

Dissertação de Mestrado

Análise de Medidas de Centralidade utilizando Distância de Hellinger

Fabiano dos Santos Brião
fabianobriao@gmail.com

Maceió/AL, maio de 2016.

Fabiano dos Santos Brião

Análise de Medidas de Centralidade utilizando Distância de Hellinger

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós Graduação em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientadora: Profa Dra Raquel da Silva Cabral.

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Bibliotecária: Janaina Xisto de Barros Lima

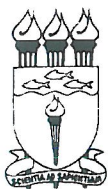
B849a Brião, Fabiano dos Santos.
Análise de Medidas de Centralidade utilizando Distância de Hellinger / Fabiano dos Santos Brião. – 2016.
72 f. : il.

Orientadora: Raquel da Silva Cabral.
Dissertação (Mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Programa de Pós-Graduação em Modelagem Computacional de Conhecimento. Maceió, 2016.

Bibliografia: f. 67–72.

1. Teoria dos grafos. 2. Redes complexas. 3. Distâncias estocásticas.
4. Perturbações. 5. Medidas de centralidade. I. Título.

CDU: 004.7



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
Avenida Lourival Melo Mota, Km 14, Bloco 09, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364



Membros da Comissão Julgadora da Dissertação de Mestrado de Fabiano dos Santos Brião, intitulada: “Análise de Medidas de Centralidade utilizando Distância de Hellinger”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas, em 24 de maio de 2016, às 15h00min, no auditório do Instituto de Computação da Ufal.

COMISSÃO JULGADORA

Raquel da Silva Cabral

Profa. Dra. Raquel da Silva Cabral

Ufal – Instituto de Computação

Orientadora

Alejandro César Frery

Prof. Dr. Alejandro César Frery

Ufal – Instituto de Computação

Examinador

Marcilio Ferreira S. Jr.

Prof. Dr. Marcilio Ferreira de Souza Júnior

Ifal – Coordenadoria do Curso de Informática

Examinador

Maceió, maio de 2016.

AGRADECIMENTOS

Há uns bons anos atrás parti para uma nova vida. Motivado pelo curso de mestrado coloquei minha vontade de vencer, coragem e uma bagagem de conhecimentos a prova. Hoje vejo que valeu tudo a pena, mas só foi possível graças a algumas pessoas que fizeram parte dessa história, as quais faço agora meus sinceros agradecimentos.

À minha mãe Anelha e minha vó Alda, obrigado pelos belos exemplos de humildade, força e dedicação em minha vida. Ao meu pai Pedro que sempre acreditou e confiou em mim.

À minha esposa Noélia que me motivou por todos os dias em que estivemos juntos nessa jornada e que dedicou seu tempo com prazer para me ajudar em todos momentos. Obrigado meu amor por ter sido a companheira que desejei para chegar na fase final deste caminho.

A todos os meus familiares e amigos que torceram por mim.

Aos companheiros de mestrado, em especial, Fernando Pinto por ter sido um grande parceiro em importantes momentos desse curso.

Ao meu amigo prof. Andreoli, que hoje não está mais conosco, por ter me motivado a buscar a realização deste projeto enquanto ainda era um sonho.

À minha orientadora Profa Dra Raquel da S. Cabral, pela sua ajuda e compreensão em todos os momentos, pelo seu tempo de dedicação a este trabalho.

Enfim, a uma pessoa que me ajudou a dar o primeiro passo rumo a esse sonho, Prof. Dr. Alejandro C. Frery. Obrigado pela oportunidade, pela confiança em mim depositada e pelos ricos ensinamentos como pessoa e profissional. É um limite infinito a imensa gratidão por sua parceria e amizade.

"...E nunca considerem seu estudo como uma obrigação, mas sim como uma oportunidade invejável de aprender, sobre a influência libertadora da beleza no domínio do espírito, para seu prazer pessoal e para o proveito da comunidade à qual pertencerá o seu trabalho futuro."

Albert Einstein

RESUMO

A teoria das redes complexas tem entre suas características a interdisciplinaridade, além de ser concisa e clara consegue envolver simultaneamente diversas áreas com ampla aplicabilidade em modelagens reais. Algumas redes, por exemplo, a Internet, as rotas de aeroportos, colaboradores científicos, de energia elétrica e de esgotos possuem uma estrutura física que lhes caracterizam como redes complexas. Por serem dinâmicas essas redes estão sujeitas a alterações de topologia. Sendo assim, é necessário conhecer suas características físicas e analíticas para melhor compreender tais acontecimentos. Em particular analisamos um conjunto de Sistemas Autônomo (ASs) formado por redes de clientes, redes universitárias, além de pontos de intercâmbios, entre outros. Sendo os vértices cada AS e arestas as ligações entre estes ASs. Este estudo teve por objetivo verificar uma rede real da Internet formada por ASs, através de seu respectivo grafo para compreender seu comportamento e particularidades por vias de medidas de centralidade baseadas em vértices. Para atingir tal objetivo utilizamos estudos estatísticos associados ao Método Monte Carlo e modificamos de forma aleatória seu grafo com perturbações para obter novos grafos modificados topologicamente. Através das frequências relativas das medidas de centralidade baseadas em vértices destes grafos, original e modificados, obtivemos por meio do quantificador distância de Hellinger diversos resultados relevantes. Acerca de perturbações, a remoção de arestas foi o tipo de perturbação que mais alterou a topologia entre as redes. Com referência as medidas de centralidade, o grau de proximidade foi o mais vulnerável de modo geral em relação as perturbações.

Palavras-chave: Teoria dos grafos. Redes complexas. Distâncias estocásticas. Perturbações. Medidas de centralidade.

ABSTRACT

The theory of complex networks has among its characteristics the interdisciplinarity, besides being concise and clear can simultaneously involve several areas with broad applicability in real modeling. Some networks, for example, the Internet, airports routes, scientific collaborators, electricity and sewage have a physical structure that characterize them as complex networks. Because they are dynamics these networks might be affected by topology changes. In this case, it is necessary to know physical and analytical characteristics to better understand such events. In particular we analyzed a set of Autonomous Systems (ASs) composed of customer networks, university networks, beyond points exchanges, and others. It is the vertices and edges each AS the connections between the ASs. This study has the objective to verify a real Internet network of ASs, through their respective graph to understand their behavior and their characteristics by way of centralized measures based on vertices. To achieve this goal we use statistical studies associated of the Monte Carlo Method and modified randomly your graph with perturbations for new graphs modified topologically. Through the relative frequencies of the centrality measures based on these graphs vertices, original and modified. We obtained through Hellinger distance quantifier many relevant results. About doing perturbations, removing edges was the most kind of perturbations that had altered the topology compared the others networking. According to the centrality measures, the degree of Closeness measure was the most vulnerable in general compared another perturbations.

Keywords: Graph teory. Complex network. Stochastic distances. Perturbations. Centrality measures.

LISTA DE FIGURAS

2.1	Desenho das sete pontes de Königsberg.	17
2.2	Representação gráfica das setes pontes de Königsberg.	17
2.3	Grafo simples, grafo com arestas paralelas e grafo com lacete	18
2.4	Grafo ponderado, dígrafo e dígrafo ponderado.	19
2.5	Grafos, conexo e desconexo.	19
2.6	Grafo com vértices rotulados e seus respectivos graus.	20
2.7	Grafos completos K_1, K_2, K_3, K_4 e K_5	20
2.8	Grafo seu complementar.	21
2.9	Grafo e seu subgrafo.	21
2.10	Grafo e seu subgrafo induzido por vértices.	21
2.11	Grafo e subgrafo conjunto independente.	22
2.12	Grafo e seu subgrafo gerador.	22
2.13	Grafo desconexo e componentes conexas.	22
2.14	Gráfico com vértices de corte.	23
2.15	Grafos ciclos e grafos passeios.	23
2.16	Grafo com duas pontes.	24
2.17	Grafo com quatro cliques maximais.	24
2.18	Distâncias geodésica e diâmetro entre dois vértices em grafo simples.	25
2.19	Grafos denso k_{16} e esparso C_{16}	25
2.20	O sociograma de Jacob Moreno.	26
2.21	Árvore de probabilidades.	27
2.22	Rede de citações de trabalhos acadêmicos e a rede WWW.	29
2.23	Redes: de colaboração científica.	30
2.24	Diagrama das relações dos tipos de redes complexas.	30
2.25	Exemplos de tipos de redes.	31
2.26	Redes aleatórias regulares com 10 vértices e $\langle k \rangle = 2,4,5$	31
2.27	Rede aleatória $G(N,M)$ com $ N = 20$ $ M = 30$	32
2.28	Duas redes aleatórias $G(N,p)$ diferentes com os mesmos $N = 20$ e $p = 0.1$	32
2.29	Histograma de densidades dos graus de $G(5000,0.01)$	33
2.30	Rede de amizades entre pessoas.	33
2.31	Construção de uma rede aleatória $G(N,p)$	34
2.32	Gráficos das densidades de distribuições de leis de potência.	35
2.33	Rede livre de escala $BA(N = 50, m = 1)$	36
3.1	Grafo com os valores $C_g(v_i)$ de seus vértices v_i s.	40
3.2	Grafo com os valores $P_c(v_i)$ de seus vértices v_i s.	41
3.3	Grafo com os valores $I_c(v_i)$ de seus vértices v_i s.	42
3.4	Operação de remoção de aresta sem remoção de vértice(s).	46
3.5	Operação de inclusão de aresta.	46
3.6	Operação de remoção de vértice e aresta(s)).	47
3.7	Operação de remoção e inclusão de aresta.	47
3.8	Tipos de perturbações que geram grafos desconexos.	48
3.9	Fluxograma das etapas do desenvolvimento.	53
4.1	dHel médias dos graus dos vértices por níveis de perturbação.	59
4.2	Comparação do grau do vértice entre os tipos de perturbações.	60

4.3	Percentual de variação da dHel das perturbações sob o grau médio.	60
4.4	dHel média dos graus de proximidade dos vértices por níveis de perturbação. . .	61
4.5	Comparação do grau de proximidade entre os tipos de perturbações.	62
4.6	Percentual de variação da dHel das perturbações sob o grau de proximidade. . .	63
4.7	dHel média dos graus de intermediação dos vértices por níveis de perturbação. .	64
4.8	Comparação do grau de intermediação entre os tipos de perturbações.	64
4.9	Percentual de variação da dHel das perturbações sob o grau de intermediação. .	65

LISTA DE TABELAS E QUADROS

4.1	Estatísticas das perturbações do grau do vértice.	59
4.2	Estatísticas das perturbações do grau de proximidade.	62

LISTA DE EQUAÇÕES

2.1	Grau médio $\langle k \rangle$.	20
2.2	Grafo completo k_n .	20
2.3	Lei de Potência.	35
3.1	Grau do vértice.	40
3.2	Grau de proximidade (grafo conexo).	40
3.3	Grau de proximidade (grafo qualquer).	41
3.4	Grau de intermediação.	42
3.5	Distância de Hellinger.	45
3.6	Fórmula de Sturges.	56

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Definição do Problema	14
1.2	Objetivos	15
1.3	Organização da Dissertação	15
2	FUNDAMENTOS TEÓRICOS	16
2.1	Teoria dos Grafos	16
2.2	Teoria das Redes	25
2.2.1	Redes Complexas	29
3	MATERIAIS E MÉTODOS	38
3.1	Medidas de Rede	38
3.1.1	Medidas de Centralidade baseadas em Vértices	39
3.2	Quantificadores Estocásticos	42
3.2.1	Distância de Hellinger (dHel)	44
3.2.2	Perturbações em Grafos	45
3.3	Ambiente Computacional	48
3.4	Modelagem	51
3.5	Validação da Metodologia	55
4	RESULTADOS E ANÁLISES	58
4.1	Grau do Vértice	58
4.2	Grau de Proximidade	61
4.3	Grau de Intermediação	63
5	CONSIDERAÇÕES FINAIS	66
	REFERÊNCIAS	67

I INTRODUÇÃO

INICIAMOS este capítulo com a seção 1.1 que contextualiza acerca deste tema e descreve a motivação por este trabalho. Na seção 1.2 temos os objetivos que buscamos alcançar e por último a seção 1.3 Organização do trabalho que detalha como essa dissertação está estruturada.

1.1 Definição do Problema

Relações entre pessoas, rotas de aviões, cadeias de proteínas, distribuições de energia elétrica, redes de esgoto, Wide Word Web (WWW) e Internet possuem arquitetura de malha, estruturas definidas como redes que permitem suas análises como grafos. Por serem representadas como um grafo é possível analisar várias características nelas presentes como, por exemplo, a formação de comunidades, a aglomeração, a densidade, a relevância dos elementos e a robustez, entre outras.

No início do século XX as ciências sociais buscavam entender como determinados efeitos coletivos aconteciam entre as pessoas, acerca desses estudos sobre as sociedades e os laços de amizade entre as pessoas surgiram as *redes sociais* (Gould, 1993). Mais recentemente na década de 1990, através da iniciativa de físicos é que os estudos sobre redes ressurgiram (Newman, 2003). Em 1999, os trabalhos seminais de Watts & Strogatz (1998) e Albert & Barabasi (1999) deram início a um novo campo da ciência, capaz de estudar não somente as redes sociais como também outros tipos de rede, conhecidas hoje como *redes complexas*.

Segundo Albert & Barabasi (1999) e Guimera & Amaral (2005) as redes complexas são redes dinâmicas por não possuírem uma topologia¹ única, ou seja, em algum momento qualquer passam a ter nova estrutura topológica gerada por perturbações como, por exemplo, remoção de vértices e/ou arestas, ocasionadas por diversos fatores como ataque de hackers e fenômenos naturais, entre outros. E como consequência da mudança topológica, certas redes passam a ter alterações de medidas associadas ao seu grafo. Até o final do século XX embora houvessem vários estudos sobre redes, não existiu na literatura investigações específicas que analisassem medidas de centralidade em grafos após perturbações na

¹estrutura física, disposição e interação como se encontram os elementos da rede.

topologia (Redner, 1998; Faloutsos et al., 1999; Maslov & Sneppen, 2002). Mas atualmente, Cabral et al. (2014) apresentaram uma série de resultados de análises acerca de medidas de centralidade para redes complexas submetidas a perturbações de topologia e inferidas por distâncias estocásticas, entre os resultados destacam-se a distância de Hellinger como um quantificador de maior precisão e a rede Livre de escala sendo de um modo geral a mais sensível entre outras redes observadas.

Logo, a motivação para o desenvolvimento dessa pesquisa surgiu da seguinte pergunta: Uma rede real possui estatísticas de distâncias estocásticas por medidas de centralidade baseadas em vértices sob perturbações semelhantes a parâmetros de uma rede teórica? Villas Boas et al. (2010) e Cabral et al. (2014) apresentaram alguns métodos para analisar distâncias estocásticas entre redes complexas sob medidas de redes que utilizaremos nesta dissertação.

1.2 Objetivos

Este trabalho tem como objetivo maior analisar se o comportamento de uma rede real qualquer submetida a perturbações possui características de medidas de centralidade baseadas em vértices utilizando a distância de Hellinger semelhantes aos parâmetros de rede obtidos por Cabral et al. (2014). Para tanto, será necessária a realização dos seguintes objetivos específicos:

- Identificar quais tipos e níveis de perturbações influenciam mais fortemente nas mudanças de topologia;
- Classificar entre as medidas de centralidade quais são mais significativas;
- Validar o modelo matemático proposto.

1.3 Organização da Dissertação

Esta dissertação está dividida em cinco capítulos com a seguinte estrutura: O presente capítulo apresentou a motivação para este trabalho, uma solução proposta e nossos objetivos. No capítulo 2 apresentaremos a fundamentação teórica para compreensão e entendimento dos temas presentes. Em seguida, no capítulo 3 é feita a apresentação dos materiais e métodos. Por sua vez no capítulo 4 são apresentados os resultados e análises. Finalizamos com as considerações finais.

O capítulo apresentou a contextualização do problema, bem como uma solução proposta e os objetivos desta pesquisa. No próximo capítulo serão apresentados os conceitos e as definições que foram a base teórica de todo trabalho.

2 FUNDAMENTOS TEÓRICOS

NESTE capítulo apresentaremos os fundamentos teóricos necessários para o entendimento deste trabalho. Abordaremos inicialmente na seção 2.1 a teoria dos grafos a partir do seu surgimento enfatizando definições e conceitos importantes relacionados ao estudo. Adiante, na seção 2.2 trataremos acerca da teoria das redes incluindo um tópico sobre redes complexas. Este capítulo utilizou essencialmente as definições dos livros "Graph Theory with Applications" de [Bondy & Murty \(1976\)](#), "Graph Theory" de [Diestel \(2010\)](#) e "Dynamical Processes on Complex Networks" de [Barrat et al. \(2008\)](#).

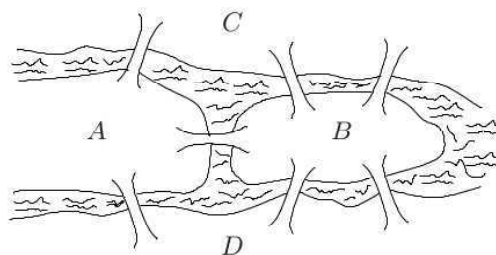
2.1 Teoria dos Grafos

Os primeiros estudos que deram origem a teoria dos grafos apareceram em meados do século XVIII quando o matemático Leonhard Paul Euler¹ (1707 – 1783) resolveu o problema conhecido como Sete pontes de Königsberg ([Euler, 1736](#)). Königsberg, hoje chamada de Kaliningrad, situada na Rússia, tornou-se famosa por ter em seu território duas grandes ilhas, que ligam à cidade por sete pontes, seis delas interligavam as ilhas às margens do Rio Pregel e uma fazia a ligação entre as duas ilhas. Consta que os habitantes gostavam de passear e atravessar todas as pontes da cidade, porém, a possibilidade de se atravessar todas as pontes apenas uma vez e retornar ao ponto de partida era uma dúvida que havia se tornado uma lenda. Até que Euler, em 1736, mostrou que não era possível realizar essa façanha. Euler resolveu este enigma através da representação de um grafo. Para solucionar o enigma Euler desenhou linhas para representar os lugares por onde as pessoas passeavam e pontos para representar os cruzamentos, criando possivelmente o primeiro grafo da história.

A Figura 2.1 abaixo mostra as denominações dadas por Euler para as cidades e ilhas, as cidades de Altstadt e Löbnicht são representadas pelo círculo C que estão localizadas na costa norte do rio, a cidade de Vorstadt por D está localizada na costa sul e as duas ilhas de Kneiphof por A e Lomse por B que estão ao centro ([Euler, 1736](#)).

¹Leonhard Paul Euler foi um dos maiores matemáticos da história, passou a maior parte de sua vida na Rússia e na Alemanha. Em sua vida fez importantes descobertas em campos variados, na área de cálculo e de grafos, entre outras.

Figura 2.1: Desenho das sete pontes de Königsberg.



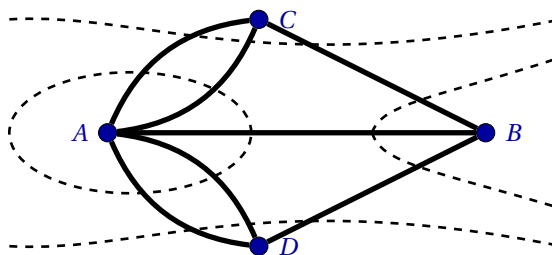
Fonte: Bondy & Murty (2008).

Mas o que é um grafo? Graficamente, um desenho no qual tem-se alguns pontos e linhas ligando pares de pontos, com todos os pontos ligados ou alguns sem ligação, ou ainda podendo haver linhas com origem e extremidade em um mesmo ponto. Entretanto, é necessário para compreender o grafo criado por Euler de algumas definições matemáticas preliminares.

Um *grafo* G ou $G(V,A)$ é um par (V,A) tal que $V(G) = \{v_1, v_2, \dots, v_n\}$ é um conjunto não vazio de n vértices ou nós de G e $A(G) = \{a_1, a_2, \dots, a_m\}$, com $m \leq n$, o conjunto de m arestas ou arcos de G . Denota-se $|V(G)|$ e $|A(G)|$ as quantidades de n vértices e de m de arestas, denotados por *ordem* e *dimensão* de G respectivamente, e define-se por *caminho* uma sequência de vértices e arestas, e por *caminho simples* ou *caminho elementar* toda a sequência de vértices e arestas sem repetição de vértices.

Diz-se que $a_i \in A$ é incidente a $v_i \in V$ se existe um $v_j \in V$ tal que a aresta v_i liga os vértices v_i e v_j . Define-se *vértices adjacentes* dois vértices ligados por uma aresta que os une e duas arestas são ditas *arestas adjacentes* se existe um vértice em comum. As *arestas* de um grafo são segmentos (v_i, v_n) ou $v_i v_n$ que fazem a conexão entre dois vértices v_i e v_n . Se uma aresta possui extremidades coincidindo num mesmo vértice é denominada de *lacete* ou *laço*, (v_i, v_i) , ou seja, é um seguimento com origem e extremidade no mesmo vértice (Figura 2.2).

Figura 2.2: Representação gráfica das sete pontes de Königsberg.



Fonte: Adaptada de Bondy & Murty (2008).

Seja um vértice v , define-se o *grau* de v ou *cardinalidade* de v por $|v_i|$ ou $deg(v_i)$, o número de arestas incidentes a v , os lacetes são contados duas vezes. A solução de Euler foi

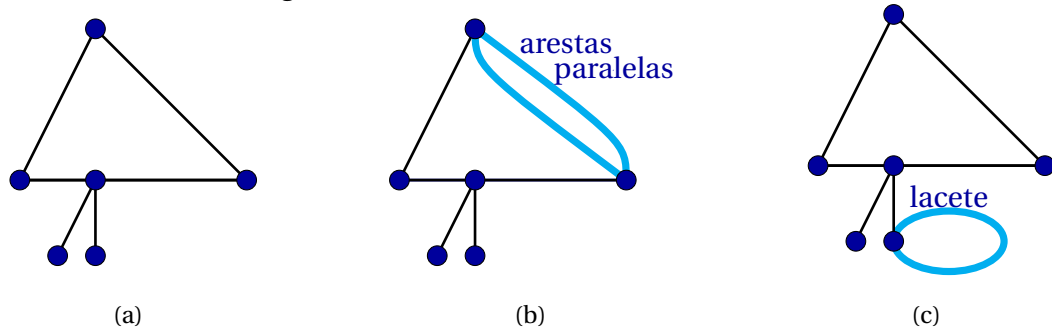
mostrar que o grafo não poderia ser percorrido se houvesse mais de dois vértices com graus ímpares, neste caso o grafo criado possuía quatro vértices ímpares. Ao estudar o problema de Königsberg Euler apresentou e demonstrou diversos resultados que são aplicáveis a todos os casos, por exemplo, se um número de vértices com grau ímpar é exatamente dois, a partir de qualquer um desses vértices com grau ímpar existe um caminho que se possa percorrer por todas as passagens apenas uma vez passando por todos os vértices. Esse caminho chama-se *Euleriano*.

Com esses conceitos, Euler demonstrou que não existia solução para o enigma das quatro pontes de Königsberg. Desde seu nascimento em 1735 a teoria dos grafos passou a responder por questões importantes como, por exemplo, a distribuição de tarefas entre pessoas e a coloração de regiões em uma mapa usando um número mínimo de cores.

Na teoria dos grafos existem muitos conceitos e definições, porém, nessa pesquisa apenas dissertamos sobre alguns dos mais revistos pela literatura e sobre os grafos e subgrafos necessários para o entendimento.

Definido anteriormente o que é um grafo, veremos agora algumas peculiaridades sobre esta teoria. A Figura 2.3 abaixo apresenta o grafo mais trivial e conhecido, o *grafo simples*, este grafo não possui *lacetes*, também não possui vértices com cardinalidade zero e nem pares de vértices que possuem mais de uma aresta ligando-os, diz-se que estas são *arestas paralelas*, além de não possuir sentido entre seus vértices e peso em suas arestas. Outros grafos comumente estudados são os grafos com arestas paralelas e com lacetes.

Figura 2.3: Representações gráficas de um grafo simples (2.3 (a)), de um grafo com arestas paralelas (2.3 (b)) e de um grafo com lacete (2.3 (c)).

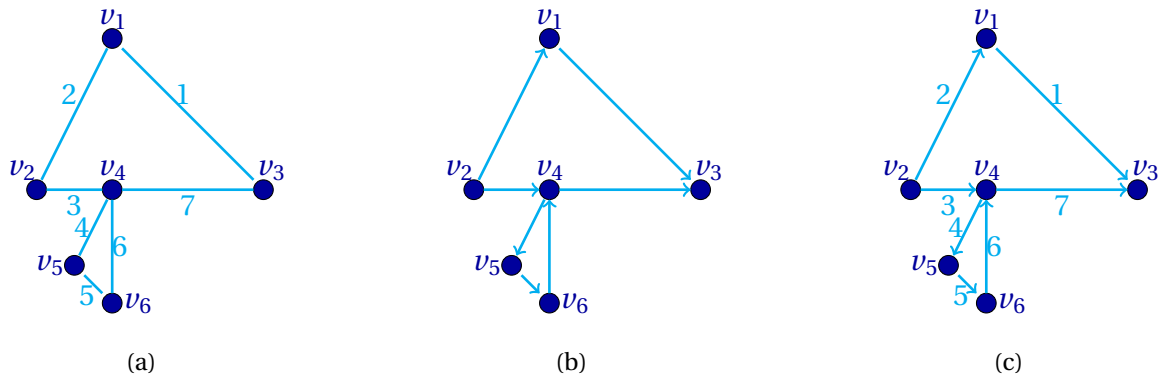


Fonte: Autor (2016).

Segundo Boccaletti et al. (2006) um grafo denomina-se *grafo não orientado* quando não possui sentido entre seus vértices. Um grafo é dito *grafo orientado*, *grafo dirigido* ou apenas *dígrafo* quando possui sentido entre seus vértices $v_i v_j$ ou $v_i v_i$. *Arco* é o nome dado a toda conexão de um dígrafo. Todo grafo não orientado construído a partir de um dígrafo chama-se *grafo fundamental*. Se um grafo possuir rótulos nos vértices e/ou arestas diz-se *grafo rotulado*, esse rótulo pode ser um nome, um número ou conjunto de letras ou números, quando as arestas são rotuladas de valores numéricos, comumente chamados de peso, então

diz-se que *grafo ponderado*. Um grafo é denominado *dígrafo ponderado* quando possui direção e peso nas arestas (Figura 2.4).

Figura 2.4: Representações gráficas de um grafo fundamental ponderado (2.4(a)), de um dígrafo (2.4(b)) e de um dígrafo ponderado (2.4 (c)).

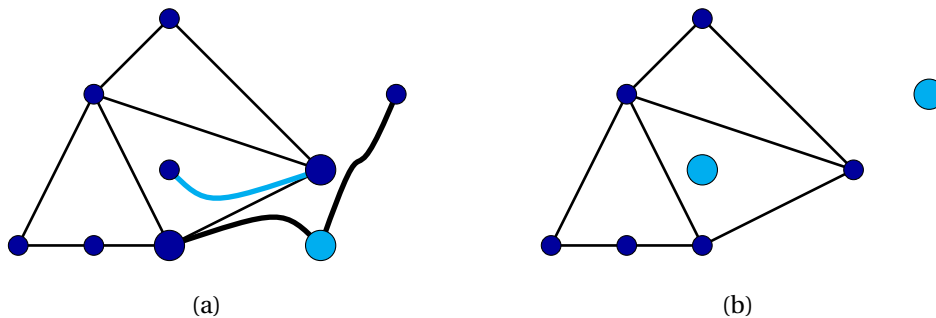


Fonte: Autor (2016).

Os grafos ponderados possuem uma outra representação, denotados comumente por $G(V,A,P)$, sendo o terceiro elemento $P = \{p_1, p_2, p_3, \dots, p_m\}$, o conjunto de pesos atribuídos as m arestas. Um arco $v_i v_j$ possui sentido de origem v_i para sua extremidade v_j . Existem também grafos que possuem direção e/ou peso em apenas algumas arestas, esses grafos são denominados por *grafos mistos*.

Um grafo $G(V,A)$ é denominado *grafo conexo* se há pelo menos um caminho entre qualquer par de vértices distintos do grafo, caso contrário denomina-se G por *grafo desconexo* (Figura 2.5).

Figura 2.5: Representações gráficas de um grafo conexo (2.5 (a)) e de um grafo desconexo (2.5 (b)).

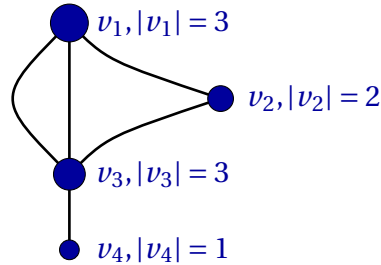


Fonte: Autor (2016).

Alguns grafos possuem todos os vértices com o mesmo grau, denominam-se daí por *grafo regular*, se um vértice tiver cardinalidade nula, isto é, o grau deste vértices é igual a zero, denota-se por $|v| = 0$ e denomina-se de *vértice isolado*. Em dígrafos o grau dos vértices tem duas outras denominações, chama-se *grau de entrada* de v o número de arcos que tem origem em v e *grau de saída* de v o número de arestas que chegam (tem extremidade) em v .

Para determinar o grau de um vértice é necessário que o grafo seja rotulado em vértices. A Figura 2.6 a seguir mostra a representação de um grafo rotulado $G(V,A)$ com seu conjunto de vértices $V(G) = \{v_1, v_2, v_3, v_4\}$ e suas cardinalidades $|v_1| = 3, |v_2| = 2, |v_3| = 3$ e $|v_4| = 1$.

Figura 2.6: Representação gráfica do grafo G com vértices rotulados e seus respectivos graus.



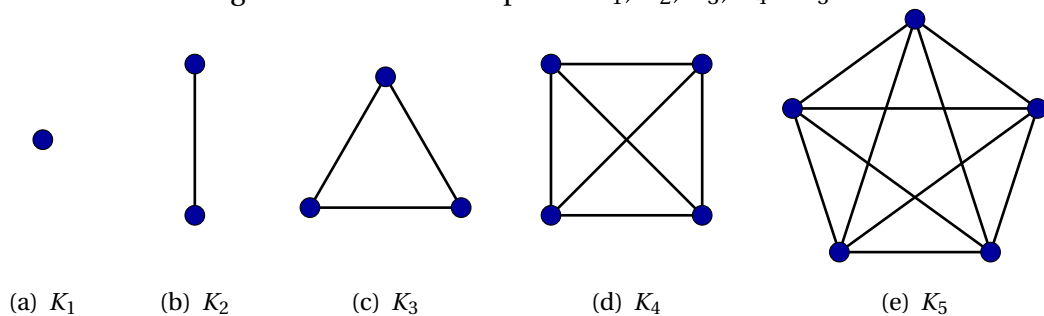
Fonte: Autor (2016).

Pode-se obter o grau médio $\langle k \rangle$ (2.1) de um grafo pela média aritmética do grau $|v_i|$ de todos os vértices. Ou seja,

$$\langle k \rangle = \frac{1}{n} \sum_{i \in V} |v_i| = \frac{2m}{n}. \tag{2.1}$$

Um grafo diz-se *grafo completo* (Figura 2.7) de ordem n se for um grafo simples e entre quaisquer dois vértices v_i, v_j existir uma aresta, comumente denota-se por K_n .

Figura 2.7: Grafos completos K_1, K_2, K_3, K_4 e K_5 .



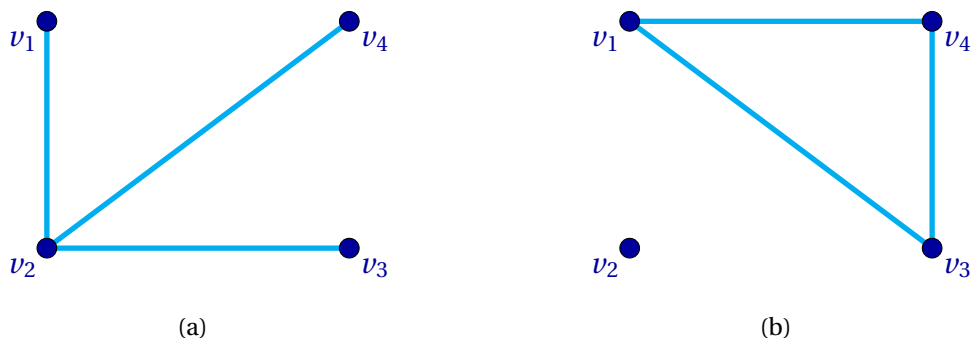
Fonte: Autor (2016).

Calcula-se a quantidade de m arestas de um grafo completo de n vértices pelo número de pares possíveis de vértices adjacentes dado pela Equação 2.2,

$$\binom{n}{2} = \frac{n!}{(n-2)!2!} = \frac{n(n-1)(n-2)!}{(n-2)!2} = \frac{n(n-1)}{2} \text{ arestas.} \tag{2.2}$$

Um grafo $G'(V, \bar{A})$ é denominado *grafo complementar* de um grafo $G(V, A)$ de um mesmo número de vértices, tal que dois vértices distintos são adjacentes em G se e somente se não são adjacentes em G' (Figura 2.8).

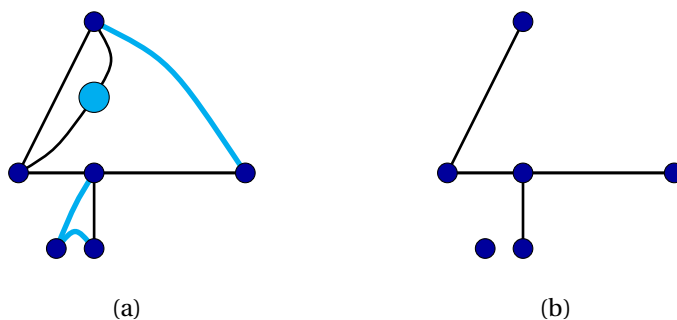
Figura 2.8: Grafo (2.8(a)) e seu grafo complementar (2.8(b)).



Fonte: Autor (2016).

Dado um subconjunto $V'(G)$ de vértices (G) , *subgrafo* $G'(V', A')$ de G é o grafo gerado por $V'(G)$ e por um subconjunto de arestas $A'(G)$ (Figura 2.9).

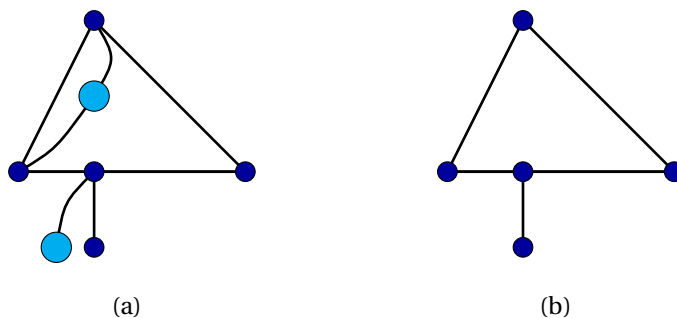
Figura 2.9: Representação de um grafo (2.9(a)) e seu subgrafo (2.9(b)).



Fonte: Autor (2016).

Um subgrafo $G'(V', A)$ ou $G'(V', A')$ de $G(V, A)$ chama-se *subgrafo induzido* de G se após remoção em G de vértices e suas arestas adjacentes quando houverem, todo vértice e aresta de G' for vértice e aresta de G (Figura 2.10).

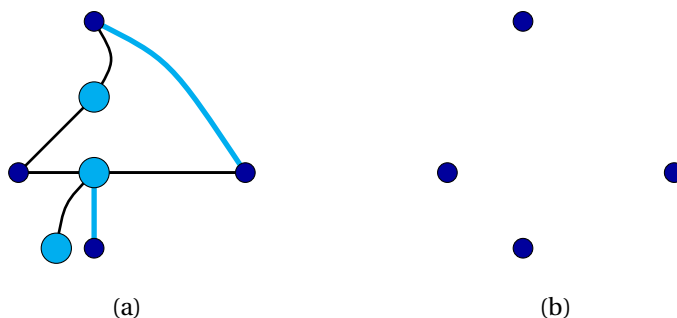
Figura 2.10: Representação de um grafo (2.10(a)) e seu subgrafo induzido (2.10(b)).



Fonte: Autor (2016).

Denomina-se *grafo vazio* $G'(V, \{\})$ (Figura 2.11(b)) todo grafo que só conter vértices isolados. Um subgrafo denomina-se *conjunto independente de vértices* $G'(V, \{\})$ de um grafo $G(V, A)$ se não existir nenhuma aresta $v_i v_j$ entre qualquer par de vértices $V(G')$ (Figura 2.11).

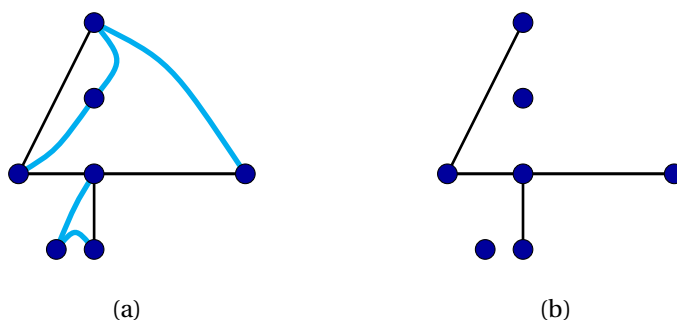
Figura 2.11: Grafo (2.11(a)) e seu subgrafo conjunto independente (2.11(b)).



Fonte: Autor (2016).

Se um subgrafo $G'(V,A')$ é um *subgrafo gerador* de $G(V,A)$, então $V(G) = V(G')$ e $A(G') = A(G) - a$, isto é, G' é obtido pela remoção de uma ou mais arestas de G (Figura 2.12).

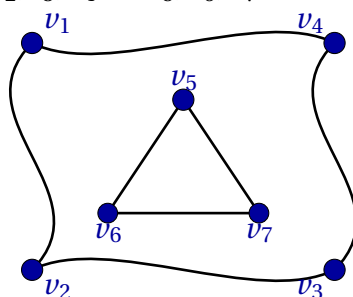
Figura 2.12: Representação de grafo (2.12(a)) e seu subgrafo gerador (2.12(b)).



Fonte: Autor (2016).

Um grafo G não conexo é a união de dois ou mais subgrafos conexos, dos quais qualquer par de subgrafos não tem vértices em comum (Figura 2.13).

Figura 2.13: Representação grafo desconexo de vértices $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ e duas componentes conexas de vértices $\{v_1, v_2, v_3, v_4\}$ e $\{v_5, v_6, v_7\}$.



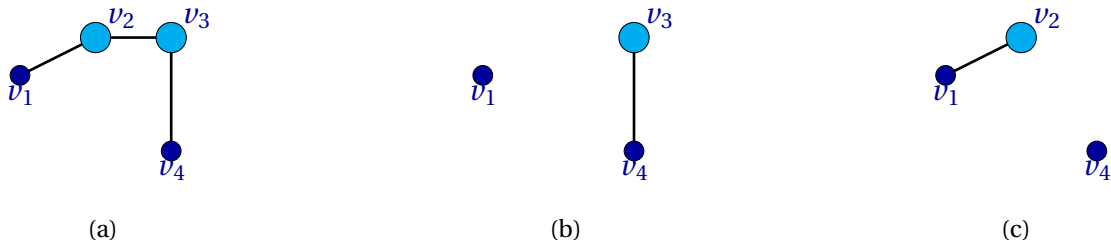
Fonte: Autor (2016).

Cada subgrafo conexo disjunto é denominado *componente conexa*.

Define-se *vértice de articulação* ou *vértice de corte* todo vértice de um grafo que ao ser removido (juntamente com as arestas a ele conectadas) não aumente o número de compo-

nentes conexas. Ou seja, toda remoção de uma vértice de corte gera um grafo desconexo (Figura 2.14).

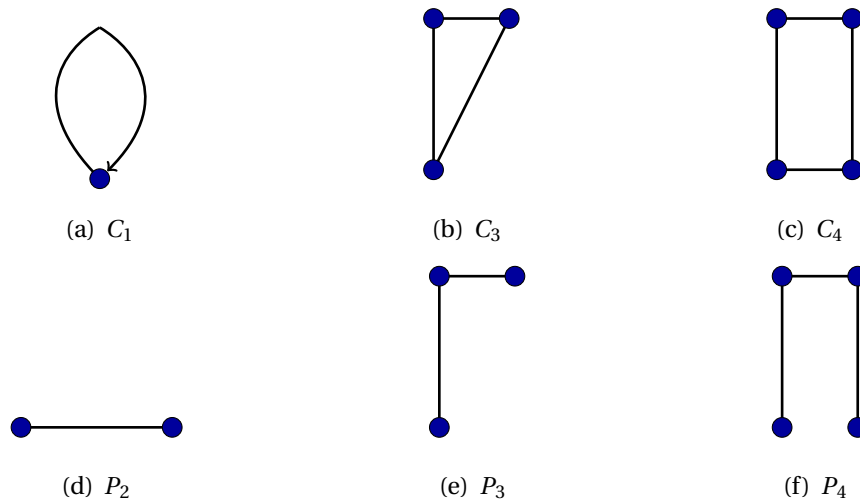
Figura 2.14: Grafo conexo com vértices de corte v_2 e v_3 (2.14(a)), grafo desconexo gerado pela remoção do vértice de corte v_2 (2.14(b)) e grafo desconexo gerado pela remoção do vértice de corte v_3 (2.14(c)).



Fonte: Autor (2016).

Em um grafo G uma *trilha*, *trajeto* ou *passeio* (*trail*) é um caminho sem arestas repetidas. Define-se por *ciclo* (*circuit*) ou *circuito* todo caminho de comprimento $n \geq 1$, em que as arestas não se repetem e o vértices inicial e final são os mesmos, ou seja, é um passeio fechado de comprimento não nulo. Se um ciclo tiver comprimento 1 então ele é um lacete. Um grafo que é um ciclo com n vértices denota-se comumente por C_n , e um grafo que é um passeio com n vértices denota-se comumente por P_n . C_3 é usualmente denominado *grafo cúbico* ou *triângulo* (Figura 2.15).

Figura 2.15: Grafos ciclos C_1 , C_3 e C_4 e grafos passeios P_2 , P_3 e P_4

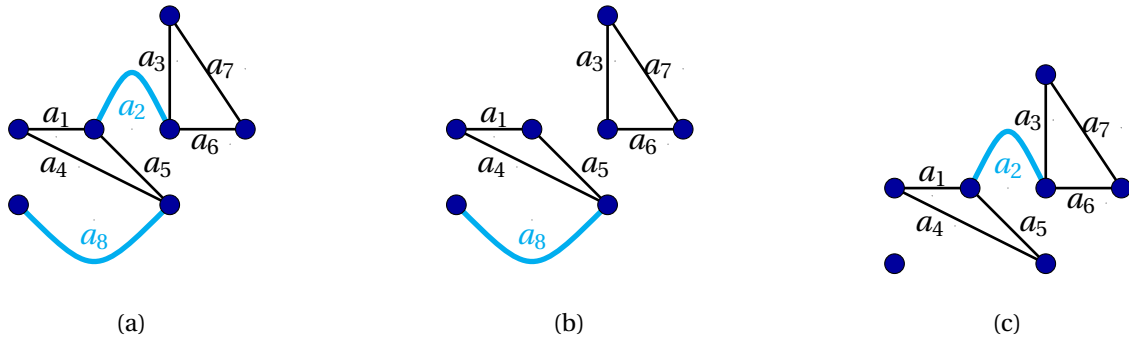


Fonte: Autor (2016).

Um *ciclo simples* é um caminho de comprimento $n \geq 3$ em que o vértice inicial e o último são iguais, mas nenhum outro vértice aparece mais de uma vez. Define-se com *cintura* o ciclo de menor comprimento de um grafo. Um grafo sem ciclos tem uma cintura de comprimento infinito.

Diz-se que uma aresta a é uma *ponte*, se e somente se, ela não está contida em qualquer ciclo. Isto é, se ao removê-la o grafo aumenta o número de suas componentes conexas (Figura 2.16).

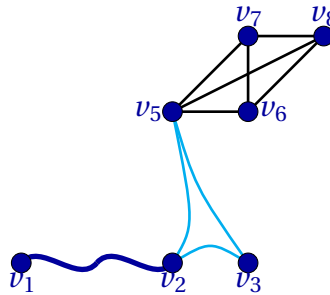
Figura 2.16: Grafo conexo com duas pontes a_2 e a_8 (2.16(a)), grafo desconexo pela remoção da ponte a_2 (2.16(b)) e grafo desconexo pela remoção da ponte a_8 (2.16(c)).



Fonte: Autor (2016).

Clique em um grafo G não orientado são os subconjuntos $V(G)$ dos vértices de G , tais que cada dois vértices do subconjunto são conectados por uma aresta. Ou seja, um clique em G é um subgrafo completo de G , e o tamanho de uma clique é igual a cardinalidade de seu conjunto de vértices. Uma *clique maximal* é um conjunto de vértices adjacentes entre si que não estão estritamente contidos em outros cliques (Figura 2.17). Em um grafo é denominado *clique máximo* o clique de maior cardinalidade, comumente denominado por $\omega(G)$

Figura 2.17: Grafo com quatro cliques maximais: $\{v_1, v_2\}, \{v_2, v_3, v_5\}$ e $\{v_5, v_6, v_7, v_8\}$ e $\omega(G) = 4$



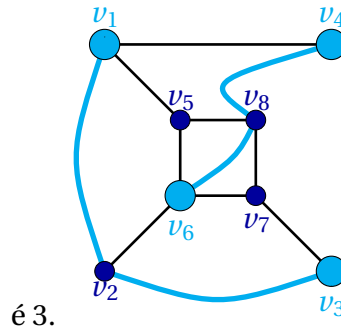
Fonte: Autor (2016).

Ao menor caminho entre dois vértices v e w denomina-se de *distância geodésica*, caso não exista um caminho entre v e w denota-se por $d(v, w) = \infty$, e para uma distância $d(v, v) = 0$. Nota-se que:

- $d(v, w) \geq 0$ e $d(v, w) = 0 \iff v = w$;
- $d(v, w) = d(w, v)$;
- $d(u, w) \leq d(u, v) + d(v, w)$.

Denomina-se de *diâmetro* $d(G)$ de um grafo G sua maior distância geodésica (Figura 2.18).

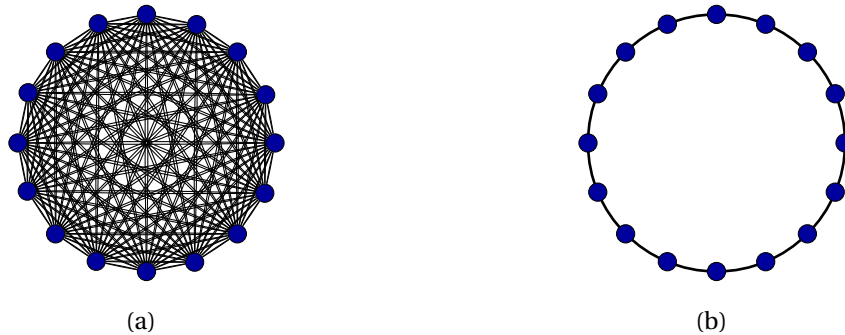
Figura 2.18: A distância geodésica entre v_1, v_3 e v_4, v_6 é 2 e o diâmetro entre $d(v_1, v_7)$ e $d(v_3, v_5)$.



Fonte: Autor (2016).

Outro aspecto importante de um grafo é a sua densidade, a densidade de um grafo é comumente analisada no estudo de redes complexas e vista com base na quantidade de ligações. Diz-se *grafo denso* $G(V,A)$ quando $|A(G)| \geq |V(G)|^2$. Por outro lado define-se um *grafo esperso* $G(V,A)$ se $|A(G)| \leq |V(G)|$ (Figura 2.19).

Figura 2.19: Representações gráficas de um grafo denso K_{16} (2.19(a)) e de um grafo esperso C_{16} (2.19(b)).



Fonte: Autor (2016).

A densidade entre os dois grafos acima pode ser calculada apenas pelas suas dimensões, pelo fato dos grafos possuírem mesma ordem. A Figura 2.19(a) mostra um grafo completo K_{16} , e pela Equação 2.2 encontramos $|A| = 90$, enquanto que a Figura 2.19(b) mostra um grafo com $|A| = 15$. Logo, pela definição de grafo denso podemos ver que o grafo com maior densidade é o grafo da Figura 2.19(a).

2.2 Teoria das Redes

Com o surgimento da teoria dos grafos nasce a teoria das redes. Teoricamente a teoria das redes difere-se da teoria dos grafos devido a três aspectos básicos, elas são associadas a

modelagem de redes reais por dados empíricos, elas são geralmente dinâmicas e se modificam com o tempo alterando sua topologia, e raramente são formadas por objetos estáticos, pois possuem estruturas que simulam processos dinâmicos.

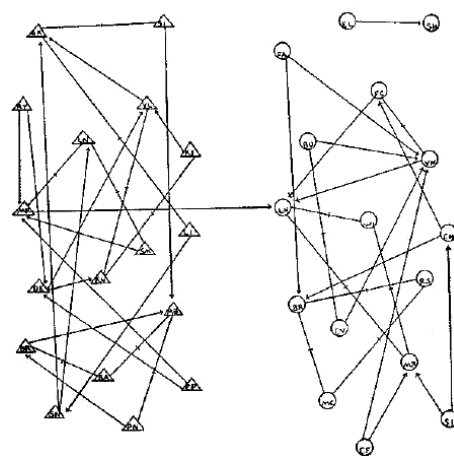
Pode-se notar que várias redes como as de computadores, de pessoas, de empresas ou de moléculas são teias ou malhas, apresentam ligações entre seus elementos. O corpo humano é um rede de interação complexa de moléculas dentro das células, a economia representa empresas e consumidores, a sociedade é uma rede de pessoas ligadas por laços de família, amizade, lazer e trabalho. Como nota-se todos os seres humanos fazem parte de uma rede social por alguma certa forma.

Freeman (1979) ressaltava alguns pesquisadores pioneiros nos estudos de redes sociais como John C. Almack (1883 – 1953) em 1922 com o trabalho "The influence of intelligence on the selection of associates. School and Society", Beth Lucy Wellman (1895 – 1952) em 1926 com o artigo "The school child's choice of companions", Eugénie Chevaleyra-Janovskaja em 1927 com "Groupements spontanés d'enfants à l'âge préscolaire", Helen Bott (1862 – 1953) em 1928 com "Genetic Psychology Monographs", Ruth Hubbard (1929 – 2011) em 1929 com o artigo "A method of studying spontaneous group formation. In Some New Techniques for Studying Social Behavior" e Elizabeth Hagman com "The companionships of preschool children" no ano de 1933.

No ano de 1929 Frigyes Karinthy (1887 – 1938), com o trabalho intitulado "Láncszemek", através de um personagem fictício faz diversas simulações para demonstrar que duas pessoas em qualquer parte do mundo podem estar ligadas por no máximo cinco ligações, através de sua obra inspirou o conceito conhecido hoje por "seis graus de separação".

No entanto, o conceito de teoria das redes sociais foi marcado com o lançamento do livro intitulado por "Who shall survive?" de Jacob Levy Moreno (1889 – 1974) e da revista "Sociometry" (Moreno, 1934). Jacob Moreno com seus estudos criou um meio para avaliar as relações grupais, dando origem a *sociometria* (Figura 2.20).

Figura 2.20: O sociograma de Jacob Moreno.



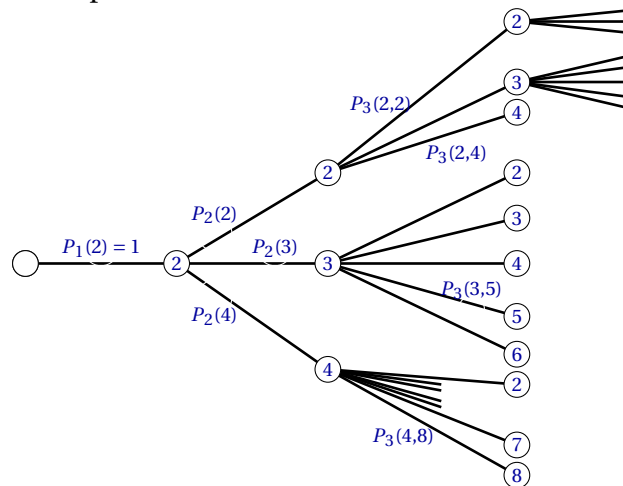
Fonte: Moreno (1934).

A sociometria é uma técnica que fornece um grafo conhecido por *sociograma* que permite analisar as escolhas e percepções das relações grupais entre indivíduos representados através de símbolos. Deste modo, consegue explicitar os vínculos/laços de influência e de preferência que existem nesse mesmo conjunto. Na Figura 2.20 acima temos o grafo sociograma de Jacob Moreno, os símbolos Δ e O representam homens e mulheres respectivamente. As linhas representam as ligações entre as pessoas nas relações de amizade.

No final da década de 40 o pesquisador Alex Bavelas (1913 – 1933) influenciado pelos estudos de Jacob L. Moreno fundou e dirigiu o "Group Networks Laboratory" do Massachusetts Institute of Technology (MIT), um grupo de estudos teóricos que estudava as redes sociais analisando as diferenças consideráveis no caráter das relações em grupos entre diversas estruturas de comunicação como outros aspectos. Tendo como contribuição especial a revelação da importância da relação entre a centralidade de um agente (vértice) e sua influência sobre o grupo (Bavelas, 1948; Deaux, 2012).

Entre outros pesquisadores pioneiros o matemático Rapoport (1911 – 2007) que no ano de 1948 publica "Cycle distributions in random nets" e Ray Solomonoff (1926 – 2009), o inventor da probabilidade algorítmica conhecido pela sua teoria geral da inferência indutiva, que em conjunto com Rapoport em 1951, publicaram "Connectivity of Random Nets" (Solomonoff, 1952). Abaixo a Figura 2.21 da árvore de probabilidades do número de ancestrais de um indivíduo de Solomonoff.

Figura 2.21: Árvore de probabilidades do número de ancestrais de um indivíduo.



Fonte: Adaptada de Solomonoff (1952).

No ano de 1954 John Arundel Barnes (1918 – 2010) publicou um artigo na revista científica antropológica britânica "Human Relations" sobre a organização social dos habitantes da cidade de Bremmer na Noruega e utilizou pela primeira vez e sistematicamente o termo rede social (*social network*). O objetivo do seu trabalho era explicar a organização social de uma pequena comunidade, através da análise de padrões de laços entre seus habitantes incorporando os conceitos tradicionalmente conhecidos. Com esse trabalho Barnes descobre que

se uma pessoa A tem laços independentes com outras duas pessoas B e C, é bem provável que exista um laço entre as pessoas B e C (Barnes, 1954).

No final da década de 50 Rapoport (1957) introduz modelos matemáticos aplicados à análise de redes sociais, destacando-se por revelar a importância da distribuição do número de ligações dessas redes, ao traçar uma visão da sociedade como uma rede de pessoas que formavam laços de amizade de forma aleatória, caracterizando laços fortes de amizade entre pessoas de comunidades próximas e laços fracos entre pessoas de comunidades distantes (Gilbert, 1959, 1960). Em 1959, os pesquisadores Paul Erdős² (1913 – 1996) e Alfréd Rényi³ (1921 – 1970) construíram dois modelos de redes denotados por $G(N, M)$ e $G(N, p)$, $G(N, M)$ é caracterizado por construir grafos aleatórios que partem de um número fixo de N vértices desconexos e M arestas, de forma que os vértices tornam-se adjacentes aleatoriamente tendo seus graus aproximadamente iguais. Enquanto que $G(N, p)$ inicia por uma quantidade fixa N de vértices desconexos onde por uma probabilidade p constante são atribuídas arestas de forma aleatória. Esse modelos foram intitulados por *Redes Aleatórias* (Erdős & Rényi, 1959). Independentemente, o modelo $G(N, p)$ de Erdős e Rényi foi proposto por Edgar Nelson Gilbert⁴ trata de construir grafos aleatórios que começam por um número fixo de N vértices desconexos e cada aresta possível ocorre de forma independente com probabilidade p .

Em 1967, Stanley Milgram (1933 – 1984) passou a se interessar pela estrutura de nossa rede social. Uma das provas definitivas de quanto o mundo é pequeno, no entanto, não veio de uma simulação matemática mas sim de um inventivo experimento em psicologia social. Stanley Milgram, ao realizar um experimento chegou a descoberta que duas pessoas no mundo estão ligadas aproximadamente por no máximo seis laços de amizade, formalizando o resultado obtido de em média 5,5 pessoas para o conceito dos “seis graus de separação” pelo qual ficou difundido. Um resultado impressionante, pois é muito próximo ao que Karinthy já havia encontrado no final dos anos 30 (Milgram, 1967; Barabási & Frangos, 2002).

Granovetter (1973) contribuiu importantemente sobre estruturas de redes sociais ao estudar os tipos de laços das pessoas dentro de grupos sociais. Entre suas descobertas nas relações de amizade está a caracterização dos laços fracos (*weak ties*) e laços fortes (*strong ties*), quando descreve que laços fortes como os de família e de amigos próximos são altamente agrupados refletindo baixa diversidade de características, ao contrário dos laços fracos como os de colegas ou amigos distantes, que podem pertencer a grupos distintos gerando maior diversidade.

²Paul Erdős foi um dos maiores autores de matemática da história, comparando-se em números apenas com Leonhard Euler e chegando a escrever mais de 1400 artigos matemáticos durante sua vida como autor co-autor (Hoffman, 1998)

³Alfréd Rényi foi um matemático brilhante que faleceu tragicamente ainda jovem aos 48 anos de idade e que dedicou a maior parte de sua vida aos estudos de teoria da probabilidade (Suzuki, 2001).

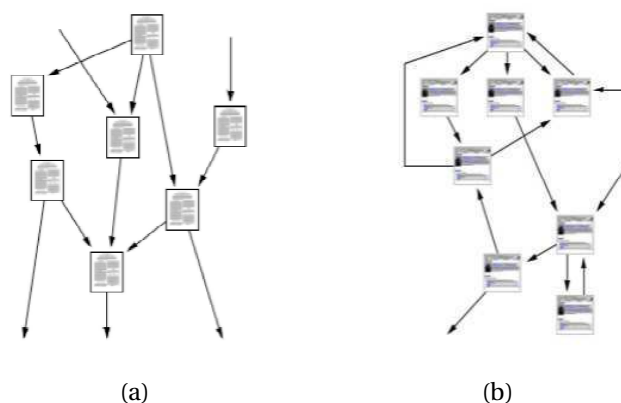
⁴Edgar Nelson Gilbert é um físico magnífico que tem suas maiores contribuições na teoria da codificação com a prova do Limite Gilbert-Varshamov-Encadeado no desenvolvimento do modelo matemático Gilbert-Elliott em 1960 e na teoria das redes complexas com o modelo de grafo aleatório $G(N, p)$.

Com os avanços tecnológicos e o surgimento dos computadores no final dos anos 60 foi possível analisar outros tipos de redes. No entanto, os primeiros computadores eram de tamanho enorme e de uso exclusivo militar. Só mais tarde no final da década de 70 que através da redução dos custos de produção e do surgimento de algumas inovações no processo de miniaturização de componentes é que foi possível a comercialização de um computador pessoal, conhecido pela expressão norte-americana (*computer personal*) (PC) (Stallings, 2010).

O advento do computador fez a comunidade científica avançar seus estudos em outras direções, pois a possibilidade de estudar redes representadas por grafos com propriedades topográficas complexas, características como alta densidade, direcionamento, milhares de vértices e peso nas arestas entre outras aumentou o interesse por parte de vários pesquisadores dando origem as *redes complexas*.

Exemplo deste tipo é o grafo que representa a *World Wide Web* (WWW), pois vértices são os sítios, arestas são os hyperlinks entre pares de sítios e o fluxo caracteriza-se pela direção. Convém esclarecer que WWW e Internet são redes distintas, a WWW é formada por sítios e documentos enquanto que a Internet é uma rede física composta por computadores que pertencem a diversas redes particulares. Podemos ver um exemplo de representação de WWW na Figura 2.22 abaixo.

Figura 2.22: Rede de citações de trabalhos acadêmicos (2.30(a)) e a rede WWW (2.30(b)).



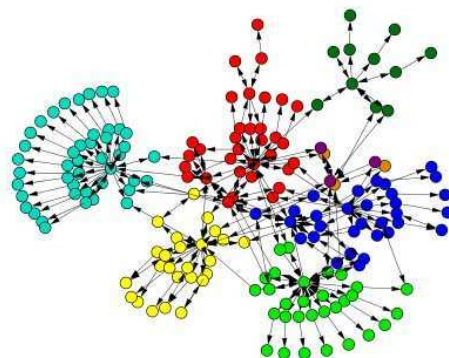
Fonte: Newman (2003).

2.2.1 Redes Complexas

Na teoria das redes, nem todo grafo pode representar uma rede complexa, por isso apenas alguns tipos de grafos são classificados como redes complexas. Alguns exemplos de redes complexas são as redes de drenagem para escoamento de água, redes sociais que mostram a relação entre os seres humanos, redes de computadores, redes de cadeia alimentar, propagação de vírus e cadeias alimentares (Newman, 2003; Newman & Girvan, 2003; Girvan & Newman, 2002). Porém, temos ainda outros exemplos como redes sociais que fornecem vários serviços do tipo comunidades, tais como o Facebook e redes profissionais como

a LinkedIn. Abaixo temos um exemplo de uma rede de colaboração entre cientistas (Figura 2.23).

Figura 2.23: Rede de colaboração científica.



Fonte: Newman & Girvan (2003).

Para caracterizar uma rede complexa através das propriedades presentes em grafos Costa et al. (2007) apresentam dois conceitos denominados *Simetrização* e *Limiarização* e com isto exemplificam diferentes tipos de redes: (i) rede sem direção e sem peso, (ii) rede sem direção e com peso, (iii) rede com direção e sem peso e (iv) rede com direção e com peso.

- **Simetrização**

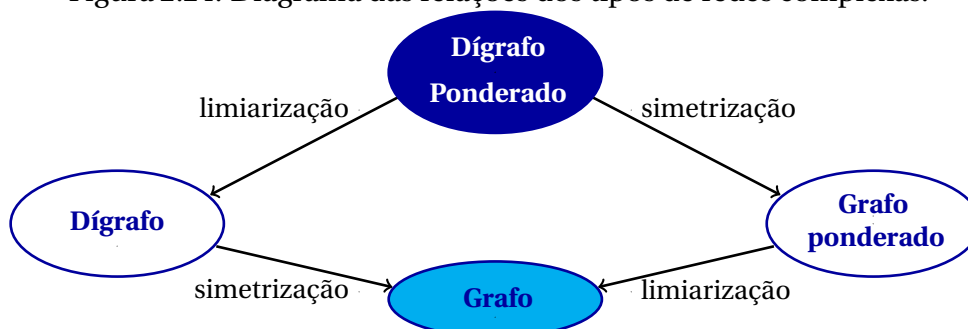
Uma simetrização entre duas redes acontece ao transformar um dígrafo e um grafo não orientado. Um exemplo é de um dígrafo e seu grafo fundamental.

- **Limiarização**

Diz-se que duas redes são limiares quando a diferença entre dois grafos representativos está apenas por um ser ponderado e o outro não. Para essa mudança de rede ponderada para não ponderada denomina-se como Limiarização.

Podemos ver na Figura 2.24 abaixo os tipos de redes e os níveis *simetrização* e *limiarização*.

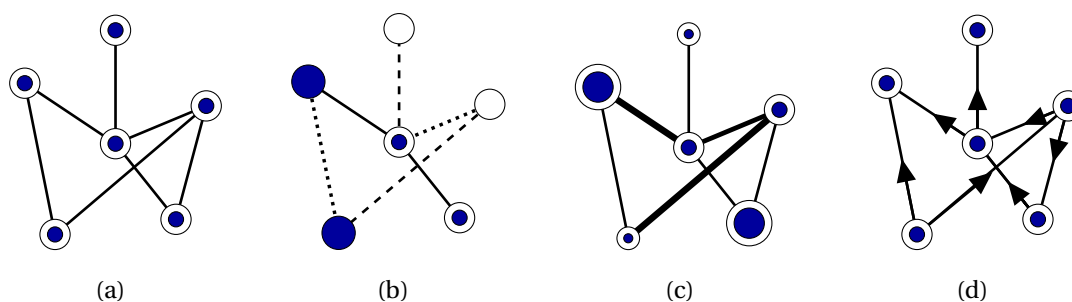
Figura 2.24: Diagrama das relações dos tipos de redes complexas.



Fonte: Adaptada de Costa et al. (2007).

Segundo Newman (2003) as redes complexas se apresentam de vários tipos diferentes, podendo variar os tipos de vértices e arestas. Essas características podem representar vértices altamente conectados, as arestas podem ser ligações que passam informações entre vértices, entre outras características (Figura 2.25).

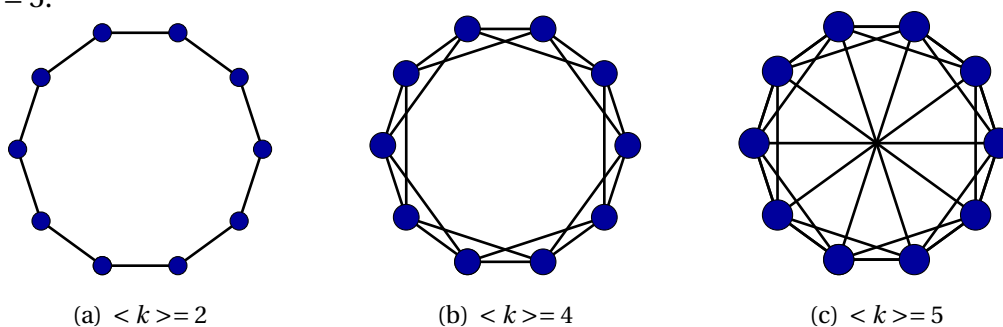
Figura 2.25: Rede não direcionada com apenas um tipo de vértice e aresta (2.34(a)); Rede não direcionada com poucos vértices e arestas distintos (2.34(b)); Rede não direcionada com diversos tipos de vértices e arestas de pesos distintos (2.34(c)); Rede direcionada em que cada nó possui uma direção diferente (2.44(d)).



Fonte: Adaptada de Newman (2003).

As redes mais triviais são as redes regulares, pois a conectividade média $\langle k \rangle$ dos seus n vértices é fixa. Basicamente são construídas começando com n vértices desconectados e a cada passo um vértice é escolhido aleatoriamente tornando-se adjacente a outros $\langle k \rangle$ vértices, não podendo gerar arestas paralelas (Figura 2.26).

Figura 2.26: Redes aleatórias regulares com 10 vértices e conectividades $\langle k \rangle = 2$, $\langle k \rangle = 4$ e $\langle k \rangle = 5$.



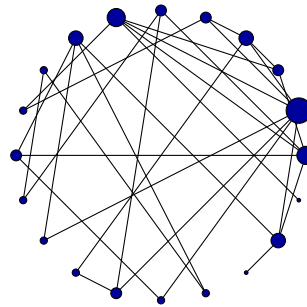
Fonte: Autor (2016).

Todas as redes complexas podem ser classificadas quanto a seus nós, arestas, estrutura física e funcionalidades entre outras características. Logo, desde o surgimento das primeiras redes complexas em meados da década de 40 que vários modelos já foram propostos, entre os principais modelos estudados estão as Redes aleatórias (*Random Graph*), as redes de mundo pequeno (*Small-World*) e as redes livres de escala (*Scale-Free*).

No ano de 1949, os pesquisadores Paul Erdős (1913 – 1996) e Alfréd Rényi (1921 – 1970) construíram dois modelos de rede que foram intitulados por *Redes Aleatórias*. O modelo

primário $G(N, M)$ (Figura 2.27) surge a partir de dois conjuntos finitos N e M , de n vértices m arestas respectivamente, com cardinalidade $|M| \leq \binom{N}{2}$. Neste modelo inicialmente todos os vértices possuem grau zero, a partir daí dois vértices são conectados aleatoriamente e sem gerar arestas paralelas finaliza-se a construção atingindo o n número de arestas (Erdős & Rényi, 1959).

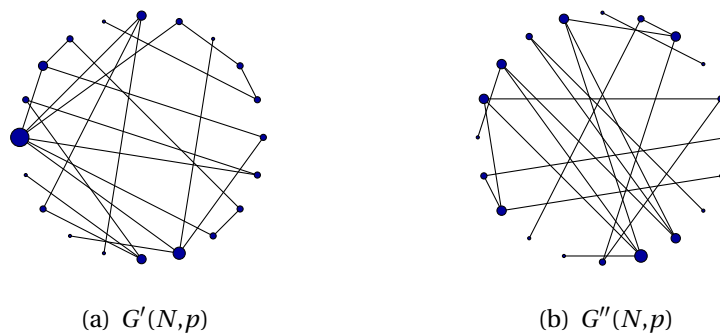
Figura 2.27: Rede aleatória $G(N, M)$ com $|N| = 20$ e $|M| = 30$, onde o grau dos vértices podem ser observados pelo tamanho dos mesmos.



Fonte: Autor (2016).

O segundo modelo $G(N, p)$ origina-se de um conjunto finito N de n vértices desconexos e de um escalar de probabilidade p , com $0 \leq p \leq 1$. Neste caso, dois vértices são escolhidos aleatoriamente por uma probabilidade p e descartando as possíveis arestas paralelas o grafo pode chegar a finalizar com até a quantidade máxima de arestas $|M| = \binom{N}{2} p$. Interessantemente, a construção de duas redes com mesmos números de vértices n e probabilidades p podem gerar redes distintas (Figura 2.28) .

Figura 2.28: Duas redes aleatórias $G(N, p)$ diferentes com os mesmos $N = 20$ e $p = 0.1$.

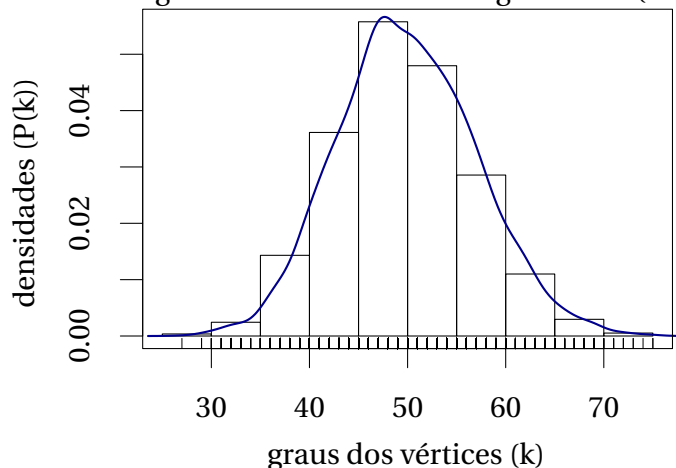


Fonte: Autor (2016).

Essas redes possuem características de grafos simples, a maioria dos vértices tem o mesmo grau e alguns poucos vértices possuem grau acima da média dos graus de todos os vértices, a distribuição de frequências dos graus dos vértices obedecem a uma distribuição binomial devido ao número p de possibilidades que são aleatórias e independentes. Por

outro lado, as distribuições de $G(N, M)$ e $G(N, p)$ tende a assumir propriedades de uma distribuição de Poisson 2.29. No entanto, quando o número de vértices N torna-se muito grande ou quando $N \rightarrow \infty$, as distribuições de $G(N, M)$ e $G(N, p)$ tendem a uma distribuição Gaussiana.

Figura 2.29: Histograma de densidades dos graus de $G(5000, 0.01)$.



Fonte: Autor (2016).

Entretanto, muitas redes reais apresentam uma propriedade que é atribuída ao chamado conceito de pequeno mundo, ou seja, qualquer vértice liga-se a um outro qualquer a partir de outros vértices por um número pequeno de arestas. No final da década de 60 o pesquisador Stanley Milgram procurou estimar o tamanho das redes sociais e quantas etapas são necessárias para que dois estranhos estabeleçam um vínculo (Figura 2.30).

Figura 2.30: Rede de amizades entre pessoas.



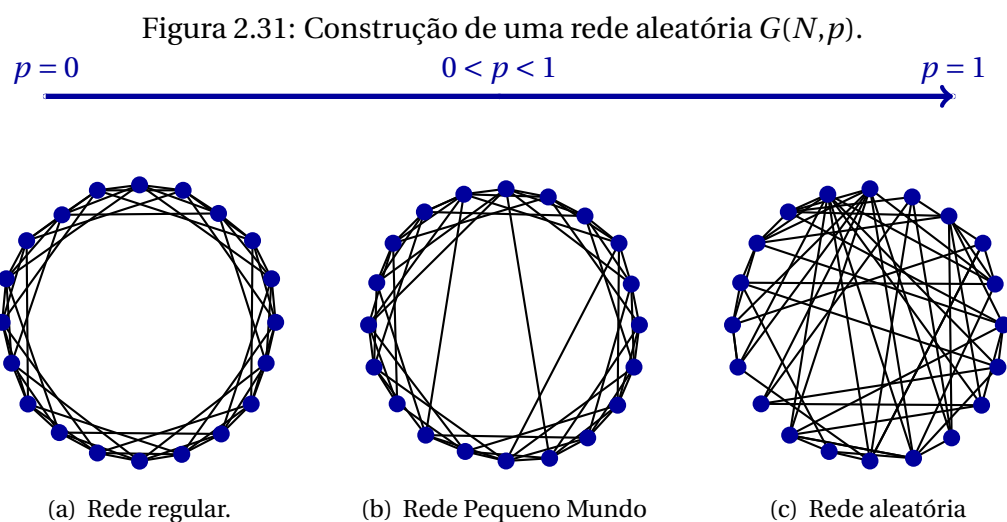
Fonte: Milgram (1967).

No seu experimento, Milgram enviou pacotes para 296 pessoas em Boston e Nebraska pedindo que reenviassem os pacotes a um destinatário em Massachusetts. Mas ele não forneceu o endereço desse destinatário, apenas o primeiro nome, a ocupação e algumas informações pessoais. As pessoas passaram a enviar os pacotes a outras que acreditavam estar mais próximas de conhecer o destinatário final. Como resultado teve-se que os pacotes chegaram ao destinatário-alvo após cinco reenvios. Isso significava que com uma postagem a

mais seria possível chegar a qualquer destinatário e descobrindo que são apenas seis graus de separação que separam qualquer pessoa na Terra, formalizando o conceito de rede "Pequeno Mundo" (Milgram, 1967; Barabási & Frangos, 2002).

Watts & Strogatz (1998) observando as redes sociais e a WWW perceberam uma relação entre as redes aleatórias e as redes de mundo pequeno, quando descobriram que a densidade de conexões de alguns vértices de muitas redes reais é tipicamente maior do que em grafo aleatórios com o mesmo número de vértices e ligações. Além disso, elevados agrupamentos e ao mesmo tempo uma distância média pequena entre os vértices, e que essas redes tem a mesma propriedade da rede social que relaciona os indivíduos, ou seja, fazem parte do fenômeno conhecido como "mundo pequeno". Watts e Strogatz propuseram o modelo de rede aleatória *Watts-Strogatz* que se situa entre redes completamente ordenadas e redes puramente aleatórias.

O algoritmo por trás do modelo consiste em começar com um grafo regular com N vértices, conectados a k vértices adjacentes em ambos os lados, totalizando $2m$ ligações mais próximas. O próximo passo é redirecionar uma ligação para qualquer outro vértice com uma probabilidade p , quando surgir arestas paralelas, exclui-se uma e continua-se o processo. Esse procedimento gera um grafo obtido ao fim do processo que é regulado pelo parâmetro p e obtemos um grafo com grau médio $\langle k \rangle = 2m$. Ao variar a probabilidade p se pode acompanhar a transição entre uma rede ordenada e uma rede totalmente aleatória, para o grafo permanecer conexo sempre considera-se $N \gg K \gg \ln(N) \gg 1$ (Figura 2.31).



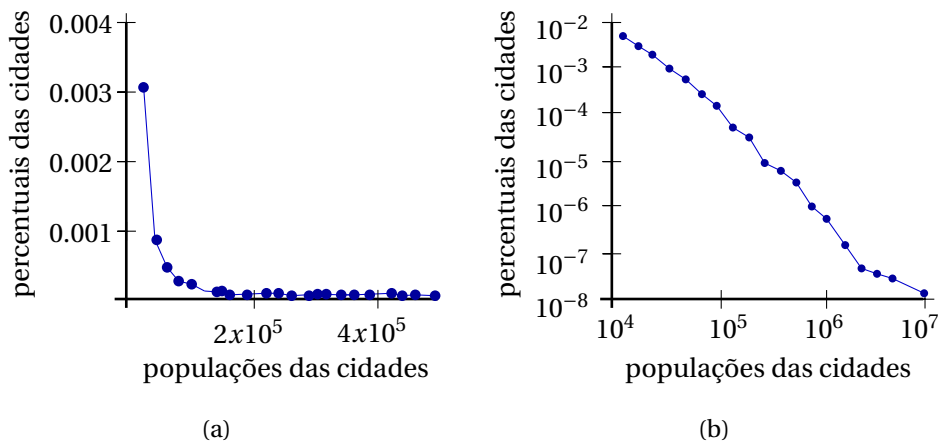
Fonte: Adaptada de Watts & Strogatz (1998).

No final da década de 90, o pesquisador Albert-László Barabási, a pesquisadora Réka Albert reunidos com outros dois pesquisadores Eric Bonadeau e Hawoong Jeong mapearam uma amostra da WWW modelando a rede como um grafo não-orientado no qual os vértices representavam as páginas e as arestas os links entre duas páginas. Neste experimento constataram que a função de distribuição de probabilidade dos graus dos vértices era semelhante

a uma distribuição exponencial de cauda pesada, ou seja, uma lei de potência⁵. Para tanto, apresentaram pela primeira vez o conceito de redes livres de escala propondo um modelo denominado Barabási-Albert(*BA*) para geração dessas redes (Albert & Barabasi, 1999).

Em Albert et al. (2000a) o modelo de redes livres de escala (*BA*) é descrito por um grafo conexo, com $|V| > 1$ inicial que cresce continuamente ao longo tempo em que há inclusão de novo vértice ligado aleatoriamente aos já existentes por uma probabilidade $P(k)$ que depende do número de k arestas de cada vértice. Ou seja, os vértices com mais ligações tem maior probabilidade de se tornarem adjacentes aos novos vértices, obviamente o contrário acontece. Essa forma de atribuir novos vértices é denominada de ligação preferencial (*preferential attachment*) e pelo fato de estarem em constante evolução como a *WWW* são caracterizadas também pelo conceito de *crescimento* (*growth*). De acordo com $P(k)$, é possível fazer uma classificação das redes complexas em duas classes, as *redes homogêneas* que a probabilidade $P(k)$ decai rapidamente e são conhecidas como redes de *cauda leve*, algumas distribuições que apresentam essa característica são a distribuição Gaussiana e a de distribuição de Poisson. A outra classe é a das *redes heterogêneas*, conhecidas como redes de *cauda pesada* (Figura 2.32).

Figura 2.32: A esquerda a densidade da população de todas as cidades norte-americanas com 10.000 habitantes ou mais. Do lado direito da figura temos a mesma distribuição de frequências plotada em escala logarítmica, podemos ver que a curva remete a equação de uma reta. Dados do Censo do ano de 2000 nos EUA.



Fonte: Adaptada de Newman (2005).

Uma distribuição segue uma lei de potência quando sua função de probabilidade $P(k)$ possui a seguinte Equação 2.3

$$P(k) \sim k^{-\gamma}, \tag{2.3}$$

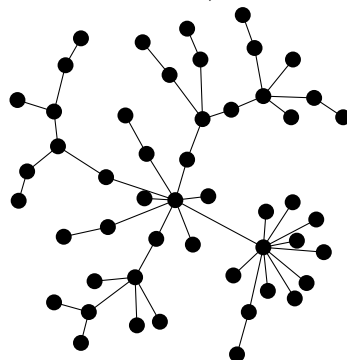
onde γ é a probabilidade da variável aleatória discreta (Boccaletti et al., 2006).

⁵do termo em inglês, power-law. Historicamente, a lei de Pareto, ou princípio 80-20, foi a primeira lei de potência, diz que: 80 das consequências advêm de 20 das causas

As redes livres de escala possuem uma topologia bem específica, pois são formadas por pouquíssimos nós altamente conectados e uma quantidade muito significativa com poucas ligações. Estes poucos nós altamente conectados são os hubs, e possuem grau muito maior que a média de toda a rede. Estas redes são muito operacionais e de boa capacidade de resistência, pois são capazes de resistir à perturbações que levem a perda de nós e/ou ligações devido a grande quantidade de nós com baixo grau. Assim, perturbações aleatórias tem baixa probabilidade de gerarem falhas de conexão. Entretanto, são extremamente passíveis de falhas de conexão quando retirados especificadamente hubs ou ligações ponte (Callaway et al., 2000; Cohen et al., 2000, 2001).

Estudos sobre ligação preferencial e distribuições com leis de potência são realizados desde a década de 20, mas apenas Price (1965) propôs um modelo que se aproximou do modelo conhecido hoje por rede livre de escala. Anos mais tarde, paralelamente a Barabási e Reka Albert, os irmãos (Faloutsos et al., 1999) apresentaram um trabalho semelhante e Broder et al. (2000) mapearam a Web numa escala com mais de 200 milhões de páginas e 1,5 bilhões de links (Figura 2.33).

Figura 2.33: Rede livre de escala, modelo $BA(N = 50, m = 1)$.



Fonte: Adaptada de Albert & Barabasi (1999).

Barabási & Bonadeau (2003) explicam que há várias redes livres de escala presentes na economia, no mundo empresarial, em redes de atores, no relacionamento entre pessoas, nas redes de estudos entre cientistas colaboradores, nas pessoas conectadas por e-mail e nas redes biológicas. Em Boguña & Pastor-Satorras (2002) tem-se um exemplo de vulnerabilidade da rede livre de escala, isto é, facilidade para desligamento de um ou mais elementos, para o caso das relações entre pessoas e epidemias. Neste caso representa-se as pessoas como sendo os vértices e as arestas a transmissão da doença, pois quando uma pessoa infectada morre ou passa a ficar de quarentena a epidemia deixa de se propagar. A teoria das redes complexas trouxe para a ciência o estudo de rotas de aviões e aeroportos (Guimera & Amaral, 2004; Li & Cai, 2004), cadeia de proteínas (Uetz et al., 2000; Maslov & Sneppen, 2002), rede de computadores (Faloutsos et al., 1999), redes de energia elétrica (Watts & Strogatz, 1998), e a relação entre pessoas (Liljeros et al., 2001; Leskovec et al., 2010) e a Internet (Albert et al.,

1999, 2000a). Em se tratando da Internet, a estrutura topológica facilita a propagação de vírus de computadores que pode levar a sérios prejuízos, por isso o estudo das ligações entre os computadores e servidores é de extrema importância para solucionar estes casos (Balthrop et al., 2004).

O capítulo apresentou o estado da arte do tema proposto, abordando as principais teorias que fundamentam o trabalho. O capítulo seguinte apresentará os materiais e métodos utilizados para o desenvolvimento da pesquisa.

3 MATERIAIS E MÉTODOS

PARA fundamentação e construção metodológica apresentaremos os tópicos 3.1 Medidas de Rede aborda em especial as medidas de centralidade baseadas em vértices. Dando continuidade a seção 3.2 com a descrição dos quantificadores de distâncias estocásticas. Para definição dos pontos-chave que contemplam esta pesquisa está o artigo "Variability analysis of complex networks measures based on stochastic distances" de Cabral et al. (2014). Entretanto, apresentaremos o Ambiente Computacional 3.3, a Modelagem 3.4 e o Validadação da metodologia 3.5 que juntos explicam os mecanismos utilizados no desenvolvimento da dissertação e o ferramental aplicado. Os itens necessários para a reprodução deste trabalho estão disponíveis em <https://sites.google.com/site/fabianobriao/>.

3.1 Medidas de Rede

As Medidas de rede são utilizadas para inferir um grafo e destacar a relevância de uma aresta, de um vértice ou definir as características e o tipo de grafo. As métricas são medidas complexas que agregam em sua fórmula duas ou mais medidas diretas. Pode-se dividir as medidas de rede em dois conjuntos, como *medidas locais* que são responsáveis por quantificar atributos para cada elemento e como *medidas globais* quantificam a rede totalmente sobre seus elementos.

As medidas mais utilizadas em redes complexas são as de conectividade, pela propriedade de serem associadas apenas aos elementos da rede individualmente e por sua simplicidade. Apesar de simples são importantes para identificação de hubs e por estarem presentes em algumas medidas como a densidade. Há também medidas de aglomeração que quantificam o nível de agrupamento dos nós indicando a probabilidade de dois vértices v_i , v_j terem um vértice v_k em comum (Watts & Strogatz, 1998; Albert et al., 2000b).

Outras medidas relacionadas a distância como *tamanho* de uma rede, calculado pela contagem de arestas, e o *diamêtro* da rede que é calculado através da distância geodésica. Essas duas medidas são fundamentais quando se trata da WWW e envio de pacotes pois uma distância muito grande acarreta em um número maior de roteadores, que pode vir a aumentar a probabilidade e ocorrer falhas e congestionamentos afetando o tráfego de informações (Tanenbaum & Wetherall, 2011). A medida *densidade* de uma rede pode ser medida pela densidade do grafo modelo que representa essa rede (Boutier et al., 2003).

Em Newman (2002) é proposta uma medida chamada de *assortatividade*, que é a correlação entre graus de vértices adjacentes, ou seja, em redes reais identifica entre os elementos tendências apresentadas em comunidades como religiosidade, raça, sexualidade, entre outras, quantificando a tendência dos elementos do mesmo tipo se conectarem.

Entre tantas medidas propostas até hoje, existem alguns tipos bem importantes e específicos relacionados aos vértices de uma rede, que esclarecem o quanto um vértice é mais importante, se comunica ou interage em relação aos demais. A esses tipos atribui-se também o nome de *Medida de Centralidade*. A seguir descreve-se sobre cada umas dessas medidas de centralidade citadas.

3.1.1 Medidas de Centralidade baseadas em Vértices

As medidas de centralidade podem ser atribuídas tanto para os vértices quanto para as arestas de uma rede. Nesse estudo analisaremos apenas a centralidade de vértices.

Segundo Bavelas (1948) um vértice pode ser exercer diversas funções em uma rede, logo quanto mais central for sua posição mais influente esse vértice pode vir a ser. O pesquisador Alex Bavelas foi um dos primeiros pesquisadores a se dedicar ao estudo do grau de centralidade dos vértices e estruturas de redes em grafos pequenos, de forma a levantar hipóteses sobre influências em comunidades. No entanto, entre nas décadas de 40 e 60 diversos pesquisadores propuseram trabalhos a cerca de medidas de centralidade de vértice, no ano de 1953 tem-se o pesquisador Alfonso Shimbel com o artigo “Structural Parameters of Communication Networks” e no ano seguinte Marvin E. Shaw publicou no The Journal of Psychology o artigo “On Group structure and the behaviour of individuals in small groups” (Freeman, 1977).

No ano de 1979 Linton Freeman analisou e distinguiu as diversas medidas de centralidade até então publicadas e definindo três destas como mais significativas, sendo elas o Grau do vértice (G_c), Grau de proximidade (P_c) e o Grau de intermediação (I_c), conhecidas também *Degree centrality*, *Closeness centrality* e *Betweenness centrality* respectivamente (Opsahl, 2010). Sendo estas as três medidas de centralidade que fazem parte deste trabalho apresentaremos cada uma dela a seguir.

Grau do vértice

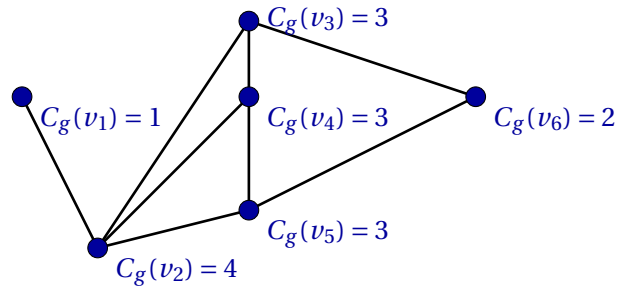
A medida mais trivial de ser calculada é o grau de um vértice, conhecida também por centralização da informação ou grau de centralidade. Por definição o $G_c(v_i)$ de um vértice i é calculado pelo número de arestas incidentes ao vértice, porém nos casos de lacetes conta-se duas vezes a mesma aresta. Outro tipo de contagem pode ser feita para grafos valorados e dirigidos (Bavelas, 1948; Freeman, 1977). Define-se o grau do vértice $G_c(v_i)$ de um vértice v_i de um grafo simples $G = (V, A)$ por

$$G_c(v_i) = \sum_{j=1}^n B_{i,j}, \quad (3.1)$$

onde $B_{i,j}$ é a matriz de adjacências de G .

A Figura 3.1 apresenta um grafo com seus vértices e respectivos graus.

Figura 3.1: Grafo com os valores $C_g(v_i)$ de seus vértices v_i s.



Fonte: Autor (2016).

Para Hanneman (2001) esta medida atribui relevância ao vértice em função do número de relações diretas que o mesmo estabelece com os demais, podendo ser aplicada em casos de fluxo de redes quando as relações indiretas entre vértices não são consideradas. Sendo assim uma medida de influência e de efeito imediato, pois acontece em um tempo $t+1$. Na área da tecnologia em redes de informações ou na saúde em redes de difusão de infecções, o grau de centralidade traduz a capacidade de receber diretamente uma informação ou de ser infectado, respectivamente.

Grau de Proximidade

A medida de centralidade grau de proximidade é baseada no número de caminhos mínimos de um vértice em relação aos outros vértices do grafo (Bavelas, 1948; Sabisussi, 1966; Moxley & Moxley, 1974). Na sua forma mais simples pode ser obtida através da soma das distâncias geodésicas de um vértice v_i aos demais vértices v_j , tal que $i \geq 1$, $j \leq n$. No entanto, o grau de proximidade também é apresentado como o inverso das somas das distâncias geodésicas (Equação 3.2). Entretanto, pode ser aplicada apenas a grafos conexos (Freeman, 1979).

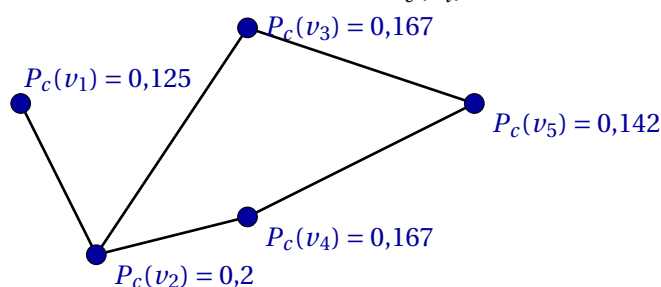
$$P_c(v_i) = \frac{1}{\sum_{j=1}^n d(v_i, v_j)} \quad (3.2)$$

Opsahl (2010) apresenta uma medida de grau de proximidade para grafos desconexos, através da soma dos inversos das distâncias geodésicas. Atribuindo o valor zero no resultado do quociente de $1/\infty$ sem perda de propriedade (Equação 3.3).

$$P_c(v_i) = \sum_{j=1}^n \frac{1}{d(v_i, v_j)} \quad (3.3)$$

Notamos que o vértice v_i está no mínimo a uma distância $P_c(v_i) = 1$ em relação a qualquer vértice v_j , para um grafo conexo com n vértices e no máximo estar conectado a $n-1$ outros v_{ks} . Logo, o maior valor que v_i pode assumir é $P_c(v_i) = \frac{1}{n-1}$. Na Figura 3.2 abaixo, podemos ver que os vértices 3 e 5 possuem maior grau de intermediação, pois eles estão na maioria dos caminhos mínimos de todos os vértices do grafo.

Figura 3.2: Grafo com os valores $P_c(v_i)$ de seus vértices v_{is} .



Fonte: Autor (2016).

Para entender melhor vamos calcular a $P_c(v_3)$ partindo das distâncias $d(v_k, v_3)$ entre v_3 e todos os outros vértices:

$$d(v_1, v_3) = 2, d(v_2, v_3) = 1, d(v_4, v_3) = 2, d(v_5, v_3) = 1$$

.

Da Equação 3.2 a $P_c(v_3)$ é dada por

$$P_c(v_3) = \frac{1}{2 + 1 + 2 + 1} = \frac{1}{6} = 0,1666\dots$$

.

Estes valores do grau de proximidade no grafo da Figura 3.2 indicam que o vértice v_2 é o mais central pois apresentou o maior valor para P_c . Essa medida está fortemente relacionado ao tempo do fluxo da informação transmitida em uma rede, abstraindo variáveis podemos medir pela P_c o tempo que uma informação levaria para passar de um vértice ao restante da rede.

Grau de Intermediação

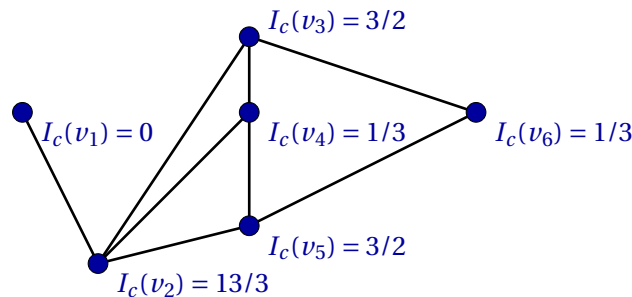
Em meados da década de 70, os pesquisadores Moxley & Moxley (1974) ao observarem as redes de pessoas descobriram que estas são sempre muito grandes e normalmente desconexas. Logo, as medidas que se conheciam não eram muito apropriadas para

redes desconexas. Posteriormente, [Freeman \(1977\)](#) motivado pela descoberta propôs algumas medidas que poderiam ser aplicadas a grafos desconexos, embora fossem baseadas em interligações entre pares de vértices. Entre estas medidas chegou a uma capaz de expressar o quanto um vértice (v_i) está nos caminhos mínimos existentes entre pares de vértices (v_j, v_k) de um grafo. Sendo n o número de vértices da rede, $k \neq i \neq j$ e $0 < j < k \leq n$, o número total de caminhos mínimos desde o vértice j ao vértice k denotado por $\sigma_{j,k}$ e o número de caminhos mínimos de v_j até v_k passando por v_i com notação $\sigma_{j,k}(v_i)$ temos a Equação 3.4) a seguir,

$$I_c(v_i) = \sum_{j < k} \frac{\sigma_{j,k}(v_i)}{\sigma_{j,k}}; k \neq i \neq j, \quad (3.4)$$

Na Figura 3.3 abaixo podemos ver o vértice v_2 possui o maior $I_c(v_2) = 13/3$ sendo um importante vértice de corte, ou articulação. Seguido dos vértices v_3 e v_5 , pois estes são os principais vértices que estão na maioria dos caminhos mínimos de todos os vértices do grafo..

Figura 3.3: Grafo com os valores $I_c(v_i)$ de seus vértices v_{is} .



Fonte: Autor (2016).

Essa medida é interessante por tratar de quantificar o quanto um vértice é importante na transmissão de informações entre outros vértices da rede, mostrando exatamente quantas vezes um vértice aparece no caminho mais curto entre dois outros vértices. No caso da Internet essas informações podem ser os pacotes de dados transmitidos entre os computadores e roteadores da rede.

3.2 Quantificadores Estocásticos

Os quantificadores estocásticos são utilizados para verificar a diferença entre duas distribuições, ou seja, uma noção relativa de distância do quanto duas distribuições se diferen-

ciam por uma medida que pode ser descrito por uma distância geométrica¹, divergência², métrica³ ou mesmo entropia⁴. Em alguns casos para se obter uma medida mais consistente é necessário usar dois ou mais quantificadores e definir novas métricas (Gibbs & Su, 2002).

Com a estatística e probabilidade aplicadas, por exemplo, à transmissão de dados, criptografia, codificação, teoria do ruído, correção de erros e compressão de dados em meados da década de XX surgiu uma nova área de estudos, a Teoria da Informação ou Teoria matemática da comunicação, que vai lidar com sistemas de comunicação. Entretanto, poucos trabalhos desta época foram apresentados à comunidade científica, apenas mais tarde em 1948 que um artigo seminal do pesquisador Claude Shannon ganha grande repercussão passando a ser marco nesta área, através de processos estatísticos Shannon introduziu pela primeira vez um modelo matemático capaz de medir a incerteza sobre um espaço desordenado, e assim descreveu os conceitos de *Entropia e Informação* (Shannon, 1948).

Inferir distâncias, divergências ou métricas entre densidades é um desafio para os todos pesquisadores e existem vários propostos desde o trabalho de Hellinger (1909) com o surgimento da conhecida distância de Hellinger. Entre outros quantificadores propostos destacaram-se anos mais tarde a distância Bhattacharyya e a divergência Jensen-Shannon (Bhattacharyya, 1943; Shannon, 1948). No início dos anos 50 a divergência Kullback-Leibler proposta por Kullback & Leibler (1951) e na década de 60 a distância de Rényi proposta por Rényi (1960). A divergência relativa Aritmética-Geométrica de Taneja (1995), Menéndez et al. (1997) com (h, Φ) -Entropy Diferencial Métrica, entre outros quantificadores presentes na literatura.

Cha & Srihari (2002) apresentam quatro propriedades que definem um quantificador quanto a ser uma métrica, distância ou divergência entre duas distribuições de probabilidades f e g pertencentes a um espaço amostral Ω , sendo estas:

1. Reflexividade: $D(f, g) = 0$;
2. Não negatividade: $D(f, g) \geq 0$;
3. Simetria: $D(f, g) = D(g, f)$;
4. Desigualdade triangular: $D(f, g) \leq D(g, h) + D(h, g); \forall h \in \Omega$.

Por estas propriedades um quantificador será definido por uma métrica se obdecer todas as quatro propriedades, definido como distância se as propriedades 1, 2 e 3 forem obedecidas, e como uma divergência se obdecer as propriedades 1 e 2.

¹É uma medida que traduz a separação entre dois pontos ou dois dados estatísticos, de duas regiões ou dois conjuntos de dados quantitativos.

²É uma medida que pode ser sinônimo de separação, discrepância, dissenção, diferença. Ou seja, a mensuração da diferença entre duas distribuições.

³É um conceito matemático relacionado a distância que pode ser tratado com um conjunto de métodos para descrever uma medida.

⁴É uma medida de desordem, pode-se entender que algo está fora de uma ordem, ou seja, uma distribuição não está da mesma forma que outra.

Podemos ver no trabalho de [Cabral et al. \(2014\)](#) uma classificação entre três quantificadores apropriados para séries estocásticas, mais especificamente distribuições de probabilidades de medidas de redes, foram a divergência de Kullback-Leibler, a distância de Hellinger e a divergência de Jensen-Shannon. Neste trabalho as séries foram construídas por medidas de redes de dois grafos, um original e outro perturbado, agrupadas para dar lugar às distribuições das frequências relativas destes grafos e assim essas novas distribuições foram comparadas. Esse estudo foi realizado utilizando o Método Monte Carlo e aplicado para as redes Livres de Escala, Pequeno Mundo e Aleatórias, para perturbações de adição de arestas, adição de vértices, remoção de arestas e remoção de vértices.

[Cabral et al. \(2014\)](#) observaram que a métrica de Hellinger mostrou-se ser um dos quantificadores mais sensíveis na análise de comparação de redes Livres de Escala. Os quantificadores Hellinger e Kullback-Leibler sob a medida de centralidade grau do vértice analisada para adição de arestas, remoção de arestas e remoção de vértices apresentaram variações significativas, mas apenas para as redes Aleatórias e Pequeno-Mundo. Também para o grau do vértice a divergência de Jensen-Shannon não mostrou variações relevantes em nenhuma das comparações. A medida grau de intermediação apenas nas redes aleatórias, sob adição e remoção de arestas apresentou oscilações, no entanto somente para distâncias de Hellinger e Jensen-Shannon.

Diante das observações chegamos a conclusão que a rede Livre de Escala é mais sensível de modo geral entre as redes analisadas, suas medidas de rede oscilam mais facilmente com as mudanças de topologia. E os quantificadores que conseguiram medir com maior precisão foram a métrica de Hellinger e a divergência de Kullback-Leibler. Para tanto, descorremos a seguir um tópico sobre a métrica de Hellinger por estar definida no escopo desta pesquisa.

3.2.1 Distância de Hellinger (dHel)

Para fins de execução deste trabalho analisaremos uma rede real da Internet que trata do tráfego de mensagens entre duas pessoas, servidor/pessoa e entre dois servidores. Pela analogia da topologia desta rede com a rede livre de escalas usaremos apenas o quantificador distância de Hellinger por ser o mais apropriado ([Cabral et al., 2014](#)).

Em 1907 o matemático Ernst David Hellinger⁵ (1883–1950) apresentou em um de seus trabalhos um tipo diferenciado da integral de Riemann que passou a ser conhecida como integral de Hellinger. Com esse trabalho descobriu um quantificador simples e de simetria clara que tem a propriedade de medir a distância entre distribuições de probabilidades discretas que ficou conhecido por distância de Hellinger ([Hellinger, 1909](#)).

Sendo assim, está definida a distância de Hellinger para distribuições de probabilidades

⁵foi um matemático alemão, nascido em 30 de setembro de 1883 em Striegau, na Silesia, Alemanha, que atualmente chama-se Strzegom no sudoeste da Polônia. Doutor pela Universidade de Göttingen no ano de 1907 com a tese intitulada “Die Orthogonalinvarianten quadratischer Formen von unendlichvielen Variablen”, Hellinger trabalhou com um dos matemáticos mais reconhecidos do século XX, David Hilber.

da seguinte forma:

Distância de Hellinger

Definição: Sejam $P = \{p_1, \dots, p_k\}$ e $Q = \{q_1, \dots, q_k\}$ duas distribuições de probabilidades de Ω , a distância de Hellinger está definida para o caso discreto como,

$$d_{Hel}(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2} \quad (3.5)$$

(Hellinger, 1909).

Cabral et al. (2014) explicam que um quantificador embora seja denominado comumente como uma distância ele também está classificado como métrica e divergência pois satisfaz todas as propriedades necessárias citadas por Cha & Srihari (2002) e a desigualdade $0 \leq DHel(P,Q) \leq 1; \forall \{P,Q\} > 0$.

3.2.2 Perturbações em Grafos

A toda operação que representa alterações sofridas em alguns dos componentes de um grafo, se define por *perturbação em grafo*. Essas mudanças na estrutura topológica podem ser nas arestas, ou nos vértices, ou nas arestas e vértices.

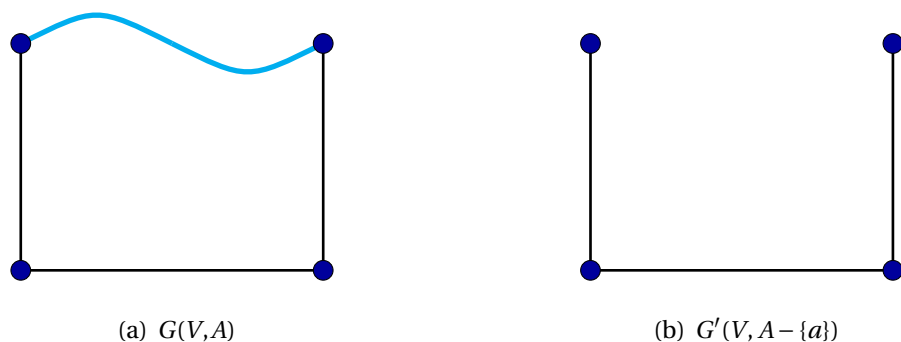
Os vértices e arestas de um grafo podem representar diversos modelos que simulam fenômenos reais, por exemplo nas redes de energia elétrica os postes e cabos podem representar os vértices e arestas respectivamente. Neste caso, a ruptura de cabos representa a retirada de arestas e a queda de postes representa a remoção de vértices, caracterizando uma modificação na estrutura topológica da rede de energia, a essa modificação chamaremos de perturbação. As operações que caracterizam essas perturbações são unárias, isto é, em cada operação existe apenas um único grafo como entrada. Porém, nem todas as perturbações são cabíveis de análise a todo tipo de medidas e redes.

Algumas redes complexas são mais sensíveis para alguns tipos de perturbações do que outras. Estas variações em alguns casos dependem também do tipo e da intensidade da perturbação (Cabral et al., 2014). Abaixo podemos ver se uma descrição detalhada acerca dos tipos de perturbações existentes.

- **Remoção de aresta sem remoção de vértice(s).**

Dado um grafo $G(V,A)$ 1.23(a) com $A(G) \neq \emptyset$, caracteriza-se um grafo modificado pela remoção de aresta, ou subgrafo gerador de G (Figura 3.4), todo grafo $G'(V, A - \{a\})$ 2.36(b).

Figura 3.4: Grafo $G(V,A)$ e seu subgrafo $G'(V, A - \{a\})$ perturbado por remoção de aresta.



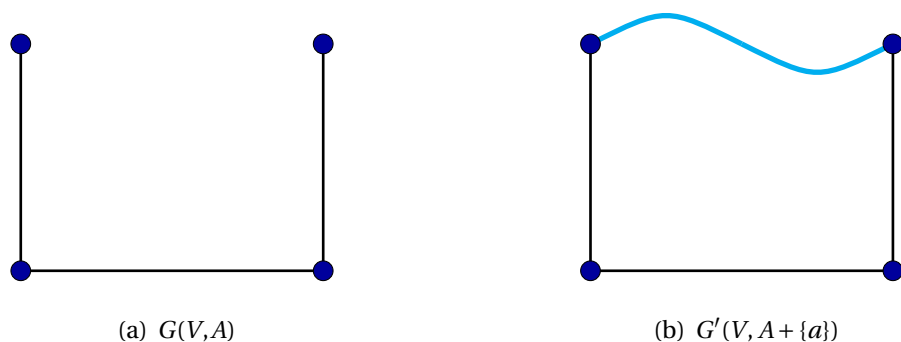
Fonte: Autor (2016).

Também considera-se como remoção de arestas sem remoção de vértices, a remoção de laços e arestas paralelas.

- **Inclusão de aresta**

A Figura 3.5 mostra a operação por perturbação do grafo $G(V,A)$, com inclusão de aresta gerando o grafo $G'(V, A + \{a\})$.

Figura 3.5: Grafo $G(V,A)$ e o grafo $G'(V, A + \{a\})$ perturbado por inclusão de aresta.



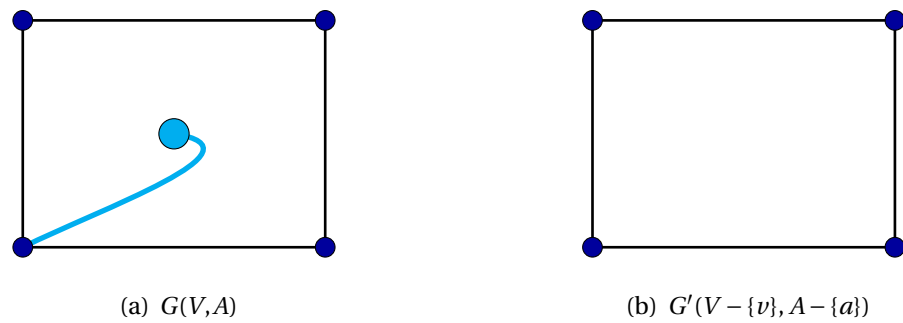
Fonte: Autor (2016).

A inclusão de laços também inclui-se nesta operação, assim como a inclusão de arestas paralelas.

- **Remoção de vértice e aresta(s)**

Dado um grafo $G(V,A)$, caracteriza-se grafo perturbado $G'(V - \{v\}, A - \{a\})$ o subgrafo induzido de G que não é gerado pela remoção de vértice de corte. A remoção de vértices em grafos conexos pode acarretar também na remoção de uma ou mais de uma arestas ligadas a ele (Figura 3.6).

Figura 3.6: Grafo $G(V,A)$ e seu subgrafo $G'(V - \{v\}, A - \{a\})$ gerado pela remoção de vértice.

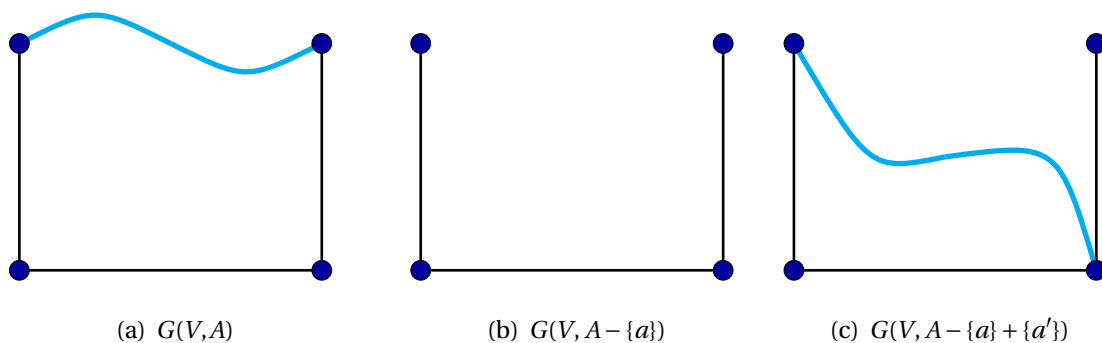


Fonte: Autor (2016).

• Remoção e inclusão aresta

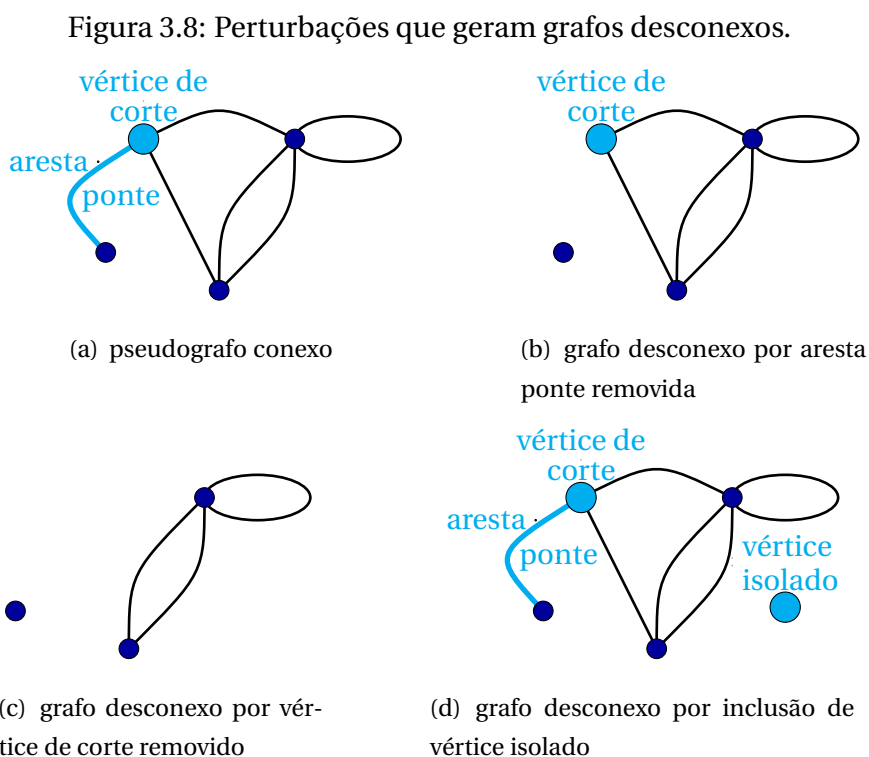
Dado um grafo $G(V,A)$ conexo, caracteriza-se $G'(V, A - \{a\} + \{a'\})$ todo grafo modificado de G através da remoção e adição de aresta (Figura 3.7).

Figura 3.7: Grafo original $G(V,A)$ 1.26(a), grafo com retirada de aresta $G(V, A - \{a\})$ 1.26(b) e o grafo $G(V, A - \{a\} + \{a'\})$ 1.26(c) gerado pela inclusão de aresta.



Fonte: Autor (2016).

A seguir temos um grafo que possui laço e arestas paralelas, denominado pseudografo. E novos grafos desconexos oriundos deste por passarem por perturbações como: remoção de aresta ponte, remoção de vértice de corte e inclusão de vértice isolado, como podemos ver na Figura 3.8 abaixo.



Fonte: Autor (2016).

3.3 Ambiente Computacional

Análise bibliográfica

Adequadamente, todo alicerce desta pesquisa está calçado em uma análise bibliográfica que tem como ponto de partida alguns dos principais repositórios, portais e revistas de literaturas reconhecidos pela comunidade científica. No entanto está se citando apenas os mais acessados como:

- ArXiv (<http://www.arxiv.org>);
- Association for Computing Machinery (ACM) (<http://www.acm.org>);
- Brazilian Journal of Probability and Statistics (BJPS) (<http://www.imstat.org>);
- Capes (<http://www.periodicos.capes.gov.br>);
- CiteSeerX (<http://citeseerx.ist.psu.edu>);
- Institute of Electrical and Electronics Engineers (IEEE) (<http://www.ieee.org>);
- ISI Web of Knowledge (<http://isiknowledge.com>);
- Science Direct (<http://www.sciencedirect.com>).

Optamos por utilizar o padrão de referências do estilo BIB_{TeX} e disponibilizá-las em <https://sites.google.com/site/fabianobriao/references>. Este estilo é um

modelo de formatação dos itens das referências (autor, título, ano, ISBN, DOI, editora, ...) e de seus tipos para bibliografia (artigo, livro, capítulo de livro, site, ...) construído sobre um arquivo de texto que geralmente tem a extensão *.bib*. O BIBTEX foi criado pelo doutor em ciência da computação Oren Patashnik em colaboração com Leslie Lamport em 1985 para facilitar a edição das referências e servir como ferramenta para processamento das mesmas em LATEX , XHTML, CSS entre outras linguagens (<http://www.bibtex.org/>).

Softwares

Segundo o conceito de RR, para garantir total transparência na reprodução desta pesquisa utilizamos somente softwares gratuitos, também conhecidos como softwares livres. Alguns outros foram utilizados, mas nem todos são citados neste escopo. Citamos aqui apenas os principais como o (1) *Sistema Operacional GNU/Linux* (<http://www.linux.com>), a (2) *Plataforma R* (Team, 2010) e a plataforma de processamento (3) LATEX (<http://www.latex-project.org>).

1. *Sistema Operacional GNU/Linux*

O Linux é simultaneamente um núcleo (ou kernel) e um sistema operacional rodando sobre este núcleo. Criado por Linus Benedict Torvalds⁶ segue o padrão POSIX, uma padronização do (IEEE). Ou seja, o Linux é parecido com sistema operacional Unix, que é um sistema operacional licenciado sob os termos da GPL, é portátil, multitarefa e multi-usuário. Com código fonte disponível no site The UNIX System (<http://www.unix.org/>) pode-se realizar instalações e cópias diversas sem nenhuma restrição, isto é, qualquer pessoa com conhecimentos da linguagem pode alterar e até comercializar o sistema, mas não pode fechá-lo, ou seja, não permitir que outros usuários o modifiquem. Entre as distribuições existentes para o Linux utilizamos a versão 12.10 Quantal Quetzal disponível no site Ubuntu (<http://www.ubuntu.com>).

2. *Plataforma R*

A Plataforma R é uma linguagem de programação voltada ao desenvolvimento de sistemas para estatística computacional e geração de gráficos de alta qualidade. Segundo Almiron et al. (2009) a Plataforma R possui excelente qualidade em suas propriedades numéricas que garantem resultados de ótima aproximação. Um projeto General Public License (GNU) (<http://www.gnu.org>) sob a licença GNU General Public License (GPL) fazendo parte dos programas de código aberto denominados Free/Libre Open Source Software (FLOSS). Usamos a versão 2.15.1 que pode ser obtida gratuitamente e atual no site The R-Project for

⁶nascido em Helsínquia/Finlândia, estudante do Departamento de Ciência da Computação da Universidade de Helsinki, no ano de 1991 criou o Linux, núcleo do sistema operacional GNU/Linux.

Statistical Computing (<http://www.cran.r-project.org>). Para o desenvolvimento da pesquisa foram utilizados na implementação os seguintes pacotes:

- Lattice
Lattice é um pacote de recursos para visualização de dados multivariados. Com o Lattice é possível refinar complexos gráficos com altíssima precisão e qualidade (Sarkar, 2008).
- Igraph
Igraph é um pacote desenvolvido para o trabalho com grafos, com ele se é capaz de gerar grafos de diversos tipos incluindo tipos específicos como redes complexas, pode-se calcular medidas e métricas de rede, além de algoritmos de alta complexidade (Csardi, 2012).
- Base
Usamos o pacote de funções base por ser útil como suporte básico de programação, e por possuir a função *set.seed* que utiliza uma expressão numérica ou literal servindo como chave para gerar os mesmos conjuntos de dados em casos de geração aleatória e utilizamos o *set.seed* para gerar o mesmo grafo modelo durante as simulações.
- TikzDevice
O pacote *tikzDevice* foi utilizado neste trabalho como um dispositivo pra transformar o código R criado para as figuras em uma saída de código \TeX para \LaTeX . Este pacote foi útil para gerar os as figuras dos grafos e de histogramas neste trabalho (<https://github.com/Sharpie/RTikZDevice>).

3. \LaTeX

O \LaTeX é um programa desenvolvido por Leslie Lamport⁷ no ano de 1994, que funciona sob o domínio da linguagem \TeX , desenvolvida por David Ervin Knuth⁸ no final da década de 70, a linguagem \TeX funciona como um compilador do \LaTeX que foi desde sua primeira versão orientado para ter objetivo facilitar e padronizar a escrita de textos científicos e fórmulas matemáticas. Entre as diversas qualidades do \LaTeX está a sua capacidade de produzir documentos com alta qualidade tipográfica (Lamport, 1994).

Ferramentas de Hardware

Seguindo o conceito de *RR* apresentamos aqui a principal ferramenta de hardware utilizada, um microcomputador notebook com as seguintes características:

⁷um norte americano graduado em matemática que teve seu maior reconhecimento pelo trabalho pioneiro da construção do sistema \LaTeX .

⁸é o criador da linguagem \TeX , entre tantos títulos possui o de Ph.D. em matemática no ano de 1963 pelo California Institute (Knuth, 1984)

- Processador Intel Core i5 2410M de 2ª Geração;
- CHIPSET HM65;
- Arquitetura de 64 bits;
- Velocidade no processamento de 2,3 GHz;
- Memória DDR3 de 8 GB.

3.4 Modelagem

A Internet é composta de milhares de empresas que fornecem seu acesso individualmente, denominados de Internet Service Provider (ISP) ou provedor de acesso à Internet (IAP). Os ISPs são empresas que fornece acesso à Internet em sua área geográfica para acesso pessoal ou comercial. Quando algumas dessas empresas ISPs trabalham sob uma gestão comum de um Sistema Autônomo (AS) possuem características e políticas de roteamento comuns. Para tanto, uma série de relações entre os servidores são necessários para se rotear o tráfego, e acordos comerciais são traduzidos em restrições de engenharia no fluxo deste trafego para que redes individuais participem do roteamento da Internet global.

Luckie et al. (2013) analisaram um conjunto de dados livres da Internet referentes as relações entre Sistemas Autônomos (ASs), como redes de clientes, redes universitárias e pontos de intercâmbio de Internet. Esta pesquisa foi realizada no Centro de Análise Aplicada de Dados da Internet⁹ (CAIDA), um empreendimento colaborativo entre organizações comerciais e governamentais formado em 1997 e localizado no San Diego Supercomputer Center (SDSC) na Universidade da Califórnia, funcionando como um centro avançado de pesquisas da internet faz investigações práticas e teóricas de grande porte em redes robustas da internet com o intuito de contribuir com pesquisas em medições, segurança, estabilidade, topologia, e análise dos dados fornecendo-os com acesso livre. Nesta pesquisa iremos discorrer sobre as relações de um conjunto de ASs fornecidos pelo CAIDA em <http://data.caida.org/datasets/2013-asrank-data-supplement/data/>.

Este trabalho teve como instrumento para geração dos resultados um algoritmo Monte Carlo¹⁰ que teve como base na sua construção o modelo utilizado por Cabral et al. (2014), no entanto apresentamos algumas modificações para concretizar este experimento. A fim de melhor ilustrar o desenvolvimento do algoritmo apresentamos a seguir um fluxograma (Figura 3.9) explicativo das fases do desenvolvimento e posteriormente o algoritmo 3.1 que demonstra a simulação. Veremos agora os conceitos e notações que utilizamos, descrições dos parâmetros e níveis de simulação, assim como as condições iniciais necessárias para o

⁹CAIDA pode ser acessado em <http://www.caida.org/home/>.

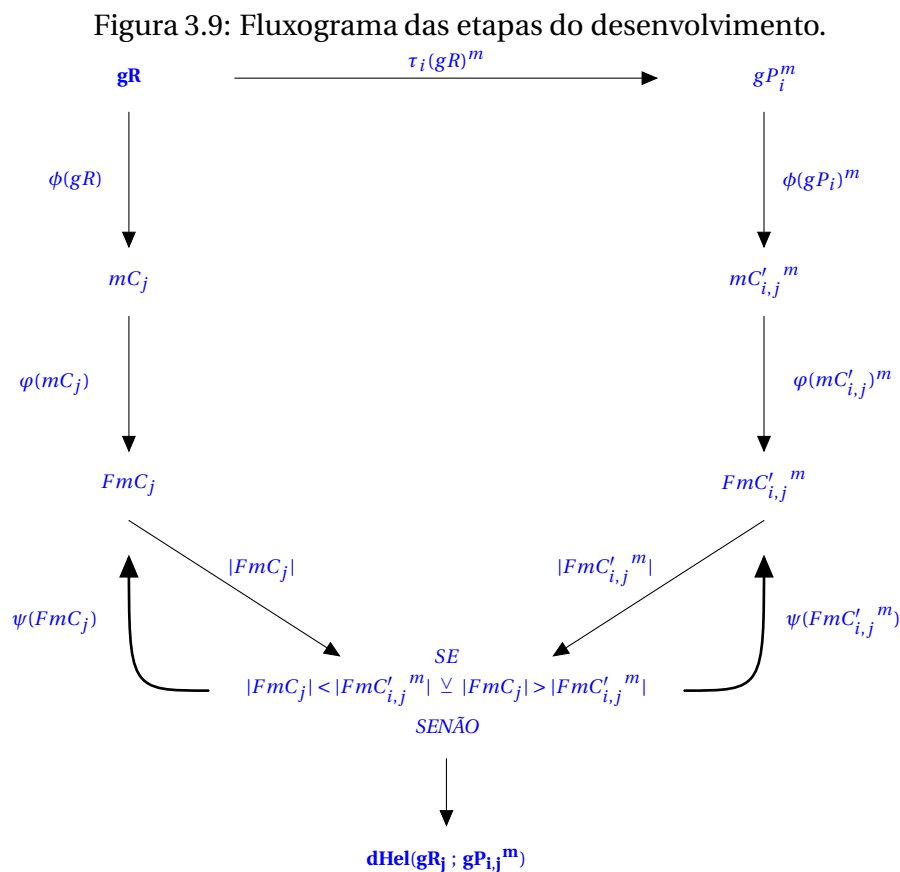
¹⁰originou-se das observações feitas em jogos de roleta no Cassino de Monte Carlo. Entre suas propriedades é um recurso fácil e poderoso para inferir parâmetros de distribuições desconhecidas ou violadas (Metropolis & Ulam, 1949; Robert & Casella, 2009).

desenvolvimento do trabalho. As notações a seguir são relativas ao fluxograma e ao algoritmo apresentados.

- \mathbf{gR} é a notação do grafo da rede real que representa a Internet, essa rede é apresentada no artigo AS Relationships, Customer Cones, and Validation que descrevemos anteriormente no capítulo da Fundamentação Teórica;
- $\mathbf{gP} = \{\mathbf{gP}_1, \mathbf{gP}_2, \mathbf{gP}_3, \mathbf{gP}_4\}$ é o conjunto de todos os grafos perturbados, sendo \mathbf{gP}_1 : grafos com vértices removidos e arestas por consequência, \mathbf{gP}_2 : grafos com arestas removidas ou arestas e vértices conseqüentemente, \mathbf{gP}_3 : grafos com adição de arestas e \mathbf{gP}_4 : grafos que sofrem uma remoção determinada de arestas aleatoriamente e logo após recebem a adição de forma aleatória desta mesma quantidade de arestas;
- $\tau_i(\mathbf{gR})$ é a notação para a aplicação de perturbação de grafos, ela transforma um grafo \mathbf{gR} e um novo grafo \mathbf{gP}_i . Definida da seguinte forma, $\tau_1(\mathbf{gR})$: remoção de vértices, $\tau_2(\mathbf{gR})$: remoção de arestas, $\tau_3(\mathbf{gR})$: adição de arestas e $\tau_4(\mathbf{gR})$: remoção e adição de arestas. Essa perturbação pode ser denominada por religação de arestas;
- **Níveis de perturbação**, cada tipo de perturbação $\tau_i(\mathbf{gR})$ possui doze níveis definidos por $m = 1\%, 2\%, 3\%, \dots, 9\%, 10\%, 15\%, 20\%$. Por exemplo, para a perturbação de remoção de vértices será retirado a quantidade de 1% dos vértices do grafo \mathbf{gR} aleatoriamente que dará origem a um grafo perturbado \mathbf{gP} com 1% a menos de vértices que \mathbf{gR} ;
- **Replicações** é a quantidade de 20 vezes que é repetida de forma aleatória a geração de grafos perturbados para cada nível de perturbação. Isto é, cada grafo \mathbf{gR} e \mathbf{gP}_i tem 20 vezes suas medidas de centralidade inferidas, sendo que cada replicação está submetida a doze níveis que são os percentuais de perturbação (Vieira & Hoffmann, 1989; Gomes, 2009);
- \mathbf{mC}_j e $\mathbf{mC}'_{i,j}$ são todos os conjuntos de medidas de centralidade ϕ_i dos respectivos grafos \mathbf{gR} e \mathbf{gP}_i ;
- $\varphi(\mathbf{mC}_j)$ e $\varphi(\mathbf{mC}'_{i,j})$ são aplicações agrupam as medidas de centralidade ϕ_i em distribuições de frequências com intervalos de classe e geram os vetores das distribuições de probabilidades das medidas de centralidade;
- \mathbf{FmC}_j são os conjuntos das distribuições de probabilidades do grafo Real e $\mathbf{FmC}'_{i,j}$ são os conjuntos das distribuições de probabilidades das medidas de centralidade dos grafos perturbados;
- $|\mathbf{FmC}_j|$ e $|\mathbf{FmC}'_{i,j}|$ são aplicações que fazem a contagem do número de classes (bins) de \mathbf{FmC}_j e $\mathbf{FmC}'_{i,j}$ respectivamente;

- $\psi(FmC_j)$ são $\psi(FmC'_{i,j})$ são uma aplicação que alteram o valor dos intervalos de classe para mais ou menos acarretando na mudança do número de classes. Por exemplo, se $|FmC_j| < |FmC'_{i,j}|$ número de classes de $|FmC_j|$ é aumentado de uma em uma classe até que $|FmC_j| = |FmC'_{i,j}|$ e vice-versa.
- $dHel(gR_j ; gP_{i,j})$ calcula a distância de Hellinger entre os dois conjuntos de distribuições de probabilidades das medidas de centralidade referentes ao grafo real com todos os seus grafos perturbados;
- $dHel^{\mu|m}$ é a distância de Hellinger média das 20 replicações de cada nível de perturbação para cada medida de centralidade e tipo de perturbação.

O fluxograma 3.9 a seguir demonstra os passos de forma intuitiva do modelo matemático para calcular as distâncias de Hellinger.



Fonte: Autor (2016).

Este fluxograma tem começo com um grafo gR de uma rede real que passa inicialmente por dois processos paralelos, um deles calcula suas medidas de centralidade mC_j , o outro processo passa por uma série de aplicações τ_i gerando novos grafos gP_i^m que posteriormente terão suas medidas de centralidade $mC'_{i,j}^m$ calculadas assim como em gR .

Diante das medidas de centralidade, para cada medida são geradas suas distribuições de frequências relativas de forma agrupada com intervalos de classe, isto é, suas distribuições de probabilidades, denotadas por FmC_j para o grafo real e $FmC'_{i,j}{}^m$ para os grafos perturbados.

Finalizados esses passos um modelo matemático implementado ao algoritmo vai calcular o número de classes de $|FmC_j|$ e $|FmC'_{i,j}{}^m|$ entre as distribuições e verificar se os dois conjuntos possuem mesmo tamanho para então calcular a $dHel[gR_j ; gP_{i,j}{}^m]$. Caso os números de classes sejam diferentes se escolhe entre os dois conjuntos o que tiver o menor número, a partir daí uma aplicação $\psi(Fm)$ aumenta o número de classes em uma unidade a cada iteração e constroi uma nova distribuição FmC_j ou $FmC'_{i,j}{}^m$, esse processo de escolher a distribuição com o menor número de classes garante mantenha-se ou melhore a qualidade dos resultados. Repetindo o processo, novamente é calculado o número de classes e se forem iguais o algoritmo avança para calcular a distância de Hellinger $dHel(gR_j ; gP_{i,j}{}^m)$. Por outro lado, se o número de classes ainda forem diferentes se repete o processo até que as duas distribuições de probabilidade tenham o mesmo tamanho.

Abaixo temos o algoritmo 3.1 utilizado nesse trabalho para realizar as simulações. Nossos dados de entrada são um grafo gR que representa a topologia de uma rede de Internet real e como resultado final temos a distância de Hellinger média de 20 replicações de cada nível de perturbação para cada medida de centralidade e tipo de perturbação que donotemos por $dHel_{i,j}^{\mu|m}$.

O algoritmo começa com o grafo gR , logo após nas linhas 2 e 3 temos duas estruturas de repetição formadas por $1 \leq m \leq \#perturbação$ e $1 \leq n \leq \#replicações$ onde $\#perturbações$ é 12 e o $\#replicações$ é 20, quantidades que o experimento irá percorrer.

Na linha 4 há uma aplicação $\tau(gR)$ que gera o conjunto de grafos perturbados $gP_i{}^{m,n}$, nesse passo são gerados a cada iteração, 3 tipos de grafos distintos pelo tipo de perturbação.

As linhas 5 e 6 apresentam as aplicações $\phi(C_j)$ e $\phi(C'_{i,j})$ que recebem os o grafos gR e $gP_i{}^{m,n}$ respectivamente. Esta aplicação calcula 3 medidas de centralidade para cada grafo, mC_j para o grafo gR e $mC'_{i,j}{}^{m,n}$ para os grafos $gP_i{}^{m,n}$. Diante dos vetores das medidas de centralidade um aplicação $\varphi(C)$ (linhas 7 e 8) gera as distribuições de frequências com dados agrupados em intervalos de classes para obtermos as frequências relativas FmC_j e $FmC'_{i,j}{}^{m,n}$ que representarão as distribuições de probabilidades utilizadas para calcular as distâncias de Hellinger.

Apartir deste passo implementamos da linha 9 em diante uma alteração ao algoritmo para que se calcula-se a distância de Hellinger caso as distribuições possuam diferentes quantidades de classes. Quando o número de classes são diferentes as distribuições FmC_j ou $FmC'_{i,j}$ passam por uma aplicação ψ que acrescenta uma classe à distribuição de menor número de classes (linhas 11 e 13). Daí, retorna-se às linhas 7 8 e repete-se novamente os passos a partir das linhas 9. Esse proceso é retido até que as distribuições tenham o mesmo tamanho.

Quando o ciclo de replicações n é encerrado calculamos o distância de Hellinger média ($dHel_{i,j}^{\mu|m}$), linha 18, para cada nível de perturbação m .

Algoritmo 3.1 Simulação de Monte Carlo.

Entrada: gR

Saída: $dHel_{i,j}^{\mu|m}$

```

1: início
2:   para  $m \leftarrow 1$  até #perturbações faça
3:     para  $n \leftarrow 1$  até #replicações faça
4:       Gere  $gP_i^{m,n}$ 
5:       Calcule  $mC_j$ 
6:       Calcule  $mC'_{i,j}{}^{m,n}$ 
7:       Gere  $FmC_j$ 
8:       Gere  $FmC'_{i,j}{}^{m,n}$ 
9:       enquanto  $|FmC_j| \neq |FmC'_{i,j}{}^{m,n}|$  faça
10:        se  $|FmC_j| < |FmC'_{i,j}{}^{m,n}|$  então
11:           $|FmC_j| \leftarrow |FmC_j| + 1$ 
12:        senão
13:           $|FmC'_{i,j}{}^{m,n}| \leftarrow |FmC'_{i,j}{}^{m,n}| + 1$ 
14:        fim se
15:      fim enquanto
16:      Calcule  $dHel(gR_j; gP_{i,j}^{m,n})$ 
17:    fim para
18:    Calcule  $dHel_{i,j}^{\mu|m}$ 
19:  fim para
20: fim.

```

O presente modelo apresentado aqui segue as propriedades do trabalho já validado e apresentado por Cabral et al. (2014) . No entanto incorporamos a este modelo uma alteração para inferir a distância de Hellinger para distribuições de probabilidades que possuem diferentes números de classes que validaremos teoricamente a seguir.

3.5 Validação da Metodologia

A distância de Hellinger tem como característica o cálculo entre as distribuições realizado aos pares. Logo, quando as distribuições possuem diferentes números de classes não é possível calcular a distância de Hellinger sem que se manipule os dados.

Ao perturbarmos um grafo este se tranforma em um grafo perturbado que terá outra topologia por sofrer alterações nas quantidades de vértices e/ou arestas, conseqüentemente

alterações nas medidas de centralidade locais. Estas alterações geram algumas vezes diferentes tamanhos de conjuntos dessas medidas e com isso possivelmente alterações no número de classes das distribuições das mesmas, dessa forma impossibilitando a verificação da distância. Para tanto, propomos uma alteração ao algoritmo de [Cabral et al. \(2014\)](#) através de um modelo matemático que viabiliza sem perda de propriedades e de forma específica um método para calcular distâncias entre as distribuições de probabilidades com diferentes tamanhos.

Para validarmos o modelo matemático utilizado entre as linhas 9 e 15 do algoritmo 3.1 descrevemos abaixo alguns conceitos e definições utilizados na literatura que provam sua consistência.

As distribuições de frequências simples ou relativas (probabilidades) são um sumário organizado e compacto dos dados tabelados que se apresentam divididos em faixas geralmente de igual tamanho que são chamadas de intervalos de classe. Geralmente se utilizam intervalos de classes quando o conjunto de dados é moderadamente grande com $n = 70, 80, 100$ ou mais dados, a quantidade de intervalos depende do tamanho do conjunto de dados, da homogeneidade e da amplitude dos dados. Geralmente com o número de classes entre 5 e 20 se pode obter uma boa representação dos dados ([Montgomery & Runger, 2012](#)).

Para construir uma distribuição de frequências é necessário calcular a amplitude da classe (h) que é obtido pela razão entre amplitude total da classe (AT), onde AT é obtido pela diferença entre o limite superior (L_{max}) e o limite inferior (L_{min}), e o número k de classes que pode ser obtido de até 3 formas, ou até mesmo por bom senso a critério do pesquisador. O número de classes calculado para obter h é um número hipotético, pois pode aumentar ou diminuir por depender de h ([Loesch, 2013](#)).

O número aproximado de classes pode ser obtido pela fórmula de Sturges

$$K = 1 + 3,322 \cdot \log_{10} n \quad (3.6)$$

[Montgomery & Runger \(2012\)](#) ainda definem um outro valor para k , sendo este um valor aproximado da \sqrt{n} .

Outra maneira de obter o número de classes é encontrar um valor para k tal que

$$\sqrt[3]{n} < k < \sqrt{n}$$

Como podemos ver o número de classes pode variar para mais ou para menos após a construção da distribuição e também variar de acordo com a fórmula utilizada. Sendo assim, nesse trabalho utilizamos a fórmula de Sturges para calcular k , pois $\sqrt[3]{n} < k < \sqrt{n}$ cresce muito para grandes conjuntos de dados, uma vez que trabalhos com redes complexas.

O presente capítulo demonstrou os materiais e métodos utilizados nesta pesquisa, destacando o algoritmo utilizado e a validação do modelo proposto. A seguir, no quarto capítulo, apresentaremos os resultados e análises dos experimentos.

4 RESULTADOS E ANÁLISES

O presente capítulo apresentará os resultados e análises das simulações aqui realizadas. Para cada uma das medidas de centralidade mostramos uma figura com gráfico de linhas, caixas e setores, além de tabelas que descrevem cada uma das perturbações, possibilitando assim compará-las de forma analítica e visual. Calculamos também, as médias da distância de Hellinger de cada nível de perturbação para cada uma das medidas de centralidade de 20 grafos perturbados, com a distância de Hellinger calculada para cada nível de perturbação entre o grafo perturbado e seu grafo original. Os resultados foram apresentados separadamente por tópicos de medidas de centralidade: grau do vértice, grau de proximidade e grau e intermediação.

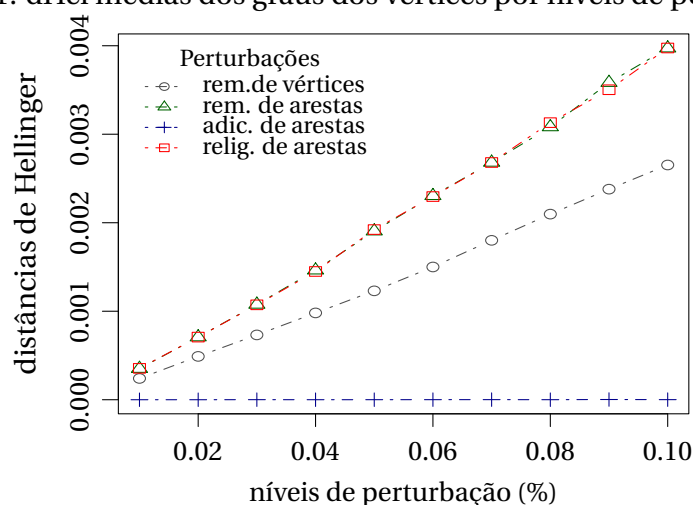
4.1 Grau do Vértice

A medida de centralidade grau do vértice frente as perturbações remoção e religação de arestas demonstrou oscilações significativas nos intervalos de 1% até 10%, com valores para a distância de Hellinger muito próximos e com mesmas amplitudes totais no valor de 0,0036 unid. de DHel. Seguidas da remoção de vértices que teve uma menor variação, e uma menor amplitude na valor de 0,0024 unid. de DHel.

Pelo gráfico de linhas (Figura 4.1) é possível ver a evolução dos tipos de perturbação em relação aos níveis e notar que a perturbação adição de arestas não causou mudanças relevantes na rede perturbada pois apresentou valores muito pequenos comparados as demais perturbações.

A religação e remoção de arestas são as mais sensíveis seguidas da remoção de vértices. Adição de arestas não apresentou diferenças relevantes entre os níveis de perturbação, indicando que o fluxo de informações enviadas diretamente de vértices não é alterado.

Figura 4.1: dHel médias dos graus dos vértices por níveis de perturbação.



Fonte: Autor (2016).

A tabela 4.1 mostra que a perturbação adição de arestas não influenciou no grau do vértice, pois sua média, desvio padrão e amplitude total apresentaram valores superiores a 4 casas decimais. No entanto, as distâncias médias oscilaram consideravelmente em relação a média mostrando-se uma série heterogêna, o que ocasionou no alto coeficiente de variação de 99,02%. A remoção de vértices teve o menor desvio padrão e amplitude total entre as perturbações. O coeficiente de variação com excessão da adição de arestas foi muito semelhante entre as outras perturbações.

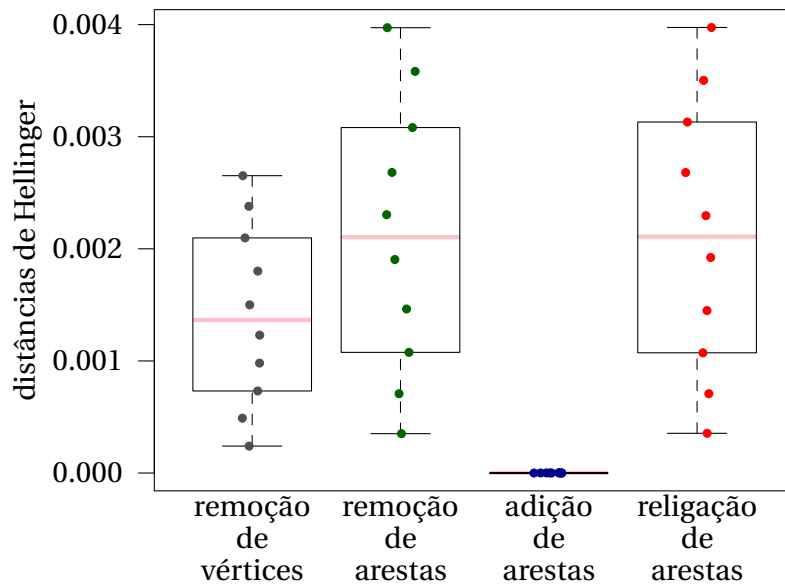
Tabela 4.1: Média, desvio padrão, mediana, amplitude total, e coeficiente de variação das medidas grau do vértice para cada tipo de perturbação.

	média	desvio padrão	mediana	ampl. total	coef. de var. (%)
remoção de vértices	0.0014	0.0008	0.0014	0.0024	57.93
remoção de arestas	0.0021	0.0012	0.0021	0.0036	58.05
adição de arestas	0	0	0	0	99.02
religação de arestas	0.002	0.0012	0.0021	0.0036	57.89

Fonte : Autor (2016).

A seguir analisando a figura 4.2 visualmente percebemos que a remoção e religação de arestas apresentam mesma amplitude total e mesma mediana de distâncias. Com a adição de arestas mostrando-se com média, mediana e amplitude iguais na visualiação. Os pontos presentes nos gráficos de caixas são as médias medidas de centralidade para cada nível de perturbação.

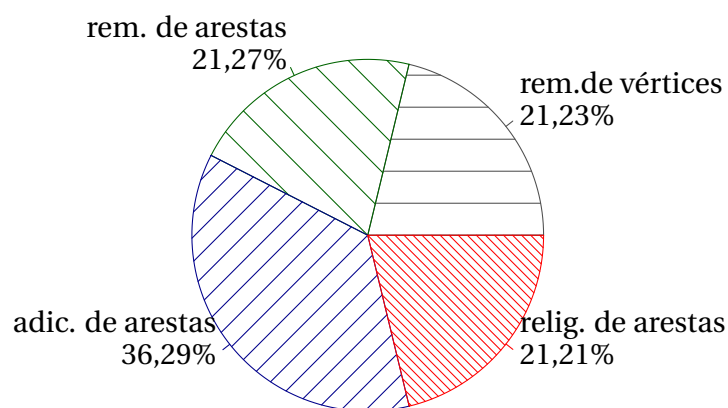
Figura 4.2: Comparação do grau do vértice entre os tipos de perturbações.



Fonte: Autor (2016).

Outro resultado relevante é a comparação da taxa de variação de cada perturbação, nesta análise podemos ver que a remoção de vértices, remoção e religação de arestas são pouco mais homogêneas e parecidas em relação a adição de arestas. Embora adição de arestas não tenha apresentado valores relevantes, as dHel variam com maior maiores intervalos em relação a sua média (Figura 4.3).

Figura 4.3: Percentual de variação das distâncias de Hellinger sob o grau médio para os tipos de perturbações.



Fonte: Autor (2016).

Comparando os resultados da rede real com os resultados obtidos da rede generalizada de Cabral et al. (2014) a perturbação de religação de arestas da rede real foi a que apresentou maior amplitude total, enquanto que na rede generalizada a religação de arestas foi a de menor amplitude total de todas as perturbações. Analisando o desvio padrão entre, a remoção de vértices apresentou o maior desvio padrão entre as perturbações das duas redes. Os

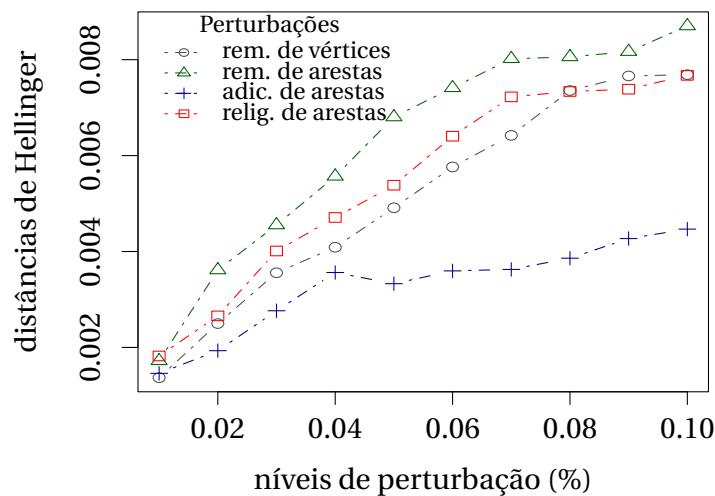
valores das distâncias de Hellinger da rede real não ultrapassaram a valor de 0.004 unid. de DHel, bem abaixo da rede generalizada que foi de aproximadamente 0.1 unid. de DHel..

4.2 Grau de Proximidade

Podemos observar na Figura 4.4 que a adição de arestas foi a perturbação que menos influenciou nos graus de proximidade, tendo um pequeno aumento até 4% e oscilando nessa faixa até 7%, logo após caindo em 8% e voltando a crescer.

Podemos dizer que a distâncias geodésicas não se alteram relevantemente com a adição de arestas. Por outro lado, as outras perturbações tiveram uma significativa influência sobre os caminhos mínimos. Para tanto, a remoção de arestas superou todas as outras perturbações, mostrando-se ser forte e importante na topologia da rede para o grau de proximidade dos vértices. Nesta rede especificamente o tempo gasto para uma informação saída de um vértice chegar ao restante dos vértices de toda a rede aumentou mais significativamente para a remoção de arestas.

Figura 4.4: dHel média dos graus de proximidade dos vértices por níveis de perturbação.



Fonte: Autor (2016).

Segundo a Tabela 4.2 o coeficiente de variação mostra que a amplitude da dHel para a remoção de arestas é a maior entre as perturbações, explicando que os caminhos mínimos são realmente afetados. Em seguida, a remoção de vértices e religação de arestas também mostraram-se com medidas que influenciam na topologia da rede. O desvio padrão da adição de arestas apresentou resultado bem inferior as demais perturbações, indicando a maior dispersão das dHel em relação a média o que mostra uma maior dispersão dos valores com isto uma distribuição mais heterogêna.

A taxa de variação mais alta foi apresentada pela remoção de arestas, levando a concluir que os caminhos mínimos foram mais alterados.

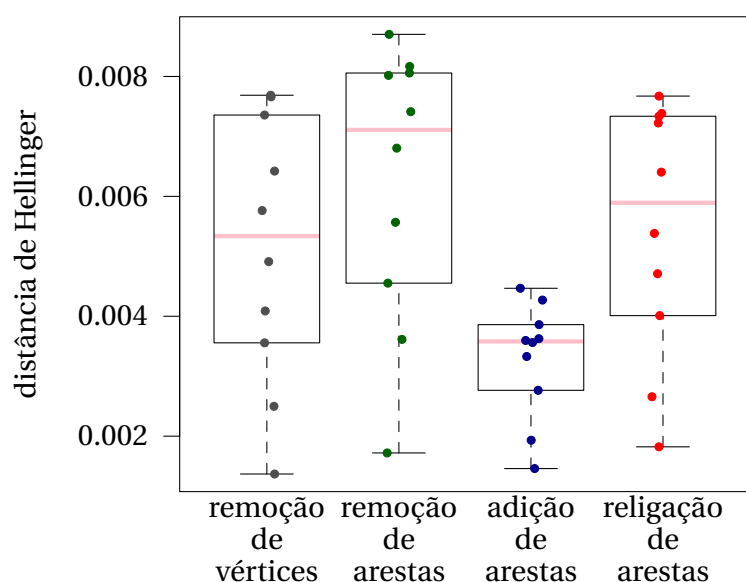
Tabela 4.2: Média, desvio padrão, mediana, amplitude total, e coeficiente de variação das medidas grau de proximidade para cada tipo de perturbação.

	média	desvio padrão	mediana	ampl. total	coef. de var. (%)
remoção de vértices	0.0051	0.0022	0.0053	0.0063	43.41
remoção de arestas	0.0063	0.0023	0.0071	0.0070	37.09
adição de arestas	0.0033	0.0010	0.0036	0.0030	29.43
religação de arestas	0.0055	0.0021	0.0058	0.0059	38.55

Fonte: Autor (2016).

A Figura 4.5 mostra claramente que a remoção de arestas teve aproximadamente 25% das dHel acima das dHel máximas das outras perturbações. Em todas as perturbações os evidência-se assimetria positiva das distribuições, mostrando que as dHel a partir de um certo percentual de perturbação tendem a se assemelhar. Podemos ver isso pela mediana acima da média com excessão da adição de arestas que teve a menor amplitude e o maior valor dHel inferior a 75% dos maiores valores dHel da remoção de arestas.

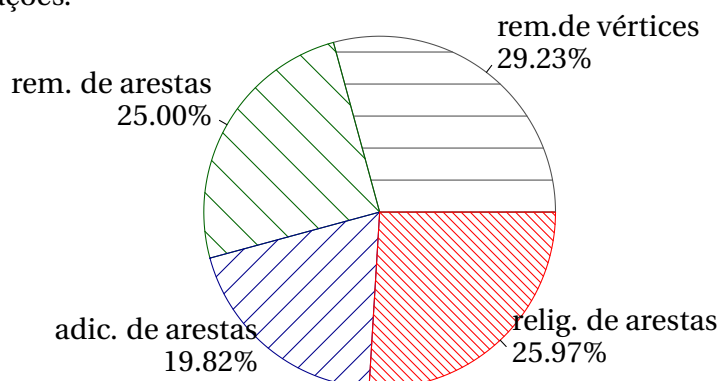
Figura 4.5: Comparação do grau de proximidade entre os tipos de perturbações.



Fonte: Autor (2016).

Analisamos o percentual de variação de cada perturbação e observamos que a adição de arestas é uma perturbação que não afeta efetivamente a rede, pois apresenta o menor percentual de dispersão comparada aos demais, menor taxa de variação, menor média, desvio padrão e amplitude. A remoção de vértices embora não tenha apresentado maior amplitude, apresentou dHel mais heterogênea (Figura 4.6).

Figura 4.6: Percentual de variação das distâncias de Hellinger sob o grau de proximidade para tipos perturbações.



Fonte: Autor (2016).

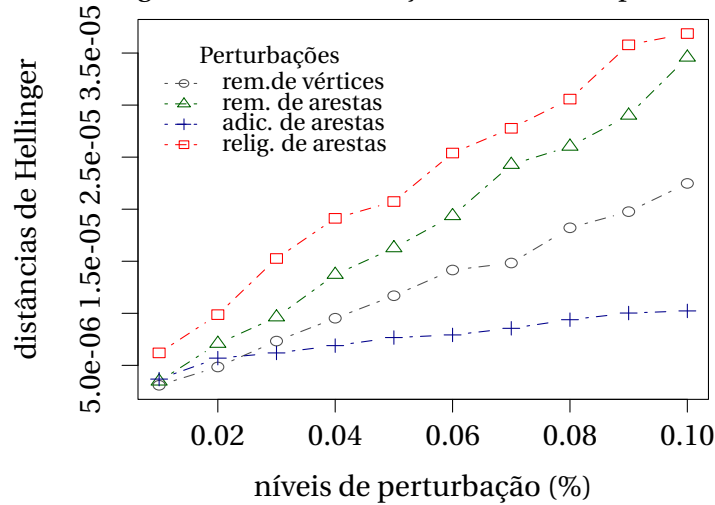
Cabral et al. (2014) apresenta resultados acerca da dHel para redes livres de escala onde adição de arestas é a perturbação mais efetiva, provocando alterações na rede de mudanças de caminhos mínimos com maior intensidade que as outras perturbações. Resultado diferente da rede real apresentada neste trabalho, onde a adição de arestas foi das perturbações de menor influência sobre dHel, significando baixíssima mudança de topologia para os caminhos mínimos.

4.3 Grau de Intermediação

Veremos agora a medida de centralidade grau de intermediação, esta medida diferentemente das demais foi a que menos sofreu com as perturbações. Para nosso padrão de arredondamento estipulado de 4 casas decimais as estatísticas de média, desvio padrão, mediana e amplitude total apresentaram valores iguais zero em todas as perturbações.

Plotamos o gráfico de linhas, Figura 4.7, da evolução das distâncias médias dos níveis de perturbação para análise visual e observamos que a adição de arestas provocou alteração mínima no grau de intermediação da topologia da rede perturbada. A religação de arestas alterou a topologia relativamente em relação a adição de arestas, pois apresentou uma amplitude total maior. Com valores acima de 4 casas decimais obtidos das distâncias de Hellinger entendemos que a rede não sofre alterações relevantes em relação aos seus vértices que fazem parte do caminho mais curto entre outros vértices.

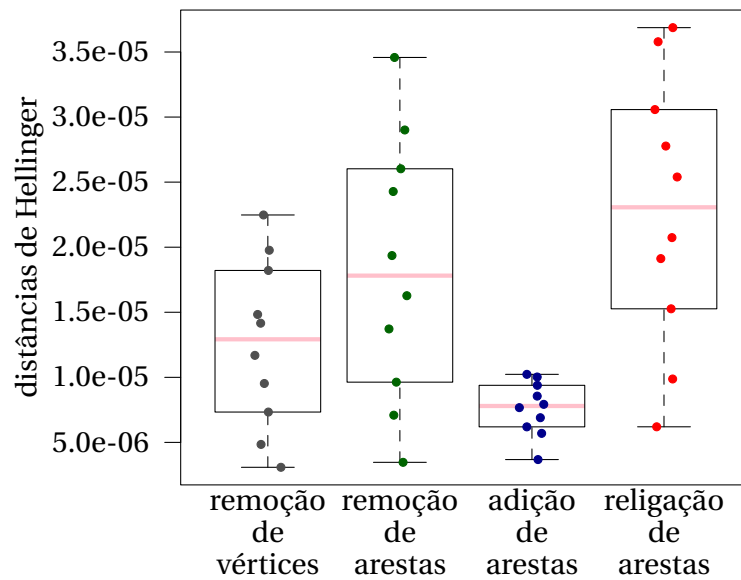
Figura 4.7: dHel média dos graus de intermediação dos vértices por níveis de perturbação.



Fonte: Autor (2016).

Como consequência dos valores das estatísticas serem todos zeros pelo arredondamento, só obtivemos os valores para os percentuais de variação de casa perturbação. Neste caso a menor dispersão foi da adição de arestas com 27.11% em relação a sua média, e a maior dispersão apresentada foi da remoção de arestas com 55.18%.

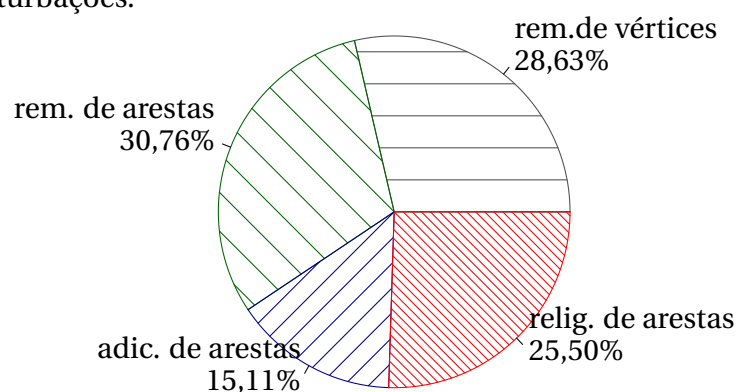
Figura 4.8: Comparação do grau de intermediação entre os tipos de perturbações.



Fonte: Autor (2016).

Ao analisarmos percentualmente qual perturbação provocou maior maior oscilação nos valores médios das distâncias de Hellinger aos níveis de perturbação, temos como distâncias mais homogêneas a adição de arestas, seguido da religação de arestas e remoção de vértices, por último a remoção de arestas com o dobro de percentual de dispersão que a adição (Figura 4.9).

Figura 4.9: Percentual de variação das distâncias de Hellinger sob o grau de intermediação para os tipos perturbações.



Fonte: Autor (2016).

O grau de intermediação foi muito pouco afetado entre as perturbações, com valores acima de 5 casas decimais em todos os níveis. A d_{Hel} apresenta-se muito pequena em todos os casos, um caso relativamente igual ao das redes apresentados por [Cabral et al. \(2014\)](#), o que leva a concluir que o grau de intermediação não foi afetado até 10% de perturbações.

Este capítulo apresentou as diferenças entre os tipos de medidas de centralidade baseadas em vértices e perturbações de topologia em grafos, que possibilitou observarmos quais destas medidas e perturbações foram mais relevantes para conhecermos o comportamento da rede original em relação a sua rede perturbada. A seguir apresentaremos as considerações finais.

5 CONSIDERAÇÕES FINAIS

ESTE trabalho teve como objetivo maior analisar uma rede real frente a medidas de centralidade baseadas em vértices após sofrer alterações de topologia por perturbações. Para realizar este estudo foi utilizada a distância de Hellinger que mostrou boa sensibilidade aos dados com valores decimais, possibilitando desta forma uma análise mais minuciosa dos resultados.

Todo o experimento foi realizado sob o método Monte Carlo permitindo o desenvolvimento do processo de geração de perturbações e grafos perturbados de forma aleatória, que representaram as novas redes reais e forneceram dados e resultados numéricos para o trabalho. A implementação do algoritmo para o código usado por [Cabral et al. \(2014\)](#) foi fundamental pois evitou as diferenças de classes entre as distribuições das redes e viabilizou utilizar a distância de Hellinger em todos os fatores e níveis.

Como conclusões acerca dos tipos e níveis de perturbações a remoção de arestas foi de todas a que apresentou maior amplitude de modo geral, ou seja, perante as medidas de centralidade houve maiores alterações na comparação com a rede original. A adição de arestas pouco influenciou nas medidas de centralidade, estando atrás das demais perturbações em todas as medidas.

Referente as medidas de centralidade, o grau de proximidade foi a medida que mais se alterou com as perturbações, com amplitude total de 0.0070 unid. de dHel, chegando a ter 94,4% a mais de amplitude total das demais medidas, neste caso o grau do vértice com 0.0036 unid. de dHel. Levando a concluir que o o grau de proximidade é a medida de centralidade baseada em vértices mais frágil perante as perturbações. Para a rede real da Internet aqui analisada concluímos que os caminhos mínimos foram fortemente afetados, o que influenciará diretamente nas transmissões de dados. Entretanto, os nós hubs não foram afetados, pois a medida grau do vértice demonstrou ser mais robusta.

Os resultados obtidos neste trabalho apresentaram algumas semelhanças aos obtidos por [Cabral et al. \(2014\)](#), a análise visual dos dados nos gráficos de linhas apresentou boa semelhança aos obtidos pela pesquisadora. Levando-nos a ressaltar que a metodologia proposta cumpriu os objetivos da pesquisa.

REFERÊNCIAS

- Albert, R. & Barabasi, A. (1999), 'Emergence of scaling in random networks', *Science* **286**, 509–512.
- Albert, R., Barabási, A. & Jeong, H. (2000a), 'Scale-free characteristics of random networks: The topology of the world wide web', *Physica A* **281**, 69–77.
- Albert, R., Jeong, H. & Barabási, A. (1999), 'The diameter of the world wide web', *Nature* **401**, 130–131.
- Albert, R., Jeong, H. & Barabási, A. (2000b), Error and attack tolerance of complex networks, *in* 'Nature', pp. 378—382.
- Almiron, M. G., Almeida, E. S. & Miranda, M. N. (2009), 'The reliability of statistical functions in four software packages freely used in numerical computation', *Brazilian Journal of Probability and Statistics* **23**(2), 107–119.
- Balthrop, J., Forrest, S., Newman, M. E. J. & Williamson, M. M. (2004), 'Technological networks and the spread of computer viruses', *Science* **304**, 527 – 529.
- Barabási, A. & Bonadeau, E. (2003), 'Scale-free network', *Scientific American* **12**(13), 64–72.
- Barabási, A. & Frangos, J. (2002), *Linked: The New Science of Networks*, 1 ed., Perseus Books Group. ISBN 978-0738206677.
- Barnes, J. A. (1954), 'Class and committees in a norwegian island parish', *Human Relations* **7**(1), 39–58.
- Barrat, A., Barthlemy, M. & Vespignani, A. (2008), *Dynamical Processes on Complex Networks*, 1 ed., Cambridge University Press, New York. ISBN 978-0521879507.
- Bavelas, A. (1948), 'A mathematical model of group structure', *Human Organizations* **7**, 16–30.
- Bhattacharyya, A. (1943), 'On a measure of divergence between two statistical populations defined by their probability distribution', . *Bulletin of the Calcutta Mathematical Society* **35**, 99–110.
- Boccaletti, S. and Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. (2006), 'Complex networks: Structure and dynamics', *Physics Reports* **424**(1-5), 175–308.
- Boguña, M. & Pastor-Satorras, R. (2002), 'Epidemic spreading in correlated complex networks', *Physical Review E* **66**, 047104.

- Bondy, J. A. & Murty, U. S. R. (1976), *Graph Theory with Applications*, 2 ed., North Holland, New York. ISBN 0444194517.
- Bondy, J. A. & Murty, U. S. R. (2008), *Graph Theory*, Springer, New York. ISBN 9781846289699.
- Boutier, J., Di Francesco, P. & Guiter, E. (2003), 'Geodesic distance in planar graphs', *Nuclear Physics* **663**(3), 535–567.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. (2000), 'Graph structure in the web', *Computer Network* **33**(1-6), 309–320.
- Cabral, R. S., Frery, A. C. & Ramires, J. A. (2014), 'Variability analysis of complex networks measures based on stochastic distances', *Physica A: Statistical Mechanics and its Applications* **415**, 73–86.
- Callaway, D. S., Newman, M. E. J. E. J., Strogatz, S. H. & Watts, D. J. (2000), 'Network robustness and fragility: Percolation on random graphs', *American Physical Society* **85**(25), 5468–5471.
- Cha, S.-H. & Srichari, S. N. (2002), 'On measuring the distance between histograms', *Pattern Recognition* **35**(6), 1355–1370.
- Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. (2000), 'Resilience of the internet to random breakdowns', *Physical Review Letters* **85**, 4626–4628.
- Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. (2001), 'Breakdown of the internet under intentional attack', *Phys. Rev. Lett.* **86**, 3682–3685.
- Costa, L. d., Rodrigues, F. A., Traverso, G. T. & Villas Boas, P. R. (2007), 'Characterization of complex networks: A survey of measurements', *Advances in Physics* **56**(1), 167–242.
- Csardi, G. (2012), *Network analysis and visualization*, 6 ed. URL <http://igraph.sourceforge.net>.
- Deaux, K. and Snyder, M. (2012), *The Oxford Handbook of Personality and Social Psychology*, 1 ed., Oxford University Press, USA. ISBN978-0195398991.
- Diestel, R. (2010), *Graph theory*, Vol. 173, fourth ed., Graduate Texts in Mathematics: Springer-Verlage.
- Erdős, P. & Rényi, A. (1959), *On random graphs. I.*, Vol. 6, Publicationes Mathematica.
- Euler, L. (1736), 'Solutio problematis ad geometriam situs pertinentis', *Comment. Academiae Sci. I. Petropolitanae* **8**, 128—140.

- Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999), 'On power-law relationships of the internet topology', *SIGCOMM Computer Communication Review* **29**(4), 251–262.
- Freeman, L. C. (1977), 'A set of measures of centrality based on betweenness', *Sociometry* **40**(1), 35–41. URL <http://www.jstor.org/stable/3033543>.
- Freeman, L. C. (1979), 'Centrality in social networks conceptual clarification', *Social Networks* **1**(3), 215–239.
- Gibbs, A. L. & Su, F. E. (2002), 'On choosing and bounding probability metrics', *International Statistical Review* **70**(3), 419–435. URL <http://dx.doi.org/10.1111/j.1751-5823.2002.tb00178.x>.
- Gilbert, E. N. (1959), 'Random graphs', *The Annals of Mathematical Statistics* **30**(4), 1141–1144.
- Gilbert, E. N. (1960), 'Capacity of a burst-noise channel', *Bell System Technical Journal* **39**, 1253–1265.
- Girvan, M. & Newman, M. E. J. (2002), 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences* **99**, 7821–7826.
- Gomes, F. (2009), *Curso de estatística experimental*, Fundação de Estudos Agrários Luiz de Queiroz, São Paulo. ISBN 978-85-7133-055-9.
- Gould, R. V. (1993), 'Collective action and network structure', *American Sociological Review* **58**(2), 182–196.
- Granovetter, M. S. (1973), 'The strength of weak ties', *American Journal of Sociology* **78**(6), 1360–1380. URL <http://www.jstor.org/stable/2776392>.
- Guimera, R. & Amaral, L. A. N. (2004), 'Modeling the world-wide airport network', *The European Physical Journal B* **38**(2), 381–385.
- Guimera, R. & Amaral, L. A. N. (2005), 'Functional cartography of complex metabolic networks', *Nature* **433**, 895–900.
- Hanneman, R. A. (2001), 'Introduction to social network methods'. URL <http://faculty.ucr.edu/>, Acessado em novembro de 2015.
- Hellinger, E. (1909), 'Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.', *Journal Reine Angewandte Mathematik* **136**, 210–271.
- Hoffman, P. (1998), *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth*, Hyperion Books, United States. ISBN 978-078-686-362-4.

- Knuth, D. E. (1984), *The TeXbook. Tomo A de Computers and Typesetting*, spi ed., Addison-Wesley Publishing Company. ISBN 0-201-13448-9.
- Kullback, S. & Leibler, R. (1951), 'On information and sufficiency', *Ann. Math. Stat.*
- Lamport, L. (1994), *LaTeX A Document Preparation System*, 2 ed., Addison-Wesley Reading, Massachusetts. ISBN 0-201-52983-1.
- Leskovec, J., Huttenlocher, D. & Kleinberg, J. (2010), Signed networks in social media, in 'Proceedings of the 28th international conference on Human factors in computing systems', CHI '10, ACM, New York, pp. 1361–1370.
- Li, W. & Cai, X. (2004), 'Statistical analysis of airport network of china', *Physical Review E* **69**(4), 1–11.
- Liljeros, F., Edling, C., Amaral, L. A. N., Stanley, H. E. & Aberg, Y. (2001), 'The web of human sexual contacts', *Nature* **411**, 907–908.
- Loesch, C. (2013), *Probabilidade e Estatística*, 1 ed., LTC Editora, Rio de Janeiro. ISBN 9788521621003.
- Luckie, M. and Huffaker, B., Dhamdhere, A., Giotsas, V. & Claffy, K. (2013), As relationships, customer cones, and validation, in 'Proceedings of the 2013 Conference on Internet Measurement Conference', pp. 243–256.
- Maslov, S. & Sneppen, K. (2002), 'Specificity and stability in topology of protein networks', *Science* **296**, 910–913.
- Menéndez, M., Morales, D., Pardo, L. & Salicrú, M. (1997), '(h, ϕ)-entropy differential metric', *Appl. Math.* **42**(2), 81–98.
- Metropolis, N. & Ulam, U. (1949), 'The monte carlo method', *Journal of the American Statistical Association* **44**(247), 335–341. URL <http://www.jstor.org/stable/2280232>.
- Milgram, S. (1967), *The small-world problem*, Vol. 1, Psychology Today, pp. 60–67.
- Montgomery, D. C. & Runger, G. C. (2012), *Estatística Aplicada e Probabilidade para Engenheiros*, 5 ed., LTC Editora, Rio de Janeiro. ISBN 9788521619024.
- Moreno, J. L. (1934), *Who shall survive?*, Beacon House, New York. ISBN 0976497689.
- Moxley, R. L. & Moxley, N. F. (1974), 'Determining point-centrality in uncontrived social networks', *Sociometry* **37**, 120–133. URL <http://www.jstor.org/stable/2786472>.
- Newman, M. E. J. (2002), 'Assortative mixing in networks', *Physical Review Letters* **89**(20), 208701.

- Newman, M. E. J. (2003), 'The structure and function of complex networks', *SIAM Review* **45**, 167–256.
- Newman, M. E. J. (2005), 'Power laws, pareto distributions and zipf's law.', *Contemporary Physics* **46**(5), 323–351.
- Newman, M. E. J. & Girvan, M. (2003), 'Finding and evaluating community structure in networks', *American Physical Society* **69**(2), 026113.
- Opsahl, T. (2010), 'Closeness centrality in networks with disconnected components', <http://toreopsahl.com>. Acessado em outubro de 2015.
- Price, D. J. d. S. (1965), 'Networks of scientific papers', *Science* **149**(3683), 510–51.
- Rapoport, A. (1957), 'Contribution to the theory of random and biased nets', *The Bulletin of Mathematical Biophysics* **19**(4), 257–277.
- Redner, S. (1998), 'How popular is your paper? an empirical study of the citation distribution', *European Physical Journal B* **4**(2), 131–134.
- Renyi, A. (1960), On measures of entropy and information, in 'Fourth Berkeley Symposium on Mathematical Statistics and Probability', Vol. 2, pp. 547–561.
- Robert, C. P. & Casella, G. (2009), *Introducing Monte Carlo Methods with R (Use R)*, 1 ed., Springer-Verlag, Heidelberg. ISBN 9781441915757.
- Sabisussi, G. (1966), 'The centrality index of a graph', *Psychometrika* **31**(4), 581–603.
- Sarkar, D. (2008), *Lattice: Multivariate Data Visualization with R*, Springer, New York. ISBN 978038775968-5.
- Shannon, C. E. (1948), 'A mathematical of theory communication', *Bell System Tec* **27**, 379–423.
- Solomonoff, R. (1952), 'An exact method for the computation of the connectivity of random nets', *Bulletin of Mathematical Biophysics* **14**(2), 153–157.
- Stallings, W. (2010), *Computer Organization and Architecture: Designing for Performance*, 8 ed., Pearson Prentice Hall. ISBN 9780136073734.
- Suzuki, J. (2001), *A History of Mathematics*, 1 ed., Pearson. ISBN 0130190748.
- Taneja, I. (1995), 'New developments in generalized information measures', *Advances in Imaging and Electron Physics* **91**, 37–135.

Tanenbaum, A. S. & Wetherall, D. J. (2011), *Computer Networks*, 5 ed., Prentice Hall. ISBN 9780132126953.

Team, R. D. C. (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, ISBN 3-900051-07-0.

Uetz, P., Giot, L., Cagney, G. & Mansfield, T. (2000), 'A comprehensive analysis of protein-protein interactions in *saccharomyces cerevisiae*', *Nature* **403**, 623–627.

Vieira, S. & Hoffmann, R. (1989), *Estatística Experimental*, Atlas.

Villas Boas, P. R., Rodrigues, F. A., Travieso, G. & Costa, L. d. (2010), 'Sensitivity of complex networks measurements.', *Journal of Statistical Mechanics: Theory and Experimental* (3), P03009.

Watts, D. J. & Strogatz, S. H. (1998), 'Collective dynamics of small-world networks', *Nature* **393**, 440–442.

Este trabalho foi redigido em \LaTeX utilizando uma modificação do estilo IC-UFAL. As referências bibliográficas foram preparadas no JabRef e administradas pelo \BIBTeX com o estilo LaCCAN. O texto utiliza fonte Fourier-GUTenberg e os elementos matemáticos a família tipográfica Euler Virtual Math, ambas em corpo de 12 pontos.