Exemplo de um grafo dirigido à esquerda onde a seta representa o sentido da conexão (aresta). À direita tem-se um grafo não dirigido com múltiplas arestas (vértices 6 e 5) e uma auto aresta (vértice 9).

Matematicamente, um grafo  $\mathbf{G}$  é composto pelo seu conjunto de vértices  $\mathbf{V}(v \in V)$  e arestas  $\mathbf{E}(e \in E)$ . A forma mais comum de representá-lo é através de uma matriz de adjacência  $\mathbf{A}(\mathbf{G})$ , essa matriz consiste em uma tabela  $\mathbf{A}$  com  $n$  linhas e  $n$  colunas, onde a entrada  $A[i, j]$  se refere ao número de arestas se conectando ao vértice  $v_i$  e  $v_j$ . Uma matriz de adjacência  $\mathbf{A}(\mathbf{G})$  é simétrica em grafos não dirigidos, pois uma aresta é representada por um par de vértices não ordenados  $e = \langle v_i, v_j \rangle = \langle v_j, v_i \rangle$  ([STEEN, 2010](#)). Essa propriedade possibilita outra forma mais eficiente de armazenamento, a lista de arestas (*edge list*), que ordena o conjunto de arestas de um grafo  $\mathbf{G}$  com cada vértice incidente sobre ela, podendo ser dirigida (*digraph*) — que indica o sentido da conexão como em uma PTN — ou não dirigida, quando esse sentido não é representado (Figura 5).

Essa formalização matemática é imprescindível para analisarmos sistemas complexos, visto que só iremos compreendê-los a partir do entendimento das redes por trás deles.

Nesse ponto, o uso das métricas da teoria de grafos desempenha um papel fundamental, permitindo que redes distintas sejam comparadas e que possíveis fenômenos gerais regendo esses sistemas possam ser buscados. Inclusive, [Barabási \(2016\)](#) observa que “por trás de cada sistema complexo, existe uma complexa rede que codifica as interações entre seus componentes”, estas redes são o cerne de algumas das tecnologias mais revolucionárias do século XXI, suportando soluções do Google ao Facebook, Cisco e Twitter continua [Barabási \(2016\)](#).

Essa abordagem vem sendo aplicada à diversas áreas da ciência e tecnologia, proporcionando uma linguagem para expressar pontos em comum entre sistemas complexos na natureza. Isso pode permitir a portabilidade de percepções e descobertas entre diferentes áreas, resultando em influência mútua. De forma auto-referencial, a própria ciência de redes desempenha um papel de *hub*, conectando disciplinas científicas distantes. Como observado por [Miller e Page \(2009\)](#), evidências foram encontradas de que o cérebro tem propriedades pequeno-mundo, onde o mesmo é visto como uma rede em diversos níveis diferentes de referência: neurônios como nós e sinapses como *links*, ou com as áreas funcionais completas como nós e as ligações de grande escala entre eles como *links*. A mesma analogia pode ser feita para redes genéticas regulatórias, redes metabólicas, na epidemiologia, na ecologia e inclusive para redes de transporte público.

### 3.1.2 Propriedades de redes complexas

Grafos regulares são altamente ordenados, todos os vértices tem o mesmo número de arestas, grau ( $k$ )<sup>1</sup> local e vizinhos<sup>2</sup>. Em outras palavras, se a quantidade mínima de arestas conectando à um vértice em um grafo for igual à quantidade máxima, o mesmo pode ser dito como regular. O grafo da figura 6 também é chamado de conectado, ou seja, todos os pares de vértices estão ligados. Se dividirmos o grafo do exemplo, suponhamos, uma parte contendo os vértices de 0 a 5 e a outra com os vértices de 6 a 9, temos então dois sub grafos, que também podem ser chamados de componentes ou *clusters*.

Um componente nada mais é do que um subconjunto de vértices em um grafo, de modo que exista um caminho entre quaisquer dois vértices que pertencem à esse componente. Temos então um componente conectado quando todos os vértices da rede são

<sup>1</sup> Número de arestas que conectam ao vértice. Ver Apêndice 3.1.3 para definição matemática.

<sup>2</sup> Quando um par de vértices é adjacente.

acessíveis através de algum caminho e um componente desconectado quando existe algum subconjunto de vértices que não podem ser acessados via arestas existentes.

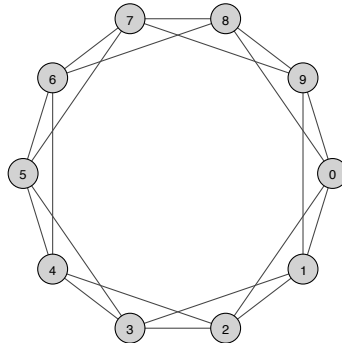


Figura 6 – Exemplo de um grafo regular em círculo conectado, possuindo 10 vértices, grau  $k = 4$  dos nós e probabilidade 0 de reconexão.

Enquanto grafos regulares são extremamente ordenados, grafos aleatórios são desordenados e regidos por probabilidades. Atribui-se à [Erdős e Rényi \(1959, 1960\)](#) seu estudo sistemático e estabelecimento como um campo na Matemática. Um grafo aleatório  $\mathbf{G}_{N,P}$ , consiste de um numero  $N$  de vértices que se conectam resultando em  $K$  arestas, de tal forma que cada par de vértice  $i, j$  tem uma aresta se conectando com uma probabilidade independente  $p$ . Apesar de ter sido muito estudado, este modelo tem graves deficiências em comparação com resultados observados em redes reais, como o grau de distribuição dos vértices que difere de resultados empíricos ([NEWMAN; WATTS; STROGATZ, 2002](#)).

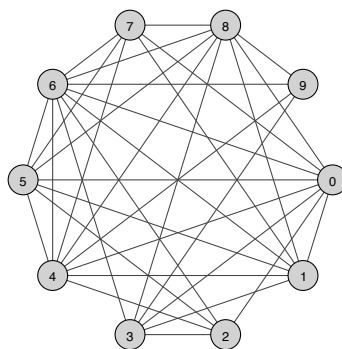


Figura 7 – Exemplo de um grafo aleatório do tipo Erdős e Rényi possuindo 10 vértices e probabilidade  $p = 0.7$  de reconexão das arestas.

Já o modelo pequeno-mundo (*small-world*) apresenta maior proximidade com algumas redes reais. Numa rede regular composta exclusivamente por conexões locais com seus vizinhos é necessário um grande numero de passos para que a informação trafegue e chegue ao destino. Numa rede pequeno-mundo, no entanto, existem conexões de longa



distância, atuando como atalhos, onde a informação pode ser transmitida entre quaisquer dois nós por em média um pequeno número de conexões. Esta propriedade mostra-se importante em sistemas de transporte, atuando por exemplo como linhas expressas conectando regiões distantes ao centro da cidade sem muitas paradas.

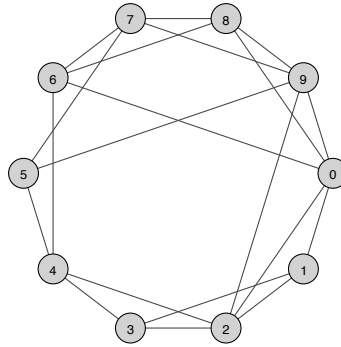


Figura 8 – Grafo pequeno-mundo de 10 vértices e probabilidade  $p$  0.2 de reconexão.

Para alternar entre grafos regulares e aleatórios, [Watts e Strogatz \(1998\)](#) propõem um procedimento aleatório de reconexão (figura 9). Começando a partir de uma estrutura circular com  $n$  vértices e  $k$  arestas por vértice, cada aresta é religada aleatoriamente com uma probabilidade  $p$ . Este método permite ajustar o grafo entre regularidade ( $p = 0$ ) e desordem ( $p = 1$ ), possibilitando explorar a região intermédia  $0 < p < 1$ , que não é completamente regular nem completamente aleatória.

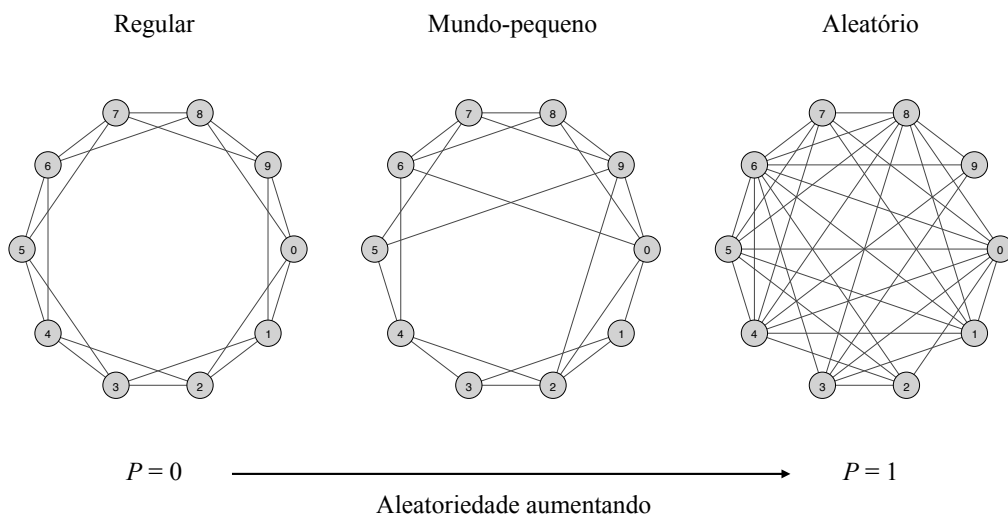


Figura 9 – Processo de reconexão aleatória a partir de uma estrutura circular regular para uma rede aleatória, sem alterar o número de vértices ou arestas do grafo. Fonte: [\(WATTS; STROGATZ, 1998\)](#)

O modelo de rede pequeno-mundo é matematicamente caracterizado pelas métricas

de comprimento do caminho e coeficiente de clusterização<sup>3</sup>, medindo respectivamente a distância típica entre dois vértices genéricos na rede e a aglomeração média de um vértice. De fato, redes do tipo pequeno-mundo são altamente aglomeradas, como em redes regulares, tendo ainda um característico comprimento curto de caminhos, como grafos aleatórios (LATORA; MARCHIORI, 2002). No caso de uma PTN, estas propriedades são observadas na presença de estações com grande número de rotas que as servem, bem como na existência de rotas que cruzam a rede e conectam nós distantes.

Nos modelos Erdős e Rényi (1959) e Watts e Strogatz (1998) apresentados anteriormente, é possível observar uma propriedade em comum, ambos tem um tamanho característico  $\langle k \rangle$  com forte dependência de  $p$  e uma distribuição probabilística da conectividade do vértice  $p(k)$  com corte exponencial (BARABÁSI; ALBERT; JEONG, 2000). Entretanto, algumas redes empíricas já estudadas apresentaram um comportamento livre de escala de  $p(k)$ , com distribuição de grau do nó em lei de potência para diferentes ordens de grandeza, como constatado por Ferber, Holovatch e Palchykov (2005) em redes de transporte.

Dois aspectos podem ser associados à essa discrepância dos modelos em relação aos dados empíricos dessas redes estudadas. Primeiramente, os modelos assumem um número fixo ( $N$ ) de vértices, que são aleatoriamente conectados ou desconectados sem alterar o  $N$ , em contraste com redes reais que tem crescimento incremental ao longo do tempo. Na sequência, os modelos de redes aleatórios assumem uma probabilidade aleatória e uniforme de conexão dos vértices, contrastando com a conectividade preferencial<sup>4</sup> observada em diversas redes reais, continua Barabási, Albert e Jeong (2000).

Tendo isso em mente, o modelo Barabási e Albert (1999) livre de escala (*scale-free*) propõe uma teoria que explica o crescimento de um grafo baseado na preferência de conexão por vértices altamente conectados, o chamado efeito “rico fica mais rico”, como pode ser visto com o vértice (0) na figura 10 que tem maior preferência de conectividade pelo seu grau em relação aos seus vizinhos. Esse feito evidencia nós altamente conectados atuando como *hubs*, que podem ser entendidos como os grandes terminais de ônibus e metrô presentes no sistema. Esses *hubs* são críticos por apresentar uma grande quantidade de linhas os conectando (*links*), podendo causar grande impacto à rede quando removidos.

<sup>3</sup> Ver Apêndice 3.1.3 para maiores informações das métricas.

<sup>4</sup> Ao entrar na rede, um novo nó escolhe o nó  $i$  ao qual irá se conectar baseado na conectividade  $k_i$ .

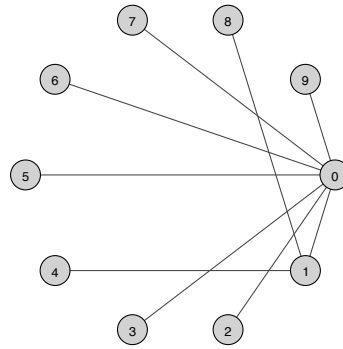


Figura 10 – Grafo livre de escala de 10 vértices com conectividade preferencial do modelo Barabasi Albert.

O crescimento<sup>5</sup> começa a partir de uma configuração inicial que é irrelevante para a estrutura do grafo quando ele já é grande, então é adicionado um novo vértice que é conectado aos vértices  $m \geq 1$  selecionados preferencialmente com uma probabilidade proporcional ao seu grau. Quando o grafo atinge um tamanho grande em  $t \rightarrow \infty$  vértices, a distribuição de grau alcança um estado estacionário em lei de potência  $P(k) \propto k^{-\alpha}$  com expoente  $\alpha = 3$  (DOROGOVTSEV, 2010).

Duas características importantes em redes livres de escala são destacadas por Barabási e Bonabeau (2003), são robustas em situações de falhas acidentais, ou seja, a maioria dos pontos tem uma pequena quantidade de linhas os conectando, sua remoção causa dano reduzido. Entretanto, são vulneráveis à ataques coordenados que miram seus elementos mais importantes (*hubs*), interrompendo uma grande quantidade de linhas imediatamente após sua remoção. Entender essas propriedades e identificar esses elementos são importantes para que se pense em medidas e soluções de proteção aos mesmos, podendo tornar o sistema mais robusto diante de interrupções.

### 3.1.3 Métricas de redes complexas

Existem centenas de métricas distintas na Ciência de Redes, sendo muitas baseadas em modelos específicos e outras para áreas específicas da ciência. Em uma rápida pesquisa da formalização matemática dessas métricas é possível encontrar divergências na forma com que muitos autores as definem e as adaptam para o propósito de seu estudo. A próxima sessão se limitará apenas às métricas utilizadas na análise empírica e modelo de simulação de ataques que este trabalho se propõe, observando as formalizações utilizadas

<sup>5</sup> Crescimento entende-se como o aumento do número de nós na rede ao longo do tempo.

na literatura de redes de transporte público como uma rede complexa.

## Grau

Uma das mais comuns e importantes métricas de um vértice é o seu grau (*degree*), que representa o número de arestas que o conectam à outros vértices. No caso de uma PTN, representa a quantidade de rotas (linhas) que atendem uma determinada estação ou ponto. Denota-se  $\mathbf{K}_i$  o grau de ordem  $i$  do vértice na rede, como no exemplo da (Figura 5) à direita, temos  $K_{4,7} = 3$ ,  $K_{0,1,2,3,8} = 4$ ,  $K_9 = 5$ ,  $K_{5,6} = 6$ .

Como mencionado anteriormente, nós que possuem alto grau em relação aos outros vértices da rede são tipicamente denominados *hubs*, que desempenhando um papel fundamental de conectividade, podendo particionar a rede quando removidos. Segundo Barabási (2016), em redes não dirigidas o número total de arestas  $\mathbf{L}$  pode ser expressado pela soma dos graus do vértice, onde o fator  $1/2$  corrige as arestas contadas duas vezes na soma:

$$L = \frac{1}{2} \sum_{i=1}^N K_i, \quad (3.1)$$

e o grau médio da rede não dirigida dado pela expressão:

$$\langle K \rangle \equiv \frac{1}{N} \sum_{i=1}^N K_i = \frac{2L}{N}. \quad (3.2)$$

No caso das redes dirigidas como na (Figura 5) à esquerda, Barabási (2016) ainda apresenta a diferenciação do grau para capturar o sentido da aresta, onde temos o grau de entrada  $K_i^{in}$  (*in-degree*) que representa o número de aresta que apontam para o vértice  $i$ , e grau de saída  $K_i^{out}$  (*out-degree*), que representando o número de arestas apontadas tendo como origem o vértice  $i$ , sendo o grau total  $\mathbf{K}$  dado pela equação:

$$K_i = K_i^{in} + K_i^{out}, \quad (3.3)$$

onde o grau médio da rede dirigida dado por:

$$\langle K^{in} \rangle = \frac{1}{N} \sum_{i=1}^N K_i^{in} = \langle K^{out} \rangle = \frac{1}{N} \sum_{i=1}^N K_i^{out} = \frac{L}{N}. \quad (3.4)$$

Uma variação essencial do grau de um vértice é a distribuição de grau  $p(k)$ , que é a probabilidade de um vértice selecionado aleatoriamente na rede ter grau  $k$ . Uma rede com  $N$  vértices tem sua distribuição de grau por um histograma normalizado

$$P_k = \frac{N_k}{N} \quad (3.5)$$

onde  $N_k$  é o número de vértices com grau  $k$ , sendo possível obter o número de vértices de grau  $k$  a partir da distribuição de grau como  $N_k = N p_k$ . A maioria das propriedades da rede requer o conhecimento da probabilidade do grau, como por exemplo o grau médio, que pode ser dado por:

$$\langle K \rangle = \sum_{k=0}^{\infty} k p_k \quad (3.6)$$

### Assortatividade

Para alguns tipos de redes é importante saber se seus elementos desempenham algum tipo de preferência na conexão com os demais, como eles se combinam e se associam. Essa propriedade é conhecida como assortatividade, que como posto por [Newman \(2003a\)](#), é a tendência dos vértices em uma rede se conectarem à outros vértices que são de alguma forma parecidos com eles (ou o contrário disso). Ou seja, se indivíduos preferem se associar com outros iguais à si, podemos dizer que a rede mostra um padrão de combinação assortativa ou preferencial. Mas se os mesmos preferem se associar com indivíduos diferentes, a combinação é disassortativa.

Essa análise é realizada através de alguma característica específica para comparação, como por exemplo bilhetagem das estações ou lotação dos veículos no caso das PTNs. Entretanto, geralmente este tipo de dado não pode ser obtido com facilidade, cabendo o uso de características mais facilmente acessíveis, como o grau dos vértices, fator escolhido para cálculo do coeficiente. Essa característica possibilita observar se pontos altamente conectados apresentam uma preferência de se associar com outros pontos de mesmo tipo.

Sendo  $e_{jk}$  a fração de arestas que conecta os vértices de grau  $j$  aos vértices de grau  $k$  e a probabilidade de um vértice selecionado aleatoriamente ter o grau  $k$  ([NEWMAN, 2003a](#)), temos o grau excedente do vértice distribuído de acordo com:

$$q_k = \frac{(k+1)p_{k+1}}{z}, \quad (3.7)$$

onde  $z = \sum_k k p_k$  é o grau médio na rede, e a distribuição  $q_k$  se relacionando com  $e_{jk}$  via:

$$\sum_j e_{jk} = q_k. \quad (3.8)$$

O coeficiente de Assortatividade pelo grau do vértice pode então ser obtido por

$$r = \frac{\sum_{jk} j k (e_{jk} - q_j^{in} q_k^{out})}{\sigma_{in} \sigma_{out}}, \quad (3.9)$$

onde  $\sigma$  é o desvio padrão da distribuição  $q_k$  e  $e_{jk}$  é a probabilidade de um link dirigido selecionado aleatoriamente se conecte à um vértice com grau de entrada  $j$  e à um vértice com grau de saída  $K$ . A função pode então ser reescrita para cálculo de elementos específicos como:

$$r = \frac{\sum_i j_i k_i - M^{-1} \sum_i j_i \sum_{i'} k_{i'}}{\sqrt{[\sum_i j_i^2 - M^{-1}(\sum_i j_i)^2][\sum_i k_i^2 - M^{-1}(\sum_i k_i)^2]}} \quad (3.10)$$

sendo  $j_i$  e  $k_i$  os graus de entrada e saída excedentes do vértice que a aresta  $i$ -ésimo leva à entrada e saída respectivamente.

### Comprimento dos caminhos e Diâmetro

Em sistemas reais, o comprimento dos caminhos são caracterizados por distâncias óbvias, como a distância em quilômetros entre duas estações de metrô. Em redes, essa distância física é substituída pelo comprimento do caminho. Quando existe a necessidade de se transportar um elemento de um ponto ao outro, encontrar o caminho mais curto é essencial para que se poupe recursos do sistema, o que torna essa métrica importantíssima para medir a eficiência da rede.

Um caminho é uma rota que percorre as arestas da rede e seu comprimento dado pelo número de arestas que essa rota contém, ou seja, um caminho entre os vértices  $i_0$  e  $i_n$  é uma lista ordenada de  $n$  arestas  $P_d = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_{n-1}, i_n)\}$ , onde  $d$  é seu comprimento. Desse modo, é possível obter a distância média dos caminhos mais curtos entre todos os pares de vértices de uma rede dirigida de  $N$  vértices pela equação:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j=1}^N d_{ij}. \quad (3.11)$$

O comprimento do caminho mais curto pode ser representado como uma matriz onde a entrada  $d_{ij}$  é o comprimento da geodésica (caminho mais curto) do vértice  $i$  ao vértice  $j$ . Para uma distância  $d_{ij} = 1$ , existe uma aresta que conecta o vértice  $i$  ao vértice  $j$  diretamente.

Tendo isso em mente, é possível definir um grafo completo quando o comprimento do caminho mais curto possui valor 1, ou seja, o vértice  $i$  está conectado à todos os outros vértices por apenas uma aresta de distância. No entanto, redes de transporte público são bastante esparsas, com a quantidade de estações e linhas seguindo a mesma ordem. Consequentemente, quanto mais extensa for a rede, maior será o valor médio obtido no comprimento do caminho.

A partir das definições anteriores, o diâmetro da rede pode ser definido como a maior distância  $\ell$  entre os caminhos mais curtos na rede, podendo ser dado por:

$$d_G = \max_{i,j} \ell_{ij}. \quad (3.12)$$

Por definição  $\ell \leq d_G$ , e, no caso de uma distribuição de comprimento dos caminhos mais curtos bem delimitada, é possível mostrar heurísticamente que, em muitos casos, as duas definições se comportam da mesma forma com o tamanho da rede.

### Clusterização

Também conhecido como transitividade, a clusterização é uma métrica básica e importante que fornece informações locais do grafo. Pode ser usada para capturar o nível de redundâncias que os pontos possuem, já que vizinhos também conectados podem oferecer caminhos alternativos no caso de uma linha interrompida. Sua definição está geralmente relacionada à quantidade de triângulos<sup>6</sup> na rede, sendo alto quando dois vértices dividindo um vizinho tem alta probabilidade de se conectarem. Captura o grau ao qual os vizinhos de um dado vértice se conectam uns com os outros, podendo ser dado por:

$$C = \frac{2L_i}{K_i(K_i - 1)}. \quad (3.13)$$

É mais comum o uso de sua medida derivada coeficiente de clusterização, que mede a probabilidade média de dois vizinhos de um vértice serem vizinhos deles mesmos. Através da observação da densidade dos triângulos na rede, é possível capturar um valor de  $\mathbf{C}$  para todos os vértices  $i = 1, \dots, N$

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^N C_i. \quad (3.14)$$

Em grafos aleatórios, o coeficiente de clusterização se apresenta baixo e tende à 0 à medida que o sistema cresce, ao contrário de grandes redes reais como rotas de trem e malhas elétricas que tem um alto coeficiente e permanece constante (COHEN; HAVLIN, 2010). Essa medida possibilita capturar como a rede se fragmenta à medida que seus elementos são removidos nas simulações de ataques, tendendo a diminuir em função dessas remoções.

### Densidade

A densidade de um grafo indica a relação entre a quantidade de arestas existentes pela quantidade máxima de arestas possíveis. Também por ser usada para capturar o nível

<sup>6</sup> Conjunto de três vértices conectados no grafo.

de redundâncias da rede, já que quanto mais densa, maior a quantidade de *links* conectando seus nós. Apesar de conceitualmente ser parecido com o coeficiente de clusterização, seu cálculo pode ser dado por:

$$D = \frac{2E}{N(N-1)}. \quad (3.15)$$

Onde E corresponde ao número de arestas no grafo. Por essa definição, um grafo é então esparso se  $D \ll 1$ .

### 3.1.4 Sistema de transporte como rede complexa

A partir das definições anteriores, é possível perceber que sistemas de transporte são de fato redes, sua estrutura, topologia e composição de pontos com respectivas linhas de serviço podem ser representadas naturalmente como um grafo. Uma importante propriedade de sistemas de transporte é a relação de redundância de linhas com o comprimento dos caminhos, que exibem o nível de alternativas providas ao usuário e o sua abrangência, podendo ambos serem medidos de forma rigorosa através das métricas de redes.

Em uma PTN, as estações, terminais ou paradas podem ser representadas como nós e suas linhas e rotas como *links*. Essa abordagem possibilita investigar alguns aspectos desses sistemas que a abordagem tradicional não consegue, como identificação dos *hubs* medidos de forma empírica, aspectos que favorecem a formação de *clusters* e os nós e *links* de maior importância na rede, principalmente quando esses elementos importantes são de alguma forma removidos do sistema.

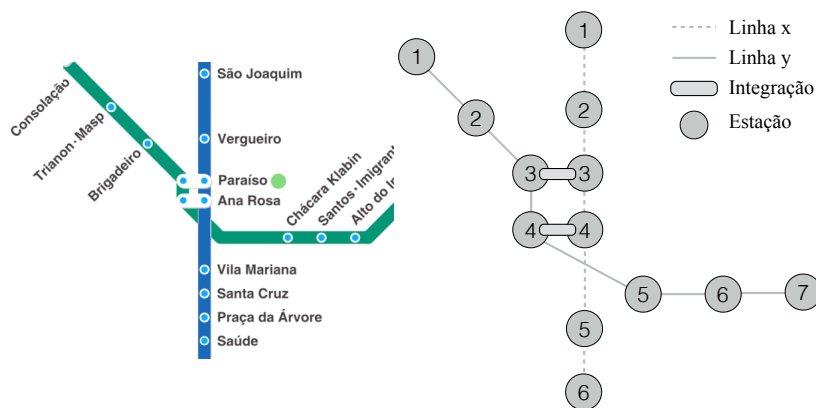


Figura 11 – Rede parcial do metrô de São Paulo e seu modelo representativo como um grafo (Fonte: adaptado da SPtrans).

A figura 11 mostra uma representação típica de uma rede PTN, que pode ser



dirigida ou não dirigida. O grafo  $\mathbf{G} = \{\mathbf{V}, \mathbf{E}\}$  dessa rede é composto pelo seu conjunto de vértices  $\mathbf{V}$  (estações e paradas) e arestas  $\mathbf{E}$  (linhas). Cada vértice  $v \in \mathbf{V}$  tem um valor  $k$  relacionado ao número de linhas que passa por ele, possibilitando o cálculo da frequência. Todos os vértices são similares para a construção do grafo, ou seja, não há distinção entre paradas de ônibus, estações de metrô ou terminais de transferência.

### 3.1.5 GTFS

Com o intuito de observar as propriedades gerais da estrutura da rede, esta precisa ser grande o suficiente em relação ao número de nós e *links*. Por esse motivo são selecionadas as redes de ônibus, trem e metrô em conjunto. Para criação do grafo que representa a PTN de São Paulo, foram obtidos dados da página de desenvolvedor da SPTrans<sup>7</sup>, que utilizam o formato *General Transit Feed Specification*<sup>8</sup> (GTFS), desenvolvido e mantido pelo Google. Esse formato possibilita que agências de trânsito publiquem seus dados e desenvolvedores criem aplicações que os consumam. É um padrão utilizado globalmente e composto por até 13 arquivos, onde os mais relevantes para a análise proposta são:

- *stops.txt* - Localizações individuais onde há embarque e desembarque de passageiros nos veículos (paradas ou pontos).
- *routes.txt* - Rotas que atendem à cidade (linhas).
- *trips.txt* - Viagens e as rotas das quais fazem parte (sequência de paradas num tempo específico).
- *stop\_times.txt* - Horários de partida e chegada individuais para cada viagem.
- *shapes.txt* - Regras para projetar as linhas no mapa representando a organização das rotas.

O arquivo *stop\_times* contém os horários de forma sequencial das chegadas e partidas individuais de um veículo para suas paradas em cada viagem. Ao analisar essa sequência de paradas, é possível obter os *links* entre as mesmas, onde por exemplo o *stop\_id* 301703 se conecta ao *stop\_id* 30003160, indicado pela coluna *stop\_sequence*, como pode ser visto na figura abaixo:

<sup>7</sup> <http://www.sptrans.com.br/desenvolvedores/>

<sup>8</sup> <https://developers.google.com/transit/>

```

"trip_id","arrival_time","departure_time","stop_id","stop_sequence"
"1015-10-0","05:00:00","05:00:00",301703,1
"1015-10-0","05:06:40","05:06:40",30003160,2
"1015-10-0","05:13:20","05:13:20",30003001,3
"1015-10-0","05:20:00","05:20:00",30003051,4
"1015-10-0","05:26:40","05:26:40",30003003,5
"1015-10-0","05:40:00","05:40:00",30003002,6
"1016-10-0","12:00:00","12:00:00",830004111,1
"1016-10-0","12:01:22","12:01:22",830004014,2
"1016-10-0","12:02:44","12:02:44",830004008,3

```

Figura 12 – Estrutura do arquivo *stop\_times* no padrão GTFS. Os *links* se dão através da sequência de pontos para o mesmo *id* de viagem, onde o ponto inicial *301703* se conecta ao próximo ponto *30003160* e assim por diante.

Sempre que muda-se o sentido da viagem ou a linha (campo *trip\_id*), o sequencial recomeça. Com base nessa premissa, foi desenvolvido um *script* em Python que, comparando o *trip\_id* da linha posicionada no arquivo com o da próxima linha, cria uma lista de arestas para todos as paradas sequentes da linha até que o sequencial recomece na posição 1 e o *trip\_id* mude, fazendo uma varredura por todo o arquivo.

Devido às características geográficas da rede e a forma de representação dos dados, foi observado que algumas linhas apresentavam *stop\_id* inicial e final da viagem diferentes, ou seja, o ciclo não é fechado, como demonstrado na figura 13. Com isso, vários nós da rede apresentaram grau 1 pois só possuíam um *link* de entrada (parada final) ou de saída (parada inicial). Esse fato é também responsável pela grande presença de graus ímpar na rede. É razoável assumir que essas paradas possuem uma conexão, visto que uma viagem possui ponto inicial/final e o funcionamento da linha se dá através de ciclos.

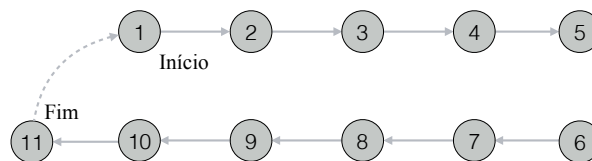


Figura 13 – Viagem com o ciclo não fechado, onde a linha contínua representa as conexões existentes e o pontilhado a conexão ausente.

Quanto ao problema dos componentes desconectados, a maioria dessas ocorrências foi encontrada para as redes ferroviárias, como por exemplo as linhas 03 vermelha do metrô e 4 amarela que são importantes fluxos de deslocamento. Também foi observado um grande volume de pontos muito próximos aglomerados (10 - 15 metros), que aparentemente são os terminais de ônibus e cada ponto pode corresponder à uma vaga de parada do veículo. Essas observações exigiram a criação do algoritmo de agrupamento, que conecta

esses componentes ao sistema de forma que ao final é obtida uma rede única. O mesmo é descrito com detalhes na sessão seguinte.

### 3.1.6 Algoritmo de agrupamento dos pontos

De forma simplista, o algoritmo de agrupamento feito em Python<sup>9</sup> transforma todos os pontos próximos – de acordo com um raio de distância – em um único ponto. Ou seja, todos os antigos *ids* de pontos recebem um novo *id* com o mesmo código. Como não é trivial definir o valor de raio ideal para transformar a rede em um único componente, são testados diferentes valores entre 0 e 200 metros até obtenção do resultado esperado. O processo se dá através da execução de diversas funções e passos como visto no diagrama da figura 3.

Primeiramente são coletadas as coordenadas de latitude e longitude de cada ponto do sistema contidos no arquivo *stops.txt*. Esses dados geográficos são utilizados como entrada na função que busca todos os pontos próximos de um determinado ponto de acordo com uma distância de raio em metros, aqui definida pela letra grega  $\rho$ . É realizada uma comparação de  $n \times n$  para todos os pontos existentes.<sup>10</sup> Ao fim do processamento dessa função, tem-se um dicionário com cada ponto e a lista de seus respectivos *vizinhos*.

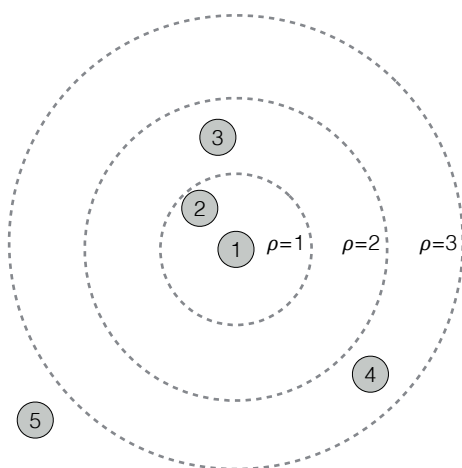


Figura 14 – Agrupamento dos nós com base em um valor de raio  $\rho$ . No exemplo, para  $\rho = 1$  o nó 1 tem apenas o nó 2 como vizinho, á medida que o raio se expande são adicionados o nó 3 e 4 para  $\rho = 2$  e  $\rho = 3$  respectivamente. O nó 5 é desconsiderado por estar fora do raio.

<sup>9</sup> <https://www.python.org/>

<sup>10</sup> O valor retornado assume a terra como sendo perfeitamente esférica. Algoritmo usado de Fonte: [http://www.johndcook.com/blog/python\\_longitude\\_latitude/](http://www.johndcook.com/blog/python_longitude_latitude/)

Com o dicionário de vizinhos, é feito então um agrupamento dos pontos, onde todos recebem um novo *id* com o mesmo código. Como exemplo, na figura 14 os pontos 1, 2, 3 e 4 recebem o mesmo *id v1* para  $\rho = 3$ . Para garantir que os *hubs* sejam formados e que os nós próximos recebam a classificação correta do novo *id*, é feita uma busca recursiva que percorre também os vizinhos de cada elemento da lista. Isso é necessário pelo fato do arquivo ser uma sequência de pontos e do código computacional funcionar de formar serializada, ou seja, o primeiro item da lista é analisado e passa-se para o segundo. Essa sequência altera a forma como os grupos são feitos. Como por exemplo, para  $\rho = 1$ , o único vizinho seria o ponto 2, mas como o ponto 3 é vizinho do ponto 2, ambos recebem o mesmo *id*, caso contrário poderia ocorrer conflito ao decidir se o ponto 2 se agruparia com o ponto 3 ou 1, onde a ordem com que eles aparecem na lista definiria o grupo.

Finalizado este processo, é então utilizado esse dicionário de vizinhos para atualizar o arquivo *stop\_times.txt*, substituindo os códigos existentes pelos novos obtidos da função anterior. Cada vez que o código original é encontrado, tem seu valor substituído pelo novo em toda a lista, garantindo integridade da estrutura de *links*. Isso possibilita criar a lista de arestas, que é composta basicamente pelo *link* de origem e de destino, tendo o código da rota (linha) como critério para definir que a sequência faz parte da mesma rota.

Nesse ponto, é criado o grafo da rede, onde é feito o cálculo das métricas<sup>11</sup> para cada valor de  $\rho$  simulado, indo de 0 a 200 metros em intervalos de 5, ou seja, são criados 41 grafos diferentes e as métricas calculadas para cada um deles. Esses diversos valores de raio podem simular uma característica importante em política de mobilidade urbana, que pode ser interpretada como a disposição do usuário em se locomover até algum ponto de parada próximo. A medida que amplia-se esse raio, aumenta-se a oferta de linhas e pontos disponíveis, que conseqüentemente podem levar a uma quantidade maior de destinos e opções para chegar ao local desejado.

### 3.1.7 Abordagem de espaço

Até o momento, os trabalhos realizados com redes de transporte público tem maior concentração na abordagem topológica de espaço bidimensional do tipo **L** (FERBER; HOLOVATCH; PALCHYKOV, 2005, 2007, 2009, 2012, LATORA; MARCHIORI, 2002) e o espaço **P** (LENG; ZHAO; XIONG, 2014; QING et al., 2013) que correspondem

<sup>11</sup> É utilizada a biblioteca gratuita *igraph* em Python. <<http://igraph.org>>

respectivamente ao foco nas paradas e ao foco nas transferências. Em uma rede baseada em paradas, um nó é uma parada, e quaisquer duas estações atendidas sucessivamente por uma rota estão conectadas por um *link*. Numa rede baseada em transferências, um nó continua sendo uma parada, mas o *link* é estabelecido se existe uma rota direta que os conecta, sem a necessidade de mudança de linha ou sistema.

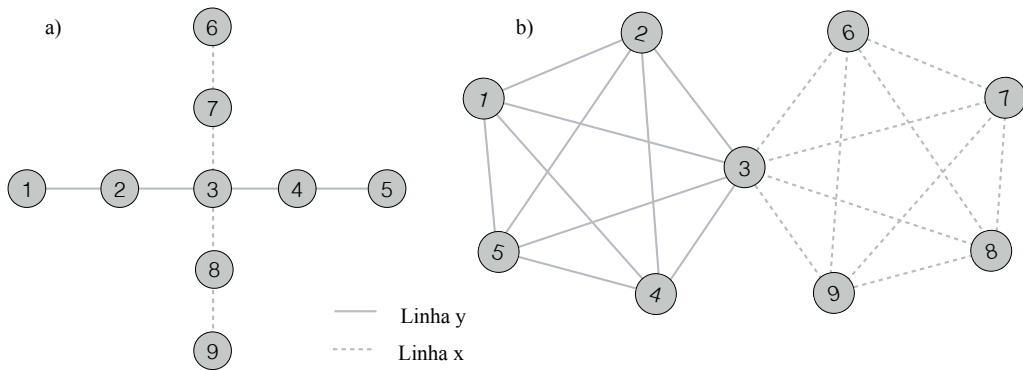


Figura 15 – Representação de duas linhas no espaço  $\mathbf{L}$  (a) e no espaço  $\mathbf{P}$  (b).

Para este trabalho é adotada a representação de espaço  $\mathbf{L}$ , nesta representação o grafo obtido se assemelha ao mapa real da rede, preservando todos os seus pontos e não somente as conexões. Apesar de não haver consenso na literatura em relação ao tipo de representação ideal para uma rede em duas dimensões, a maioria dos trabalhos tende a usar a abstração que possibilita explorar da melhor forma o fenômeno que se quer observar. A abordagem em espaço  $\mathbf{L}$  possibilita o uso das informações geográficas de forma mais precisa e correspondente ao sistema real, atributo necessário ao modelo de ataques, que visa remover nós ou *links* na forma com que se encontram no sistema real.

## 3.2 Modelo exploratório de simulação de ataques

Com o intuito de testar a robustez e vulnerabilidades da PTN de São Paulo, é utilizado um modelo computacional exploratório, com parâmetros definidos a partir dos resultados obtidos na análise empírica. Nesse contexto, *ataque* tem sentido de simular possíveis interrupções no sistema. Também referidos como cenários ou estratégia, os mesmos podem representar:

- Falha mecânica: Queda de energia, veículos com problemas operacionais, manutenção emergencial;

- Acidentes de veículos: Acidentes na via, acidentes envolvendo veículos do sistema, acidentes com passageiros dos veículos;
- Obstrução da via: Obstruções planejadas ou não planejadas devido à construções, via fechada temporariamente;
- Desastre natural: Tempestades, alagamentos, deslizamentos, tremores, afundamento de solo.

De maneira geral, esses cenários simulam remoções de linhas de ônibus ou metrô – no caso de ataque à *links* – e os terminais e paradas que as compõem, para ataque aos nós. Um maior detalhamento será feito adiante no âmbito de rede complexa, onde métricas são utilizadas para definir como os elementos serão removidos e seu impacto calculado.

### 3.2.1 Cenários de ataques

Robustez é uma propriedade importante em redes complexas do tipo PTN<sup>12</sup>, pois envolvem o transporte de elementos e um ponto crítico de sua operação é continuar transferindo esses elementos à seus respectivos destinos diante de perturbações. O estudo dessa propriedade é necessário devido ao papel essencial que sistemas de transporte desempenham nas cidades.

Para verificar a robustez da rede diante de ataques simulados, são utilizados diferentes cenários, realizados em um momento na estratégia aleatória<sup>13</sup> e em outro na estratégia de alvos selecionados por sua importância. A escolha do alvo é realizada de duas formas, sendo determinístico ou probabilístico. Estes cenários, podem ser descritos como:

1. **Ataques a nós alvos - determinístico:** A remoção de nós pelo grau decrescente reflete no tamanho dos componentes da rede, que podem diminuir. A medida que é obtido o nó de grau máximo da rede, as métricas são calculadas e este nó é removido, como exemplo o nó 33 de maior grau da figura 16.
2. **Ataques a nós alvos - probabilístico:** A remoção de nós pela probabilidade do mesmo ser selecionado em função de seu grau na rede. O nó tem seu *id* repetido em

<sup>12</sup> Rede do tipo tecnológica.

<sup>13</sup> Pseudoaleatório, visto que todo número aleatório gerado computacionalmente é determinístico.

uma lista de nós proporcionalmente ao seu grau, ou seja, o nó de grau 100 aparece 100 vezes enquanto o de grau 1 uma única vez.

3. **Ataques a nós aleatórios:** A remoção aleatória a partir da lista de todos os nós permite comparar a estratégia com nós alvos. A medida que é obtido um nó via escolha aleatória da lista de nós da rede, as métricas são calculadas e este nó é removido.
4. **Ataques a *links* alvos - determinístico:** Seleção de *links* alvos com peso baseado em sua multiplicidade na rede, refletindo o quanto aquele caminho é usado na rede como os *links* (12, 13) e (32, 33) na figura 16 que são usados por duas linhas. A medida que é obtido o *link* de índice maior, as métricas são calculadas e este *link* é removido.
5. **Ataques a *links* alvos - probabilístico:** Remoção de *links* pela probabilidade do mesmo ser selecionado em função de seu índice de multiplicidade. Da mesma forma como em nós alvos probabilístico, uma lista reflete a probabilidade do *link* com repetição do *id* proporcional ao índice de multiplicidade.
6. **Ataques a *links* aleatórios:** Seleção de *links* aleatórios da lista de todos os *link* da rede, permitindo comparar com a estratégia de *links* alvos. A medida que é obtido o *link* via escolha aleatória, as métricas são calculadas e o mesmo é removido.
7. **Pontos de articulação:** Identificação dos nós necessários para dividir a rede em partes. Para todos os valores de  $\rho$ , retorna os nós necessários para dividir a rede em pelo menos dois componentes.

Todos os cenários juntos constituem um processo de simulação ataques que se dá de forma independente como visto no diagrama da figura 3. O primeiro cenário de nós alvo - determinístico é executado calculando-se as métricas da rede em seu estado sem perturbações, então é selecionado o nó de grau máximo para remoção. Se houver mais de um nó com o mesmo grau, ambos são removidos um de cada vez, a mesma regra é usada para o *link* alvo - determinístico com seu respectivo critério de seleção. Esse cenário representa um ataque coordenado, quando tem-se a intenção de remover elementos importantes da rede para causar o maior dano possível.

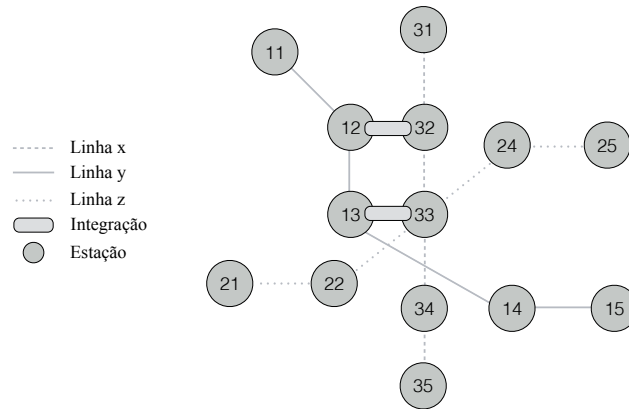


Figura 16 – Modelo simbólico no espaço L para representação dos possíveis cenários de ataque nas estratégias de seleção aleatória, alvo determinístico ou probabilístico.

No segundo cenário de nós alvo - probabilístico, é criada uma lista com todos os nós da rede a ser testada, repetindo-se o *id* do nó de acordo com seu grau, isso define uma maior probabilidade para os nós mais importantes. O nó é então selecionado aleatoriamente na lista, as métricas da rede são calculadas e todas as entradas repetidas desse nó são removidas. O processo é executado da mesma forma para o cenário de *link* alvo - probabilístico, considerando-se o respectivo critério de probabilidade definido para este cenário.

Para os cenários nas estratégias aleatório e probabilístico não existe o problema de conflito em qual nó ou *link* remover, visto que um elemento é selecionado por vez via algoritmo de escolha aleatória. O processo é executado da mesma forma para ambos, com a diferença de que na estratégia aleatório não há repetições de elementos na lista.

O processo para nós e *links* se repete por 200 interações cada um, deste modo, 200 nós ou *links* são removidos em cada cenário executado. Para os pontos de articulação, é feita apenas uma execução que processa o algoritmo de busca dos nós para cada rede de  $\rho$ . Tanto para os quatro primeiros cenários quanto o último, a execução é realizada apenas uma vez, desta forma os valores resultantes para os cenários que possuem aleatoriedade são apenas um dos possíveis resultados, já que não é realizada uma média desses valores.



## 4 Resultados

### 4.1 Análise empírica da rede de transporte de São Paulo

Com o intuito de melhor visualizar o estado do sistema de transporte, é feita uma análise espacial da rede no QGIS<sup>1</sup> considerando apenas a localização dos pontos no mapa, mostrando as linhas aparentemente desconectadas em seu estado original, que correspondem ao raio de distância  $\rho = 0$ . Sua distribuição pode ser vista na figura abaixo:

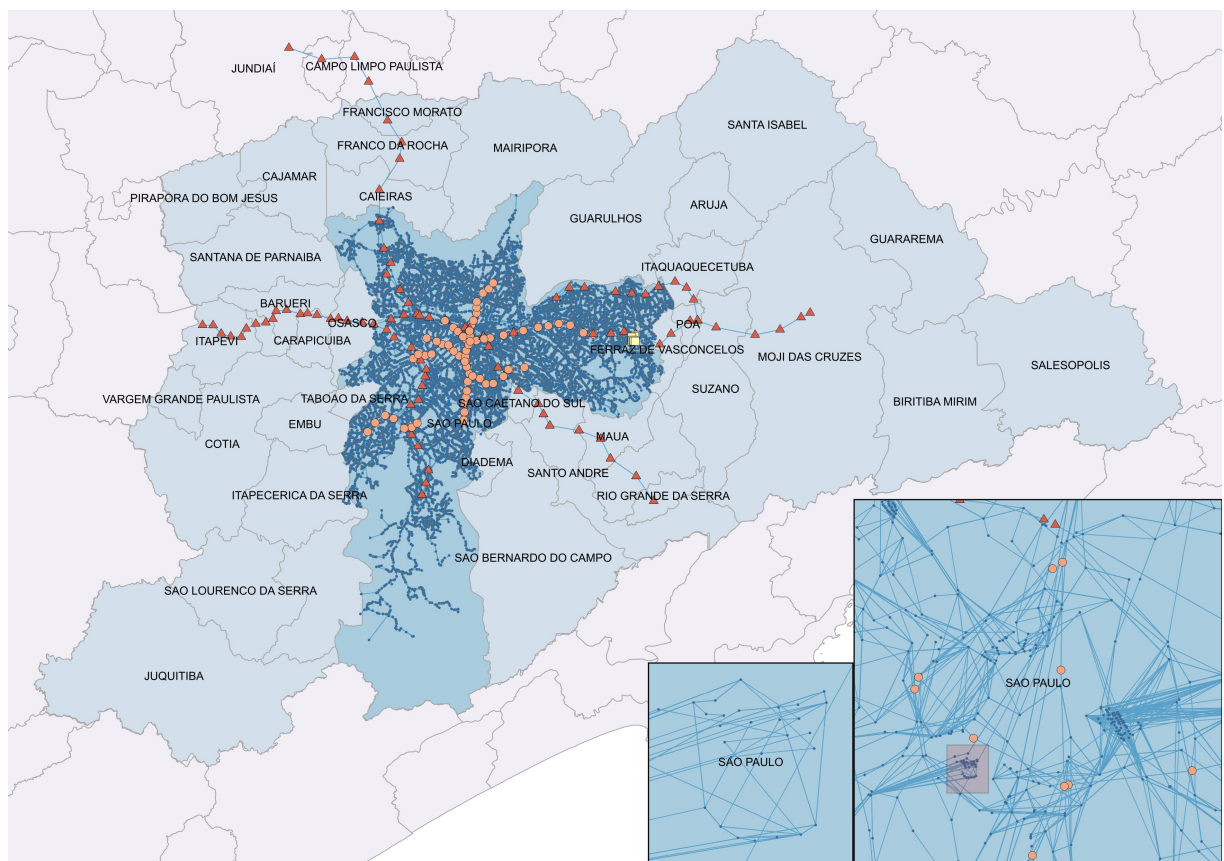


Figura 17 – Representação do sistema de transporte de São Paulo com os componentes desconectados. Em destaque tem-se os trens da CPTM (triângulos), linhas de metrô (círculos) e uma linha de ônibus (quadrados). Os demais pontos em azul correspondem ao restante da estrutura com suas linhas. No zoom à direita tem-se terminais com um exemplo ampliado, onde percebe-se que não há linhas entre todos os pontos. A área maior demarca a RMSP tendo o município de São Paulo em destaque ao centro.

Foram detectados 12 componentes, ou sub-redes, que estão destacados na figura 17

<sup>1</sup> *Open Source Geographic Information System.* <<http://www.qgis.org/en/site/>>

pelos pontos maiores que foram agrupados por tipo. Os trens da CPTM são representados pelos triângulos em vermelho, as linhas de metrô pelos círculos em laranja e a única linha de ônibus pelos quadrados em amarelo. Os demais pontos de tamanho menor em azul correspondem ao restante da estrutura com suas respectivas linhas. A área maior no mapa representa a RMSP composta por seus 39 municípios, destacando-se ao centro em tom mais escuro o município de São Paulo. É possível perceber maior concentração da infraestrutura de transporte em São Paulo, onde a expansão além desta se dá através do sistema sob trilhos.

As linhas desconectadas indicam uma característica diferente da realidade, pois mesmo que não exista um *link* entre uma estação de metrô ou terminal de ônibus próximo, o usuário consegue realizar a transferência se deslocando a pé. No zoom à direita inferior da figura 17 é possível visualizar com mais detalhes a região central do município de São Paulo, onde as aglomerações correspondem à alguns de seus terminais, o mesmo pode ser visto com mais detalhes no box ao lado com a ampliação da área demarcada em vermelho. Tem-se um terminal onde a distância média entre os pontos é de 14 metros, mas como pode ser visto, não existe uma linha que conecta todos os pontos entre si. É possível que esses pontos próximos correspondam às vagas nos terminais onde os veículos estacionam para embarque e desembarque de passageiros.

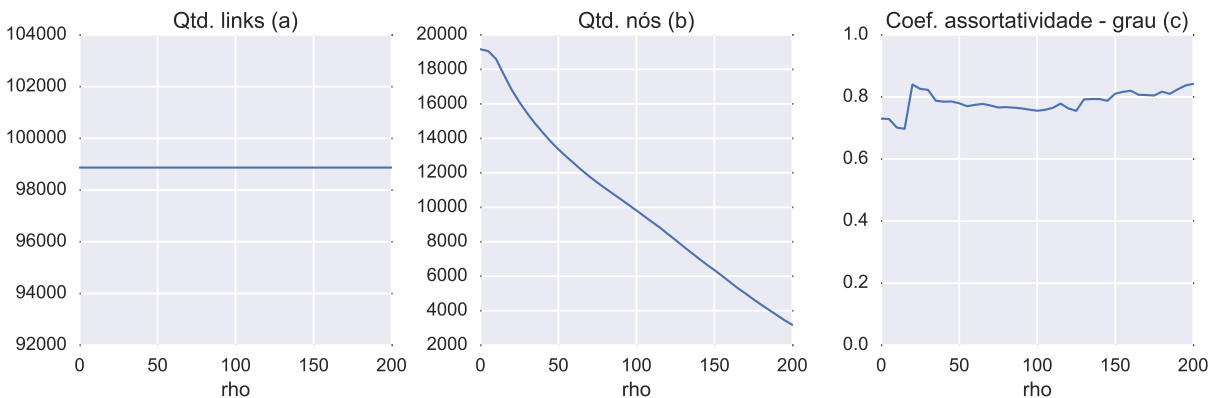


Figura 18 – Quantidade de *links* estável para cada rede em função de  $\rho$  na figura (a). Quantidade de nós em função de  $\rho$  na figura (b) com comportamento decrescente. Coeficiente de assortatividade em função de  $\rho$  na figura (c).

Diante do exposto, tem-se a necessidade de integrar esses componentes ao sistema para uma análise do mesmo como um todo, o que é feito pela execução do algoritmo de agrupamento onde é possível acompanhar o comportamento da rede a medida que os valores de raio são ampliados. A quantidade de *links* em função de  $\rho$  pode ser observada na

figura 18a, que encontra-se em estado estacionário para todas as distâncias testadas. Isso ocorre devido a preservação dos *links* durante o agrupamento dos pontos, que são mantidos como auto *links* (*self-loops*). Já a quantidade de nós apresenta uma curva com declive acentuado, com uma redução de 83,4% em função do aumento de  $\rho$ , caindo de 19.169 para 3.180 no maior  $\rho$ . Este resultado é previsto, já que vários pontos são transformados em um no agrupamento, como o terminal mencionado anteriormente na ampliação da figura 17.

Já o coeficiente de assortatividade, que basicamente mede o quanto nós estão conectados com outros nós semelhantes, teve uma leve queda para 0,69 em  $\rho = 15$  e um pico de 0,84 em  $\rho = 20$ , apresentando pequenas flutuações para os demais valores. Este resultado é totalmente inesperado para uma rede tecnológica, mostrando que a mesma é altamente assortativa, ou seja, apresenta um certo tipo de preferência de conexão com nós do mesmo tipo. Uma possível explicação para este fato pode ser a grande concentração de nós com baixa conectividade e esses nós estarem conectados a outros nós do mesmo tipo. Um ponto de ônibus está ligado geralmente a somente seu ponto sucessor e antecessor no sentido da linha. A frequência de pontos altamente conectados é baixa, como pode ser visto na cauda da distribuição de grau que é exibida adiante na figura 24. Entretanto, resultados deste tipo são mais comum em redes sociais, inclusive, Newman (2003a) encontrou valores negativos (ou nulos) para redes do mesmo tipo, como a internet e redes de energia, requerendo uma melhor investigação do índice obtido.

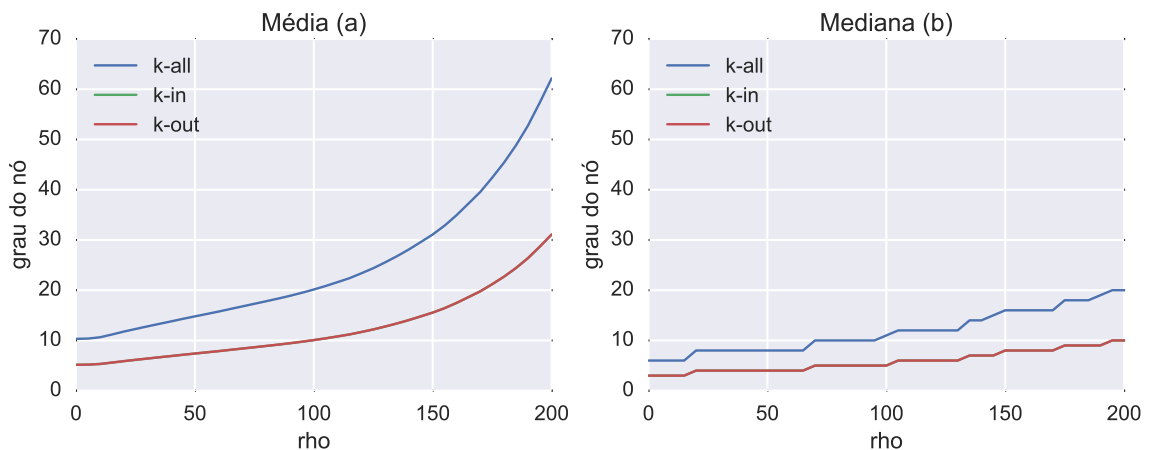


Figura 19 – Média dos graus da rede em função de  $\rho$  para grau total (*k-all*), grau de entrada (*k-in*) e grau de saída (*k-out*) na figura (a). (*k-in*) e (*k-out*) aparecem sobrepostos por não haver variação. A figura (b) apresenta a mediana dos mesmos índices.

Contudo, obtém-se uma correlação positiva entre o coeficientes de assortatividade para a média e mediana dos graus dos nós, que são respectivamente 0,62 e 0,65. Para

reduzir possíveis erros do algoritmo de cálculo dessa métrica, a mesma foi testada em redes criadas aleatoriamente com as mesmas proporções da rede original (19.000 nós e 100.000 *links*). Em 100 execuções para cada rede do tipo aleatória, pequeno-mundo e livre de escala, foram obtidos coeficientes em média negativos para assortatividade, sendo respectivamente -0,0008, -0,0004 e -0,0002. Uma indicação de que as redes criadas empiricamente muito provavelmente exibem o comportamento obtido.

Mas para um resultado mais conclusivo seria interessante a comparação dessa métrica com outro índice de heterogeneidade, visto que a característica usada para cálculo da assortatividade é o grau dos nós, conseqüentemente os agrupamentos realizados com os nós alteram a distribuição de grau. É possível observar inclusive a curva da mediana dos graus aumentando para 4 no mesmo valor de  $\rho$  (Figura 19). Contudo, a variância e desvio padrão dos graus é muito grande, não sendo uma métrica confiável para classificação desse tipo de resultado, tem-se por exemplo  $\sigma^2 = 124,13$  e  $\sigma = 11,14$  para a rede de  $\rho = 0$  quando o grau médio total é 10, ou mais extremo,  $\sigma^2 = 3.3870,96$  e  $\sigma = 184,04$  na rede  $\rho = 200$  onde o grau médio total é 62.

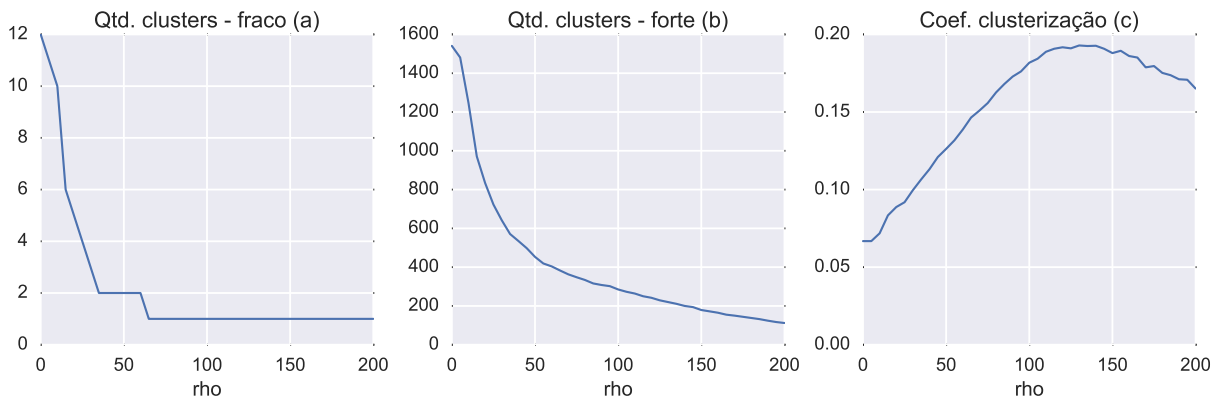


Figura 20 – Quantidade de *clusters* da rede em função de  $\rho$  nas figuras (a) e (b). Coeficiente de clusterização na figura (c) em função de  $\rho$ , tendo pico de 0,19 em 140.

A métrica mais importante para avaliar a efetividade do agrupamento é o *cluster* (figura 20a). Busca-se aqui juntar os componentes desconectados da rede com o algoritmo de agrupamento, sendo obtido com sucesso em  $\rho = 65$ , que mantém-se estável para os valores seguintes, já que a rede tende a conectar-se cada vez mais. O *cluster* forte é uma variação dessa métrica, mas leva em consideração que grupos dentro da rede estejam conectados por uma grande quantidade de *links*. Já o coeficiente de clusterização – que basicamente mede a probabilidade dos vizinhos de um nó estarem também conectados –

aumenta em função de  $\rho$ , tendo um pico em 0,19 para  $\rho = 140$ , quando aparentemente começa-se uma tendência de diminuição.

Apesar disso, o coeficiente de clusterização foi calculado como uma rede não dirigida<sup>2</sup>. Esta limitação não condiz com a realidade das PTNs, já que o sentido do *link* que liga o ponto A ao B representa o sentido da via, que não necessariamente é verdade para o *link* inverso. Entretanto, para cálculo da probabilidade de conexão dos vizinhos este aspecto é irrelevante, já que a existência do *link* já é suficiente.

Conforme esperado, o comprimento médio dos caminhos (figura 21a) apresenta uma diminuição em função de  $\rho$ , saindo de  $\approx 55$  para  $\approx 19$  no último valor, podendo estar relacionado com a redução da quantidade de pontos. Como os *links* são preservados, conseqüentemente reduz-se a quantidade de passos já que a rede está mais conectada, podendo gerar caminhos mais curtos. Na mesma linha, o diâmetro da rede – que mede a distância máxima entre dois pontos – apresentou o mesmo tipo de comportamento em função do aumento de  $\rho$ , reduzindo de seu índice máximo de 170 passos para 81 ao final. Esses índices globais reduzindo refletem no quanto a rede se torna mais conectada, que conseqüentemente pode aumentar sua robustez em relação ao *link*, já que redundâncias reduzem a probabilidade de um *link* causar grande impacto na rede ao ser removido.

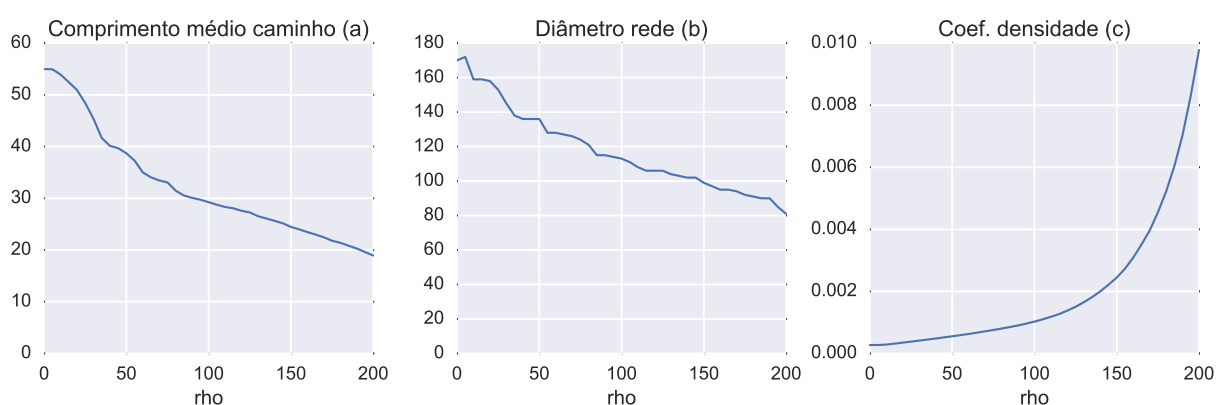


Figura 21 – Comprimento do caminho médio em função do aumento de  $\rho$  (a). Diâmetro da rede diminuindo em proporção parecida (b). Densidade da rede que aumenta à medida que os nós são agrupados (c).

De fato, percebe-se que a rede se torna mais densa em função de  $\rho$  como pode ser observado na figura 21c, ou seja, a relação entre o número de *links* e o número de pares de nós aumenta, visto que a rede parte de um estado de  $\approx 99$  mil *links* com  $\approx 19$  mil nós

<sup>2</sup> Devido à limitações do algoritmo usado.

para  $\approx 3$  mil nós com o mesmo número de *links*. Esse resultado lembra a recomendação de [Derrible e Kennedy \(2010\)](#) em seu estudo com 33 diferentes sistemas de metrô pelo mundo, onde recomendam a ampliação de redundâncias no sistema para as regiões periféricas ao centro, o que pode tornar a rede mais robusta. Uma possível forma de se aplicar e medir essa recomendação seria pelo aumento da densidade da rede.

Continuando com a análise de distâncias, é feita uma seleção de 5 redes com diferentes valores de  $\rho$  como pode ser visto na figura 22. Tem-se os histogramas para cada rede  $\rho$  com a distribuição do comprimento médio dos caminhos em função de sua frequência. O sistema aparenta sair de um estado de alta frequência com um grande intervalo de caminhos ( $\rho = 0$ ) para um estado onde os caminhos médios são bem menores e menos frequentes ( $\rho = 200$ ). Essa série representa como o sistema transita entre estados de conectividade em função de  $\rho$ , com cada histograma representando uma rede distinta.

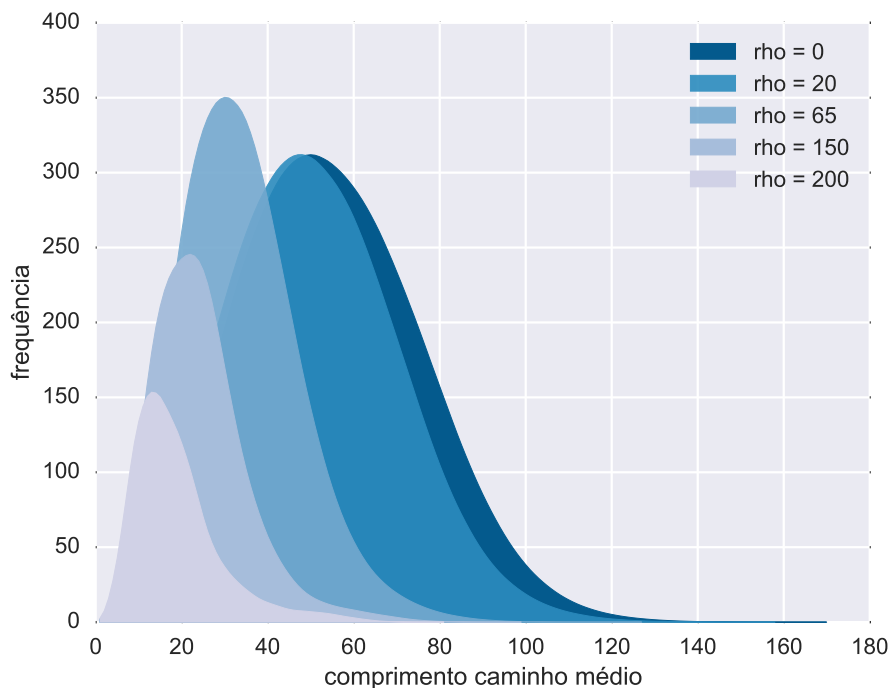


Figura 22 – Histograma com a distribuição da frequência do caminho médio em função de seu tamanho para todos as distâncias de  $\rho$ . Normalizado pelos nós da rede.

Os dados foram normalizados dividindo-se a frequência pela quantidade de nós na rede. Esses valores podem ser observados no gráfico (b) da figura 18. A área sob a curva parece diminuir com o aumento do  $\rho$ , uma possível explicação é a diminuição na quantidade de nós que é  $\approx 12$  vezes menor em  $\rho = 200$  comparado à  $\rho = 0$  (figura 18b). Essa redução faz com que menos passos em média sejam necessários para alcançar dois

pontos aleatórios na rede, como também pode ser observado no comprimento médio dos caminhos (figura 21a) e diâmetro da rede (figura 21b).

Tendo os resultados prévios em mente, é possível propor uma medida simplista de eficiência para o agrupamento. Tem-se então a razão do comprimento do caminho médio pelo raio de distância  $\rho$ . Essa medida pode ser interpretada como uma relação entre o esforço do usuário ao caminhar para o ponto mais próximo em função do benefício de obter mais linhas e destinos possíveis à partir da nova posição.

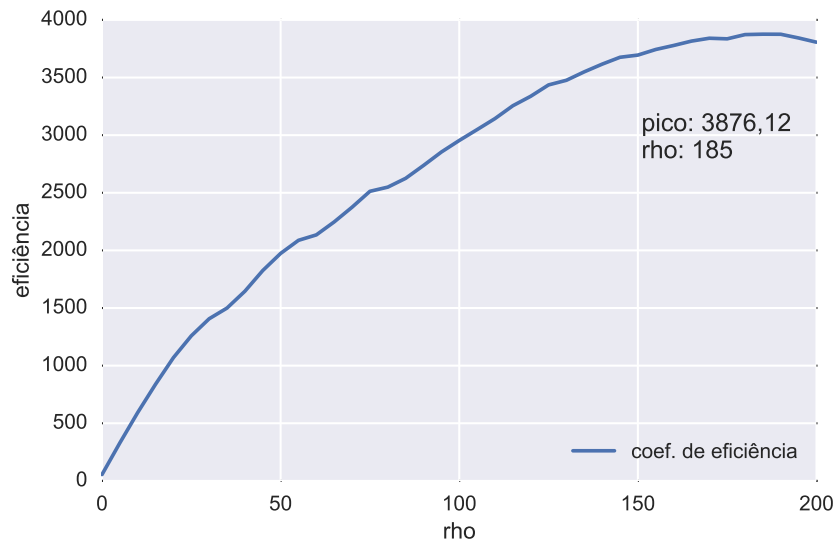


Figura 23 – Eficiência das redes em função de  $\rho$  calculada como a razão entre  $\rho$  e o comprimento do caminho médio. Tem-se o pico em  $\rho = 185$ , onde aparentemente começa-se uma queda.

Na figura 23 é possível observar o pico do gráfico em  $\rho = 185$  com o índice 3876,12 obtido. Aparentemente começa-se uma queda a partir do pico do coeficiente obtido, seria interessante continuar o processo de agrupamento para mais valores de  $\rho$  e observar o comportamento ao longo do processo. Entretanto, o comprimento do caminho médio tem limite em 0, enquanto  $\rho$  tende ao  $\infty$ , é razoável pressupor que essa curva tende a cair e se manter próxima de zero à medida que amplia-se essa distância. Uma variação interessante dessa métrica seria incorporar pesos às duas variáveis representando a equivalência entre a distância até ao próximo ponto (ou o tempo, incluindo espera no ponto) e a sua proporção em relação a distância percorrida utilizando o transporte público .

Uma análise típica em redes complexas é verificar se a distribuição dos graus dos nós segue uma lei de potência (*power law*), que pode ser descrito pela forma:

$$p(k) \propto k^\alpha. \quad (4.1)$$

Leis de potência são interessantes por apresentarem uma probabilidade ainda relativamente alta para a cauda direita da distribuição (*heavy-tailed*). A constante  $\alpha$  é conhecida como o expoente ou parâmetro de escala, que tipicamente figura no intervalo  $2 < \alpha < 3$ . Na prática, poucos fenômenos obedecem lei de potência para todos os valores de  $k$ , sendo mais aplicados a partir de valores maiores que algum  $k$  mínimo, nesses casos somente a cauda da distribuição segue uma lei de potência (CLAUSET; SHALIZI; NEWMAN, 2009).

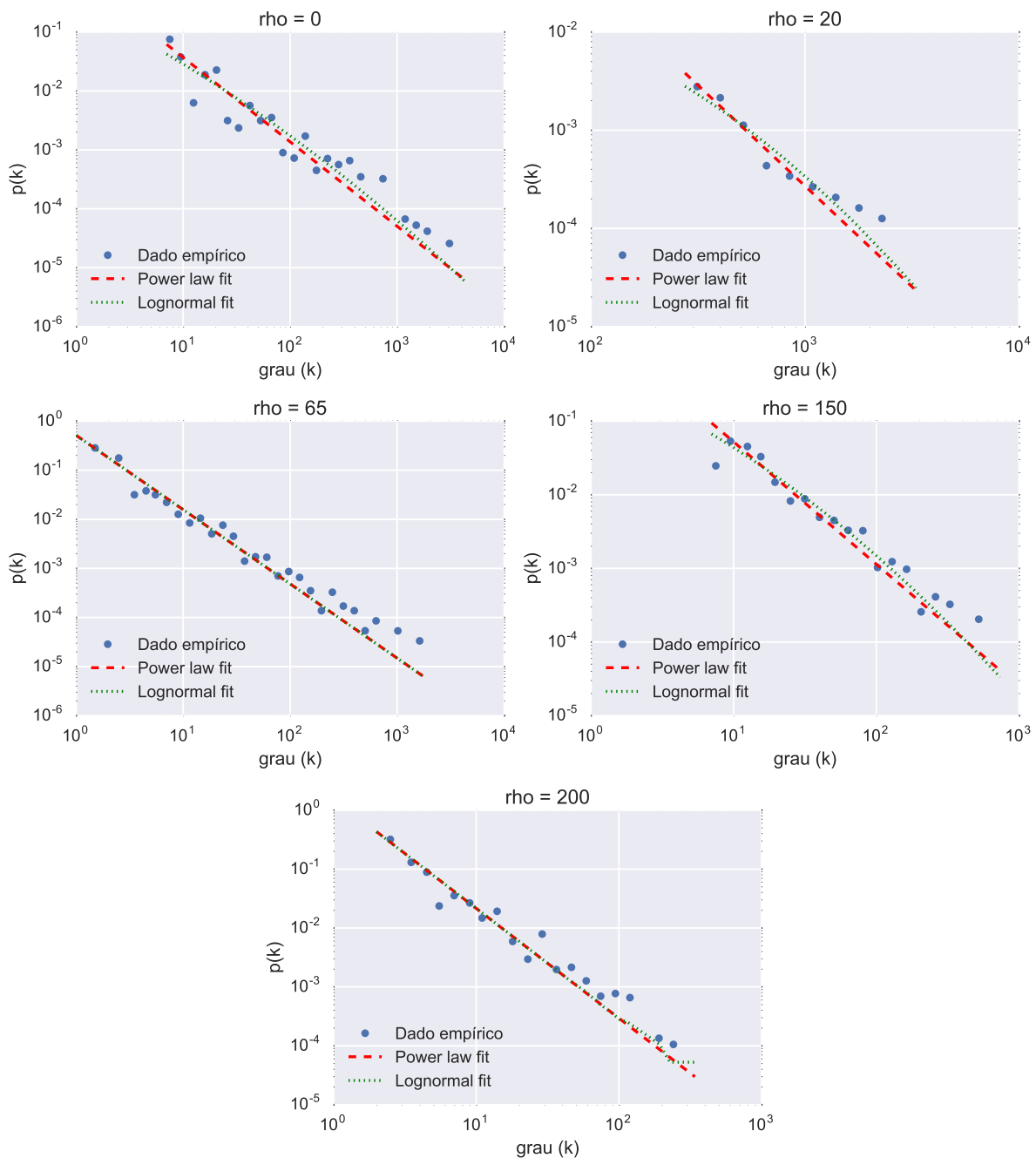


Figura 24 – Distribuição de grau dos nós em escala logarítmica base 10 com fit dos dados em power law e lognormal. O fit é feito para 5 diferentes redes  $\rho$ . As séries podem ser descritas pelas duas distribuições comparadas.



A dificuldade em se analisar leis de potência acaba inviabilizando a comparações de resultados em diferentes trabalhos, já que muitas vezes os métodos usados não são claros ou não seguem um padrão comum. Por esse motivo é utilizado o pacote *powerlaw* em Python (ALSTOTT; BULLMORE; PLENZ, 2014), que é gratuito e possibilita o uso de diferentes tipos de distribuição para teste dos dados, como pode ser visto nos gráficos da figura 24.

Os dados de entrada são criados no formato de histograma e valores com zero são eliminados automaticamente no *fit*, já que os eixos seguem escala logarítmica. A análise é feita para as 5 redes  $\rho$  escolhidas anteriormente, onde percebe-se um resultado indistinguível entre lei de potência ou exponencial, ou seja, ambas as distribuições podem ser usadas para descrever o comportamento da reta.

Na tabela 1 são apresentados os valores de  $\alpha$  obtidos com o respectivo desvio padrão, onde apenas na rede  $\rho = 20$  é obtido um valor interessante que pode indicar a propriedade livre de escala. Na mesma tabela tem-se R que corresponde a razão da probabilidade entre as duas distribuições testadas. Um valor de R positivo indica que os dados são mais prováveis na primeira distribuição (*power law*) e negativo se os dados são mais prováveis na segunda distribuição (*lognormal*), onde p é o valor de significância dessa relação. Como pode ser visto, todas as distribuições aparentemente podem ser descritas por uma função Lognormal.

Redes	$\alpha$	$\sigma$	R	p
$\rho = 0$	1,433	0,058	-1,992	0,236
$\rho = 20$	2,051	0,271	-0,401	0,555
$\rho = 65$	1,512	0,040	-0,007	0,440
$\rho = 150$	1,661	0,073	-2,594	0,136
$\rho = 200$	1,863	0,066	-0,820	0,217

Tabela 1 – Tabela com os resultados obtidos no *fit* da distribuição de grau dos nós nas 5 redes  $\rho$  analisadas.

A discussão a respeito de uma série de dados seguir ou não uma lei de potência é extensa, mesmo a utilização de alguns testes de comparação de melhor *fit* é muitas vezes insuficiente para uma conclusão. Sistemas encontrados no mundo real têm ruído, dessa forma alguns fenômenos empíricos onde espera-se uma lei de potência com a perfeição de uma distribuição teórica raramente ocorrem. Sempre haverá uma distribuição com melhor *fit* dos dados, até que seja encontrada alguma que descreva exatamente os valores e frequências obtidos.

O teste se uma rede é pequeno-mundo ou não também não é trivial, uma das formas possíveis é comparar o comprimento do caminho médio e o coeficiente de clusterização com uma rede aleatório do tipo Erdős e Rényi (ER) de mesmo número de nós  $n$  e *links*  $m$ . Formalmente, [Humphries e Gurney \(2008\)](#) apresentam como um dos possíveis métodos que a rede tenha um comprimento do caminho médio  $L$  similar à da rede aleatória  $L_{rand}$ , mas um coeficiente de clusterização  $C$  maior que na rede aleatória  $C_{rand}$ . As mesmas 5 redes foram analisadas usando este critério e seus resultados podem ser visualizados na [tabela 2](#).

Redes	$n$	$m$	$C$	$C_{rand}$	$L$	$L_{rand}$
$\rho = 0$	19169	98869	0,0667	0,0005	54,99	6,19
$\rho = 20$	16825	98869	0,0887	0,0006	50,93	5,69
$\rho = 65$	12150	98869	0,1464	0,0014	34,05	4,71
$\rho = 150$	6359	98869	0,1879	0,0049	24,46	3,50
$\rho = 200$	3180	98869	0,1651	0,0195	18,93	2,72

Tabela 2 – Tabela com os resultados do teste de pequeno-mundo para as 5 redes  $\rho$  comparadas com redes aleatórias do tipo Erdős e Rényi.

Nas redes empíricas,  $C$  é consideravelmente maior para todos as redes aleatórias, mas o comprimento do caminho médio varia drasticamente e em nenhum dos grafos aleatórios se aproxima das redes reais. Reproduzir o critério de  $L$  é extremamente difícil já que o modelo ER é construído atribuindo unicamente cada *link* à um par de nós com probabilidade uniforme, desconsiderando o quanto a rede é esparsa. Isso faz com que a rede se aproxime muito de um componente conectado pela proporção de nós e *links*, como pode ser visto com o baixo  $L_{rand}$  na rede  $\rho = 200$ . Com apenas um critério atendido, não há indícios suficiente para caracterizar se as redes são do tipo pequeno-mundo.

Como mencionado anteriormente, tem-se o objetivo de juntar as linhas fragmentadas do sistema em um único componente, resultado obtido na rede de  $\rho = 65$  como visto na [figura 20a](#). Hipoteticamente, nesse estado o sistema de transporte apresenta maior proximidade com a realidade, pois os *links* que não existiam anteriormente entre estações e terminais próximos (ampliação da [figura 17](#)) agora foram estabelecidos, representando a possibilidade de transferências que o usuário de fato possui.

Na [tabela 3](#) tem-se o *ranking* com os pontos de maior conectividade na rede, que aparecem em maioria relacionados aos terminais de ônibus com integração ao sistema ferroviário. As demais localidade que não tem conexão com o sistema ferroviário são regiões periféricas distantes do centro, como o terminal V. Nova Cachoeirinha, Terminal A. E.

Carvalho e pontos no Estr. Do M'boi Mirim. A importância dos pontos apresentados é proporcional ao seu grau, ou seja, quanto mais conexões o mesmo apresentar maior a quantidade de linhas de serviço interrompidas quando removido. Esse impacto pode ser observado nos resultados das simulações de ataques que serão apresentados a seguir, onde esses terminais são os primeiros a serem removidos no cenário alvo determinístico.

Qtd. pontos	ID	Grau	Descrição do ponto / Região
42	v900	1005	Terminal Metrô Santana (ZN)
24	v663	563	Terminal V. Nova Cachoeirinha (ZN)
23	v2109	446	Terminal A. E. Carvalho (ZL)
8	v3148	369	Estr. Do M'boi Mirim com João De Lima (ZS)
38	v6405	326	Terminal Barra Funda (ZO)
28	v1011	322	Terminal Estação CPTM Guaianazes (ZL)
8	v2708	316	Estr. Do M'boi Mirim com Nicolino Léo (ZS)
40	v3621	312	Terminal Lapa (ZO)
23	v7598	300	Terminal Cidade Tiradentes (ZL)
16	v183	242	Terminal Penha (ZL)

Tabela 3 – Tabela com *ranking* dos 10 pontos de maior grau e sua descrição/localidade correspondente na rede de  $\rho = 65$ .

Na mesma tabela, tem-se na coluna “Qtd. pontos” a quantidade de pontos (paradas de ônibus e estações) que foram agrupados para resultar no novo ID de nó gerado. O grau máximo de 1005 obtido é a soma de todos os graus desses nós originais. Conseqüentemente, todas as linhas que passavam por esses pontos anteriores passam a estar ligados ao novo ID, gerando a grande quantidade de auto *links* que pode ser observada na figura 25.

O grafo da PTN de São Paulo é tão esparsa que não é possível obter muita informação prática a partir de sua imagem. Os padrões obtidos não representam sua forma no espaço físico, trata-se apenas do resultado das forças de gravidade e repulsão do algoritmo de layout utilizado<sup>3</sup>. Contudo, é possível observar visualmente a presença de *hubs* altamente conectados. Os grandes laços em laranja ao redor dos nós representam a quantidade de auto *links* para aquele nó, que refletem naturalmente o *ranking* apresentado na tabela anterior.

Com isso, observa-se vários caminhos destacados em relação aos demais, resultado que pode estar associado à assortatividade da rede, já que visualmente nota-se que esses nós mais conectados estão ligados à outros nós de mesmo tipo, formando de certa forma uma trilha. O mesmo acontece para os nós pobremente conectados, que formam de certa forma um padrão de conexão com nós do mesmo tipo, mas também se conectam as trilhas

<sup>3</sup> Layout processado via Cytoscape: Algoritmo AllegroLayout.

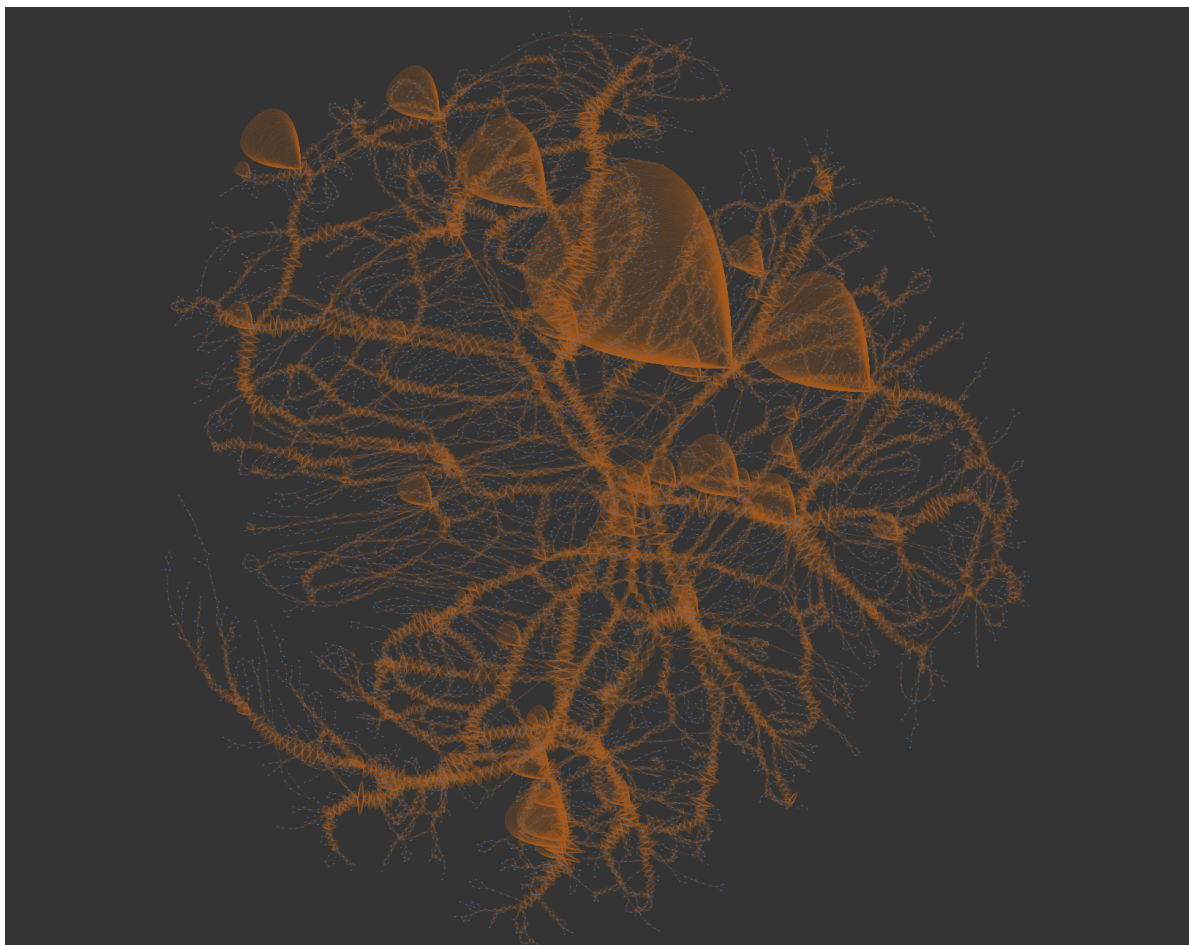


Figura 25 – Grafo da rede de São Paulo para  $\rho = 65$ . A forma da figura representa o resultado das forças de gravidade e repulsão do algoritmo de layout e não necessariamente sua forma geográfica.

de maior conectividade, o que lembra o sistema irrigatório cardiovascular com sua estrutura em ramificações. Esse padrão que pode ser interpretado como a oferta de linhas para essas regiões do grafo, o que poderia ser melhor investigado se espacializado com as coordenadas dos nós, podendo talvez indicar relações de oferta de transporte com aspectos de uso do solo e características sociais dos residentes daquela área.

## 4.2 Simulação de ataque à rede de transporte de São Paulo

Como demonstrado na sessão anterior, 41 diferentes grafos foram criados para os distintos valores de  $\rho$ . Desses, 5 grafos foram selecionados para execução dos ataques e testes de robustez. O primeiro grafo em  $\rho = 0$  foi selecionado por ser o estado inicial do sistema, sem nenhum tipo de agrupamento executado. Em  $\rho = 20$  ocorre uma mudança brusca no coeficiente de assortatividade, como pode ser visto na figura 18c. O valor de

raio em que a rede se torna um componente único ocorre em  $\rho = 65$ , motivo pelo qual é selecionado. Em  $\rho = 150$  tem-se um declínio mais acentuado na curva do coeficiente de clusterização (figura 20c) e também é um valor intermediário entre o último selecionado e o raio máximo,  $\rho = 200$ , escolhido por esse motivo.

Os resultados apresentados mostram comparações de diversas métricas da rede diante da estratégia de nós aleatórios, nós alvos determinístico e nós alvos probabilístico para capturar seus respectivos impactos e qual resulta em maior dano. Primeiramente é feita a análise das métricas diante da remoção de nós, as mesmas estratégias para *links* serão avaliadas separadamente na sequência. Apesar do componente único ser obtido a partir de  $\rho = 65$ , é importante testar a rede em estados anteriores (com componentes desconectados) para investigar se o agrupamento influencia de alguma forma no aumento da robustez do sistema.

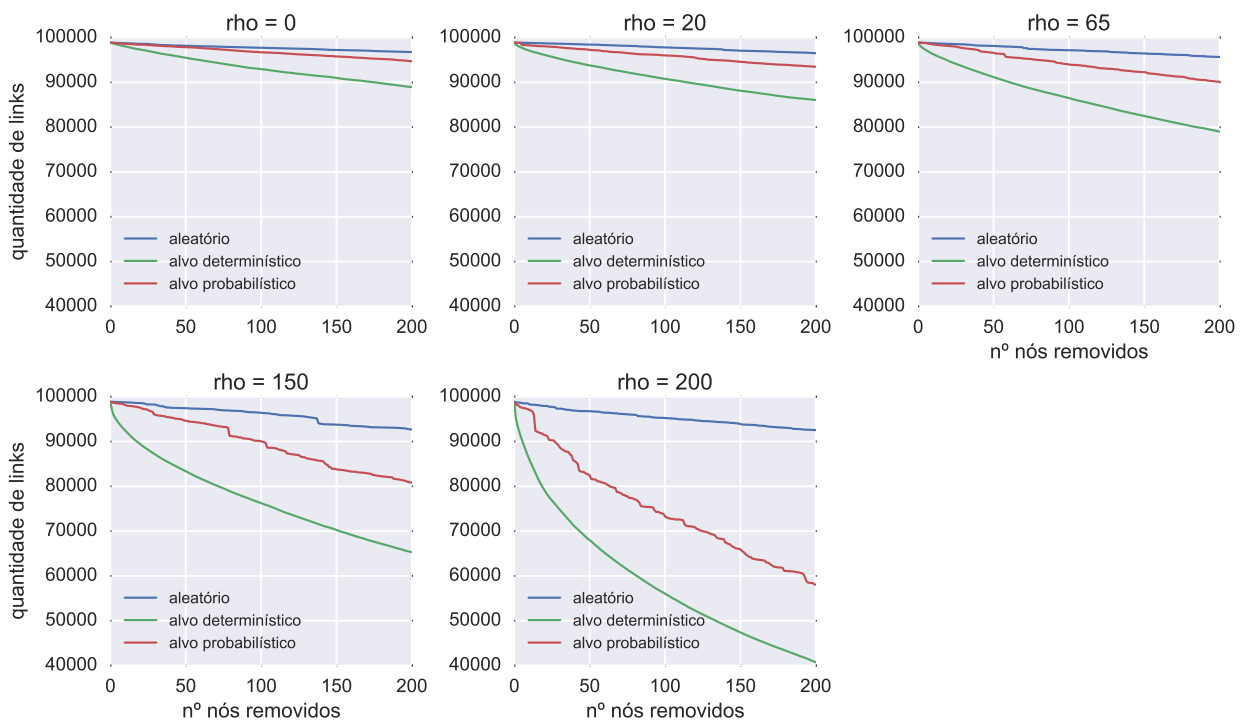


Figura 26 – Impacto na quantidade de *links* em função dos nós removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

A primeira característica analisada na remoção dos nós é o impacto na conectividade da rede, ou seja, a estrutura de seus *links* à medida que nós são removidos. Para isso são comparadas as três estratégias de remoção de nós para os valores de  $\rho$  selecionados. É feito um corte no gráfico da figura 26 para facilitar a visualização do comportamento da curva. Aparentemente a quantidade de *links* diminui em função dos nós removidos no

cenário aleatório, percebendo-se um modesto aumento do impacto nas redes de  $\rho$  maior. Entretanto, é nítido o quanto a estratégia alvo determinístico causa maiores danos à rede, como pode ser observado em  $\rho = 200$  onde 60% dos *links* chegam a ser eliminados.

Percebe-se também que a estratégia alvo probabilístico apresenta um resultado mais próximo do aleatório pra redes de  $\rho$  menor (0, 20 e 65), mas à medida que o agrupamento aumenta se aproxima de um valor intermediário, tendendo para o determinístico como pode ser visto em  $\rho = 200$ . É importante lembrar que a estratégia aleatório é apenas um dos possíveis resultados, já que a o efeito apresentado não é uma média de várias execuções. Entretanto, a estratégia probabilística serve como um bom indicador dos possíveis resultados aleatórios.

Esse comportamento pode estar ocorrendo devido à remoção dos nós ricamente conectados e que possuem uma grande quantidade de auto *links*, como em  $\rho = 200$  que chega a ter nós de grau maior que 6000, consequentemente os primeiros a serem removidos. Inclusive, percebe-se que a maior porção de *links* já é removida nas 50 primeiras interações para os 5 redes  $\rho$  analisados. Essa curva de impacto vai ficando cada vez mais acentuada à medida que a rede é mais densa.

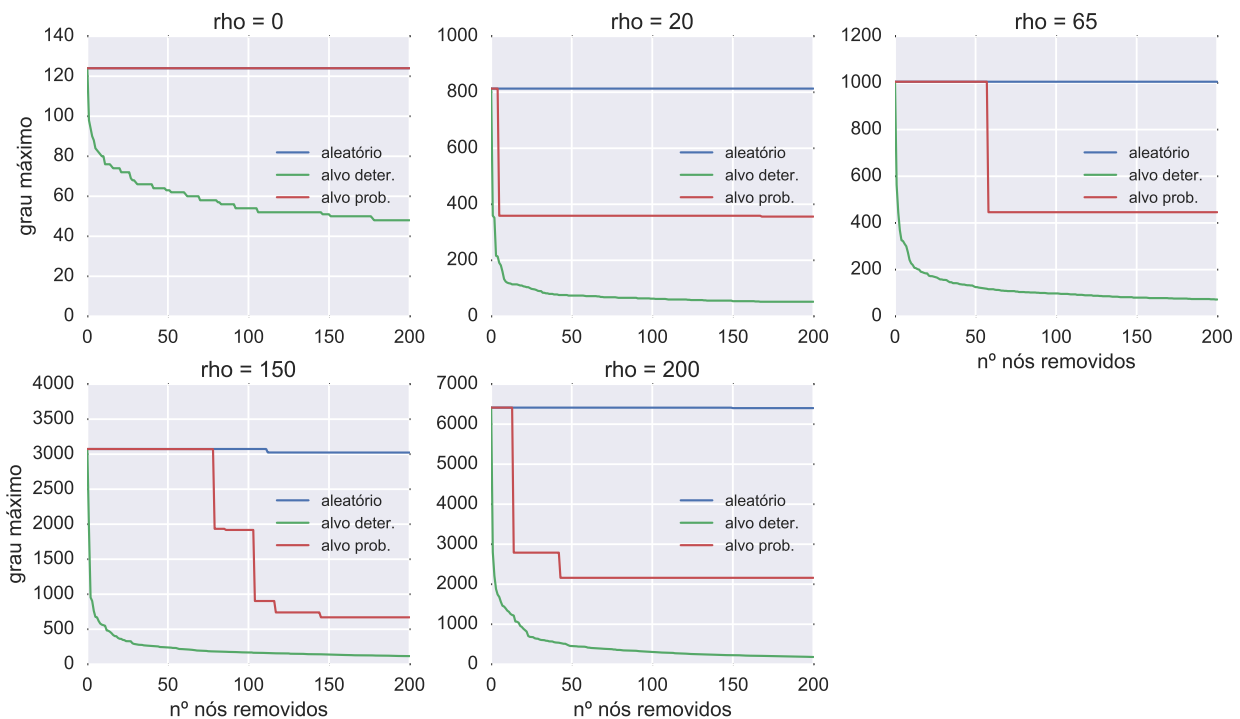


Figura 27 – Impacto no grau máximo da rede em função dos nós removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

Apesar do grau máximo na rede ser diferente para cada  $\rho$ , percebe-se que o impacto ao remover os nós mais conectados é bem maior em relação à estratégia aleatória. Em  $\rho = 0$  o grau máximo cai pela metade nas 50 primeiras interações na estratégia determinístico, enquanto que nas demais redes tem-se uma redução de em média 6 vezes seu valor original (figura 27). Na estratégia aleatório essa redução é quase imperceptível e na probabilístico tem-se quedas bruscas tendendo a acompanhar os resultados determinísticos. Como visto nas distribuições de grau na sessão anterior (figura 24), nós ricamente conectados tem frequência bem menor em relação aos demais, uma possível explicação para a curva permanecer relativamente estável após a remoção desses nós de grau alto (*hubs*).

Nos componentes da figura 28 percebe-se o mesmo tipo de resultado quanto à um maior impacto para nós alvos em comparação à aleatórios. Em  $\rho = 0$  e 20 as redes já iniciam num estado fragmentado, tem-se como resultado o quanto esse estado se amplia, enquanto nas demais redes parte-se de uma componente único. Percebe-se que a estratégia alvo probabilístico margeia o aleatório de forma mais consistente, exceto na rede de  $\rho = 200$  onde ocorre um descolamento e tem-se um resultado intermediário diante das demais estratégias.

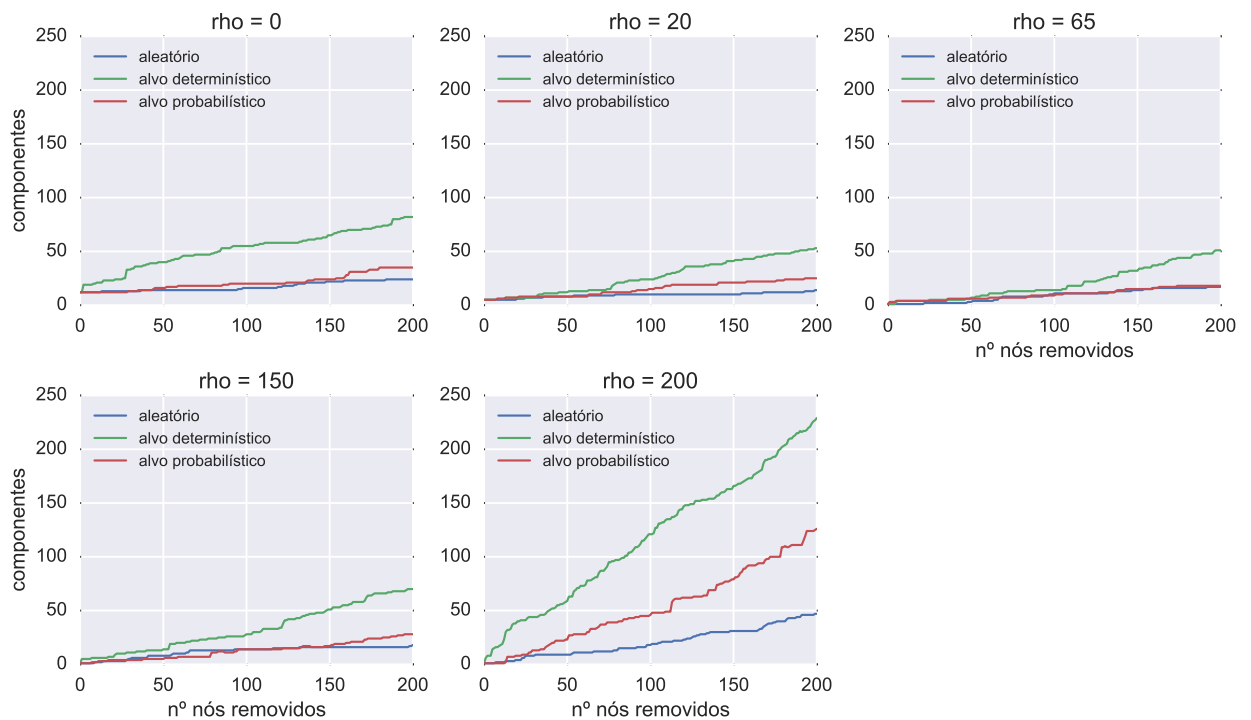


Figura 28 – Impacto na quantidade de componentes da rede em função dos nós removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

Se definida a fragmentação como critério de robustez, as redes de  $\rho = 20, 65$  e  $150$  aparentam ser mais robustas em relação às demais, sendo possível observar que durante as 50 primeiras interações, a quantidade de componentes permanece com fragmentação mínima, principalmente em  $\rho = 65$ . A rede de  $\rho = 200$  que é a mais conectada e intuitivamente poderia ser mais robusta foi a que apresentou fragmentação mais rápida, podendo ser resultado da menor quantidade de nós, já que a proporção de 200 nós removidos de um total de 3180 é bem maior que em  $\rho = 65$  por exemplo onde tem-se 12.150 nós.

Outra propriedade importante da rede diz respeito ao comprimento de seus caminhos médios entre dois pontos aleatórios. Na figura 29 percebe-se que esse caminho médio aumenta à medida que nós são removidos. Mais uma vez a estratégia alvo determinístico apresenta maior dano na rede diante de aleatório. A estratégia probabilística margeia a aleatória nas três primeiras redes, se descolando em  $\rho = 150$  e  $200$  quando apresenta novamente um resultado intermediário.

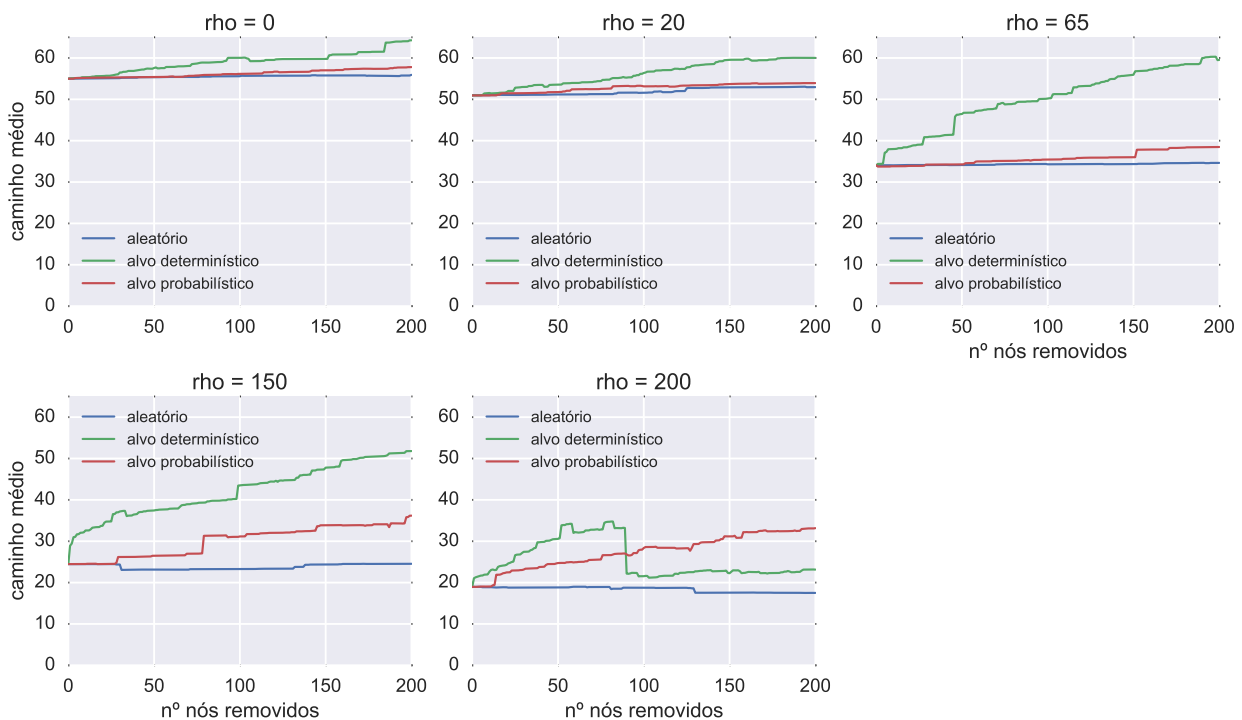


Figura 29 – Impacto no comprimento médio dos caminhos da rede em função dos nós removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

Contudo, tem-se uma queda brusca no caminho médio na rede  $\rho = 200$  próximo de 90 interações, que se mantém com pequenas flutuações ao longo dos demais nós removidos. Na mesma rede a estratégia probabilística mantém uma curva de aclave contínuo como



esperado, resultado similar da rede de  $\rho = 150$ . Como alguns nós na rede tem o papel de pontes entre componentes, é intuitivo pensar que o comprimento dos cominhos tende a aumentar à medida que esses nós são removidos, já que um novo caminho pode resultar em mais passos. O comportamento observado é bastante inesperado, já que a rede mais densa oferece mais redundâncias em relação à quantidade de *links* disponíveis.

Talvez este resultado esteja relacionado com a quantidade de componentes da rede, que como visto anteriormente na figura 28 fragmenta-se bem mais rápido em comparação às demais. Outra possível influência é a grande presença de nós altamente conectados, como o algoritmo de agrupamento transformou diversos pontos próximos em um só, teoricamente aqueles caminhos existentes viraram um auto *link*, que não contribuem para os caminhos da rede, apenas para o grau do nó. Desse modo, talvez o sistema esteja indicando que exista um valor de equilíbrio para os agrupamentos, sendo que o mesmo feito de forma excessiva não contribui para a robustez da rede.

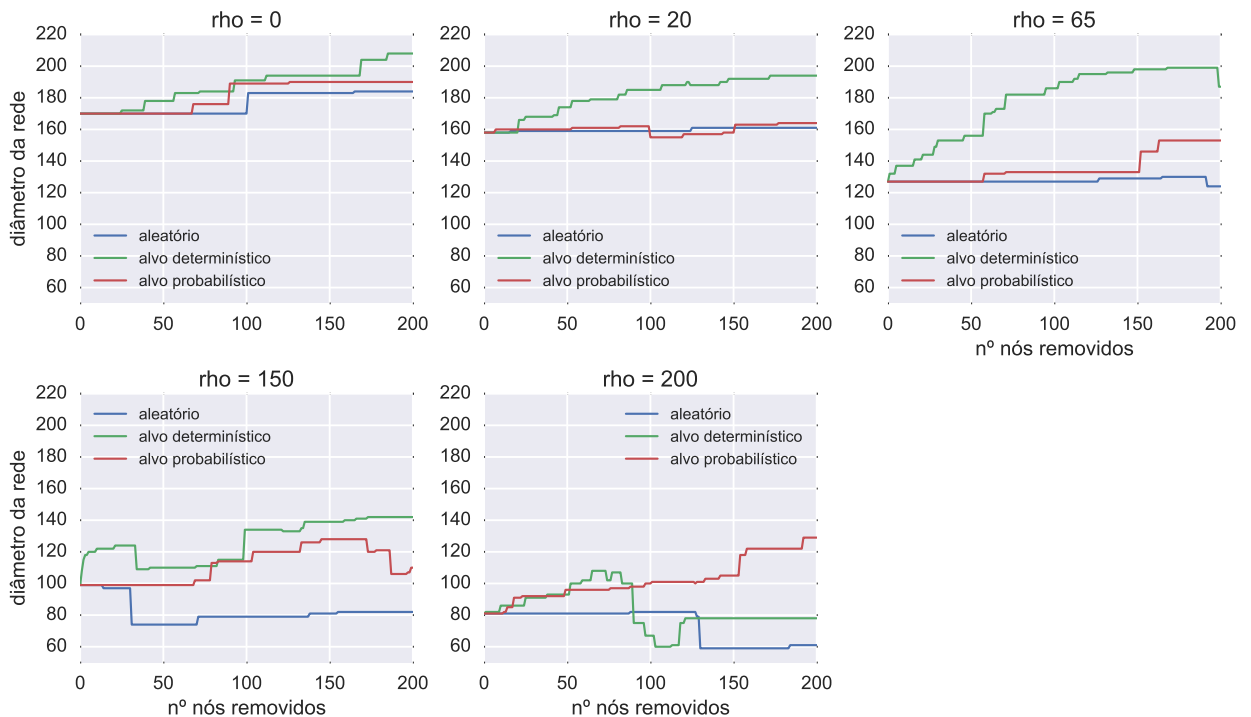


Figura 30 – Impacto no diâmetro da rede em função dos nós removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

Associado diretamente ao comprimento dos caminhos médios, o diâmetro captura a maior distância entre dois pontos na rede (figura 30). Espera-se que a distância aumente a medida que nós são removidos, mas a remoção de muitos nós pode causar impacto inverso pois consequentemente diminui-se também o tamanho da rede em relação à quantidade

de nós e *links*. Pelo corte na série no gráfico, obtém-se a mesma diferença de impacto da estratégia aleatório em relação à alvo encontrada nos casos anteriores, inclusive com a estratégia probabilística margeando a aleatória.

Em  $\rho = 200$  é observado um resultado inconclusivo, tem-se um aumento no diâmetro até  $\approx 60$  nós removidos com uma queda brusca nas próximas interações, seguido de outro aumento e uma provável estabilização. Talvez este resultado esteja relacionado com o comprimento médio dos caminhos, já que um tipo parecido de comportamento é observado no mesmo intervalo de interações como visto anteriormente na figura 29. De fato, o coeficiente de correlação de Pearson entre o comprimento médio dos caminhos e o diâmetro é positivo, sendo 0.90 para a rede  $\rho = 200$  e 0.96 para a série completa de dados com todas as redes.

Entretanto, para obtenção de um resultado mais confiável, seria interessante executar o processo de remoção por mais etapas e observar seu comportamento até uma proporção maior de nós removidos. Isso possibilita entender como essa curva se comporta, pois pode ser simplesmente uma questão de escala e o intervalo analisado pequeno demais para concluir se há uma tendência. Ainda vale salientar que em  $\rho = 200$  a quantidade de nós da rede é bem menor que nas demais, esses nós ricamente conectados podem gerar um resultado imprevisível quando removidos.

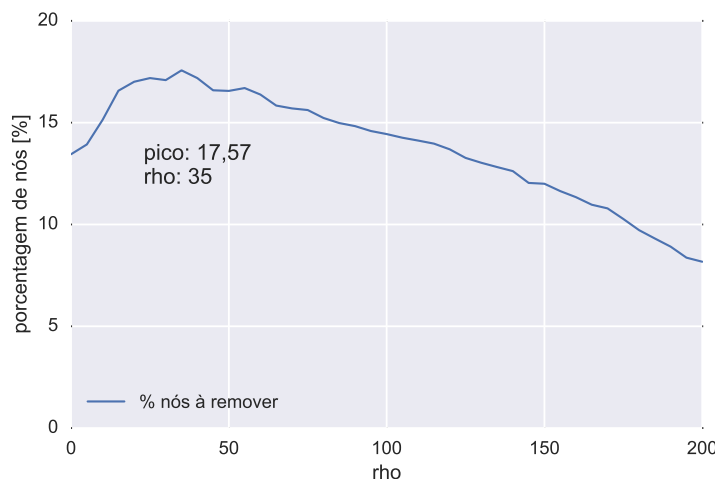


Figura 31 – Percentual de nós necessários para dividir a rede em dois ou mais componentes em função de  $\rho$ . Cada valor de  $\rho$  é uma rede diferente.

Finalizando a análise dos cenários de ataques aos nós, observa-se na figura 31 a proporção em percentual de nós que precisam ser removidos para que a rede seja dividida

em pelo menos dois componentes em função de  $\rho$ . Isto é, para cada uma das 41 redes criadas, a quantidade de nós que precisam ser removidos para dividir a rede em pelo menos duas partes. Tem-se o pico dessa quantidade na rede de  $\rho = 35$ , onde é necessário remover 17,16% de seus nós, que corresponde à 937 nós do total de 16.103 que essa rede possui.

Se usado este índice como critério de robustez, quanto maior a proporção, mais robusta é a rede. Com isso, é possível afirmar que em  $\rho = 35$  tem-se a rede mais robusta para os agrupamentos realizados, sendo a rede de  $\rho = 200$  o pior dos cenários, onde apenas 8,17% de seus nós são necessários para divisão da rede. Já em  $\rho = 65$ , rede de maior interesse nessa análise, a proporção de nós necessários é de 15,84%, valor relativamente próximo do máximo obtido. Os melhores resultados foram observados para o intervalo das redes de  $\rho = 10$  à  $\rho = 80$  onde as proporções de nós a serem removidos são maiores. A tabela 4 apresenta os valores obtidos na execução do cenário de pontos de articulação.

$\rho$	nós	cortes	$\rho$	nós	cortes	$\rho$	nós	cortes	$\rho$	nós	cortes
<b>0</b>	19169	1425	<b>50</b>	13366	807	<b>100</b>	9817	680	<b>150</b>	6359	530
<b>5</b>	19061	1368	<b>55</b>	12946	775	<b>105</b>	9485	665	<b>155</b>	6017	517
<b>10</b>	18606	1229	<b>60</b>	12550	766	<b>110</b>	9149	648	<b>160</b>	5661	499
<b>15</b>	17698	1068	<b>65</b>	12150	767	<b>115</b>	8829	632	<b>165</b>	5310	484
<b>20</b>	16825	989	<b>70</b>	11778	750	<b>120</b>	8462	618	<b>170</b>	4998	463
<b>25</b>	16103	937	<b>75</b>	11432	732	<b>125</b>	8110	611	<b>175</b>	4664	454
<b>30</b>	15452	904	<b>80</b>	11102	729	<b>130</b>	7740	594	<b>180</b>	4354	448
<b>35</b>	14865	846	<b>85</b>	10784	720	<b>135</b>	7383	576	<b>185</b>	4049	435
<b>40</b>	14333	834	<b>90</b>	10468	706	<b>140</b>	7031	557	<b>190</b>	3750	421
<b>45</b>	13819	833	<b>95</b>	10143	695	<b>145</b>	6681	555	<b>195</b>	3449	412
									<b>200</b>	3180	389

Tabela 4 – Tabela com a quantidade de nós necessários para dividir a rede em pelo menos dois componentes no cenário de pontos de articulação.

A seguir, tem-se o mesmo tipo de gráfico comparando as estratégias aleatório, alvo determinístico e alvo probabilístico no contexto de cenários de ataque aos *links*. As mesmas métricas e métodos são usadas como descrito na sessão de cenários de ataque com o intuito de, além de observar a diferença entre as duas estratégias, compreender também os tipos de elementos que causam mais dano à rede.

O mesmo comportamento observado no grau máximo para o cenário de nós se repete em *links* (figura 32). Para as 5 redes analisadas a estratégia de alvo determinístico apresenta maior impacto em relação à aleatório e probabilístico, sendo mais nítido nas redes de  $\rho = 20$  e 65, mas quase imperceptível em  $\rho = 0$ . A estratégia probabilística continua mantendo a tendência de margear a aleatória. Percebe-se que o grau máximo se

mantem estável para a estratégia aleatória como nos cenários de nós, mas em compensação a redução desse índice é bem menor que a obtida nos cenários de nós. É possível que este resultado esteja associado à grande quantidade de *links* conectados aos nós a partir de  $\rho = 20$ , principalmente aos auto *links*, que quando removidos não geram impacto nenhum na rede, apenas na distribuição de grau. Uma variação interessante desse cenário seria desconsiderar os auto *links* durante a remoção, usando-os apenas para cálculo do peso.

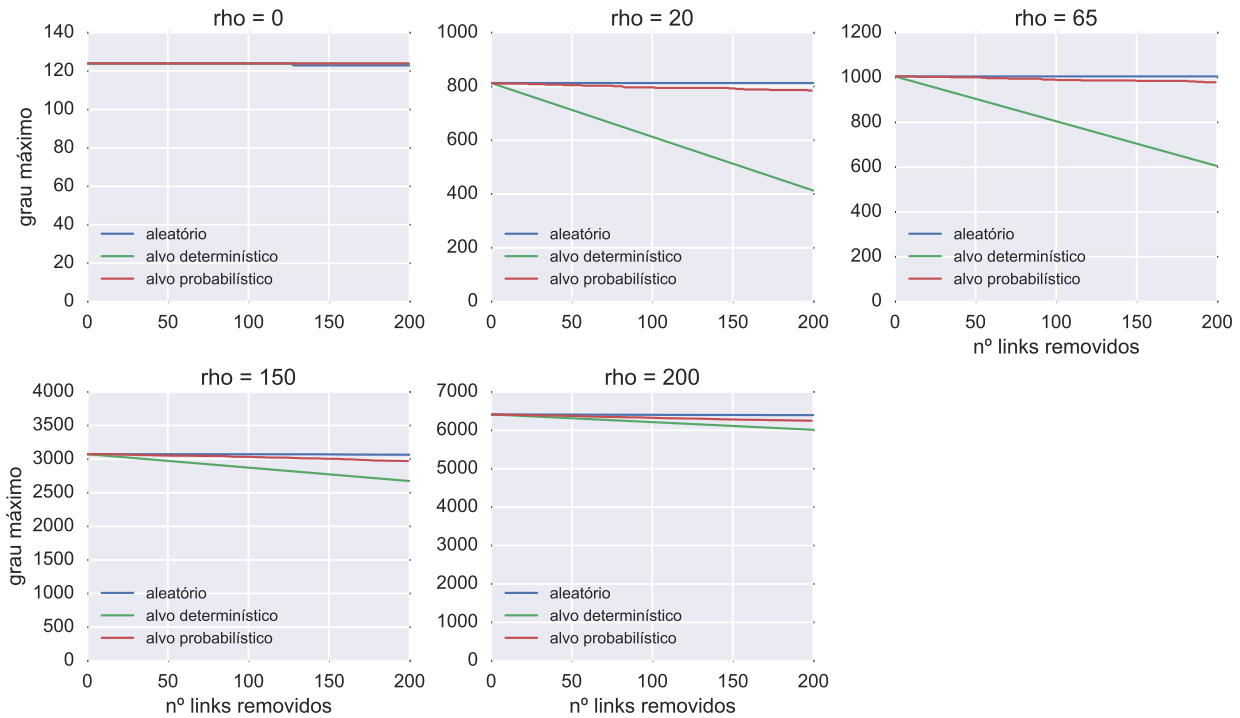


Figura 32 – Impacto no grau máximo da rede em função dos *links* removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

É importante também esclarecer que o peso do *link* aqui atribuído refere-se a sua multiplicidade na rede, isto é, a quantidade de vezes que aquele caminho entre dois pontos é usado pelas linhas do sistema. Esse peso é um cálculo simples que se baseia em frequência, visto que dados de bilhetagem, lotação dos veículos e capacidade do sistema não estão facilmente disponíveis. No entanto, este peso reflete satisfatoriamente a importância do *link* no contexto de robustez analisado, visto que a remoção do mesmo poderá influenciar diretamente na quantidade de linhas que seu índice obtido reflete.

As variações foram bem sutis na quantidade de componentes como pode ser visto na figura 33, sendo as redes  $\rho = 65$  e 200 as mais robustas se usado este critério, mantendo-se como em componente único durante todo o processo de ataques. Nas demais redes, tem-se um resultado parecido com o cenário de nós, só que numa escala de fragmentação

bem menor nas redes  $\rho = 0$  e 20. No entanto, a estratégia aleatória acaba gerando mais fragmentação de forma geral, por mais que pequena, se destacando até da probabilística. É possível que haja uma relação com a imensa quantidade de *links*, que é invariável ao longo de  $\rho$  como visto previamente na figura 18a. Como sugerido na análise do grau máximo, talvez fosse interessante também remover os auto *links* e executar o cenário de ataques para mais interações, já que a escala é de  $\approx 100.000$  *links* e a remoção de apenas 200 tem impacto pequeno se levado em consideração que grande parte são auto *links*.

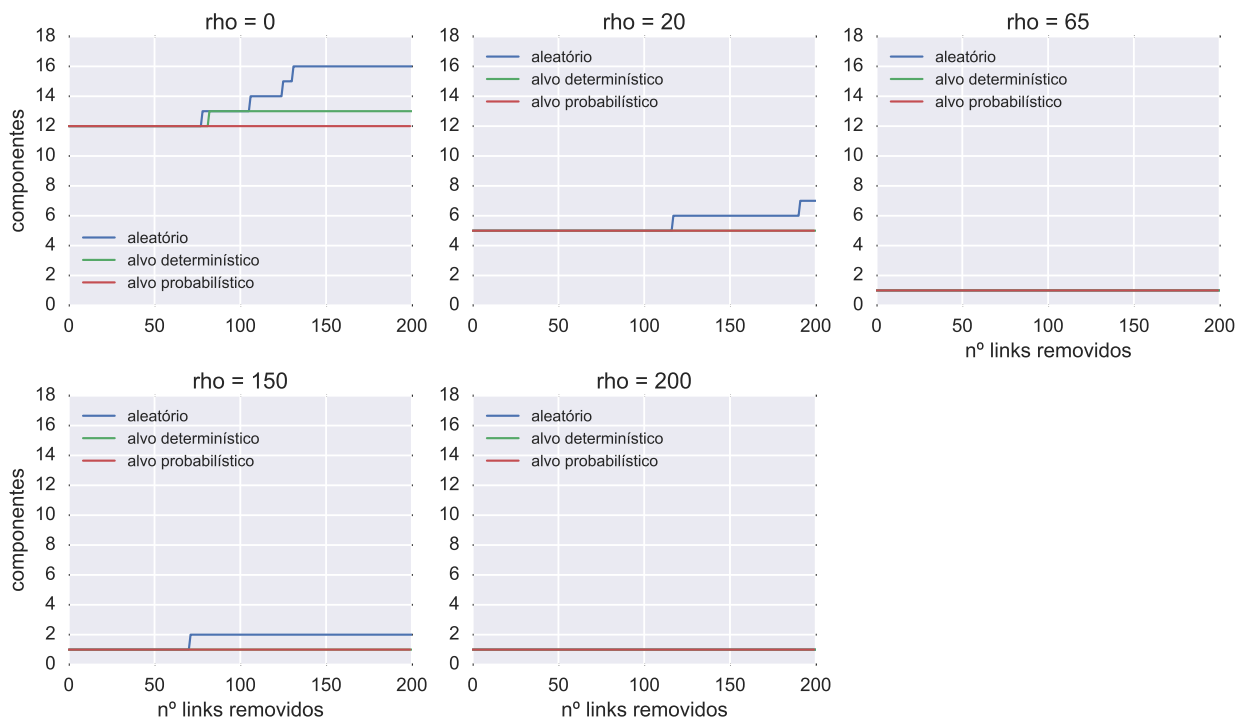


Figura 33 – Impacto na quantidade de componentes da rede em função dos *links* removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

Ao que parece, o caminho médio manteve-se estável para as redes analisadas, sendo necessário fazer um corte no gráfico da figura 34 para visualização das pequenas flutuações que ocorrem. A única alteração maior ocorre na rede  $\rho = 65$  na estratégia aleatória, mas assim como nos componentes, trata-se de apenas um dos possíveis resultados, já que a estratégia aleatória foi executada apenas uma vez.

A estrutura de *links* é um fator importante no cálculo dessa métrica, provavelmente a quantidade removida foi insuficiente para qualquer dano à rede mesmo removendo seus *links* de maior frequência. A grande presença de linhas de ônibus e metrô combinados torna o sistema relativamente robusto mesmo removendo seus caminhos mais usados,

ou seja, a distância média não aumenta com a ausência dos mesmos. É provável que este comportamento esteja relacionado a topologia pequeno-mundo que apresenta um alto índice de agrupamento, consequentemente resultando em redundância de caminhos. Entretanto, para um melhor entendimento do impacto dos *links* nesse cenário, talvez seja interessante alterar a métrica usada para uma que capture a centralidade das conexões, já que a multiplicidade não mostrou-se relevante como fator de impacto na conectividade.

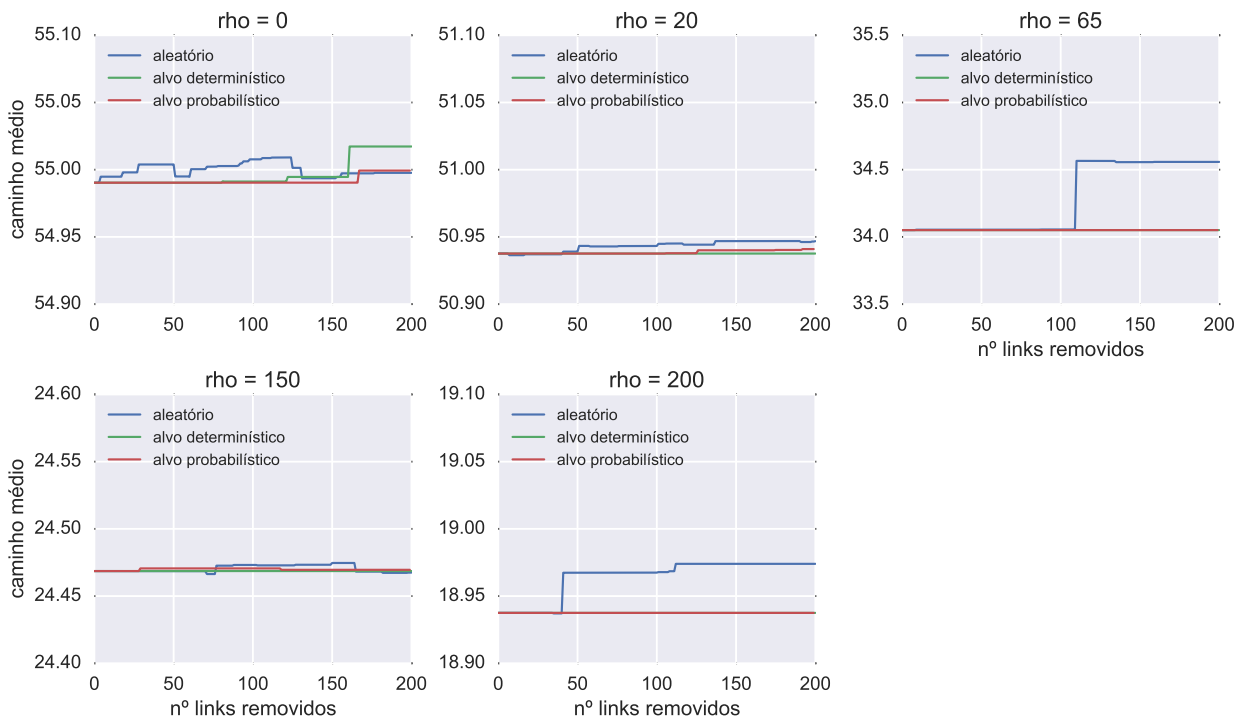


Figura 34 – Impacto no comprimento médio dos caminhos da rede em função dos *links* removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  selecionadas.

O diâmetro da rede é invariável ao longo das remoções dos *links*, exceto na rede de  $\rho = 65$  na estratégia aleatória que tem um leve aumento de 127 para 129 (figura 35). Como essa métrica representa a distância máxima entre dois pontos na rede, não é observada nenhuma contribuição do agrupamento na robustez da mesma, muito menos algum impacto na remoção dos *links*. O resultante é similar ao caminho médio já que essas métricas estão intrinsecamente ligadas, assim como a distância média não é alterada, remover caminhos importantes de uma linha não amplia a distância máxima. De fato, é possível afirmar que as redes analisadas são invariáveis em diâmetro para a proporção de *links* removidos, mas não é possível afirmar se existe alguma influência na robustez, visto que pequenas alterações ocorreram em todos os cenários de ataque à *links*.

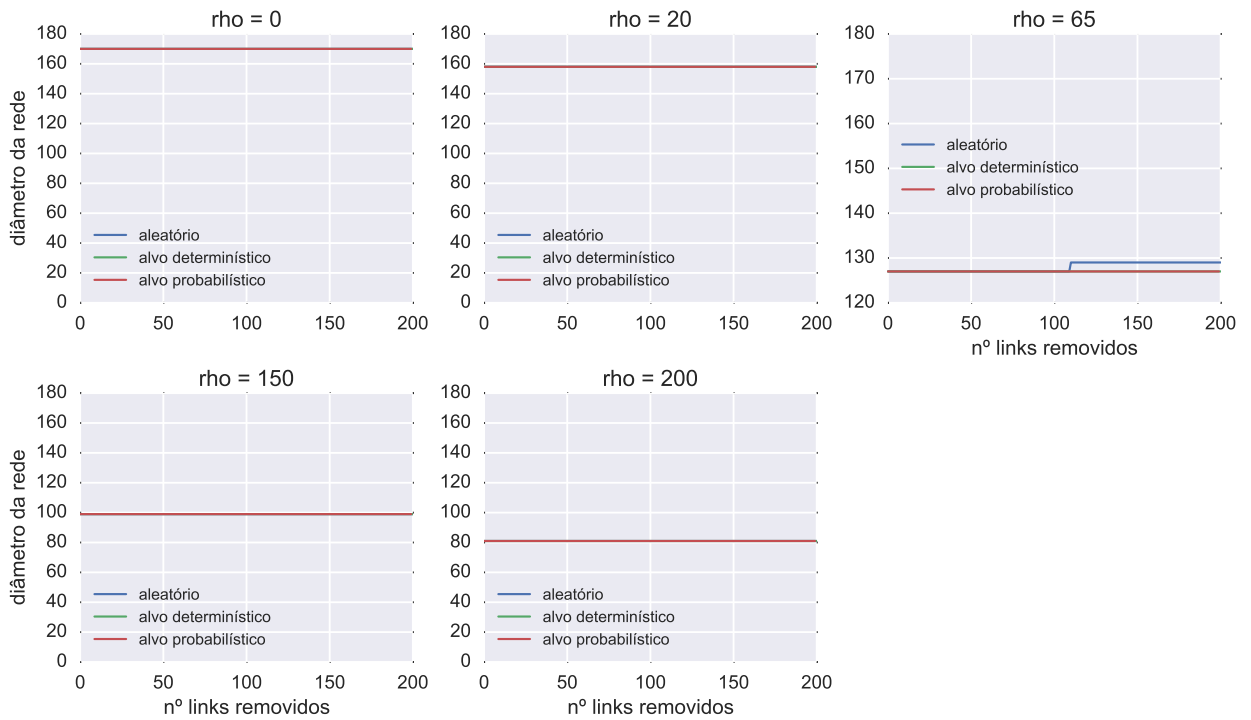


Figura 35 – Impacto no diâmetro da rede em função dos *links* removidos nas estratégias aleatório, alvo determinístico e alvo probabilístico para as 5 redes  $\rho$  seleccionadas.

Como mencionado anteriormente, a rede  $\rho = 65$  é a de maior interesse por ser o primeiro valor de raio à obter-se um componente único. É realizado um paralelo entre os cenários de nós e *links* nas estratégias alvo determinístico e probabilístico por apresentarem maior influência na estrutura da rede nas análises anteriores. Na figura 36 tem-se o grau máximo em função da remoção de nós e *links* (elementos), quanto mais reta ou inclinada positivamente a curva, melhor o resultado.

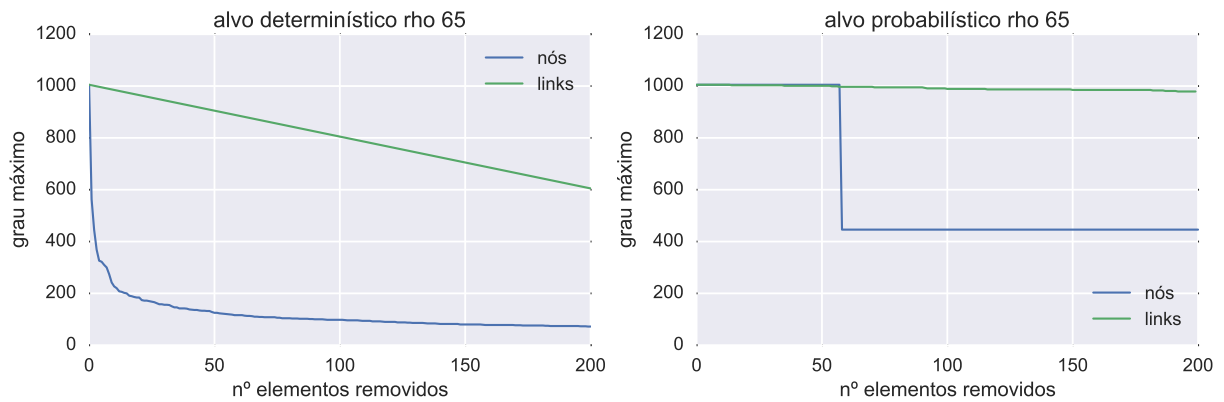


Figura 36 – Comparação do grau máximo em função dos nós e *links* removidos nas estratégias alvo determinístico e probabilístico na rede  $\rho = 65$ .

Ao que tudo indica, os *links* apresentam maior robustez em relação aos nós na

estratégia determinística, o grau máximo tem uma redução de apenas 1,9% nas 10 primeiras interações, enquanto o decaimento se dá de forma bem mais acentuada para os nós, reduzindo 77,4% no mesmo intervalo. Na estratégia probabilística, esse resultado é ainda mais evidente, o grau máximo mantém-se relativamente estável à medida que os *links* são removidos, tendo uma redução de apenas 2,58% para todas as rodadas de remoção. No caso dos nós, é observado um resultado semelhante aos *links* até o 58º elemento removido, onde tem-se uma queda brusca de 55,6% no grau máximo. Muito provavelmente o nó de maior grau (1005) foi removido nesta etapa devido à sua alta probabilidade.

Continuando a comparação, tem-se na figura 37 a quantidade de componentes, comprimento do caminho médio e diâmetro da rede em função da remoção de nós e *links* na estratégia determinística. Ao contrário do caso anterior, neste contexto quanto mais reta ou inclinada negativamente a curva melhor o resultado. Mais uma vez nota-se o nó causando maior impacto na rede em relação ao *link*. Percebe-se uma fragmentação que resulta em 51 componentes no ultimo nó removido, enquanto a rede mantém-se íntegra no caso dos *links*. O mesmo ocorre com o caminho médio e o diâmetro da rede, que aumentam 42,3% e 32% respectivamente, enquanto esses índices permaneceram estáveis para o *link*.

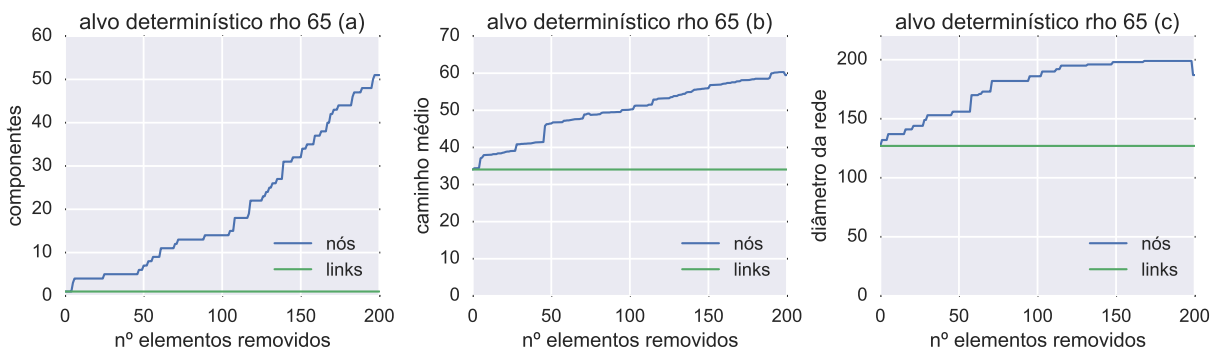


Figura 37 – Comparação da quantidade de componentes (a), comprimento do caminho médio (b) e diâmetro (c) em função dos nós e *links* removidos na estratégia alvo determinístico na rede  $\rho = 65$ .

Na estratégia probabilística (figura 38), apesar dos nós continuarem causando mais impacto à rede, a diferença é mais sutil, principalmente no comprimento do caminho médio e diâmetro da rede, onde as dois tipos mantem-se praticamente acoplados nos 50 primeiros elementos removidos. Há um descolamento maior somente a partir de 150 interações, ficando mais evidente no diâmetro. A fragmentação dos componentes segue um comportamento parecido com a estratégia determinística, mas com uma inclinação bem



menos acentuada.

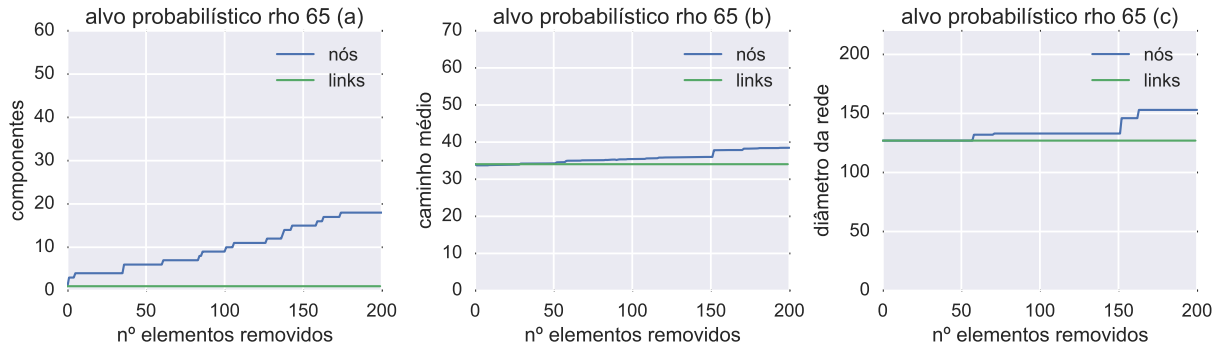


Figura 38 – Comparação da quantidade de componentes (a), comprimento do caminho médio (b) e diâmetro (c) em função dos nós e *links* removidos na estratégia alvo probabilístico na rede  $\rho = 65$ .

Tendo essas comparações em mente, é possível compreender esses resultados como a remoção dos grandes terminais de ônibus e estações de metrô próximas, que nesse  $\rho$  foram agrupados. O que faz sentido se comparado com o sistema real, já que removendo esses *hubs* dezenas de linhas serão interrompidas, conseqüente os pontos próximos também podem ser afetados, já que muitas vezes tem como destinos esses terminais. Diferente de interromper uma via ou uma linha do sistema, onde o usuário teoricamente pode se deslocar para o ponto mais próximo e utilizar alguma alternativa, inclusive, as próprias linhas de ônibus tem relativa flexibilidade para contornar obstáculos, diferente dos pontos que são estáticos.

É importante lembrar que os dois elementos comparados encontram-se distribuídos em proporções distintas na rede, numa escala de  $\approx 5$  *links* para cada nó. Para uma comparação mais equilibrada, seria necessário remover os elementos em termos de proporção, como por exemplo 20% de nós e *links*. Entretanto, tratando-se de uma rede tão grande, são necessários recursos computacionais superiores a equipamentos pessoais para tais cálculos, visto que a quantidade de *links* removidos salta de 200 para 20.000 considerando 20% como critério. Cada interação dessas 20 mil requer cálculos de caminho médio e diâmetro que computam valores médios numa matriz de 100.000 por 100.000 *links*, inviabilizando o processamento em computadores pessoais.



## 5 Conclusões

O algoritmo de agrupamento mostrou-se um método eficaz para o objetivo proposto, obtendo a junção da rede em um componente único no raio de 65 metros. Percebeu-se também que este processo de agrupamento possibilita testar políticas públicas de mobilidade, onde a distância pode ser usada para simular a predisposição do usuário em se locomover e acessar diferentes opções de transporte em relação a sua posição original. De fato, aumentar essa disponibilidade gera grande redução nas distâncias e passos necessários para atingir destinos, o que ficou evidente no caminho médio e diâmetro que reduziram. Entretanto, obtêm-se um ponto máximo dos benefícios desse agrupamento em  $\rho = 185$  metros, onde a relação de esforço em andar diante da redução das distâncias começa a diminuir de acordo com a medida de eficiência proposta. Ao agrupar pontos próximos, diminui-se a quantidade de saltos que um veículo faz para completar uma linha, o que hipoteticamente pode reduzir tempo de viagem e custos operacionais, mas um raio de distância muito grande pode fazer o passageiro se locomover além da sua disposição, perdendo tempo e conforto andando.

A análise estatística realizada mostra que a rede de transporte público de São Paulo tem um comprimento médio de caminho relativamente curto diante do seu diâmetro, possui grau médio dos nós baixo que aumenta em função de  $\rho$  e distribuição de grau que pode ser descrita tanto por lei de potência quanto lognormal. Os coeficientes  $\alpha$  obtidos nas 5 redes analisadas só apresentaram um valor interessante em  $\rho = 20$ , onde obtêm-se  $\alpha = 2.05$  indicando mecanismos diferentes da aleatoriedade para descrever o crescimento dessa rede (BARABÁSI; ALBERT, 1999). A lei de potência também evidencia a presença de *hubs* (terminais) nas redes e uma estrutura indiferente à escala, ou seja, mesmo reduzindo-se a quantidade de pontos e linhas a topologia de certa forma se mantém similar.

Apesar de redes PTN serem normalmente caracterizadas como pequeno-mundo, no teste realizado não foram obtidos indícios suficientes para confirmar tal propriedade, já que o método de comparação com redes aleatórias não atendeu os critérios definidos por Watts e Strogatz (1998). Entretanto, o coeficiente de clusterização obtido nas 5 redes testadas é significativamente maior comparado às redes aleatórias, mostrando uma probabilidade alta dos vizinhos de um ponto também serem conectados por uma linha, o que sugere

redundâncias.

A estratégia de ataque a alvo determinístico mostrou-se mais impactante na maioria dos cenários testados nas 5 redes escolhidas, quando não, apresentou no mínimo impacto igual. Essa estratégia representa uma ação que tem por objetivo interromper o funcionamento do sistema atacando seus elementos mais importantes. Se considerado a fragmentação como critério de robustez, as redes com raio de distância  $\rho = 20, 65$  e  $150$  foram mais robustas nos cenários de nós (paradas e estações), enquanto todas as demais nos cenários de *links* (linhas de serviço), exceto a rede  $\rho = 0$ . Ou seja, agrupar pontos próximos nesses raios de distância pode fazer com que o sistema de transporte apresente menos linhas desconectadas se removido um ponto ou caminho.

Ser mais conectado não necessariamente induz maior robustez como visto na rápida fragmentação da rede de raio  $\rho = 200$  no cenário nó alvo. O agrupamento contribui para a robustez até uma certa distância, visto que no cenário de pontos de articulação as redes com maior proporção de nós removidos para sua divisão foram obtidas para os raios no intervalo  $\rho = 10$  à  $\rho = 80$  metros (pico do gráfico 31). Na rede  $\rho = 65$  de maior interesse, percebe-se que os nós mais importantes são compostos por terminais e estação ferroviária integrada a estes. Os resultados quanto a sua robustez, relação de eficiência e distâncias mostraram-se frequentemente entre os melhores obtidos, sendo uma proposta de agrupamento para análises futuras em trabalhos similares.

Por fim, os resultados obtidos confirmam duas principais características em redes deste tipo: possuem um elevado grau de robustez em relação à falhas aleatórias, principalmente quando selecionados caminhos das linhas como visto nas simulações de ataque aos *links*. Entretanto, são particularmente vulneráveis a ataques que têm como alvo elementos importantes, especialmente os terminais nos cenários de simulação de ataque aos nós.

## Referências<sup>1</sup>

ALSTOTT, J.; BULLMORE, E.; PLENZ, D. powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, Public Library of Science, v. 9, n. 1, p. e85777, 2014. Citado na página 63.

ÂNTICO, C. Deslocamentos pendulares na região metropolitana de são paulo. *São Paulo em Perspectiva*, SciELO Brasil, v. 19, n. 4, p. 110–120, 2005. Citado na página 26.

ARBEX, R. O.; CUNHA, C. B. da. Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, Pergamon, v. 81, p. 355–376, 2015. Citado na página 20.

BAGLER, G. Analysis of the airport network of india as a complex weighted network. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 387, n. 12, p. 2972–2980, 2008. Citado na página 19.

BARABÁSI, A.-L. *Network Science*. [S.l.]: Cambridge University Press, 2016. Citado 3 vezes nas páginas 35, 37 e 42.

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *science*, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999. Citado 3 vezes nas páginas 19, 40 e 81.

BARABÁSI, A.-L.; ALBERT, R.; JEONG, H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 281, n. 1, p. 69–77, 2000. Citado na página 40.

BARABÁSI, A.-L.; BONABEAU, E. Scale-free. *Scientific American*, 2003. Citado na página 41.

BOCCALETTI, S. et al. Complex networks: Structure and dynamics. *Physics reports*, Elsevier, v. 424, n. 4, p. 175–308, 2006. Citado 2 vezes nas páginas 19 e 36.

BOCCARA, N. *Modeling complex systems*. [S.l.]: Springer Science & Business Media, 2010. Citado 2 vezes nas páginas 30 e 31.

CETESB. *1º Inventário de Emissões Antrópicas de Gases de Efeito Estufa Diretos e Indiretos do Estado de São Paulo*. [S.l.], 2011. Citado 2 vezes nas páginas 26 e 27.

CINTRA, M. A crise do trânsito em são paulo e seus custos. *GVexecutivo*, v. 12, n. 2, p. 58–61, 2013. Citado 2 vezes nas páginas 21 e 25.

CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. Power-law distributions in empirical data. *SIAM review*, SIAM, v. 51, n. 4, p. 661–703, 2009. Citado na página 62.

COHEN, R.; HAVLIN, S. *Complex networks: structure, robustness and function*. [S.l.]: Cambridge University Press, 2010. Citado na página 45.

---

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

DERRIBLE, S.; KENNEDY, C. The complexity and robustness of metro networks. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 389, n. 17, p. 3678–3691, 2010. Citado 3 vezes nas páginas 19, 24 e 60.

DOROGOVTSEV, S. N. *Lectures on Complex systems*. [S.l.]: Oxford University Press, 2010. Citado na página 41.

DOROGOVTSEV, S. N.; MENDES, J. F. *Evolution of networks: From biological nets to the Internet and WWW*. [S.l.]: Oxford University Press, 2013. Citado na página 19.

DUNNE, J. A.; WILLIAMS, R. J.; MARTINEZ, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology letters*, Wiley Online Library, v. 5, n. 4, p. 558–567, 2002. Citado na página 19.

EPSTEIN, J. M. Why model? *Journal of Artificial Societies and Social Simulation*, v. 11, n. 4, p. 12, 2008. ISSN 1460-7425. Disponível em: <<http://jasss.soc.surrey.ac.uk/11/4/12.html>>. Citado na página 32.

ÉRDI, P. *Complexity Explained*. [S.l.]: Springer, 2008. Citado 2 vezes nas páginas 30 e 36.

ERDŐS, P.; RÉNYI, A. On random graphs i. *Publ. Math. Debrecen*, v. 6, p. 290–297, 1959. Citado 2 vezes nas páginas 38 e 40.

ERDŐS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, Citeseer, v. 5, p. 17–61, 1960. Citado na página 38.

FERBER, C. V. et al. A tale of two cities: Vulnerabilities of the london and paris transit networks. *Journal of Transportation Security*, Springer, v. 5, n. 3, p. 199–216, 2012. Citado 3 vezes nas páginas 19, 24 e 50.

FERBER, C. V.; HOLOVATCH, T.; HOLOVATCH, Y. Attack vulnerability of public transport networks. *ArXiv e-prints*, Springer, p. 721–731, 2007. Citado 2 vezes nas páginas 19 e 24.

FERBER, C. V. et al. Network harness: Metropolis public transport. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 380, p. 585–591, 2007. Citado 3 vezes nas páginas 19, 23 e 50.

FERBER, C. V. et al. Public transport networks: empirical analysis and modeling. *The European Physical Journal B-Condensed Matter and Complex Systems*, Springer, v. 68, n. 2, p. 261–275, 2009. Citado 3 vezes nas páginas 19, 23 e 50.

FERBER, C. V.; HOLOVATCH, Y.; PALCHYKOV, V. Scaling in public transport networks. *arXiv preprint cond-mat/0501296*, 2005. Citado 4 vezes nas páginas 19, 23, 40 e 50.

FOLKE, C. Resilience: The emergence of a perspective for social–ecological systems analyses. *Global environmental change*, Elsevier, v. 16, n. 3, p. 253–267, 2006. Citado na página 23.

FRIGG, R.; HARTMANN, S. Models in science. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Fall 2012. [S.l.: s.n.], 2012. Citado na página 30.

GUIMERA, R.; AMARAL, L. A. N. Modeling the world-wide airport network. *The European Physical Journal B-Condensed Matter and Complex Systems*, Springer, v. 38, n. 2, p. 381–385, 2004. Citado na página 19.

HUANG, A. et al. Cascading failures in weighted complex networks of transit systems based on coupled map lattices. *Mathematical Problems in Engineering*, Hindawi Publishing Corporation, v. 2015, 2015. Citado na página 20.

HUMPHRIES, M. D.; GURNEY, K. Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence. *PloS one*, Public Library of Science, v. 3, n. 4, p. e0002051, 2008. Citado na página 64.

KOSSINETIS, G.; WATTS, D. J. Empirical analysis of an evolving social network. *Science*, American Association for the Advancement of Science, v. 311, n. 5757, p. 88–90, 2006. Citado na página 19.

LATORA, V.; MARCHIORI, M. Is the boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 314, n. 1, p. 109–113, 2002. Citado 4 vezes nas páginas 19, 24, 40 e 50.

LENG, B.; ZHAO, X.; XIONG, Z. Evaluating the evolution of subway networks: Evidence from beijing subway network. *EPL (Europhysics Letters)*, IOP Publishing, v. 105, n. 5, p. 58004, 2014. Citado 2 vezes nas páginas 20 e 50.

LI, W.; CAI, X. Statistical analysis of airport network of china. *Physical Review E*, APS, v. 69, n. 4, p. 046106, 2004. Citado na página 19.

MA, K. et al. Power law and small world properties in a comparison of traffic city networks. *Chinese Science Bulletin*, SP Science China Press, v. 56, n. 34, p. 3731–3735, 2011. ISSN 1001-6538. Disponível em: <<http://dx.doi.org/10.1007/s11434-011-4769-4>>. Citado 2 vezes nas páginas 19 e 24.

MILGRAM, S. The small world problem. *Psychology today*, New York, v. 2, n. 1, p. 60–67, 1967. Citado na página 35.

MILLER, J. H.; PAGE, S. E. *Complex adaptive systems: an introduction to computational models of social life*. [S.l.]: Princeton university press, 2009. Citado na página 37.

MITCHELL, M. *Complexity: A guided tour*. [S.l.]: Oxford University Press, 2009. Citado 2 vezes nas páginas 29 e 36.

MOURONTE, M. L.; BENITO, R. M. Structural analysis and traffic flow in the transport networks of madrid. *Networks and Heterogeneous Media*, v. 10, n. 1, p. 127–148, 2015. ISSN 1556-1801. Disponível em: <<http://aimsciences.org/journals/displayArticlesnew.jsp?paperID=10843>>. Citado na página 20.

NEWMAN, M. Resource letter cs–1: Complex systems. *American Journal of Physics*, American Association of Physics Teachers, v. 79, n. 8, p. 800–810, 2011. Citado na página 30.

NEWMAN, M. E. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 98, n. 2, p. 404–409, 2001. Citado na página 35.

NEWMAN, M. E. Mixing patterns in networks. *Physical Review E*, APS, v. 67, n. 2, p. 026126, 2003. Citado 2 vezes nas páginas 43 e 57.

NEWMAN, M. E. The structure and function of complex networks. *SIAM review*, SIAM, v. 45, n. 2, p. 167–256, 2003. Citado na página 34.

NEWMAN, M. E.; BARABASI, A.-L.; WATTS, D. J. *The structure and dynamics of networks*. [S.l.]: Princeton University Press, 2006. Citado na página 35.

NEWMAN, M. E.; WATTS, D. J.; STROGATZ, S. H. Random graph models of social networks. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 99, n. suppl 1, p. 2566–2572, 2002. Citado 2 vezes nas páginas 19 e 38.

PIETRANTONIO, H.; STRAMBI, O.; GUALDA, N. D. F. Integração entre políticas de uso de solo e de transportes: dificuldades e necessidades. Disponível em: <<http://sites.poli.usp.br/p/hugo.pietrantonio/UsoSolo3.pdf>>. 1996. Citado na página 26.

QING, X. et al. Space p-based empirical research on public transport complex networks in 330 cities of china. *Journal of Transportation Systems Engineering and Information Technology*, Elsevier, v. 13, n. 1, p. 193–198, 2013. Citado 3 vezes nas páginas 20, 24 e 50.

RODRÍGUEZ-NÚÑEZ, E.; GARCÍA-PALOMARES, J. C. Measuring the vulnerability of public transport networks. *Journal of transport geography*, Elsevier, v. 35, p. 50–63, 2014. Citado na página 23.

ROLNIK, R.; KLINTOWITZ, D. (im) mobility in the city of são paulo. *Estudos Avançados*, SciELO Brasil, v. 25, n. 71, p. 89–108, 2011. Citado na página 20.

SIENKIEWICZ, J.; HOLYST, J. A. Statistical analysis of 22 public transport networks in poland. *Physical Review E*, APS, v. 72, n. 4, p. 046127, 2005. Citado 2 vezes nas páginas 19 e 23.

SPTRANS. *Cronologia do Transporte Coletivo em São Paulo*. [S.l.], 2006. Citado na página 27.

STEEN, M. V. *Graph Theory and Complex Networks*. [S.l.]: Maarten Van Steen, 2010. Citado na página 36.

STRAMBI, O.; BILT, K.-A. V. D. Trip generation modeling using chaid, a criterion-based segmentation modeling tool. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, n. 1645, p. 24–31, 1998. Citado na página 20.

SUSTENTABILIDADE, I. S. e. *Avaliação do impacto da poluição atmosférica no Estado de São Paulo sob a visão da saúde*. [S.l.], 2013. Citado na página 27.

UNITED, N. *World Urbanization Prospects: The 2014 Revision Highlights*. [S.l.], 2014. Citado na página 20.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. *nature*, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. Citado 4 vezes nas páginas 19, 39, 40 e 81.



YANG, Y.; HUANG, A.; GUAN, W. Statistic properties and cascading failures in a coupled transit network consisting of bus and subway systems. *International Journal of Modern Physics B*, World Scientific, v. 28, n. 30, 2014. Citado na página [20](#).