



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Escola de Engenharia  
Programa de Pós-Graduação em Engenharia Elétrica

# **ESTUDO DA VARIABILIDADE DE MEDIDAS EM REDES COMPLEXAS**

Raquel da Silva Cabral

Belo Horizonte  
Novembro de 2013

**RAQUEL DA SILVA CABRAL**

**ESTUDO DA VARIABILIDADE DE MEDIDAS EM  
REDES COMPLEXAS**

Tese apresentada ao Programa de Pós-  
-Graduação em Engenharia Elétrica da  
Universidade Federal de Minas Gerais  
como requisito parcial para a obtenção  
do grau de Doutor em Engenharia Elé-  
trica.

**ORIENTADOR:**  
**JAIME A. RAMIREZ**  
**ALEJANDRO O. FRERY**

Belo Horizonte  
Novembro de 2013

*Dedico a minha vizinha Anael (in memoriam) e minha linda filha Maria Júlia.*

# Agradecimentos

Este espaço é dedicado àqueles que deram a sua contribuição para que este trabalho fosse realizado. A todos eles deixo aqui o meu agradecimento sincero.

O espaço dessas poucas páginas não são suficientes para agradecer a todas as pessoas que estiveram comigo, serei eternamente grata! Obrigada a todas as pessoas que contribuíram para meu sucesso, meu crescimento pessoal, intelectual e moral neste processo de doutoramento. Com certeza, este trabalho é resultado da confiança e da força de cada um de vocês.

Agradeço aos meus orientadores, Jaime e Alejandro, que me acolheram e me orientaram com excelência. Vocês são os melhores!

Agradeço a minha família, minha Júlia por aguentar uma mãe a beira de um ataque de nervos. A minha mãe Nazaré por corrigir meu texto, meu pai Cabral e meu irmão Robério que sempre estiveram na torcida.

Um obrigada especial aos meus queridos amigos Josirene, Marcela, Márcio e Luciana que arrumaram um tempinho para estarem comigo no momento da defesa. Minha amiga Eunice, que teve uma enorme paciência de revisar meu texto. Aos meus amigos de Manaus e de Belo Horizonte aos quais tive a imensa felicidade de conhecer neste processo. E claro, agradeço aos meus amigos de Maceió.

Obrigada aos colegas do LaCCAN e da UFMG, em especial a todos os professores que contribuíram para minha formação.

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil. Agradeço esta instituição pelo apoio financeiro e, principalmente pelo interesse nesse estudo.

*“De tudo, ficam três coisas:  
A certeza de que estava sempre começando;  
A certeza de que era preciso continuar;  
A certeza de que enfrentaríamos obstáculos pela frente.”*  
(Fernando Sabino)

## Resumo

As redes complexas podem modelar a estrutura e dinâmica de vários tipos de sistemas, sendo comum a caracterização dessas redes através de medidas. Neste trabalho, avalia-se a variabilidade de medidas de redes complexas quando as redes são submetidas a perturbações e, para isso, as perturbações são controladas e seus efeitos quantificados. Foram avaliadas as *redes aleatórias*, *de mundo pequeno* e *livres de escala* em relação às medidas grau do vértice, coeficiente de agrupamento local, *betweenness centrality*, *comprimento dos caminhos mínimos* e *comprimento médio dos caminhos mínimos*. Foram consideradas redes com 1000, 5000 e 10000 vértices. Em tal análise, utilizou-se três quantificadores: a divergência de Kullback-Leibler, a distância de Jensen-Shannon e a distância de Hellinger. A sensibilidade das medidas foi analisada com respeito às seguintes perturbações: adição de arestas, remoção de arestas, religação de arestas e remoção de nós, todas elas aplicadas em diferentes intensidades. Adicionalmente, foram feitos testes de hipótese para verificação do comportamento da distribuição dos graus e identificar que perturbações e seus níveis levam à quebra dessa propriedade. Dentre as perturbações aplicadas, a religação de arestas causou menor variação nos quantificadores estocásticos. O grau do vértice não é alterado com esta perturbação, pois a religação preserva seus valores; enquanto que, o comprimento dos caminhos mínimos apresentou uma pequena variação nas redes aleatórias e de mundo pequeno e uma variação maior nas redes livres de escala. A adição e remoção de arestas atinge diretamente o comportamento das medidas locais, já que estas alteram sempre seus valores causando variações nos quantificadores estocásticos. Especificamente, as redes livres de escala são mais sensíveis a adição de arestas, já que os quantificadores atingiram os valores mais altos, 0.807 para no nível de perturbação 10% em redes com 10000 vértices. Em relação a sensibilidade dos quantificadores utilizados, a distância de Hellinger apresentou os maiores valores na maioria dos casos, seguida pela divergência de Kullback-Leibler e a distância de Jensen-Shannon. O coeficiente de agru-

pamento local não apresentou variação para as perturbações de arestas, assim pode-se dizer que essa medida é robusta ou insensível a estas perturbações.

Além disso, foi realizada uma caracterização da estrutura de comunicação de redes de sensores sem fio usando redes complexas. Essas redes de sensores são uma tecnologia importante para coleta de dados em ambientes distribuídos e hostis. Entre os desafios, está a forma como os dados que elas coletam são transportados na rede, pois sua estrutura é bastante dinâmica. A topologia operacional desses dispositivos pode muitas vezes ser descrita por redes complexas. Foi avaliada a variação do comprimento dos caminhos mínimos, uma medida comumente empregada na caracterização de redes de sensores sem fio. Foram consideradas quatro estratégias de propagação de dados que seguem os modelos: geométrico, aleatório, de mundo pequeno, livre de escala. A sensibilidade desta medida foi analisada em relação a perturbações: inserção e remoção de vértices na estratégia geométrica; e inserção, remoção e religação de arestas nos demais modelos. A avaliação foi realizada em redes com 1000 vértices por meio da divergência de Kullback-Leibler normalizada e da distância Hellinger. Os resultados revelaram que as medidas avaliadas são influenciadas pelas perturbações. Tanto a adição como a remoção de vértices alteram o comprimento dos caminhos mínimos no modelo de propagação geométrico. A variabilidade dos quantificadores tem dependência com o nível de perturbação aplicado; eles aumentam quando o nível de perturbação aumenta, sendo a distância de Hellinger a mais sensível as perturbações. A remoção de nós teve um impacto maior sobre o comprimento dos caminhos mínimos comparada à adição de nós. Nas estratégias de propagação baseadas em redes complexas o comprimento do caminho mínimo é sensível à adição e remoção de arestas. A distância de Hellinger é consistente e significativamente maior do que a divergência Kullback-Leibler por um fator de, aproximadamente, dois, em todas as situações em que a adição da ligação e remoção são aplicadas.

**Palavras-chave:** Redes complexas, medidas, distâncias estocásticas.

# Abstract

Complex networks are able to model the structure and dynamics of different types of systems. It has been shown that they can be characterized by certain measures. In this work, we evaluate the variability of complex networks measures face to perturbations and, for this purpose, we impose controlled perturbations and quantify their effect. We analyze the *random*, *small-world* and *scale-free* models, along with the *vertex degree*, *local cluster coefficient*, *betweenness centrality*, *shortest path length* and *average shortest path length* measures. We consider networks with 1000, 5000 e 10000 vertices. We employ three quantifiers for the change of vector-valued measures: the Kullback-Leibler divergence, and the Jensen-Shannon and Hellinger distances. The sensitivity of these measures was analyzed with respect to the following perturbations: edge addition, removal, rewiring and node removal, all of them applied at different intensities. Additionally, hypotheses tests were performed to verify the behavior of the degree distribution to identify the intensity of the perturbations that leads to the breakdown of these properties. Among the perturbations applied, the rewiring of edges caused less variation in stochastic quantifiers. The degree of vertices is not changed with this perturbation because the rewiring preserves its values. The shortest paths length showed a small variation in random and small-world networks and a greater variation in scale-free networks. The edge addition and removal affects the behavior of local measures, since they always change their values causing variations in stochastic quantifiers. Specifically, the scale-free networks are more sensitive to edge addition, since the quantifiers have the highest values, 0.807 to the level of perturbation 10% in networks with 10.000. Regarding the sensitivity of quantifiers used, Hellinger distance showed the highest in most cases, followed by Kullback-Leibler divergence and the Jensen-Shannon distance. The local clustering coefficient did not change for the disruption of edges, so it can be said that this measure is robust or insensitive to these perturbations.

Furthermore, we characterized the communication structure of wireless



sensor networks as complex network. These sensor networks are an important technology for making distributed autonomous measures in hostile or inaccessible environments. Among the challenges they pose, the way data travel among them is a relevant issue since their structure is quite dynamic. The operational topology of such devices can be often described by complex networks. We assess the variation of the shortest paths, a measures commonly employed in the complex networks to characterize wireless sensor networks. Four data propagation strategies were considered: geometric, random, small-world, and scale-free models. The sensitivity of these measures was analyzed with respect to perturbations: insertion and removal of nodes in the geometric strategy; and insertion, removal and rewiring of links in the other models. The assessment was performed in networks with 1000 vertices using the normalized Kullback-Leibler divergence and the Hellinger distance, both stemming from the Information Theory framework. The results reveal that the evaluated measures are influenced by these perturbations. Both the node addition and removal change the shortest path length in geometric strategy. The variability of the quantifiers is dependent on the level of disturbance applied, they increase when the level of perturbations increases, with the distance from the Hellinger more sensitive to perturbation. Node removal had a greater impact on the shortest paths length compared to node addition. In strategies based on complex networks the shortest path length is sensitive to the edges addition and removal. The Hellinger distance is significantly higher than the Kullback-Leibler divergence by a factor of about two, in all cases where edge addition and removal are applied.

**Keywords:** Complex networks, measures, stochastic distances.

# Lista de Figuras

1.1	Exemplos de redes complexas. . . . .	1
2.1	Caracterização e/ou representação de uma rede complexa . . . . .	10
2.2	Grafo com $N = 9$ vértices e $M = 10$ arestas. . . . .	10
2.3	Três grafos distintos com $N = 5$ vértices. . . . .	11
2.4	Exemplos de grafos com coeficiente de agrupamento $C = 0$ . . . . .	12
2.5	Ilustração da cidade de Königsberg em 1652 e o grafo correspondente ao problema . . . . .	14
2.6	Exemplo de geração de grafos aleatórios usando-se $\mathcal{G}_{er}(N, p_c)$ . . . . .	18
2.7	Exemplo da geração de uma rede de mundo pequeno. . . . .	20
2.8	Exemplos de grafos $G_{ws}$ e $G_{er}$ . . . . .	21
2.9	Histograma de proporções para distribuição dos graus de uma rede $G_{ws}$	22
2.10	Exemplo de uma rede $G_{ba}$ . . . . .	24
2.11	Exemplo de grafos com $N = 10$ vértices . . . . .	27
3.1	Representação da análise da variabilidade de medidas para redes com- plexas. . . . .	30
3.2	Exemplos de perturbações aplicadas em um grafo. . . . .	34
3.3	Etapas na análise da variabilidade de medidas. . . . .	36
4.1	Variação da divergência de Kullback-Leibler ( $D_{KL}$ ) para o grau do vér- tice ( $\kappa_i$ ) nos modelos teóricos. . . . .	41
4.2	Variação da divergência de Kullback-Leibler ( $D_{KL}$ ) para o compri- mento dos caminhos mínimos ( $\ell_{(i,j)}$ ) nos modelos teóricos. . . . .	42
4.3	Variação dos quantificadores para o grau do vértice $\kappa_i$ nas redes alea- tórias. . . . .	46
4.4	Variação dos quantificadores para o grau do vértice $\kappa_i$ nas redes de mundo pequeno. . . . .	47

4.5	Variação dos quantificadores para o grau do vértice $\kappa_i$ nas redes livres de escala. . . . .	48
4.6	Variação dos quantificadores para o coeficiente de agrupamento local $C_i$ em redes aleatórias. . . . .	50
4.7	Variação dos quantificadores para o coeficiente de agrupamento local $C_i$ em redes de mundo pequeno. . . . .	51
4.8	Variação dos quantificadores para o coeficiente de agrupamento local $C_i$ em redes livres de escala. . . . .	52
4.9	Variação dos quantificadores para o <i>betweenness centrality</i> $B_i$ em redes aleatórias. . . . .	53
4.10	Variação dos quantificadores para o <i>betweenness centrality</i> $B_i$ em redes de mundo pequeno. . . . .	54
4.11	Variação dos quantificadores para o <i>betweenness centrality</i> $B_i$ em redes livres de escala. . . . .	55
4.12	Variação dos quantificadores para o comprimento dos caminhos mínimos $\ell_{(i,j)}$ em redes aleatórias. . . . .	56
4.13	Variação dos quantificadores para o comprimento dos caminhos mínimos $\ell_{(i,j)}$ em redes de mundo pequeno. . . . .	57
4.14	Variação dos quantificadores para o comprimento dos caminhos mínimos $\ell_{(i,j)}$ em redes livres de escala. . . . .	58
4.15	Exemplo de uma rede livre de escala. . . . .	59
4.16	Variação da diferença absoluta para o comprimento médio dos caminhos mínimos $L$ em redes aleatórias. . . . .	61
4.17	Variação da diferença absoluta para o comprimento médio dos caminhos mínimos $L$ em redes de mundo pequeno. . . . .	62
4.18	Variação da diferença absoluta para o comprimento médio dos caminhos mínimos $L$ em redes livres de escala. . . . .	63
4.19	Variação da distância de Hellinger para o grau do vértice $\kappa_i$ nos modelos teóricos. . . . .	65
4.20	Variação da distância de Hellinger para o coeficiente de agrupamento $C_i$ nos modelos teóricos. . . . .	66
4.21	Variação da distância de Hellinger para o <i>betweenness centrality</i> $B_i$ nos modelos teóricos. . . . .	67
4.22	Variação da distância de Hellinger para o comprimento dos caminhos mínimos $l_{(i,j)}$ nos modelos teóricos. . . . .	68
4.23	$P$ -valores obtidos do teste de Kolmogorov-Smirnov. . . . .	71
4.24	$P$ -valores obtidos do teste $t$ de Student. . . . .	73

4.25 Ilustração de uma rede de sensores sem fio. . . . .	76
4.26 Exemplo do estabelecimento de uma conexão em um grafo geométrico. . . . .	77
4.27 Exemplos de estratégias de propagação de dados em RSSFs. . . . .	78
4.28 Variação dos quantificadores para o comprimento dos caminhos mínimos em RSSFs usando-se inundação. . . . .	81
4.29 Variação dos quantificadores para o comprimento dos caminhos mínimos em RSSFs usando-se redes complexas. . . . .	83

## Lista de Tabelas

4.1	Média dos valores dos quantificadores para o nível de perturbação mais intenso (10%) quando aplicado ao grau do vértice $\kappa_i$ . . . . .	49
4.2	Média dos valores dos quantificadores para o nível de perturbação mais intensa (10%) para o comprimento do caminho mínimo $\ell_{(i,j)}$ . . . .	60
4.3	Média dos valores da diferença para o nível de perturbação mais intensa (10%) para o comprimento médio dos caminhos mínimos $L$ . . . .	64
4.4	Porcentagem de $p$ -valores $\geq 0.05$ obtidas do teste de Kolmogorov-Smirnov. . . . .	72
4.5	Porcentagem de $p$ -valores $\geq 0.05$ obtidas do teste de $t$ de Student. . . .	74
4.8	Mecanismos de propagação vs. perturbações . . . . .	80
4.9	Média dos valores dos quantificadores para o nível de perturbação mais intensa (10%) para o comprimento do caminho mínimo $\ell_{(i,j)}$ . . . .	84
4.6	Comportamento geral dos quantificadores para todos os modelos e perturbações avaliadas. . . . .	86
4.7	Comportamento geral dos modelos teóricos para as perturbações avaliadas. . . . .	87

## Lista de símbolos

$\mathcal{G}$	Grafo não direcionado
$\mathcal{G}_{er}$	Rede aleatória
$\mathcal{G}_{ws}$	Rede de mundo pequeno
$\mathcal{G}_{ba}$	Rede livre de escala
$\mathcal{G}_{gr}$	Rede geométrica
$\mathcal{V}$	Conjunto de vértices de $\mathcal{G}$
$v$	Elemento de $\mathcal{V}$
$\mathcal{E}$	Conjunto de arestas de $\mathcal{G}$
$e$	Elemento de $\mathcal{E}$
$N$	Número de elementos de $\mathcal{V}$
$M$	Número de elementos de $\mathcal{E}$
$\mathcal{A}$	Matriz de adjacências de $\mathcal{G}$
$a$	Elemento da matriz de adjacências $\mathcal{A}$
$\kappa$	Grau de um vértice $i$ de $\mathcal{G}$
$K$	Número de vizinhos de um vértice $v_i$ em $\mathcal{G}$
$\mathcal{L}$	Caminho em $\mathcal{G}$
$\mathcal{T}$	Conjunto de métricas
$t$	Elemento de $\mathcal{T}$
$T$	Número de elementos de $\mathcal{T}$
$p(\kappa)$	Distribuição dos graus de uma rede
$C$	Coefficiente de agrupamento médio da rede
$C_i$	Coefficiente de agrupamento do nó $v_i$
$c$	Grau médio
$L$	Comprimento médio dos caminhos mínimos
$B$	Betweenness centrality
$\ell$	$\ell_{(v_i, v_j)}$ é caminho mínimo entre $v_i$ e $v_j$ de comprimento $ \ell_{(v_i, v_j)} $
$p$	Probabilidade de conexão de arestas
$p_c$	Probabilidade de conexão de arestas em $\mathcal{G}_{er}$

$p_r$	Probabilidade de religação de arestas em $\mathcal{G}_{ws}$
$m$	Número de arestas adicionadas a cada passo na construção de uma rede $\mathcal{G}_{ba}$
$\lambda$	Parâmetro de escala
$\Pi$	Conjunto de perturbações aplicadas a $\mathcal{G}$
$\pi$	Elemento de $\Pi$
$P$	Número de perturbações aplicado a $\mathcal{G}$
$r$	$r$ -ésima perturbação aplicada a $\mathcal{G}$
$\psi$	Função mapeamento entre entrada e saída
$G$	Número máximo de grafos de uma classe
$\eta$	$\eta_{ij}$ é o estado de um grafo $i$ após a $j$ -ésima perturbação
$X$	Variável aleatória discreta
$Y$	Variável aleatória discreta
$\Omega$	Espaço amostral
$\xi$	Elemento do espaço amostral
$S$	Entropia de Shannon
$\mathcal{D}$	Quantificador estocástico
$\mathcal{D}_{KL}$	Divergência de Kullback-Leibler
$\mathcal{D}_{KL}$	Divergência de Kullback-Leibler normalizada
$\mathcal{D}_{JS}$	Distância de Jensen-Shannon
$\mathcal{D}_H$	Distância de Hellinger
$\pi_{ae}$	Perturbação do tipo adição de aresta
$\pi_{re}$	Perturbação do tipo remoção de aresta
$\pi_{rw}$	Perturbação do tipo religação de arestas
$\pi_{av}$	Perturbação do tipo adição de vértice
$\pi_{rv}$	Perturbação do tipo remoção de vértice
$F$	Função distribuição acumulada
$\mathcal{H}$	Histograma de proporções
$\delta$	Parâmetro que é adicionado as distribuições $p$ e $q$
$g$	Contador de grafos gerados na simulação
$R$	Raio de comunicação de um nó de uma RSSF

# Sumário

<b>Agradecimentos</b>	<b>vi</b>
<b>Resumo</b>	<b>viii</b>
<b>Abstract</b>	<b>x</b>
<b>Lista de Figuras</b>	<b>xii</b>
<b>Lista de Tabelas</b>	<b>xv</b>
<b>Lista de Símbolos</b>	<b>xvi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Problema . . . . .	3
1.2 Objetivos da tese . . . . .	4
1.3 Contribuições . . . . .	4
1.4 Organização do trabalho . . . . .	5
<b>2 Fundamentos</b>	<b>7</b>
2.1 Grafos . . . . .	7
2.2 Medidas . . . . .	9
2.2.1 Distribuição dos graus . . . . .	10
2.2.2 Coeficiente de agrupamento . . . . .	11
2.2.3 Comprimento médio dos caminhos mínimos . . . . .	12
2.2.4 <i>Betweenness centrality</i> . . . . .	13
2.3 Modelos de redes complexas . . . . .	13
2.3.1 Introdução histórica . . . . .	13
2.3.2 Grafos aleatórios de Erdős Rényi . . . . .	17
2.3.3 Redes de mundo pequeno de Watts e Strogatz . . . . .	19
2.3.4 Redes livres de escala de Barabási e Albert . . . . .	21



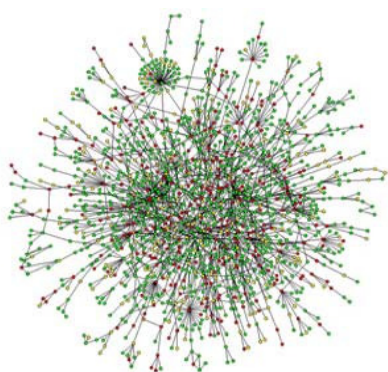
2.4	Quantificadores estocásticos . . . . .	23
2.5	Conclusões . . . . .	28
<b>3</b>	<b>Análise de medidas</b>	<b>29</b>
3.1	Introdução . . . . .	29
3.2	Definição do problema . . . . .	32
3.3	Arcabouço estatístico . . . . .	34
3.4	Metodologia para a análise de medidas . . . . .	35
3.5	Conclusões . . . . .	38
<b>4</b>	<b>Resultados</b>	<b>39</b>
4.1	Validação da metodologia . . . . .	40
4.2	Análise dos modelos teóricos . . . . .	43
4.2.1	Parâmetros de simulação e condições iniciais . . . . .	43
4.2.2	Análise dos modelos teóricos I . . . . .	44
4.2.3	Análise dos modelos teóricos II . . . . .	63
4.2.4	Testes de hipótese para a distribuição dos graus . . . . .	68
4.2.5	Discussão geral dos resultados . . . . .	74
4.3	Análise em redes de sensores sem fio . . . . .	75
4.3.1	Definições . . . . .	75
4.3.2	Topologias de propagação de dados em RSSFs . . . . .	77
4.3.3	Discussão dos resultados . . . . .	79
4.4	Conclusões . . . . .	85
<b>5</b>	<b>Conclusões</b>	<b>88</b>
	<b>Anexo A Artigo aceito para publicação</b>	<b>90</b>
	<b>Referências Bibliográficas</b>	<b>103</b>

# Capítulo 1

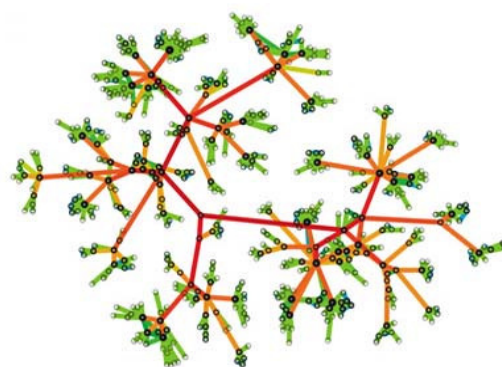
## Introdução

Uma rede, na sua forma mais simples, é definida como um conjunto de pontos ligados aos pares por retas. Tecnicamente, os pontos são chamados de vértices (ou nós) e as retas de arestas (ou ligações).

Uma rede complexa pode ser representada como uma rede de estrutura irregular e abrangente que progride dinamicamente com o tempo. É o que se compreende a partir do conceito de Boccaletti et al. [2006]. Como exemplos, pode-se observar as Figuras 1.1(a) e 1.1(b). A primeira representa uma rede de proteínas, sendo os vértices as proteínas e as arestas a interação entre elas; e a segunda ilustra a estrutura de uma rede de comunicação móvel, na qual cada nó corresponde a um usuário e as arestas representam as chamadas entre cada dois usuários da mesma rede.



(a) Rede de interação entre proteínas



(b) Rede de comunicação móvel

**Figura 1.1.** Exemplos de redes complexas. Fontes [Jeong, 2010; Onnela et al., 2007].

Nos últimos anos, inúmeras medidas vêm sendo elaboradas com o intuito de

quantificar a estrutura e o comportamento das redes complexas. Essas medidas definem um arcabouço que permite analisar e modelar diversos sistemas complexos, bem como, refletir suas diferentes características tais como: as distâncias, a conectividade, ciclos, etc.

Alguns exemplos dessas medidas são:

- O comprimento médio dos caminhos mínimos (*average shortest path length*), o qual quantifica a média dos comprimentos dos caminhos mínimos entre dois nós quaisquer da rede;
- O coeficiente de agrupamento (*clustering coefficient*), o qual indica a probabilidade de que dois nós vizinhos tenham um outro vizinho em comum;
- A vulnerabilidade, a qual mede o quão uma rede pode ser propensa a falhas;
- A centralidade de componentes, a qual identifica vértices/arestas que desempenham um papel vital dentro da rede.

Dessa forma, a rede é mapeada em um vetor de características ou medidas. Contudo, em muitos casos o mapeamento não é completo, ou não descreve seguramente a estrutura real da rede [Costa et al., 2008; Strogatz, 2001]. A escolha de medidas apropriadas para a caracterização de uma rede é um procedimento realizado através da avaliação de seu comportamento e depende, principalmente, de três fatores:

1. Disponibilidade dos dados: as redes complexas possuem milhares, ou até bilhões de nós e mudam sua estrutura constantemente. Esse dinamismo em seu comportamento junto com seu crescimento desregulado podem dificultar a coleta de informações para sua caracterização. Como exemplo, pode-se citar a *World Wide Web* (WWW), cujos vértices são as páginas Web e as arestas são as ligações que apontam de uma página para outra [Almiron et al., 2010]. Em 1999 a WWW tinha um número estimado em 800 milhões de páginas [Albert et al., 1999], e atualmente esse número chega pelo menos a 14,08 bilhões [de Kunder, 2013];
2. Capacidade de armazenamento e processamento: apesar do crescente aumento na velocidade de processamento e capacidade de armazenamento dos computadores, o processo de caracterização de uma rede com bilhões de nós não é uma tarefa trivial. Geralmente, utiliza-se apenas uma amostra da

rede para estudo do seu comportamento de acordo com os recursos e disponibilidade dos dados [Boas, 2008];

3. Comportamento de interesse: uma rede complexa pode conter informações sobre diferentes aspectos de uma aplicação; dependendo do seu comportamento um conjunto diferente de medidas pode ser escolhido. A caracterização da WWW, por exemplo, pode estar associada a medidas que identifiquem a vulnerabilidade de nós ou a eficiência do envio de informações.

Dada as restrições anteriores para caracterização de uma rede, torna-se necessária a avaliação do desempenho das medidas, quando ocorrem mudanças inesperadas na rede. Por exemplo, se uma determinada rede perde um aglomerado de arestas (ou vértices), o vetor de características definido, a priori, continua descrevendo sua estrutura? Uma forma de abordar tal problema é comparar seus diferentes estados, controlando-se as mudanças em sua estrutura, ao invés de apenas observá-las.

Dessa forma, objetiva-se investigar o uso de métodos derivados da Teoria da Informação e, em particular, o conceito de distâncias e divergências. Estes quantificadores são utilizados para comparar distribuições de probabilidade e vêm exibindo boas propriedades discriminatórias em diversos problemas, tais como, o processamento de imagens e o estudo da evolução de redes complexas [Carpi et al., 2011; Nascimento et al., 2010].

A seção que segue descreve o problema tratado neste trabalho.

## 1.1 Problema

Seja uma classe de redes, um vetor de características (ou medidas) e um conjunto de perturbações. A problemática deste trabalho está relacionada ao estudo do comportamento das medidas aplicadas a uma classe de redes quando esta é submetida a uma série de mudanças. Assim sendo, pode-se formular a seguinte questão:

*“As perturbações, aplicadas na classe de rede considerada, são capazes de destruir suas características intrínsecas?”*

Para responder a esta pergunta, é necessário quantificar as variações das medidas em diferentes estados da rede decorrentes das perturbações. Esta análise pode ser feita sobre escalares ou vetores. Os escalares podem ser comparados

com a diferença, ao passo que os vetores são comparados utilizando-se quantificadores estocásticos.

## 1.2 Objetivos da tese

Tendo posto o problema, o objetivo geral desse trabalho é investigar o comportamento de medidas usadas na caracterização de redes complexas. Partindo disto, tem-se os seguintes objetivos específicos:

1. Definir uma metodologia para avaliação do desempenho das medidas utilizadas na caracterização de uma rede considerando os seguintes aspectos: medidas e modelos a serem avaliados; quantificadores utilizados para comparar diferentes estados da rede; e mudanças que podem ocorrer na rede.
2. Investigar os efeitos provocados pelas perturbações que possam alterar as propriedades (ou medidas) de uma rede complexa. Algumas medidas são sensíveis a pequenas intensidades de perturbações aplicadas a rede, como a adição de algumas arestas. O valor da medida pode mudar de valor consideravelmente, de forma a não representar mais a estrutura da rede em questão. Já outras medidas são robustas, pois sofrem pequenas alterações (ou nenhuma alteração) quando expostas a perturbações na rede. A avaliação dessas medidas tem como objetivo encontrar medidas que mantenham o compromisso entre estabilidade e confiabilidade.
3. Implementar, testar e validar a metodologia em uma rede real: a aplicação da metodologia a um cenário real proporciona uma visão abrangente e aprofundada da relação entre sua estrutura e funcionalidade. Ademais, ao considerar diferentes tipos de perturbações em determinada aplicação, torna-se possível prever condições inesperadas e, assim, projetar soluções mais robustas a tais mudanças.

## 1.3 Contribuições

Este trabalho contribui para área de redes complexas no que tange sua caracterização e modelagem. As contribuições principais são:

- Uma análise do comportamento de diferentes medidas que englobam vários aspectos da topologia de uma rede. As medidas avaliadas são: (i) grau

do vértice, (ii) coeficiente de agrupamento local, (iii) *betweenness centrality*, (iv) comprimento dos caminhos mínimos e (v) comprimento médio dos caminhos mínimos. As medidas (i), (ii) e (iii) são calculadas para cada vértice da rede e a medida (iv) é calculada para cada par de vértices, nesses casos tem-se um vetor de valores; e, por fim, a medida (v) fornece uma informação global sobre a rede, assim tem-se um valor escalar.

- A comparação de três diferentes quantificadores estocásticos com respeito a perturbações: a divergência de Kullback-Leibler e as distâncias de Jensen-Shannon e Hellinger.
- A caracterização de redes de sensores sem fio como uma rede complexa e a análise do comprimento médio dos caminhos, medida fundamental no mecanismo de propagação de dados dessas redes.

Os estudos realizados resultaram nos seguintes artigos científicos:

- Cabral, R. S.; Aquino, A. L. L.; Frery, A. C.; Rosso, O. A. & Ramírez, J. A. (2013a). Structural changes in data communication in wireless sensor networks. *Central European Journal of Physics*.
- Cabral, R. S.; Frery, A. C. & Ramírez, J. A. (2013b). Variability analysis of complex networks measures based on stochastic distances. Submetido ao *Physica A: Statistical Mechanics and its Applications*.

A metodologia utilizada compara também diferentes quantificadores, além de identificar que tipos e intensidades de perturbações conduzem a alterações significativas das suas propriedades. Além disso, foi possível determinar quais medidas descrevem melhor o comportamento de uma rede, permitindo assim uma caracterização mais eficaz.

## 1.4 Organização do trabalho

A parte seguinte do trabalho está organizado como segue.

O capítulo 2 apresenta as definições e conceitos necessários ao entendimento deste trabalho: os conceitos básicos de teoria dos grafos, a definição das medidas avaliadas, os conceitos acerca dos modelos de redes complexas e a definição dos quantificadores estocásticos.

O capítulo 3 apresenta uma visão geral sobre a análise de variabilidade de medidas em redes complexas e a revisão da literatura acerca deste problema.

Além disso, na seção 3.2, o problema é definido formalmente e na seção 3.4 apresenta-se a metodologia utilizada.

Com base nos conceitos, na definição do problema e da metodologia de análise, o capítulo 4 apresenta os resultados encontrados neste trabalho. Este capítulo está organizado da seguinte forma: a seção 4.1 apresenta a validação da metodologia aplicada por meio da comparação dos resultados encontrados neste trabalho com os resultados de Boas et al. [2010]; a seção 4.2 apresenta a análise das medidas (grau do vértice, coeficiente de agrupamento, *betweenness centrality*, caminho mínimo e comprimento médio dos caminhos mínimos) quando aplicadas aos modelos aleatório, de mundo pequeno e livre de escala. Além disso, foram realizados testes de hipótese para a distribuição dos graus, propriedade fundamental das redes complexas; e a seção 4.3 apresenta os resultados da análise do comprimento dos caminhos mínimos aplicado às redes de sensores sem fio. Esta seção consta das definições de redes de sensores sem fio, das diferentes abordagens utilizadas como modelos para a propagação de dados nessas redes: a propagação baseada em inundação e a propagação baseada em redes complexas.

Por fim, o capítulo 5 mostra as conclusões do trabalho.

# Capítulo 2

## Fundamentos

Este capítulo tem por objetivo apresentar os conceitos básicos necessários para um bom entendimento deste trabalho. Tais conceitos estão relacionados a grafos, medidas e modelos de redes complexas. Apresentam-se aqui apenas os conceitos mais gerais, deixando aspectos específicos para serem explicados e referenciados quando necessários. A seção 2.1 apresenta os principais conceitos sobre grafos, a seção 2.2 define as medidas avaliadas, a seção 2.3 apresenta os modelos teóricos de redes complexas, a seção 2.4 define os quantificadores estocásticos e, por fim, a seção 2.5 apresenta as conclusões do capítulo.

### 2.1 Grafos

A teoria dos grafos é um arcabouço natural para o tratamento matemático de redes. As medidas utilizadas em redes complexas são formalizadas matematicamente a partir de um grafo. Dessa forma, são apresentados aqui alguns conceitos relacionados a grafos que serão utilizados ao longo do texto.

Um grafo pode ser definido como [Ahuja et al., 1993]:

**Definição 1 (Grafo)** *Um grafo  $\mathcal{G}$  é dado por um conjunto de vértices (ou nós)  $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$  e um conjunto de arestas (ou ligações)  $\mathcal{E} = \{e_1, e_2, \dots, e_M \mid e_m = (v_i, v_j) : 1 \leq i, j \leq N, v_i \neq v_j\}$ .*

O grafo  $\mathcal{G}$  pode ser representado pela matriz de adjacência  $\mathcal{A}_{N \times N}$  com elementos  $a_{ij}$ , em que  $i, j = 1, \dots, N$ . A definição precisa das entradas da matriz varia de acordo com as propriedades do grafo que se deseja representar, porém de forma geral o valor  $a_{ij}$  guarda informações sobre como os vértices  $v_i$  e  $v_j$  estão



relacionados. Cada elemento  $a_{ij}$  tem um valor constante quando existe a aresta entre  $v_i$  e  $v_j$ , e zero caso contrário.

Neste trabalho são considerados grafos  $\mathcal{G}$  não direcionados (ou não orientados), não ponderados e com um único componente conectado.

**Grafo não direcionado (ou não orientados)** : As arestas  $e_i = (v_i, v_j)$  e  $e_j = (v_j, v_i)$  são consideradas como uma única aresta, i.e.,  $e_i = e_j$ . Nesse caso, a relação de adjacência é simétrica, i.e.,  $a_{ij} = a_{ji}$ .

**Não ponderados** : Um grafo pode ser ponderado ou não-ponderado. Se todas as arestas apresentarem um mesmo peso (ou custo), diz-se que o grafo é não ponderado e considera-se todas as arestas com custo igual a 1. O grafo será ponderado se diferentes arestas possuírem custos distintos

**Grafo conexo** : Ou com um único componente conectado, quaisquer dois vértices em  $\mathcal{G}$  estão ligados por pelo menos um caminho  $\mathcal{L}$ .

Dado um subconjunto  $\mathcal{V}'$  de vértices em  $\mathcal{V}$ , chama-se subgrafo de  $\mathcal{G}$  gerado por  $\mathcal{V}'$  ao grafo  $H = (\mathcal{V}', \mathcal{E}')$ , em que  $\mathcal{E}'$  é formado por arestas de  $\mathcal{E}$  com ambas as extremidades em  $\mathcal{V}'$ . Neste caso tem-se  $\mathcal{V}' \subset \mathcal{V}$ ,  $\mathcal{E}' \subset \mathcal{E}$  e  $H \subset \mathcal{G}$ .

Um grafo  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  possui, entre outras, as seguintes propriedades fundamentais:

**Grau do vértice** O grau  $\kappa_i$  do vértice  $v_i$  é o número de arestas incidentes nele, sendo definido em termos da matriz de adjacência  $\mathcal{A}_{N \times N}$  como:

$$\kappa_i = \sum_{j \in \mathcal{V}} a_{ij}. \quad (2.1)$$

**Grau médio** É a média, por vértice, de arestas de  $\mathcal{G}$ . É calculado a partir da média do grau de todos vértices da rede:

$$c = \frac{1}{N} \sum_{i \in \mathcal{V}} \kappa_i \quad (2.2)$$

Muitas vezes é conveniente expressar essa razão diretamente em função do número de vértices e arestas de  $\mathcal{G}$ :

$$c = \frac{2M}{N} \quad (2.3)$$

**Caminho**  $\mathcal{L}_{(v_i, v_j)}$  entre dois vértices  $v_i$  e  $v_j$  em  $\mathcal{G}$  é uma sequência de arestas que se inicia em  $v_i$  e termina em  $v_j$ , i.e.,  $\mathcal{L} = (e_1, e_2, \dots, e_m)$ , em que  $e_1 = (v_i, \cdot)$  e  $e_m = (\cdot, v_j)$ . O tamanho do caminho é definido pelo número de arestas da sequência, i.e.,  $|\mathcal{L}_{(v_i, v_j)}| = m$ .

**Caminho mínimo ou geodésico**  $\ell_{(v_i, v_j)}$  é qualquer caminho  $\mathcal{L}$  que conecta os vértices  $v_i$  e  $v_j$  com a distância mínima. Seu comprimento é denotado por  $|\ell_{(v_i, v_j)}|$  ou  $|\ell_{(i, j)}|$ .

**Ciclo** Um ciclo em teoria de grafos é um caminho de comprimento mínimo três, em que o primeiro e o último vértice coincidem, mas nenhum outro vértice é repetido.

Uma discussão aprofundada sobre teoria dos grafos pode ser encontrada em Bollobás [1998].

## 2.2 Medidas

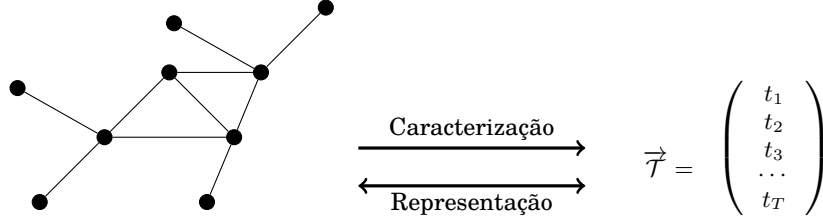
Pode-se considerar a teoria de redes complexas como uma extensão da teoria dos grafos, adicionada de mecanismos estatísticos que permitem descrever melhor seu comportamento. De acordo com Rodrigues [2007], essa teoria difere da tradicional em três aspectos básicos: (i) ela está relacionada com a modelagem de redes reais, por meio de análise de dados empíricos; (ii) as redes estudadas não são estáticas, mas evoluem no tempo; (iii) as redes, na maioria das vezes, não são consideradas apenas objetos topológicos, mas constituem estruturas de processos dinâmicos, como a propagação de doenças e opiniões.

A caracterização (ou análise estrutural) de uma rede pode ser feita através de uma inspeção visual dos grafos correspondentes, mas essa estratégia funciona apenas para redes com algumas centenas de vértices e arestas [Boas, 2008]. Para uma rede com bilhões de nós, como a WWW, é impossível determinar suas características topológicas dessa maneira.

A solução adotada para esse tipo de rede consiste em mapeá-las em termos de medidas topológicas capazes de expressar suas características mais relevantes, tais como, conectividade, densidade, graus, distâncias, centralidade, formação de comunidades, dinâmica do fluxo de informação, etc.

A Figura 2.1 ilustra a caracterização de uma rede complexa em termos do seu correspondente vetor de medidas  $\vec{T} = (t_{i-1}, t_i, t_{i+1}, \dots, t_T)$ , em que  $t_i$  é a  $i$ -ésima medida que qualifica a rede e  $T$  a quantidade de medidas utilizadas. De

acordo com Costa et al. [2008], caso seja possível a obtenção da rede original a partir do vetor de medidas, o mapeamento é uma representação.



**Figura 2.1.** Caracterização e/ou representação de uma rede complexa em termos de um vetor de medidas.

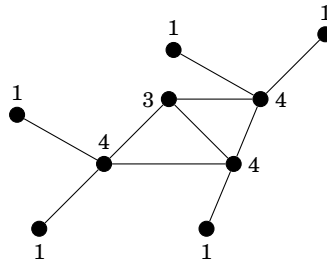
Apesar de existirem inúmeras medidas que quantificam as propriedades de uma rede complexa, algumas delas são amplamente utilizadas para sua classificação e caracterização.

A seguir são apresentadas as medidas estudadas neste trabalho.

### 2.2.1 Distribuição dos graus

A distribuição dos graus  $p(\kappa)$  é definida como a probabilidade de um vértice  $v_i \in \mathcal{V}$ , escolhido aleatoriamente, ter grau  $\kappa$ , i.e., qual a fração de nós no grafo tem grau  $\kappa$ .

Como exemplo, considere o grafo da Figura 2.2 com nove vértices ( $N = 9$ ) e dez arestas ( $M = 10$ ). A contagem de diferentes graus é: cinco vértices com grau um, um vértice com grau três e três vértices com grau quatro. Os valores de  $p(\kappa)$  para  $\kappa = 0, \dots, \kappa_{MAX}$ , com  $\kappa_{MAX} = 4$ , são respectivamente:  $p(0) = \frac{0}{9}$ ,  $p(1) = \frac{5}{9}$ ,  $p(2) = \frac{0}{9}$ ,  $p(3) = \frac{1}{9}$  e  $p(4) = \frac{3}{9}$ .



**Figura 2.2.** Grafo com  $N = 9$  vértices e  $M = 10$  arestas e próximo a cada vértice o seu respectivo grau  $\kappa$ .

A distribuição dos graus é uma propriedade fundamental de qualquer rede e se utilizada em conjunto com outras medidas pode determinar a classe a que

uma rede pertence, como será discutido nas seções que seguem. Além disso, a distribuição dos graus é definida analiticamente para alguns modelos importantes de grafos, como será discutido na seção 2.3.

### 2.2.2 Coeficiente de agrupamento

Em redes complexas, o conceito de agrupamento foi introduzido como uma forma de quantificar a transitividade das conexões, i.e., a presença de triângulos conectados na rede [Newman & Watts, 1999]. O coeficiente de agrupamento pode ser uma medida local, se definido para cada vértice do grafo, ou global, quando refere-se ao grafo. O coeficiente de agrupamento local  $C_i$  é definido como o número de arestas  $ne_i$  que conectam os vizinhos (vértices adjacentes) do vértice  $v_i$  em  $\mathcal{G}$ . Matematicamente:

$$C_i = \frac{2ne_i}{\kappa_i(\kappa_i - 1)}, \quad (2.4)$$

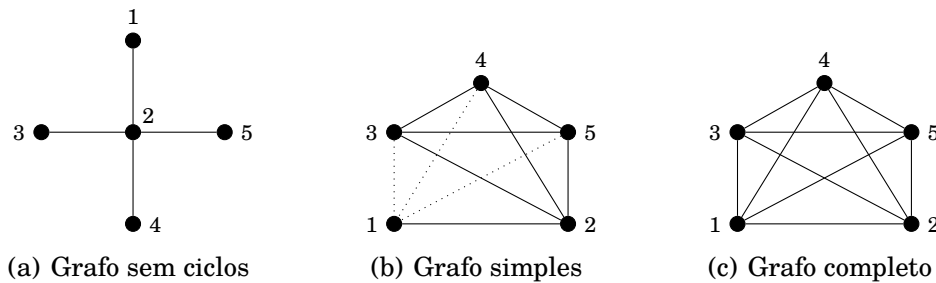
em que  $\kappa_i(\kappa_i - 1)$  é o número máximo possível de arestas no subgrafo de vizinhos de  $v_i$  e, por definição,  $0 \leq C_i \leq 1$ .

O coeficiente de agrupamento global é definido como a média de todos os valores de  $C_i$  em  $\mathcal{G}$ , matematicamente tem-se:

$$C = \frac{1}{N} \sum_{i \in \mathcal{V}} C_i, \quad (2.5)$$

em que,  $0 \leq C \leq 1$ .

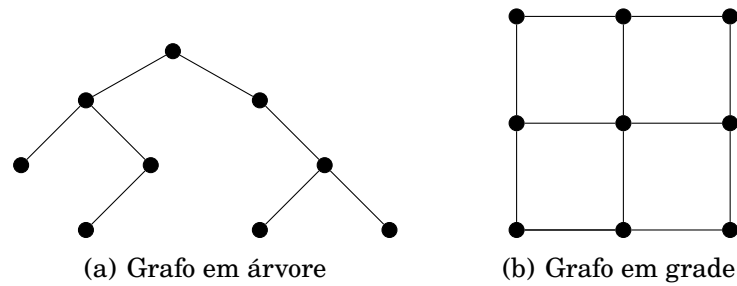
Para ilustrar o cálculo de  $C_i$  e  $C$ , considere os três grafos da Figura 2.3. Deseja-se calcular o valor do coeficiente de agrupamento local  $C_2$  do vértice  $v_2$  e o coeficiente de agrupamento global  $C$  dos grafos.



**Figura 2.3.** Três grafos distintos com  $N = 5$  vértices.

No grafo da Figura 2.3(a)  $C_2 = 0$ , pois o número de arestas que conectam

os vizinhos de  $v_2$  é igual a zero; e como não possui nenhum ciclo  $C = 0$ . Esta é uma característica presente nos grafos em árvore e grafos com topologias onde os nós são arranjados em grade e conectados apenas aos seus vizinhos imediatos, como ilustrado nas Figura 2.4(a) e 2.4(b), respectivamente. Já no grafo da Figura 2.3(c), o valor de  $C_2 = 1.0$  e  $C = 1.0$ , isso acontece pelo fato deste grafo ser completo, i.e., todos os nós são vizinhos dos demais. Quando  $C = 1.0$  tem-se transitividade perfeita. Para o grafo da Figura 2.3(b) os valores são intermediários,  $C_2 = 0.5$  e  $C = 0.7$ .



**Figura 2.4.** Exemplos de grafos com coeficiente de agrupamento  $C = 0$ .

As redes sociais têm valores altos para o coeficiente de agrupamento quando comparadas a outros tipos de rede, por exemplo, uma rede de colaboração de atores no cinema tem  $C = 0.79$ , na WWW  $C = 0.11$  e em uma rede de roteadores  $C = 0.03$  [Boccaletti et al., 2006].

O uso dessa medida é extremamente útil quando se deseja detectar a presença de diferentes comunidades em uma mesma rede.

### 2.2.3 Comprimento médio dos caminhos mínimos

O comprimento médio dos caminhos mínimos, do inglês *average shortest path length*, é uma medida relacionada a distância. Calculada como a média dos caminhos mínimos entre todos os pares de vértices de  $\mathcal{G}$ :

$$L = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{V}, i < j} \ell_{(i,j)} \quad (2.6)$$

Um aspecto importante dessa definição é sua divergência se existirem vértices desconectados ( $\kappa_i = 0$ ). Esse problema é contornado considerando-se para o cálculo apenas os vértices com  $\kappa_i \geq 1$ . De acordo com Costa et al. [2008], assim evita-se a divergência, embora provoque uma distorção para redes com muitos vértices desconectados.

### 2.2.4 Betweenness centrality

O *betweenness centrality*, ou *betweenness*, é a medida que quantifica a centralidade de um vértice na rede. As medidas de centralidade são utilizadas para classificar, sob algum aspecto, os elementos (vértices ou arestas) mais importantes da rede [Almiron et al., 2010].

As medidas de centralidade de um vértice têm como objetivo estimar sua importância topológica dentro da rede, i.e., vértices que concentram o fluxo de informação dentro da rede geralmente possuem importância vital para seu funcionamento.

O *betweenness* é definido em função do número de caminhos mínimos que passam por um vértice  $v_u$ :

$$B_u = \sum_{i,j \in \mathcal{V}} \frac{\sigma(i, u, j)}{\sigma(i, j)}, \quad (2.7)$$

em que  $\sigma(i, u, j)$  é o número de caminhos mínimos entre  $v_i$  e  $v_j$  que passam por  $v_u$  e  $\sigma(i, j)$  é o número total de caminhos mínimos entre  $v_i$  e  $v_j$ .

Considere a Figura 2.3 para ilustrar o cálculo do *betweenness*. No grafo da Figura 2.3(a),  $B_2 = 6$ , pois todos os caminhos mínimos passam pelo vértice  $v_2$ ; já no grafo da Figura 2.3(b),  $B_2 = 3$ , um valor intermediário; por fim no grafo completo da Figura 2.3(c)  $B_2 = 0$ , pois nenhum caminho mínimo passa pelo vértice  $v_2$ .

As definições apresentadas aqui não esgotam a literatura, mostram apenas as medidas utilizadas neste trabalho. Um amplo estudo de medidas utilizadas para caracterização de redes pode ser encontrado no trabalho de Costa et al. [2008].

## 2.3 Modelos de redes complexas

Dada a definição das principais medidas de redes complexas, esta seção enfoca em seus modelos teóricos compondo uma breve introdução histórica e uma descrição desses modelos.

### 2.3.1 Introdução histórica

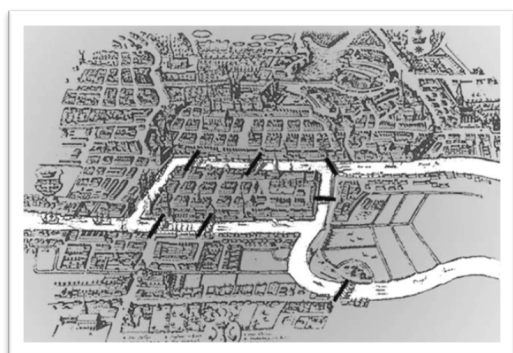
A teoria dos grafos não é uma área nova: a mesma iniciou-se quando Leonhard Euler, em meados de 1736, resolveu o problema das sete pontes de Königsberg,

uma cidade da Rússia, cortada pelo rio Pregel, formando duas ilhas. A Figura 2.5(a) ilustra o mapa da cidade, o rio Pregel está destacado em branco e as pontes em preto.

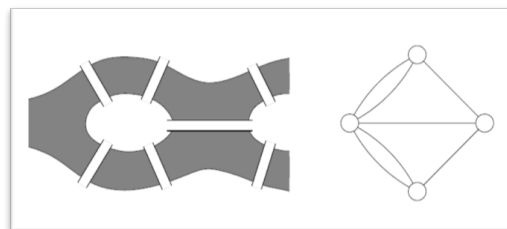
O problema consistia em atravessar todas as pontes sem repetir nenhuma delas. Euler conseguiu provar que não existia “caminho” possível para isso, por meio de um raciocínio muito simples: transformou os “caminhos” em retas e suas intersecções em pontos, criando, possivelmente, o primeiro grafo com 4 vértices e 7 arestas da história, como é mostrado na Figura 2.5(b).

Euler chegou a algumas conclusões: (i) o grafo pode ser percorrido saindo e voltando ao mesmo ponto se possuir apenas vértices com grau par; (ii) se o grafo contém, no máximo, dois vértices de grau ímpar, ele também pode ser atravessado, mas não é possível voltar ao ponto de partida; e (iii) se o grafo contém  $2n$  vértices ímpares, onde  $n$  é um número inteiro qualquer, para atravessá-lo será necessário  $n$  passagens distintas por uma mesma aresta.

No grafo da Figura 2.5(b), que representa o passeio pelas sete pontes de Königsberg, os quatro vértices têm grau ímpar, não sendo possível realizar o referido caminho [Bollobás, 1998].



(a) Ilustração da cidade de Königsberg



(b) Grafo correspondente

**Figura 2.5.** Ilustração da cidade de Königsberg em 1652 e o grafo correspondente ao problema. Fonte Newman et al. [2006]

Atualmente a teoria dos grafos é uma área da matemática bem desenvolvida destinada a diversas aplicações, que incluem problemas de transporte, planejamento e escalonamento de tarefas, análise de redes de energia elétrica, etc. A grande maioria desses problemas são de fácil definição e compreensão, embora muitas vezes difíceis de serem resolvidos, pois envolvem diversas variáveis (internas e externas ao problema) e na sua maioria buscam uma solução otimizada.

Além disso, os problemas tornam-se cada vez maiores e mais complexos e exigem o uso de técnicas eficientes para sua modelagem e caracterização.

Em meados de 1941 uma nova área de pesquisa direcionada ao estudo de redes complexas começou a emergir, definida como uma intersecção entre a teoria de grafos e mecanismos estatísticos [Costa et al., 2008]. Os primeiros trabalhos abordaram percolação em grafos\*, é o que consta em Rapoport [1951], Flory [1952] e Erdős & Rényi [1959, 1960]. Desde então, emprega-se um grande esforço na definição de modelos que retratem de forma fidedigna redes reais, bem como a definição de medidas que possam claramente caracterizá-las. A grande popularidade dessa área deve-se ao fato de detectar possibilidades de uso de medidas na caracterização de redes não contempladas pelos modelos tradicionais.

A seguir, abordam-se alguns trabalhos relacionados às inúmeras vertentes no estudo de redes complexas, tais como, definição de modelos e medidas, caracterização, otimização da robustez, problemas de sincronização e mecanismos de construção de redes complexas.

Alguns modelos teóricos foram desenvolvidos com o objetivo de caracterizar aplicações reais. Existem três deles que são mais discutidos na literatura: o modelo aleatório, de mundo pequeno e livres de escala (definidos na seção 2.3).

As primeiras análises estruturais em redes foram introduzidas por Erdős & Rényi [1959], definindo o modelo aleatório. Mais tarde, em 1960, os mesmos autores propuseram um modelo de crescimento para esses grafos e acreditavam que muitas redes reais, tais como, redes de comunicação, de ferrovias e de distribuição de energia elétrica seguiam esse modelo.

As redes de mundo pequeno surgiram do experimento de Milgram [1967], no qual foram enviadas cartas de Omaha, Nebraska e Wichita, Kansas (cidades origens) para Boston, Massachusetts (cidade destino) por meio de intermediários. Estas cidades foram escolhidas por estarem geograficamente distantes. As cartas foram inicialmente enviadas para indivíduos, escolhidos aleatoriamente, em Omaha ou Wichita. Estas incluíam detalhes do propósito do estudo, informações sobre o destinatário final em Boston e instruções para o reenvio da correspondência. Ao final do experimento, foi observado que o número de intermediários entre a origem e o destino das cartas era aproximadamente seis. Assim, Milgram concluiu que o caminho médio que separava duas pessoas quaisquer nos Estados Unidos era aproximadamente “6”, de onde surgiu a expressão “Seis graus de separação”.

---

\*A teoria de percolação examina o comportamento de grafo finito ou infinito eliminando extremidades (ou ligações) de forma aleatória.



Passaram-se muitos anos para o estudo de redes complexas se concretizar. No início os sistemas de informação eram deficientes e também não existia o poder computacional atual. A ausência desses fatores dificultava a coleta e processamento de dados.

Em 1998, Watts & Strogatz perceberam que nem todas as redes poderiam ser consideradas regulares (em que todos os vértices possuem o mesmo grau) ou aleatórias, definindo assim um modelo intermediário. Este, chamado de modelo de mundo pequeno de Watts e Strogatz, possuía um coeficiente de agrupamento alto, como nos grafos regulares, e um comprimento médio dos caminhos mínimos baixo, como nas redes aleatórias. As redes sociais, de e-mails e de atores são consideradas redes de mundo pequeno [Newman & Watts, 1999].

Em 1999, Albert et al. investigaram a conectividade e as propriedades estruturais da WWW. Eles coletaram dados e determinaram as probabilidades de um documento ter  $k$  arestas de entrada e saída para uma determinada página. Com esse experimento, perceberam que a conectividade da WWW segue uma lei de potência. Posteriormente, desenvolveram-se trabalhos destinados a definição de modelos para a WWW [Barabási et al., 2000; Broder et al., 2000; Faloutsos et al., 1999; Pastor-Satorras et al., 2001; Zhou & Mondragón, 2004].

No mesmo ano, descobre-se que os graus das conexões entre os sistemas autônomos que formam a Internet também distribuem-se segundo uma lei de potência [Faloutsos et al., 1999]. Um modelo de crescimento para as redes livres de escala é proposto por Barabási & Albert [1999], baseando-se em dois princípios: crescimento contínuo e ligação preferencial. O primeiro refere-se à adição constante de novos vértices no grafo; ao passo que o segundo determina que esses novos vértices tendem a conectar-se com vértices de maior grau. Outros modelos de ligação preferencial são encontrados em Barabási & Bonabeau [2003] e Wang & Chen [2003].

Posteriormente, define-se o modelo livre de escala, que é base para diversos sistemas complexos como as redes sociais [Carchiolo et al., 2008], redes de interação entre proteínas [Yook et al., 2004], redes de transporte aéreo [Lee, 2007; Reggiani et al., 2010], redes de operações de transferências entre instituições financeiras [Soramäki et al., 2007], redes semânticas [Steyvers & Tenenbaum, 2005], etc. A principal característica desse modelo é a presença de alguns vértices com um alto grau de conexão, denominados *hubs*.

Em conjunto com o surgimento de diferentes modelos para redes complexas veio a definição de medidas que as caracterizassem. Percebe-se a importância dessa caracterização tomando como exemplo a Internet. A caracterização da sua

topologia é fundamental para a análise de propagação de vírus entre computadores e o desenvolvimento de técnicas de bloqueio de sua infestação [Balthrop et al., 2004]. Diversas redes têm sido caracterizadas e modeladas através de medidas, a exemplo dos circuitos eletrônicos [Cancho et al., 2001], rede de e-mails [Ebel et al., 2002], redes de distribuição de energia elétrica [Bai et al., 2006; Chassin & Posse, 2005; Crucitti et al., 2004; Holmgren, 2006; Liu & Gu, 2007; Rosas-Casals, 2010], redes urbanas [Jiang & Claramunt, 2004; Wu et al., 2004], sistemas de rodovias [Boas et al., 2009], redes de telecomunicações [Onnela et al., 2007; Schintler et al., 2005; Singh & Gupte, 2005], redes móveis [Sharma & Mazumdar, 2005; Wu & Yang, 2008], o cérebro humano [Reijneveld et al., 2007], aplicações de engenharia de software [Wen et al., 2009], sistemas de refinaria [Liu et al., 2010], etc.

Uma vez apresentado um breve histórico a respeito de redes complexas, nas próximas seções são abordados os conceitos acerca de modelos de grafos (ou redes)<sup>†</sup> tratados neste trabalho: os grafos aleatórios (*random graphs*), redes de mundo pequeno (*small-world networks*) e redes livres de escala (*scale-free networks*). Para a explanação que segue, considere um grafo  $\mathcal{G}$  como na definição 1 (seção 2.1, p. 7).

### 2.3.2 Grafos aleatórios de Erdős Rényi

Os grafos aleatórios de Erdős Rényi ( $\mathcal{G}_{er}$ ) são grafos com igual probabilidade de conexão  $p_c$  para todos os pares de vértices  $(v_i, v_j)$  [Erdős & Rényi, 1959].

No processo de formação desses grafos os vértices interligam-se aleatoriamente, tendo aproximadamente a mesma quantidade de arestas, ou a mesma probabilidade de receber novas arestas. Assim, a distribuição de conectividade segue uma distribuição de Poisson da forma:

$$p(\kappa) = \frac{e^{-c} c^\kappa}{\kappa!}, \quad (2.8)$$

em que  $c = p_c(N - 1) \cong p_c N$ ,  $0 \leq p_c \leq 1$ . Embora nem todos os vértices sejam conectados com o mesmo grau, a maior parte tem um número de conexões oscilando em torno de  $c$ . Erdős & Rényi [1960] propuseram duas maneiras de construção de um grafo  $\mathcal{G}_{er}$ :

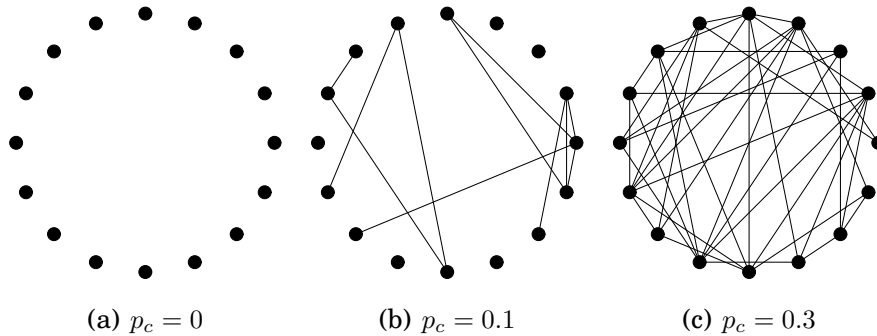
---

<sup>†</sup>Os termos grafo e rede serão utilizados neste trabalho alternadamente com o mesmo propósito.

- $\mathcal{G}_{er}(N, M)$ : o grafo é construído com  $N$  vértices e exatamente  $M$  arestas, distribuídas uniformemente entre as  $N(N-1)$  possibilidades. O algoritmo começa com um grafo totalmente desconexo e garante que não existam auto-conexões nem múltiplas arestas entre os pares de vértices.
- $\mathcal{G}_{er}(N, p_c)$ : a construção é iniciada com um grafo totalmente desconexo e então cada par de vértices  $(v_i, v_j)$  é ligado com probabilidade  $p_c$ , assim a probabilidade de se observar  $0 \leq \kappa \leq N-1$  em cada vértice segue uma distribuição Binomial com  $N-1$  ensaios e probabilidade de sucesso  $p_c$ , conduzindo a  $c = p_c(N-1)$ . O número de arestas pode ser diferente para cada grafo gerado com  $N$  vértices. A probabilidade de serem gerados grafos com exatamente  $N$  vértices e  $M$  arestas é  $p_c^M (1-p_c)^{\frac{N(N-1)}{2}-M}$  [Boccaletti et al., 2006].

Pela lei dos grandes números <sup>‡</sup> [Bernoulli, 1713], qualquer grafo do conjunto  $\mathcal{G}_{er}(N, p_c)$  tem aproximadamente  $M = \binom{N}{2} p_c$  arestas, já que o número esperado de arestas tende ao infinito. Portanto, quando  $p_c N^2$  tende ao infinito,  $\mathcal{G}_{er}(N, p_c)$  comporta-se de forma semelhante a  $\mathcal{G}_{er}(N, M)$ .

Na prática, o segundo modelo de construção é mais utilizado, pois permite a geração de grafos com um número variável de arestas. A Figura 2.6 ilustra exemplos de grafos aleatórios gerados usando-se  $\mathcal{G}_{er}(N, p_c)$  com diferentes probabilidades de conexão.



**Figura 2.6.** Exemplo de geração de grafos aleatórios usando-se  $\mathcal{G}_{er}(N, p_c)$

É importante observar que nos dois caminhos de construção o grafo resultante pode ser desconexo dependendo dos valores de  $M$  e  $p_c$ . Neste trabalho utiliza-se o modelo  $\mathcal{G}_{er}(N, p_c)$ .

<sup>‡</sup>A Lei dos Grandes Números (LGN) é um conceito fundamental em probabilidade, que declara: “Se um evento de probabilidade  $p$  é observado repetidamente em ocasiões independentes, a proporção da frequência observada deste evento em relação ao total número de repetições converge em direção a  $p$  à medida que o número de repetições se torna arbitrariamente grande.”

Os grafos aleatórios são caracterizados por um comprimento médio dos caminhos mínimos  $L$  pequeno e um coeficiente de agrupamento  $C$  baixo ( $C \ll 1$ ), definidos analiticamente como:

$$L \sim \frac{\ln N}{\ln c} \quad (2.9)$$

$$C = p_c \quad (2.10)$$

### 2.3.3 Redes de mundo pequeno de Watts e Strogatz

As redes de mundo pequeno de Watts e Strogatz ( $\mathcal{G}_{ws}$ ) apresentam coeficiente de agrupamento alto e comprimento médio dos caminhos mínimos baixo, dessa forma são consideradas redes entre os grafos regulares e os aleatórios [Watts & Strogatz, 1998].

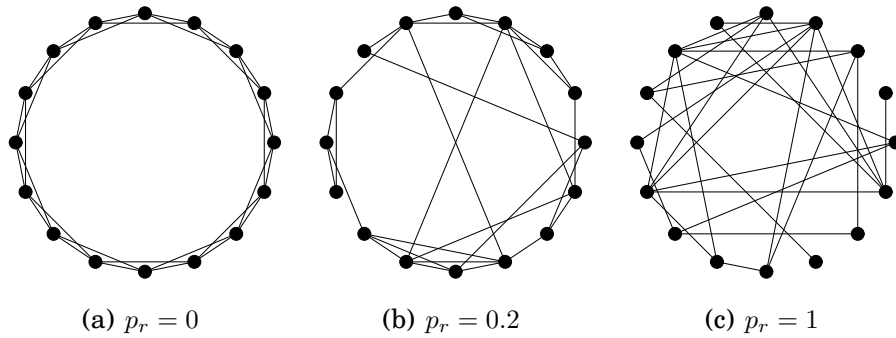
De acordo com Watts & Strogatz, o seu processo de construção dá-se a partir de uma rede regular e circular com  $N$  vértices conectados com seus  $K$  vizinhos mais próximos em cada direção (direita ou esquerda nessa topologia). Em seguida, as arestas são “reposicionadas” com probabilidade de religação  $p_r$  ( $0 < p_r \ll 1$ ). Este parâmetro controla a interpolação entre os grafos regulares e aleatórios, i.e., se  $p_r = 0$ , as arestas não são reposicionadas e tem-se um grafo regular e para  $p_r = 1$  um grafo aleatório. Para valores intermediários de  $p_r$  são geradas redes com características de mundo pequeno.

A Figura 2.7 ilustra um exemplo da construção de uma rede de mundo pequeno com  $N = 16$  e  $K = 4$ . Na Figura 2.7(a), com  $p_r = 0$  nenhuma aresta do grafo foi reposicionada obtendo-se um grafo regular. Na Figura 2.7(b), com  $p_r = 0.2$  algumas arestas foram reposicionadas, criando-se os conhecidos “atalhos”, que diminuem o caminho médio mínimo entre os vértices e mantém o coeficiente de agrupamento alto. Por fim, na Figura 2.7(c), o valor  $p_r = 1$  insere caráter completamente aleatório a rede, pois um grande número de arestas foram reposicionadas.

Entretanto, como saber se uma rede gerada com probabilidade intermediária entre zero e um é realmente de mundo pequeno, já que sua definição diz que estas estão entre os grafos regulares e aleatórios?

O processo de reconhecimento de características de mundo pequeno em um grafo  $\mathcal{G}$  é dado pelas seguintes etapas:

1. Cria-se um grafo aleatório  $\mathcal{G}_{er}$  com o mesmo número de vértices e arestas



**Figura 2.7.** Exemplo da geração de uma rede de mundo pequeno segundo o modelo de Watts e Strogatz com  $N = 16$  e  $K = 4$ .

que o grafo original  $\mathcal{G}_{ws}$ , o qual se quer mostrar possuir características de mundo pequeno.

2. Encontram-se os valores do coeficiente de agrupamento e comprimento médio dos caminhos mínimos para os dois grafos, denotados  $C_{er}$  e  $L_{er}$  para  $\mathcal{G}_{er}$  e  $C_{ws}$  e  $L_{ws}$  para  $\mathcal{G}_{ws}$ .
3. Em seguida, as condições  $L_{ws}/L_{er} \sim 1$  e  $C_{ws} \gg C_{er}$  são verificadas.

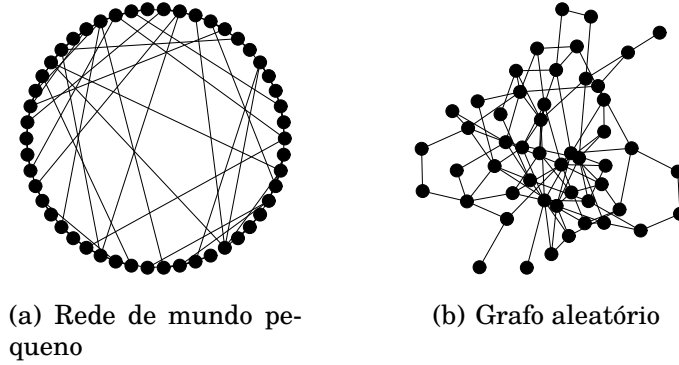
Uma rede é considerada de mundo pequeno se as condições testadas na etapa 3 são verdadeiras, ou seja, se o comprimento médio de caminhos mínimos é próximo em relação ao grafo aleatório equivalente; entretanto, essa rede possui valores de coeficiente de agrupamento muito maior.

Para ilustrar esse procedimento, considere um grafo  $\mathcal{G}$  com  $N = 50$ ,  $M = 97$ ,  $K = 4$  e  $p_r = 0.2$ . A Figura 2.8(a) representa o grafo  $\mathcal{G}_{ws}$  que supõe-se ter características de mundo pequeno e a Figura 2.8(b) o grafo aleatório  $\mathcal{G}_{er}$  correspondente, com o mesmo número de vértices e arestas que o grafo original  $\mathcal{G}_{ws}$ . Os valores de  $C$  e  $L$  para os dois grafos são:  $C_{ws} = 0.18$ ,  $C_{er} = 0.06$ ,  $L_{ws} = 3.26$  e  $L_{er} = 2.95$ ; assim,  $0.18 \gg 0.06$  e  $L_{ws}/L_{er} = 1.10 \sim 1$ , validando a condição dada em 3. Pode-se dizer então que o grafo da Figura 2.8(a) possui características de mundo pequeno.

O comprimento médio dos caminhos mínimos e o coeficiente de agrupamento para redes de mundo pequenos são definidos analiticamente como:

$$L_{ws} \sim p_r^{-\tau} f(Np_r^{-\tau}), \quad (2.11)$$

$$C_{ws} = \frac{3(K-1)}{2(2K-1)}(1-p_r)^3 \quad (2.12)$$



**Figura 2.8.** Grafos  $G_{ws}$  e  $G_{er}$  gerados com  $N = 50$ ,  $M = 97$ ,  $p_r = 0.2$  e  $K = 4$ . Tem-se,  $C_{ws} = 0.18$ ,  $C_{er} = 0.06$  ( $C_{ws} \gg C_a$ ) e  $L_{ws} = 3.26$ ,  $L_{er} = 2.95$  ( $L_{ws}/L_a = 1.10 \sim 1$ ).

em que  $f(x)$  é uma função de escala universal que obedece a:  $f(x) \sim x$ , se  $x \ll 1$  e  $f(x) \sim \ln x$ , se  $x \gg 1$  e  $\tau$  uma constante [Barthélemy & Amaral, 1999; Newman, 2000].

A distribuição dos graus em um grafo regular é uma função Delta de Dirac centrada em  $c$  (grau médio dos vértices do grafo) e a limitante é uma distribuição de Poisson, correspondente aos grafos aleatórios. Assim, para redes de mundo pequeno tem-se:

$$p(\kappa) = \sum_{1 \leq i \leq f(\kappa, K)} \binom{K}{i} (1 - p_r)^i p_r^{K-i} \frac{(p_r K)^{\kappa - K - i}}{(\kappa - K - i)!} e^{-p_r K}, \kappa \geq K, \quad (2.13)$$

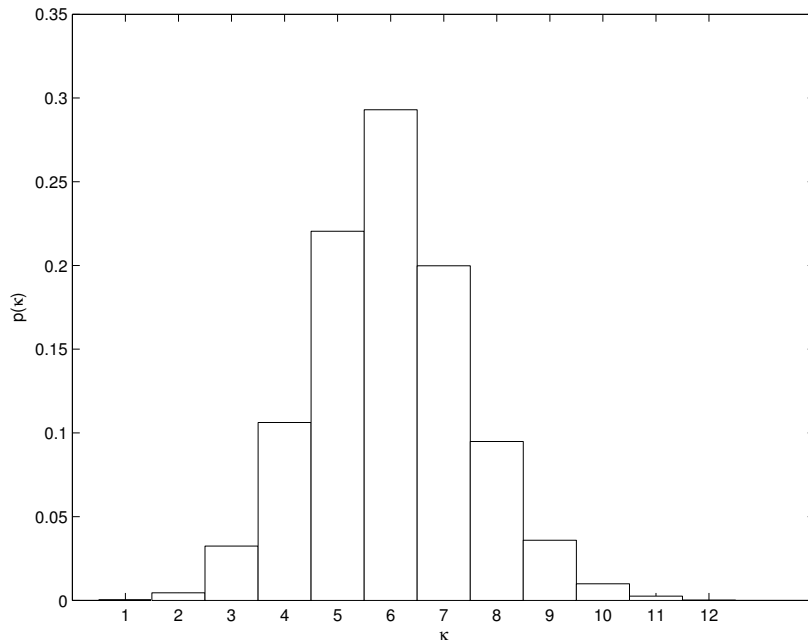
em que  $f(\kappa, K) = \min(\kappa - K, K)$  e  $K = c/2$ , pois o grau médio é  $c = 2K$ . Se  $\kappa < K$ ,  $p(\kappa) = 0$ .

A Figura 2.9 ilustra a forma usual de um histograma de proporções para a distribuição dos graus em redes de mundo pequeno.

#### 2.3.4 Redes livres de escala de Barabási e Albert

Redes livres de escala de Barabási e Albert ( $\mathcal{G}_{ba}$ ) são redes cuja distribuição dos graus segue uma lei de potência, acarretando poucos vértices com grau elevado, chamados de *hubs*, e muitos vértices com baixo grau [Barabási & Albert, 1999; Barabási & Bonabeau, 2003].

Uma quantidade  $X$  qualquer segue uma lei de potência se, dado a sua função distribuição acumulada  $F(\xi) = P(X \leq \xi)$  e seu complemento  $\bar{F}(\xi) = P(X > \xi)$  tem-se:  $\bar{F}(\xi) \approx \varsigma \xi^{-\lambda}$  para alguma constante  $0 < \varsigma < \infty$  e  $\lambda > 0$ . Para  $1 < \lambda < 2$ ,  $F$



**Figura 2.9.** Histograma de proporções para distribuição dos graus de uma rede de mundo pequeno com  $N = 10000$  vértices,  $K = 4$  vizinhos e probabilidade de religação  $p_r = 0.2$ .

tem variância infinita e média finita, porém em  $0 < \lambda \leq 1$  ambos momentos são infinitos. Pode-se escrever a distribuição dos graus para redes livres de escala como:

$$p(\kappa) \sim \kappa^{-\lambda} \quad (2.14)$$

em que  $\lambda$ , conhecido como parâmetro de escala, assume valores no intervalo  $]2, 3[$ .

Um ponto importante a se destacar é o comportamento da distribuição dos graus quando se aplica o logaritmo na equação (2.14), i.e.,  $\ln p(\kappa) \approx -\lambda \ln \kappa$ . Nesse caso, o gráfico de  $p(\kappa)$ , com escalas logarítmicas, é uma linha reta e  $\lambda$  é a inclinação da reta.

Além da distribuição dos graus seguir uma lei de potência, essas redes são caracterizadas por um comprimento médio dos caminhos mínimos pequeno, definido analiticamente como:

$$L \sim \frac{\log N}{\log(\log N)} \quad (2.15)$$

O coeficiente de agrupamento é variável, sendo muito maior do que em redes

aleatórias, mas seu valor depende de outros detalhes da topologia da rede [Antoniou & Pitsillides, 2007]. De acordo com Boccaletti et al. [2006], para redes com muitos vértices e com  $\lambda > 2.3$ ,  $C$  tende a zero e para  $\lambda < 2.3$ ,  $C$  cresce com o tamanho da rede, ou seja:

$$C \sim N^{(3\lambda-7)(\lambda-1)} \quad (2.16)$$

As propriedades das redes livres de escala estão presentes em um grande número de aplicações, incluindo a Internet [Faloutsos et al., 1999], a *World Wide Web* [Albert et al., 1999], redes de colaboração científica [Newman, 2001], etc.

O primeiro modelo de construção dessas redes foi proposto por Barabási & Albert [1999], que é ainda o melhor modelo conhecido na literatura, baseando-se em dois conceitos: (i) **crescimento** (*growth*), o que significa que o número de vértices de uma rede cresce ao longo do tempo, a exemplo da WWW que cresce exponencialmente ao longo do tempo com a adição de novas páginas; (ii) os vértices obedecem a um mecanismo de **ligação preferencial** (*preferential attachment*), i.e., quanto maior o grau de um vértice, mais alta a probabilidade dele receber novas ligações, determinando o aparecimento de *hubs* na rede.

A geração de uma rede com  $N$  vértices, segundo esse modelo, dá-se de acordo com o seguinte algoritmo:

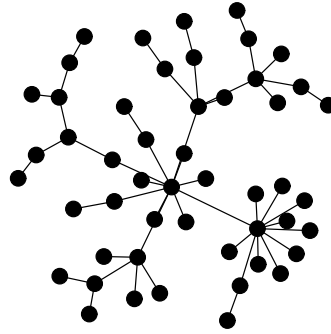
1. Cria-se uma rede com  $n_0$  vértices isolados, sendo  $n_0 \geq 2$ .
2. A cada intervalo de tempo  $\tau = 1, 2, 3, \dots, (N - n_0)$  um novo vértice  $v_j$  é adicionado e ligado com os demais com  $m$  arestas, sendo  $m \leq n_0$ . A probabilidade de um novo vértice  $v_j$  conectar-se a um vértice  $v_i$  existente na rede é proporcional ao grau de  $v_i$ .

Como todos os novos vértices adicionados a rede possuem  $m$  arestas, esta terá  $N = n_0 + \tau$  vértices no tempo  $\tau$  e  $M = m\tau$  arestas, que corresponde ao grau médio da rede  $c = 2m$  (para redes com  $N$  muito grande) [Boccaletti et al., 2006]. A Figura 2.10 ilustra uma rede livre de escala construída com esse modelo.

## 2.4 Quantificadores estocásticos

Neste trabalho, as medidas utilizadas na caracterização de uma rede são apresentadas como vetores de probabilidades que representam diferentes estados da rede. Com isso, objetiva-se comparar essas probabilidades por meio de algum quantificador.





**Figura 2.10.** Uma rede com  $N = 50$  vértices e  $m = 1$  gerada segundo o modelo de Barabási & Albert [1999].

Em Probabilidade e Estatística, métricas, divergências e distâncias são quantificadores comuns para comparação entre distribuições de probabilidade e vêm exibindo boas propriedades discriminatórias em vários problemas [Boas et al., 2010; Carpi et al., 2011]. Adicionalmente, a Teoria da Informação [Cover & Thomas, 1991] fornece medidas de divergência, baseadas no conceito de entropia, para discriminar estatisticamente distribuições estocásticas. A divergência normalizada de Kullback-Leibler, a distância de Jensen-Shannon e a métrica de Hellinger são três quantificadores apropriados para esse fim e são utilizados neste trabalho para a análise da variabilidade de medidas, além de seu desempenho ser aferido com respeito as perturbações. Os três quantificadores são definidos a seguir.

Considere as variáveis aleatórias discretas  $X$  and  $Y$  definidas no mesmo espaço amostral  $\Omega = \{\xi_1, \xi_2, \dots, \xi_n\}$ . As distribuições são caracterizadas por suas funções de probabilidade  $p, q: \Omega \rightarrow [0, 1]$ , em que  $p(\xi_i) = \Pr(X = \xi_i)$ ,  $q(\xi_i) = \Pr(Y = \xi_i)$  e  $\sum p(\xi) = \sum q(\xi) = 1$ . Uma métrica entre duas distribuições  $p$  e  $q$  é um quantificador que obedece as seguintes propriedades [Ullah, 1996]:

1.  $\mathcal{D}(p, p) = 0$ , reflexividade;
2.  $\mathcal{D}(p, q) \geq 0$ , não negatividade;
3.  $\mathcal{D}(p, q) = \mathcal{D}(q, p)$ , simetria;
4.  $\mathcal{D}(p, q) \leq \mathcal{D}(p, r) + \mathcal{D}(r, q)$ , desigualdade triangular para qualquer função de probabilidade  $r$  definida no mesmo espaço de probabilidades.

Uma “distância” não necessariamente satisfaz a propriedade 4, e uma “divergência” apenas precisa satisfazer as propriedades 1 e 2 [Cha & Srihari, 2002].

**Divergência de Kullback-Leibler** ( $D_{KL}$ ), assumindo  $q(\xi) > 0$  para todo evento  $\xi \in \Omega$  é definida como:

$$D_{KL}(p, q) = \sum_{\xi \in \Omega} p(\xi) \log \frac{p(\xi)}{q(\xi)} \quad (2.17)$$

Ela é sempre positiva, e igual a zero se e só se,  $p(\xi) = q(\xi)$  para todo  $\xi \in \Omega$ . O caso  $q(\xi) = 0$  e  $p(\xi) > 0$  tratado como divergente, assim  $D_{KL} = \infty$ . Para contornar esse problema, adiciona-se um valor  $\delta \in ]0, 1]$  a cada elemento de  $p$  e  $q$  e faz-se uma normalização tal que,  $\sum p(\xi) = 1$  e  $\sum q(\xi) = 1$  [Boas et al., 2010; Cabella et al., 2008]. Vale ressaltar que a adição de  $\delta$  pode mudar o comportamento da distribuição, e consequentemente alterar o valor de  $D_{KL}$ .

Essa divergência, muito referenciada na literatura como Entropia Relativa, foi introduzida por Kullback & Leibler [1951] e é interpretada como uma medida da perda de informação, quando  $q$  é utilizada para aproximar  $p$ . Tipicamente,  $p$  é a distribuição inicial dos dados, observações ou uma distribuição teórica e  $q$  é uma aproximação de  $p$ . No contexto de redes complexas,  $p$  e  $q$  são diferentes estados de uma rede em relação a uma medida  $t_k$ .

Apesar de Vedral [2002] considerar  $D_{KL}(p, q)$  uma “distância” entre distribuições, ela nem sempre atende a propriedade de simetria. Uma solução possível é a sua simetrização:

$$\mathcal{D}_{KL}(p, q) = \frac{1}{2} [D_{KL}(p, q) + D_{KL}(q, p)] \quad (2.18)$$

A divergência de Kullback-Leibler não possui limites para seu valor máximo, i.e.,  $0 \leq D_{KL} \leq \infty$ , porém é necessário torná-la comparável aos demais quantificadores utilizados nesse trabalho, dessa forma utiliza-se sua normalização com valores definidos no intervalo  $[0, 1]$ .

**Divergência de Kullback-Leibler normalizada** ( $\mathcal{D}_{KL}$ ) é definida como:

$$\mathcal{D}_{KL}(p, q) = 1 - \exp\{-D_{KL}(p, q)\}. \quad (2.19)$$

**Distância de Jensen-Shannon** ( $\mathcal{D}_{JS}$ ), é definida como:

$$\mathcal{D}_{JS}(p, q) = \frac{1}{2} [\mathcal{D}_{KL}(p, m) + \mathcal{D}_{KL}(q, m)] \quad (2.20)$$

em que,  $m = \frac{1}{2}(p + q)$  e com limites definidos no intervalo  $0 \leq \mathcal{D}_{JS} \leq 1$  para o logaritmo na base dois [Endres & Schindelin, 2003].

A distância de Jensen-Shannon pode também ser apresentada em função da Entropia de Shannon  $\mathcal{S}(p)$  [Lin, 1991]:

$$\mathcal{D}_{JS}(p, q) = \mathcal{S}\left(\frac{p+q}{2}\right) - \frac{1}{2}\mathcal{S}(p) - \frac{1}{2}\mathcal{S}(q) \quad (2.21)$$

em que

$$\mathcal{S}(p) = - \sum_{\xi \in \Omega} p(\xi) \log p(\xi) \quad (2.22)$$

por convenção,  $0 \ln 0 = 0$ .

A entropia é uma medida de incerteza associada a uma variável aleatória, ou uma função que indica a perda de informação de um sistema [Auder & Iooss, 2009]. Note que,  $\mathcal{S}$  assume valor máximo para a distribuição uniforme. No contexto de redes complexas, se  $p(\xi)$  é a distribuição dos graus, pode-se dizer que  $\mathcal{S}$  quantifica a heterogeneidade da rede em relação ao grau médio, sendo igual a zero para redes regulares (o grau médio é conhecido), e valor máximo em redes aleatórias uniformes [Carpi et al., 2011; Wang et al., 2006].

As definições para a distância de Jensen-Shannon dadas pelas equações (2.20) e (2.21) são equivalentes.

O quantificador de Jensen-Shannon é uma distância, pois é reflexiva, não negativa e comutativa. Além disso se tomarmos a raiz quadrada da equação (2.18) a mesma passa a ser uma métrica, pois obedece a propriedade de desigualdade triangular.

**Distância de Hellinger** ( $\mathcal{D}_H$ ) é definida como [Diaconis & Zabell, 1982]:

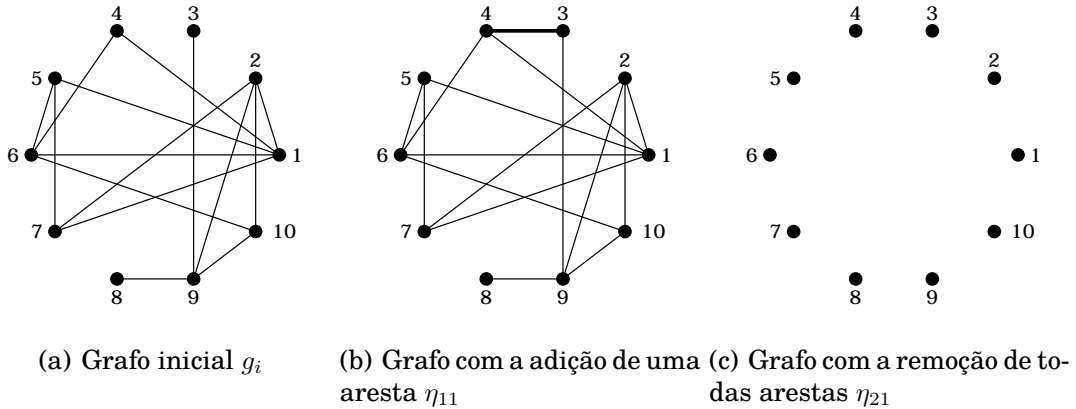
$$\mathcal{D}_H(p, q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{\xi \in \Omega} \left( \sqrt{p(\xi)} - \sqrt{q(\xi)} \right)^2} \quad (2.23)$$

Ao contrário dos quantificadores  $\mathcal{D}_{KL}(p, q)$  e  $\mathcal{D}_{JS}(p, q)$ ,  $\mathcal{D}_H(p, q)$  satisfaz a desigualdade  $0 \leq \mathcal{D}_H \leq 1$ , para  $p(\xi) = 0, \forall q(\xi) > 0$  (e vice-versa). Este quantificador é uma métrica, pois satisfaz as propriedades 1 a 4, além disso não envolve razões ou logaritmos, assim a ocorrência de zeros não acarreta em problemas numéricos.

Um tratamento moderno e unificado de distâncias estocásticas pode ser encontrado nos trabalhos de Basseville [2013], Pardo et al. [1995] e Salicrú et al. [1994].

Os grafos da Figura 2.11 são usados para exemplificar o cálculo das distâncias relativas ao grau do vértice  $\kappa$ . Estes grafos possuem  $N = 10$  vértices. Para se obter as probabilidade  $p$  e  $q$  relativas a  $\kappa$  (um vetor), primeiro calcula-se os graus de todos os vértices, dos grafos inicial ( $g_1$ ) e perturbado ( $\eta_{i1}$ ), e então obtém-se os histogramas de proporções ( $\mathcal{H}(g_1) = p$  e  $\mathcal{H}(\eta_{i1}) = q$ ). Assim sendo:

- No grafo inicial  $g_1$  da Figura 2.11(a), tem-se  $\kappa_i = (5, 4, 1, 2, 3, 4, 3, 1, 4, 3)$  e a partir da contagem de  $\kappa_i$  obtém-se a distribuição de probabilidade  $p = (0.0, 0.2, 0.1, 0.3, 0.3, 0.1)$ .
- Com a adição da aresta  $e = (3, 4)$  a  $g_1$ , o grafo perturbado  $\eta_{11}$  da Figura 2.11(b) é obtido com  $\kappa_i(\eta_{11}) = (5, 4, 2, 3, 3, 4, 3, 1, 4, 3)$  e  $q = (0.0, 0.1, 0.1, 0.4, 0.3, 0.1)$ . Nesse caso, os valores dos quantificadores são:  $\mathcal{D}_{KL}(p, q) = 0.065$ ,  $\mathcal{D}_{JS}(p, q) = 0.016$  e  $\mathcal{D}_H(p, q) = 0.012$ .
- Na Figura 2.11(c) todas as arestas são removidas obtendo-se um grafo desconectado  $\eta_{21}$ , i.e.,  $\mathcal{E} = \emptyset$ . Nesse caso,  $\kappa_i(\eta_{11}) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$  e  $q = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$  e os valores dos quantificadores são:  $\mathcal{D}_{KL}(p, q) = 0.988$ ,  $\mathcal{D}_{JS}(p, q) = 0.826$  e  $\mathcal{D}_H(p, q) = 1$ .



**Figura 2.11.** Grafos com  $N = 10$  vértices para o cálculo dos quantificadores  $\mathcal{D}_{KL}$ ,  $\mathcal{D}_{JS}$  e  $\mathcal{D}_H$ .

Como esperado, no primeiro caso, em que apenas uma aresta foi adicionada ao grafo, alterando levemente os graus dos vértices, os valores dos quantificadores são baixos. Já no segundo caso, ao se removerem todas as arestas, os graus de

todos os vértices são alterados e consequentemente, os quantificadores atingem valores altos, com  $\mathcal{D}_H$  tendo valor máximo.

Os valores altos/baixos dos quantificadores podem ser interpretados de duas formas: (i) a perturbação efetuada na rede foi ou não significativa, (ii) a medida avaliada é ou não sensível ao tipo de perturbação ocorrida. Pode-se identificar ainda a capacidade dos quantificadores captarem as mudanças ocorridas na rede.

## 2.5 Conclusões

Este capítulo apresentou as definições e conceitos necessários para realização deste trabalho. O escopo deste trabalho limita-se à análise de quatro medidas: grau do vértice, comprimentos dos caminhos mínimos, coeficiente de agrupamento local e comprimento médio dos caminhos quando utilizadas na caracterização das redes aleatórias, de mundo pequeno e livres de escala. Além disso, foram definidos os quantificadores estocásticos que são utilizados para mensurar a diferença entre os estados que uma rede complexa pode assumir.

Esses conceitos também são de suma importância para o estudo realizado em redes de sensores sem fio mostrado na seção 4.3. Nesse estudo, o mecanismo de propagação de dados em redes de sensores sem fio é modelado com as topologias de redes complexas descritas anteriormente e o comportamento do comprimento dos caminhos mínimos é avaliado com o uso dos quantificadores estocásticos.

De posse desses conceitos, pode-se definir o problema de análise de variabilidade de medidas e a metodologia utilizada neste trabalho no Capítulo 3.

## Capítulo 3

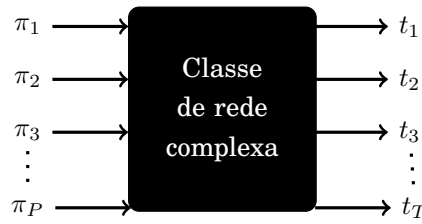
### Análise de medidas

Este capítulo tem como objetivo apresentar formalmente o problema de análise de variabilidade de medidas em redes complexas, bem como a metodologia proposta para sua resolução. Para isso, a seção 3.1 introduz os conceitos de análise de variabilidade e apresenta uma breve revisão da literatura. A seção 3.2 apresenta a definição formal do problema, o que inclui as perturbações aplicadas as redes. A seção 3.3 apresenta uma breve definição dos conceitos estatísticos utilizados na análise da variação de medidas. A seção 3.4 apresenta a metodologia para a análise da variabilidade de medidas e a estrutura do algoritmo utilizado e, por fim, a seção 3.5 apresenta as conclusões do capítulo.

#### 3.1 Introdução

A análise da variabilidade de medidas em uma determinada classe de redes complexas tem como objetivo classificar o conjunto de perturbações, de acordo com o seu grau de interferência no comportamento da rede.

Nesse caso, tem-se um mapeamento determinístico de um conjunto de entradas (perturbações)  $\Pi = \{\pi_1, \pi_2, \pi_3, \dots, \pi_P\}$  para uma ou mais saídas (medidas)  $\mathcal{T}_k = \{t_1, t_2, t_3, \dots, t_T\}$ . Para cada saída  $\mathcal{T}_k$  existe um mapeamento  $\psi_k: (\pi_1, \dots, \pi_P) \mapsto \mathcal{T}_k$ , ou seja, dependendo das perturbações aplicadas ao modelo de rede, tem-se um subconjunto específico de medidas, que conseguem caracterizá-lo de forma eficaz. A Figura 3.1 ilustra essa situação. Normalmente, a relação entre entrada e saída não está disponível de forma explícita, mas pode ser avaliada numericamente para qualquer conjunto de entradas [Lüdtke et al., 2008].



**Figura 3.1.** Representação da análise da variabilidade de medidas para redes complexas.

Objetiva-se, então, entender o comportamento das medidas do conjunto  $\mathcal{T}_k = \{t_1, t_2, t_3, \dots, t_T\}$  frente às perturbações  $\Pi = \{\pi_1, \pi_2, \pi_3, \dots, \pi_P\}$ . Considerando que as medidas utilizadas na caracterização de uma rede são apresentadas como vetores de probabilidades que representam diferentes estados da rede, uma forma de comparar essas probabilidades é por meio de algum quantificador. Neste trabalho, utilizou-se os quantificadores definidos na seção 2.4 quantificadores: a divergência normalizada de Kullback-Leibler, a distância de Jensen-Shannon e a métrica de Hellinger.

A seguir é apresentada a literatura que explora o uso de Teoria da Informação e quantificadores estocásticos para analisar o comportamento de redes complexas.

A heterogeneidade da distribuição dos graus dos vértices de uma rede é um ponto bastante explorado e o conceito de Entropia vem sendo empregado para mensurar o grau de “desordem” de um sistema. Nesse caso, o sistema é totalmente ordenado em redes regulares e desordenado em redes aleatórias [Solé & Valverde, 2004; Wang et al., 2006].

Carpi et al. [2011] propuseram o uso de um quantificador baseado na Teoria da Informação para analisar a evolução de redes de mundo pequeno. O quantificador, chamado de complexidade estatística, é utilizado para computar as mudanças topológicas relacionadas à distribuição dos graus dos vértices. É definida como o produto da Entropia pela distância de Jensen-Shannon, assim, além de mensurar a diferença entre os estágios de evolução da rede, captura a desordem da distribuição dos graus. Esse quantificador requer o uso de uma distribuição de probabilidade como referência para computar a distância de Jensen-Shannon. Os referidos autores usaram três distribuições: Poisson, uniforme, e a distribuição correspondente a uma rede regular (Dirac Delta centrada no grau médio da rede). A escolha destas está associada ao fato de que as redes de mundo pequeno encontram-se entre as redes regulares e as redes aleatórias. Dessa forma, seu

uso é restrito às medidas para as quais se conhece, a priori, a distribuição.

Uma prática comum na análise de redes complexas é utilizar apenas uma amostra da rede ao invés de considerá-la inteiramente, economizando memória e tempo de processamento. Boas et al. [2010] questionaram a adequação do uso dessas amostras e utilizaram a divergência de Kullback-Leibler como uma distância para comparar diferentes estados da rede. Eles aplicaram três tipos de perturbações: adição, remoção e religação de arestas. Contudo, esta divergência não pode ser considerada uma distância, uma vez que não possui a propriedade de comutatividade.

A análise de perturbações em redes também vem sendo estudada sob dois aspectos: sua vulnerabilidade a ataques e a identificação de elementos da rede cuja falha levam-na a um colapso (*breakdown*).

A vulnerabilidade de uma rede está associada à queda de seu desempenho na ocorrência de mudanças estruturais, que podem ser originadas pela remoção aleatória (falha) ou direcionada (ataque) de vértices. As medidas relacionadas a essa propriedade são comumente definidas em função do comprimento médio dos caminhos mínimos e o tamanho dos componentes conectados do grafo. A ideia central na análise da vulnerabilidade é inserir propositalmente uma sequência de falhas (ou ataques) na rede e avaliar seu comportamento [Albert et al., 2000; Arianos et al., 2009; Crucitti et al., 2003; Demetrius & Mankeentropy, 2005; Ghamry & Elsayed, 2012; Gol'dshtein et al., 2004; Mishkovski et al., 2011; Wang et al., 2009; Zhang et al., 2012].

O fenômeno denominado “colapso de uma rede” refere-se a um tipo de processo em cascata, em que a falha de alguns vértices pode resultar na quebra da sua funcionalidade. Pode-se citar como exemplo, as redes de distribuição de energia elétrica, onde a falha de um nó obriga que sua carga seja redistribuída para outros nós, provocando uma sobrecarga na rede e eventualmente outras falhas [Huang et al., 2008; Paul et al., 2005; Wang et al., 2012].

Como supracitado, este trabalho estende os trabalhos mencionados anteriormente em duas vertentes: uma análise do comportamento de diferentes medidas que englobam vários aspectos da topologia de uma rede, bem como a comparação de três distâncias estocásticas com respeito as perturbações.



### 3.2 Definição do problema

Uma rede complexa pode ser descrita como um grafo não direcionado  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , onde  $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$  é o conjunto de vértices (ou nós), e  $\mathcal{E} = \{e_1, e_2, \dots, e_M \mid e_m = (v_i, v_j) : 1 \leq i, j \leq N, v_i \neq v_j\}$  é o conjunto de arestas (ou ligações) entre os vértices\*.

Desta forma, o problema abordado neste trabalho, pode ser definido como:

**Definição 2 (Problema)** *Seja  $\mathcal{G}$  uma classe de redes complexas,  $\mathcal{T} = \{t_1, t_2, \dots, t_T\}$  um vetor de características de  $\mathcal{G}$  e  $\Pi = \{\pi_1, \pi_2, \dots, \pi_P\}$  um conjunto de perturbações aplicadas a  $\mathcal{G}$ . O objeto de interesse é o comportamento de  $\mathcal{T}$  aplicado a  $\mathcal{G}$  perante  $\Pi$ ; em outras palavras, deseja-se conhecer como as perturbações alteram as características intrínsecas da classe de redes em questão.*

Para realizar essa análise tomam-se elementos de  $\mathcal{G}$ , digamos  $g_1, g_2, \dots, g_G$  e submetem-se os mesmos a uma série de perturbações. Por exemplo,  $g_j$  é transformado em  $\eta_{1,j} = \pi_1(g_j), \eta_{2,j} = \pi_2(g_j), \dots, \eta_{i,j} = \pi_i(g_j)$ .

Dessa forma, tem-se  $\Pi: \mathcal{G} = (\mathcal{V}, \mathcal{E}) \rightarrow \mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ , em que  $\mathcal{V}'$  e  $\mathcal{E}'$  são os novos conjuntos de vértices e arestas, definidos de acordo com o tipo de perturbação aplicada a  $\mathcal{G}$ . Os seguintes tipos de perturbações são considerados:

1. **Adição de aresta ( $\pi_{ae}$ ):** uma aresta  $e_m = (v_i, v_j) \notin \mathcal{E}$  é adicionada a  $\mathcal{G}$ , definindo  $\mathcal{E}' = \mathcal{E} \cup (v_i, v_j)$ , i.e.,  $\pi_{ae}: \mathcal{E} = \{e_1, e_2, \dots, e_M\} \rightarrow \mathcal{E}' = \{e_1, e_2, \dots, e_m, e_{M+1}\}$ .
2. **Remoção de aresta ( $\pi_{re}$ ):** uma aresta  $e_m = (v_i, v_j) \in \mathcal{E}$  é removida de  $\mathcal{G}$ , então  $\mathcal{E}' = \mathcal{E} \setminus (v_i, v_j)$  i.e.,  $\pi_{re}: \mathcal{E} = \{e_1, \dots, e_{m-1}, e_m, e_{m+1}, \dots, e_M\} \rightarrow \mathcal{E}' = \{e_1, \dots, e_{m-1}, e_{m+1}, \dots, e_{M-1}\}$ .
3. **Religação de arestas ( $\pi_{rw}$ ):** duas arestas de  $\mathcal{G}$ ,  $(e_1 = (v_i, v_j), e_2 = (v_k, v_l)) \in \mathcal{E}$  são substituídas por  $(e'_1 = (v_i, v_l), e'_2 = (v_j, v_k)) \notin \mathcal{E}$ , então  $\mathcal{E}' = \mathcal{E} \setminus ((v_i, v_j), (v_k, v_l)) \cup ((v_i, v_l), (v_j, v_k))$ , i.e.,  $\pi_{rw}: \mathcal{E} = \{e_1, e_2, \dots, e_{m-1}, e_m, e_{m+1}, \dots, e_M\} \rightarrow \mathcal{E}' = \{e'_1, e'_2, \dots, e_{m-1}, e_m, e_{m+1}, \dots, e_M\}$ .
4. **Adição de vértice ( $\pi_{av}$ ):** um vértice  $v_i \notin \mathcal{V}$  é adicionado a  $\mathcal{G}$ , definindo  $\mathcal{V}' = \mathcal{V} \cup v_i$  e  $\mathcal{E}' = \mathcal{E} \cup (v_i, v_j), v_j \in \mathcal{V}$ , i.e.,  $\pi_{av}: \mathcal{V} = \{v_1, v_2, \dots, v_N\} \rightarrow \mathcal{V}' = \{v_1, v_2, \dots, v_N, v_{N+1}\}$  e  $\pi_{av}: \mathcal{E} = \{e_1, e_2, \dots, e_M\} \rightarrow \mathcal{E}' = \{e_1, e_2, \dots, e_M, e_{M+1}\}$ .

---

\*A definição é a mesma dada para grafos não direcionados na seção 2.1, p. 7

5. Remoção de vértices ( $\pi_{rv}$ ): um vértice  $v_j \in \mathcal{V}$  é removido de  $\mathcal{G}$ , fazendo  $\mathcal{V}' = \mathcal{V} \setminus v_j$  e  $\mathcal{E}' = \mathcal{E} \setminus (v_j, v_i)$ , em que  $v_i$  são os vértices ligados a  $v_j$ , i.e.,  $\pi_{rv}: \mathcal{V} = \{v_1, \dots, v_{n-1}, v_n, v_{n+1}, \dots, v_N\} \rightarrow \mathcal{V}' = \{v_1, \dots, v_{n-1}, v_{n+1}, \dots, v_{N-1}\}$  e  $\pi_{rv}: \mathcal{E} = \{e_1, \dots, e_{m-1}, e_m, e_{m+1}, \dots, e_M\} \rightarrow \mathcal{E}' = \{e_1, \dots, e_{m-1}, e_{m+1}, \dots, e_{M-1}\}$ .

As três primeiras perturbações foram aplicadas anteriormente por Boas et al. [2010]. Essas perturbações foram consideradas por ocorrerem com frequência em redes reais. A escolha de perturbações em detrimento a outras pode variar de acordo com a funcionalidade da rede ou a sua aplicação.

Perturbações que envolvem a adição de vértices também poderia ser avaliada, contudo este tipo de perturbação depende do modelo de rede avaliada, i.e., os nós são adicionados de acordo com o modelo de crescimento da rede. Esse fato pode tornar o estudo complexo e extenso, então optou-se por não considerar esse tipo de perturbação.

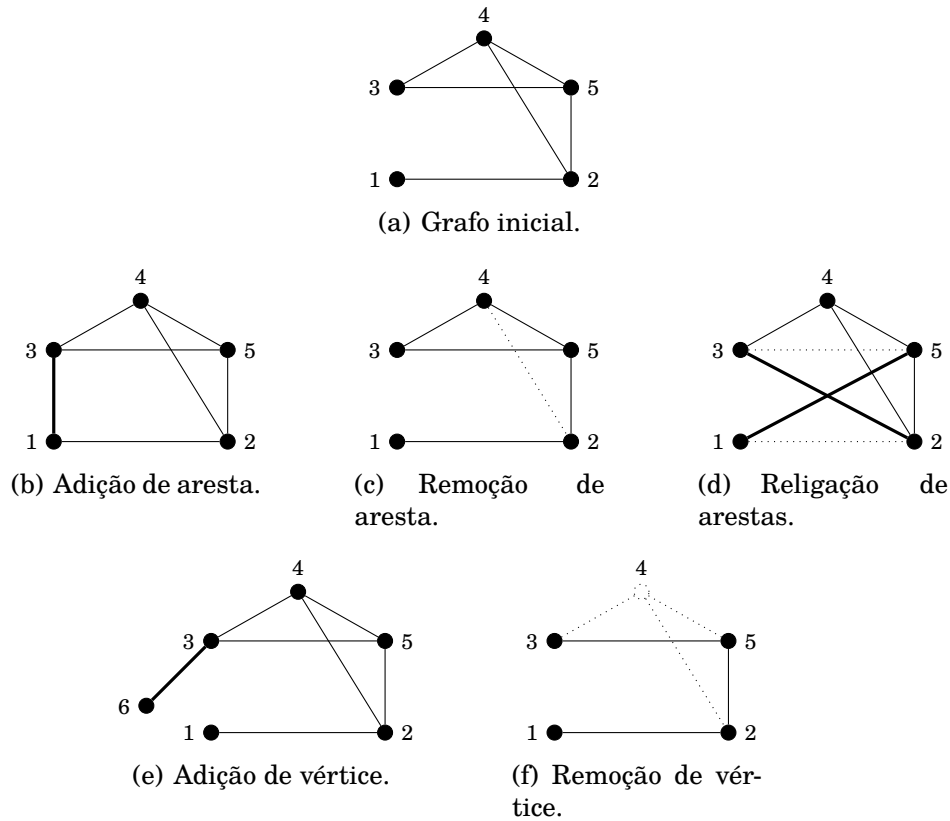
A adição de nós não é considerada uma perturbação

As Figuras 3.2(b) a 3.2(f) ilustram essas perturbações aplicadas ao grafo  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  da Figura 3.2(a), em que  $\mathcal{V} = \{1, 2, 3, 4, 5\}$  e  $\mathcal{E} = \{(1, 2), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)\}$ .

As perturbações de arestas aplicadas a  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  geram um novo grafo  $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ , i.e., apenas o conjunto de vértices é preservado. Na Figura 3.2(b), a aresta  $e = (1, 3)$  foi adicionada a  $\mathcal{G}$ , obtendo-se  $\mathcal{E}' = \mathcal{E} \cup (1, 3)$ ; na Figura 3.2(c), a aresta  $e = (2, 4)$  foi removida de  $\mathcal{G}$ , obtendo-se  $\mathcal{E}' = \mathcal{E} \setminus (2, 4)$ ; e na Figura 3.2(d), as arestas  $e = (1, 2)$  e  $e = (3, 5)$  foram religadas, resultando em  $\mathcal{E}' = \mathcal{E} \setminus ((1, 2), (3, 5)) \cup ((1, 5), (2, 3))$ .

As perturbações de vértices aplicadas a  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  modificam tanto o conjunto de vértices quanto o de arestas resultando em um grafo  $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ . Na Figura 3.2(e), o vértice  $v_6$  foi adicionado, resultando em  $\mathcal{V}' = \mathcal{V} \cup v_6$  e  $\mathcal{E}' = \mathcal{E} \cup (6, 3)$ ; na Figura 3.2(f), o vértice  $v_4$  foi removido, resultando em  $\mathcal{V}' = \mathcal{V} \setminus v_4$  e  $\mathcal{E}' = \mathcal{E} \setminus ((2, 4), (3, 4), (4, 5))$ .

Após as perturbações, as quantidades amostrais (medidas) são avaliadas. As medidas podem ser representadas por um escalar (p.ex.: comprimento médio dos caminhos mínimos), ou um vetor de valores (p.ex.: grau do vértice). Essas duas classes são avaliadas de formas diferentes. As medidas escalares são comparadas usando-se a diferença e as demais através de quantificadores estocásticos.



**Figura 3.2.** Exemplos de perturbações aplicadas ao grafo inicial  $\mathcal{G}$  com  $N = 5$  vértices e  $M = 6$  arestas.

### 3.3 Arcabouço estatístico

Esta seção apresenta uma breve introdução dos conceitos estatísticos utilizados na análise de medidas: os histogramas de proporções e a simulação de Monte Carlo.

Neste trabalho, os histogramas de proporções são utilizados para estimar a distribuição de probabilidade das medidas de redes complexas. A ideia é representar a densidade dos dados através da contagem do número de observações em uma sequência de intervalos consecutivos (classes ou *bins*), i.e., representar a proporção da amostra com um determinado valor. De acordo com Cha & Srihari [2002], pode-se definir formalmente um histograma como a seguir: seja  $X = \{x_0, x_1, \dots, x_{b-1}\}$  o conjunto de classes do histograma e  $A = \{a_1, a_2, \dots, a_n\}$  o conjunto de  $n$  observações, sendo que todas contêm as quantidades de  $x$ , em que  $a_j \in X$ . O histograma do conjunto  $A$  para a quantidade  $x$ , denotado por  $\mathcal{H}(x, A) = \mathcal{H}(A)^\dagger$ , é uma lista ordenada do número de ocorrências dos valores dis-

<sup>†</sup>Neste trabalho os histogramas são utilizados para uma mesma medida  $x$ , dessa forma,  $\mathcal{H}(A)$

cretos de  $x$  dentre os valores  $a_i$ . Se  $\mathcal{H}_i(A)$ ,  $0 \leq i \leq b-1$ , representa o número de elementos de  $A$  que tem valor  $x_i$ , então  $\mathcal{H} = [\mathcal{H}_0(A), \mathcal{H}_1(A), \dots, \mathcal{H}_{b-1}(A)]$ , em que:

$$\mathcal{H}_i(A) = \sum_{j=1}^n c_{ij}, \text{ em que } c_{ij} = \begin{cases} 1 & \text{se } a_j = x_i \\ 0 & \text{c.c.} \end{cases} \quad (3.1)$$

Se  $Pr_i(A)$  denota uma probabilidade então  $Pr_i(A) = \mathcal{H}_i(A)/n$ .

A simulação de Monte Carlo é um método empírico, baseado na realização de múltiplas avaliações sob o conjunto de entradas aleatórias  $\mathcal{T}_k = \{t_1, t_2, t_3, \dots, t_T\}$  (Figura 3.1). Os resultados destas avaliações são utilizados tanto para determinar a incerteza na previsão do modelo utilizado, quanto para o cálculo da contribuição de cada fator de entrada no comportamento das saídas [Saltelli et al., 2008]. Neste contexto, a ideia básica é utilizar amostras aleatórias para explorar e avaliar o comportamento das medidas em redes complexas. De uma forma geral, no experimento de Monte Carlo realizado aqui, são geradas replicações (pseudo-)aleatórias dos diferentes modelos de redes complexas em questão e o valor esperado e a variância da variável aleatória  $Y$  são estimados, respectivamente por:

$$\hat{E}(Y) = \hat{\mu} = \frac{1}{N} \sum_{i=1}^N y_i, \quad (3.2)$$

$$\hat{V}(Y) = \hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N [y_i - \hat{E}(Y)]^2, \quad (3.3)$$

A seção que segue apresenta a metodologia utilizada para análise das medidas com a aplicação dos conceitos definidos anteriormente.

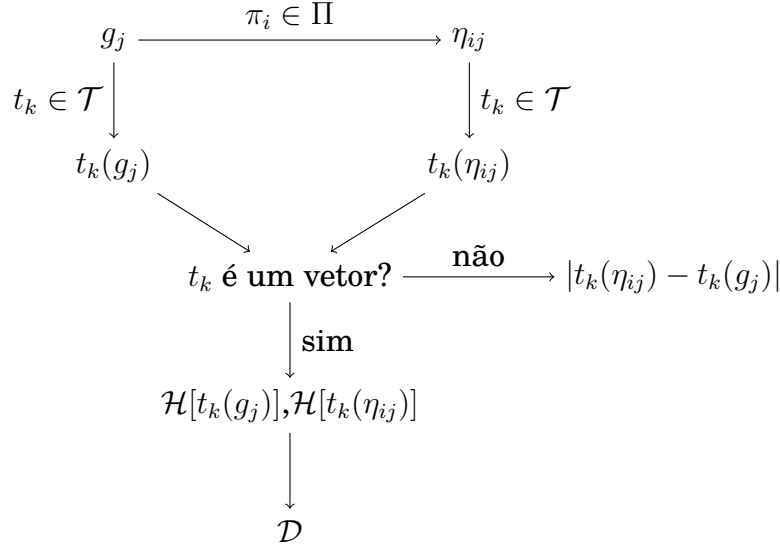
### 3.4 Metodologia para a análise de medidas

Os passos necessários para a avaliação de medidas são ilustrados na Figura 3.3 e descritos a seguir:

Dada uma rede inicial  $g_j$ , calcula-se a medida  $t_k$  obtendo-se  $t_k(g_j)$ . A perturbação  $\pi_i$  é aplicada a  $g_j$  gerando a rede  $\eta_{ij} = \pi_i(g_j)$ . A medida  $t_k$  é calculada para a rede  $\eta_{ij}$  obtendo-se  $t_k(\eta_{ij})$ . Se  $t_k$  é um escalar, a diferença entre  $t_k(g_j)$  e  $t_k(\eta_{ij})$  é

---

pode ser usado no lugar de  $\mathcal{H}(x, A)$ .



**Figura 3.3.** Etapas na análise da variabilidade de medidas.

computada; se  $t_k$  é um vetor, os histogramas de proporções  $\mathcal{H}[t_k(g_j)]$  e  $\mathcal{H}[t_k(\eta_{ij})]$  são obtidos a partir de  $t_k(g_j)$  e  $t_k(\eta_{ij})$ , respectivamente. Estes são empregados como estimativas da densidade de probabilidade de cada medida.  $\mathcal{H}[t_k(g_j)]$  e  $\mathcal{H}[t_k(\eta_{ij})]$  são comparadas com os quantificadores estocásticos: divergência de Kullback-Leibler ( $\mathcal{D}_{KL}$ ), distância de Jensen-Shannon ( $\mathcal{D}_{JS}$ ) e distância de Hellinger ( $\mathcal{D}_H$ ) (definidos na seção 2.4).

O Algoritmo 1 apresenta a simulação de Monte Carlo utilizada neste trabalho para a análise da variabilidade de medidas. O experimento obtém um estimador para o quantificador utilizado para comparação dos diferentes estados de uma rede.

As entradas do algoritmo são: uma classe de grafos  $\mathcal{G} = \{g_1, g_2, \dots, g_j, \dots, g_G\}$ , o vetor de medidas  $\mathcal{T} = \{t_1, t_2, \dots, t_k, \dots, t_T\}$ , o conjunto de perturbações  $\Pi = \{\pi_1, \pi_2, \dots, \pi_i, \dots, \pi_P\}$  e o número máximo de perturbações ( $r_{\max}$ ) aplicadas a cada elemento de  $\mathcal{G}$ . A saída é a média ( $\hat{\mu}$ ) e variância ( $\hat{\sigma}^2$ ) amostrais da diferença ou distâncias estocásticas ( $\mathcal{D}_{KL}(p, q)$ ,  $\mathcal{D}_{JS}(p, q)$  e  $\mathcal{D}_H(p, q)$ ) para cada elemento de  $\mathcal{T}$ .

O algoritmo começa com uma estrutura de repetição que percorre o vetor de medidas  $\mathcal{T}$ , na linha 2. Na linha 3, a estrutura de repetição determina o número de grafos (ou redes) de uma classe  $\mathcal{G}$  que serão gerados para análise. Na linha 4, a medida  $t_k$  é calculada para o grafo  $g_j$ . As diferentes perturbações  $\pi_i$  aplicadas a  $g_j$  são controladas na estrutura de repetição da linha 5. A estrutura de repetição que controla o número de replicações por grafo  $g_j$  para cada tipo de perturbação

**Algoritmo 1:** Simulação de Monte Carlo.

---

**Entrada:**  $\mathcal{G} = \{g_1, g_2, \dots, g_j, \dots, g_G\}$ ,  $\mathcal{T} = \{t_1, t_2, \dots, t_k, \dots, t_T\}$ ,  
 $\Pi = \{\pi_1, \pi_2, \dots, \pi_i, \dots, \pi_P\}$  e  $r_{\max}$ .  
**Saída:**  $\mu$  e  $\sigma^2$  para  $\mathcal{D}_*$  ( $\mathcal{D}_{KL}$ ,  $\mathcal{D}_{JS}$  e  $\mathcal{D}_H$ ).

```

1 inicio
2   for  $1 \leq t_k \leq t_T$  do
3     for  $1 \leq g_j \leq g_G$  do
4       Calcule  $t_k(g_j)$ ;
5       for  $1 \leq \pi_i \leq \pi_P$  do
6         for  $1 \leq r \leq r_{\max}$  do
7           Gere a rede  $\eta_{ij}^r = [\pi_i(g_j)]^r$ ;
8           Calcule  $t_k(\eta_{ij}^r)$ ;
9           if  $t_k$  é um vetor then
10            Calcule  $p = \mathcal{H}[t_k(g_j)]$  e  $q = \mathcal{H}[t_k(\eta_{ij}^r)]$ ;
11            Calcule  $\mathcal{D}_*^r(p, q)$ ;
12          else
13            Calcule  $|t_k(\eta_{ij}^r) - t_k(g_j)|$ ;
14          end
15        end
16      end
17    end
18  end
19  Calcule  $\mu$  e  $\sigma^2$  para a diferença ou  $\{\mathcal{D}_*^1, \mathcal{D}_*^2, \dots, \mathcal{D}_*^{r_{\max}}\}$ ;
20 fin

```

---

$\pi_i$  é iniciada na linha 6, tendo número máximo  $r_{\max}$ . Assim, na linha 7, a perturbação  $\pi_i$  é aplicada ao grafo  $g_j$  gerando o grafo  $\eta_{ij}^r = [\pi_i(g_j)]^r$  na replicação  $r$ ; e na linha 8 a medida  $t_k$  é calculada para o grafo perturbado  $\eta_{ij}^r$ . Na linha 9, a condição verifica se a medida  $t_k$  é um vetor de valores; se satisfeita, os histogramas de proporções são obtidos para a medida  $t_k$  da rede inicial  $g_j$  e sua respectiva perturbação  $\eta_{ij}^r$  (linha 10). Esses valores são as distribuições de probabilidades  $p$  e  $q$  necessárias ao cálculo das distâncias, denotados por  $\mathcal{H}[t_k(g_j)] = p$  e  $\mathcal{H}[t_k(\eta_{ij}^r)] = q$ . Seguindo, na linha 11 as distâncias estocásticas  $\mathcal{D}_{KL}(p, q)$ ,  $\mathcal{D}_{JS}(p, q)$  e  $\mathcal{D}_H(p, q)$  são calculadas. Caso a condição não seja satisfeita, tem-se um medida global representada por um escalar; então a diferença absoluta entre  $t_k(\eta_{ij}^r)$  e  $t_k(g_j)$  é calculada na linha 13. Por fim, na linha 19 calculam-se a média ( $\mu$ ) e a variância ( $\sigma^2$ ) do resultado final, que pode ser a diferença ou as distâncias estocásticas ( $\mathcal{D}_{KL}(p, q)$ ,  $\mathcal{D}_{JS}(p, q)$  e  $\mathcal{D}_H(p, q)$ ) para cada elemento de  $\mathcal{T}$ .

Em suma, para cada medida avaliada são gerados  $G$  grafos distintos de uma determinada classe. Para cada um deles,  $P$  tipos de perturbações foram

aplicados diversas vezes ( $r_{\max}$ ). A distribuição das medidas é estimada por meio de um histograma de proporções, e os quantificadores estocásticos são calculados para computar a variação das distribuições.

### 3.5 Conclusões

Este capítulo apresentou uma revisão da literatura acerca do problema de análise da variabilidade de medidas de redes complexas utilizando quantificadores estocásticos e a metodologia usada para a análise de medidas. A seção 3.2 apresentou a definição formal do problema, incluindo as perturbações aplicadas às redes, que foram: adição de arestas, remoção de arestas, religação de arestas, adição de vértices e remoção de vértices. Além disso, apresentou-se a metodologia aplicada à análise da variabilidade de medidas e a estrutura do algoritmo de simulação utilizado. Partindo dos conceitos apresentados aqui, com o objetivo de validar a metodologia apresentada, a seção 4.1 do capítulo 4 apresenta a comparação dos resultados encontrados neste trabalho com resultados conhecidos na literatura. Além disso, o capítulo discute os resultados da análise obtidos para os modelos teóricos (seção 4.2) e para as redes de sensores sem fio (seção 4.3).

# Capítulo 4

## Resultados

Este capítulo apresenta a avaliação do comportamento das medidas ( $\mathcal{T}$ ) aplicadas a uma classe de grafos ( $\mathcal{G}$ ) submetidas a um conjunto de perturbações ( $\Pi$ ).

As medidas avaliadas foram: (i) o grau do vértice ( $\kappa_i$ ), (ii) coeficiente de agrupamento local ( $C_i$ ), (iii) *betweenness centrality* ( $B_i$ ), (iv) comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ) e (v) comprimento médio dos caminhos mínimos ( $L$ ) aplicadas às classes de grafos: (i) aleatória, (ii) mundo pequeno e (iii) livre de escala. A comparação entre os estados da rede foi realizada através dos quantificadores estocásticos: (i) distância de Hellinger, (ii) distância de Jensen-Shannon e (iii) divergência de Kullback-Leibler.

Além disso, realizou-se uma caracterização do mecanismo de propagação de dados em redes de sensores sem fio como uma rede complexa e a análise da medida do comprimento médio dos caminhos.

Este capítulo está organizado como segue:

- A seção 4.1 mostra a comparação dos resultados encontrados neste trabalho com resultados encontrados no trabalho de Boas et al. [2010], com o objetivo de validar a metodologia descrita no capítulo 3.
- A seção 4.2 apresenta os resultados da análise nos modelos teóricos. A seção 4.2.1 apresenta a descrição dos parâmetros e cenários de simulação para análise dos modelos teóricos. Os resultados dessa avaliação são apresentados em duas seções: a seção 4.2.2 apresenta os resultados analisando o comportamento de cada medida aplicada a cada modelo, bem como a comparação dos quantificadores; e a seção 4.2.3 apresenta os resultados comparando-se o comportamento de cada modelo em relação às perturbações aplicadas usando-se a distância de Hellinger ( $\mathcal{D}_H$ ), dessa forma pode-



se comparar a sensibilidade entre os diferentes modelos avaliados. A seção 4.2.4 apresenta testes de hipótese realizados com a motivação de identificar que perturbações e seus níveis levam à quebra da distribuição dos graus de redes complexas nos modelos teóricos. A seção 4.2.5 apresenta uma discussão geral acerca da análise dos resultados para os modelos teóricos.

- A seção 4.3 discorre sobre a propagação de dados em redes de sensores sem fio caracterizada como uma rede complexa, bem como a avaliação do comprimento dos caminhos mínimos frente às perturbações. Nesta seção são mostrados os conceitos básicos a respeito dessas redes, como também alguns trabalhos relacionados (seção 4.3.1). A seção 4.3.2 apresenta as diferentes topologias de propagação de dados e por fim, a seção 4.3.3 apresenta o estudo da análise de medidas.
- Por fim, a seção 4.4 apresenta as conclusões do capítulo.

## 4.1 Validação da metodologia

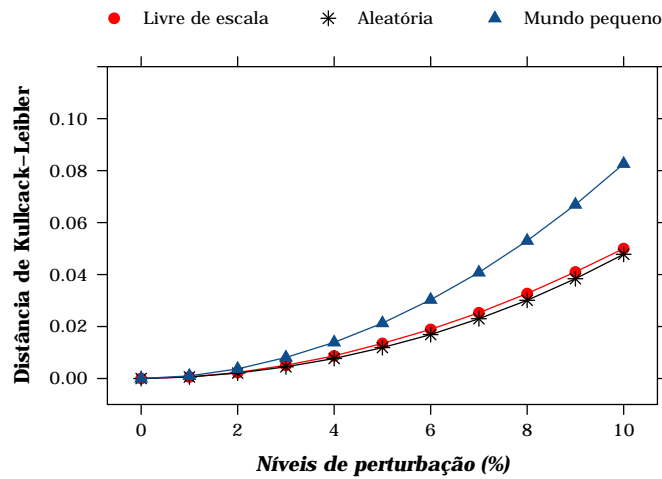
Esta seção tem como objetivo validar a metodologia descrita no capítulo 3. Como citado anteriormente, Boas et al. [2010] usaram esta metodologia anteriormente para estudar o problema de amostragem em redes complexas utilizando a divergência de Kullback-Leibler frente a perturbações de arestas (adição, remoção e religação de arestas).

Os autores avaliaram nove modelos de redes complexas, dentre os quais seis são teóricos e três são redes reais. Os modelos teóricos foram as redes aleatórias de Erdős Rényi, redes de mundo pequeno de Watts e Strogatz, redes livre de escala de Barabási e Albert, o modelo geográfico de Waxman, o modelo livre de escala limitado e o modelo de ligação preferencial não linear, todos com 10000 vértices e grau médio igual a 6; as redes reais foram a rede de distribuição de energia dos EUA, a rede de colaboração *high-energy*, a rede de palavras de um livro Japonês.

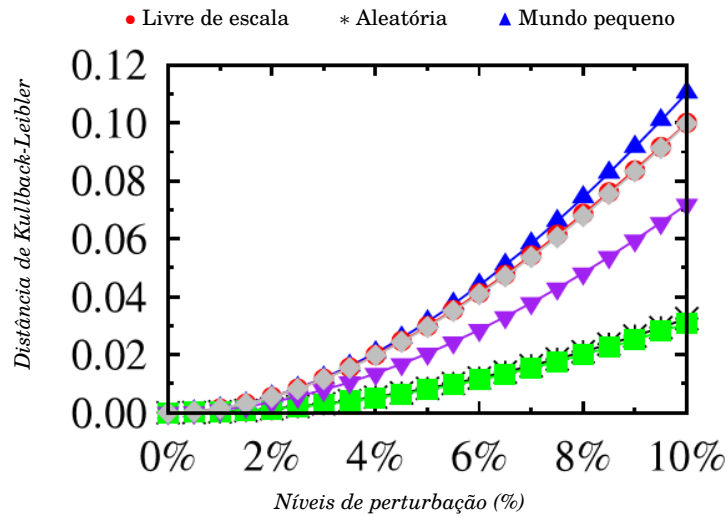
Neste trabalho, além de estudar o comportamento de diferentes modelos de rede, foram avaliados diferentes quantificadores estocásticos e perturbações de vértices. Para fins de comparação e validação da metodologia considerou-se:

- Redes com 10000 vértices e grau médio 6.

- Os modelos teóricos de Erdős Rényi, redes de mundo pequeno de Watts e Strogatz e redes livre de escala de Barabási e Albert.
- O quantificador de Kullback-Leibler.
- O grau do vértice ( $\kappa_i$ ) sujeita a adição de arestas e o comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ) sujeito a remoção de arestas.



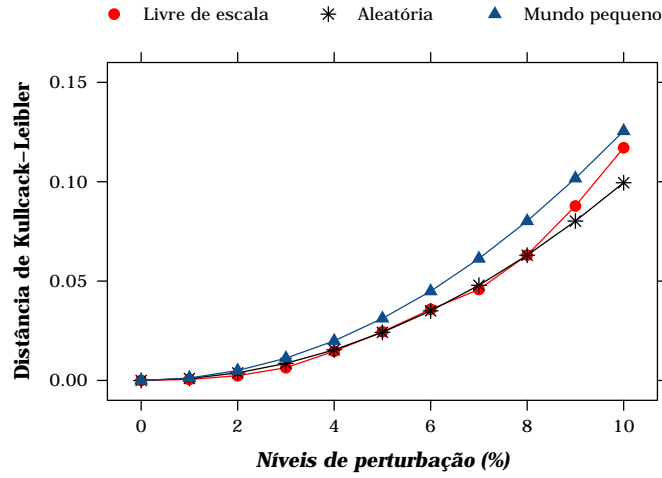
(a) Resultados encontrados neste trabalho



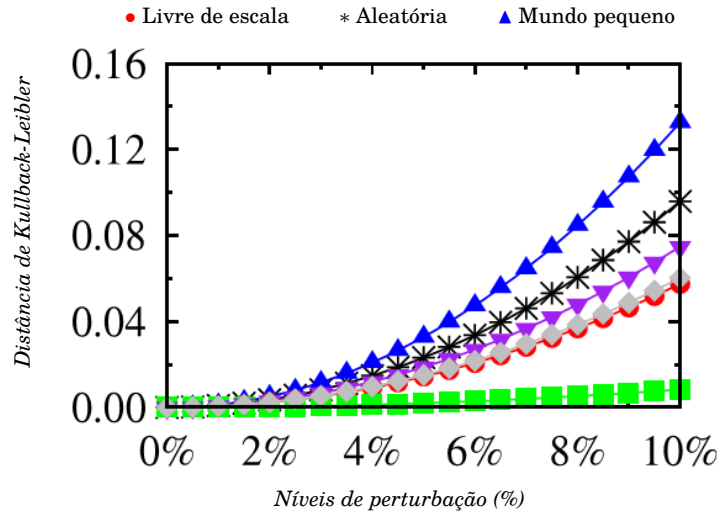
(b) Resultados retirados de Boas et al. [2010]

**Figura 4.1.** Variação da divergência de Kullback-Leibler ( $D_{KL}$ ) para o grau do vértice ( $\kappa_i$ ) aplicado às redes livres de escala, aleatórias e de mundo pequeno sujeitas a perturbação do tipo adição de arestas ( $\pi_{ae}$ ).

A Figura 4.1 apresenta os resultados para o grau do vértice ( $\kappa_i$ ) e a Figura 4.2 para o comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ). As Figuras 4.1(a)



(a) Resultados encontrados neste trabalho



(b) Resultados retirados de Boas et al. [2010]

**Figura 4.2.** Variação da divergência de Kullback-Leibler ( $D_{KL}$ ) para o comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ) aplicado às redes livres de escala, aleatórias e de mundo pequeno sujeitas a perturbação do tipo remoção de arestas ( $\pi_{re}$ ).

e 4.2(a) ilustram os resultados encontrados neste trabalho para os três modelos de rede avaliados e as Figuras 4.1(b) e 4.2(b) apresentam os resultados encontrados por Boas et al. [2010]. Observa-se que, nos casos avaliados, as perturbações inseridas incrementam a divergência de Kullback-Leibler, i.e., seus valores aumentam progressivamente quando o nível de perturbação aumenta.

Vale ressaltar que os grafos de entrada utilizados no presente trabalho não foram os mesmos utilizados por Boas et al., pois os mesmos não foram dispo-

nibilizados pelos autores. Desse modo não foi possível sincronizar os gráficos no eixo das ordenadas. Porém como ambos resultados têm o mesmo comportamento, pode-se dizer, baseando-se na inspeção visual dos gráficos, que a metodologia foi aplicada corretamente neste trabalho.

## 4.2 Análise dos modelos teóricos

Esta seção apresenta a discussão dos resultados para os modelos teóricos. O estudo apresentado aqui mostra a aplicação da metodologia descrita anteriormente estando limitado ao seguinte escopo:

- As medidas avaliadas foram o grau do vértice ( $\kappa_i$ ), o coeficiente de agrupamento local ( $C_i$ ), *betweenness centrality* ( $B_i$ ), o comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ) e o comprimento médio dos caminhos mínimos ( $L$ ), assim  $\mathcal{T} = \{\kappa_i, C_i, B_i, \ell_{(i,j)}, L\}$  (definidas na seção 2.2);
- Foram examinadas as classes de redes aleatórias ( $\mathcal{G}_{er}$ ), de mundo pequeno ( $\mathcal{G}_{ws}$ ) e livre de escala ( $\mathcal{G}_{ba}$ ) (definidas na seção 2.3).
- As perturbações aplicadas foram a adição, remoção e religação de arestas e remoção de nós, assim  $\Pi = \{\pi_{ae}, \pi_{re}, \pi_{rw}, \pi_{rv}\}$  (definidas na seção 3.2).

O comportamento das medidas  $\mathcal{T}$  foi avaliado de duas formas diferentes: fora utilizados quantificadores estocásticos  $\mathcal{D}_{KL}$ ,  $\mathcal{D}_{JS}$  e  $\mathcal{D}_H$  para vetores ( $\kappa_i$ ,  $C_i$ ,  $B_i$  e  $\ell_{(i,j)}$ ) e a diferença absoluta para a medida escalar ( $L$ ).

### 4.2.1 Parâmetros de simulação e condições iniciais

As medidas do conjunto  $\mathcal{T}$  foram avaliadas nos seguintes cenários:

**Parâmetros para geração das redes:** foram consideradas redes com  $N = \{1000, 5000, 10000\}$  vértices e grau médio  $c = 6$ . As probabilidades de conexão  $p_c$  para geração das redes aleatórias foram 0.06, 0.012 e 0.0006 para 1000, 5000 e 10000 vértices, respectivamente. As redes de mundo pequeno foram geradas com  $K = 3$  vizinhos e uma probabilidade de religação  $p_r = 0.3$ . As redes livres de escala foram geradas com uma aresta adicionada a cada passo da construção da rede, i.e.,  $m = 1$ .

**Perturbações:** foram executadas em intensidades de  $\{1\%, 2\%, 3\%, \dots, 10\%\}$  em relação ao total de arestas e vértices dos grafos.

**Histogramas:** os histogramas de proporções  $\mathcal{H}(\cdot)$  foram obtidos com 200 classes de mesma largura. Como foi mencionado na seção 2.4, a divergência de Kullback-Leibler ( $\mathcal{D}_{KL}$ ) diverge para  $q(\xi) = 0$  e  $p(\xi) \neq 0$ , então deve-se adicionar um parâmetro  $\delta$  a cada classe do histograma. De acordo com Cabella et al. [2008] o valor de  $\delta$  deve ser 0.001. Já para o cálculo das distâncias de Jensen-Shannon ( $\mathcal{D}_{JS}$ ) e Hellinger ( $\mathcal{D}_H$ ) é utilizado o histograma original.

**Parâmetros de simulação:** para cada classe de rede considerada foram geradas 10 amostras diferentes e cada uma delas foi perturbada 20 vezes, assim tem-se  $G = 10$  e  $r_{\max} = 20$ . Cada replicação refere-se ao estado da rede  $g_j$  após a perturbação  $\pi_i$ , denotada por  $\eta_{ij}$ . Dessa forma, os resultados puderam ser apresentados com intervalo de confiança assintótico e simétrico com nível de significância de 95%.

**Recursos computacionais:** o algoritmo foi implementado na linguagem R [R Development Core Team, 2011] com o uso da biblioteca *igraph*\* para manipulação dos grafos [Csardi & Nepusz, 2006]. As simulações foram executadas em um computador Intel(R) Core(TM) i5 CPU 760 2.80 GHz com 8 GB RAM e sistema operacional Ubuntu 12.04 (64 bits).

#### 4.2.2 Análise dos modelos teóricos I

Esta seção apresenta os resultados da análise do comportamento de cada medida aplicada a cada modelo teórico e a comparação dos quantificadores. Os resultados foram sumarizados e apresentados em quinze figuras mostradas na seguinte sequência: as Figuras 4.3 a 4.5 são os resultados para grau do vértice  $\kappa_i$ , as Figuras 4.6 a 4.8 os resultados para coeficiente de agrupamento local  $C_i$ , as Figuras 4.9 a 4.11 os resultados para *betweenness centrality*, as Figuras 4.12 a 4.14 os resultados para comprimento médio dos caminhos mínimos  $\ell_{(i,j)}$  e as Figuras 4.16 a 4.18 os resultados para comprimento médio dos caminhos mínimos  $L$ . Cada figura apresenta o resultado para uma medida  $t_k$  aplicada a uma classe de rede complexa  $\mathcal{G}$ .

Cada uma dessas figuras é um conjunto de doze gráficos, ou uma grade de gráficos de tamanho  $3 \times 4$ , totalizando 180 gráficos. As linhas da grade corres-

---

\*Igraph é um pacote de software livre para criação e manipulação de grafos. Inclui implementações de algoritmos clássicos da teoria dos grafos, como árvore geradora mínima e algoritmos de fluxo, e também implementa algoritmos recentes utilizados na análise de redes complexas, como detecção de comunidades. Além disso, contém funções para geração de grafos regulares, aleatórios, livres de escala, sendo possível também o cálculo de propriedades estruturais.

pondem aos três tamanhos de grafos estudados (de baixo para cima) 1000, 5000 e 10000 vértices; enquanto as colunas da grade descrevem os quatro tipos de perturbações aplicadas a estes modelos (da esquerda para a direita) adição, remoção e religação de arestas e remoção de vértices.

Cada gráfico do conjunto apresenta a variação dos quantificadores utilizados (eixo das ordenadas) em relação ao nível de perturbação aplicada à rede (eixo das abscissas). Para as medidas  $\kappa_i$ ,  $C_i$ ,  $B_i$  e  $\ell_{(i,j)}$  (vetores) apresenta-se a variação dos três quantificadores estocásticos: divergência de Kullback-Leibler  $\mathcal{D}_{KL}$  ( $\triangle$ ), distância de Hellinger  $\mathcal{D}_H$  ( $\square$ ) e distância de Jensen-Shannon  $\mathcal{D}_{JS}$  ( $\circ$ ). Para  $L$ , um escalar, tem-se a variação da diferença.

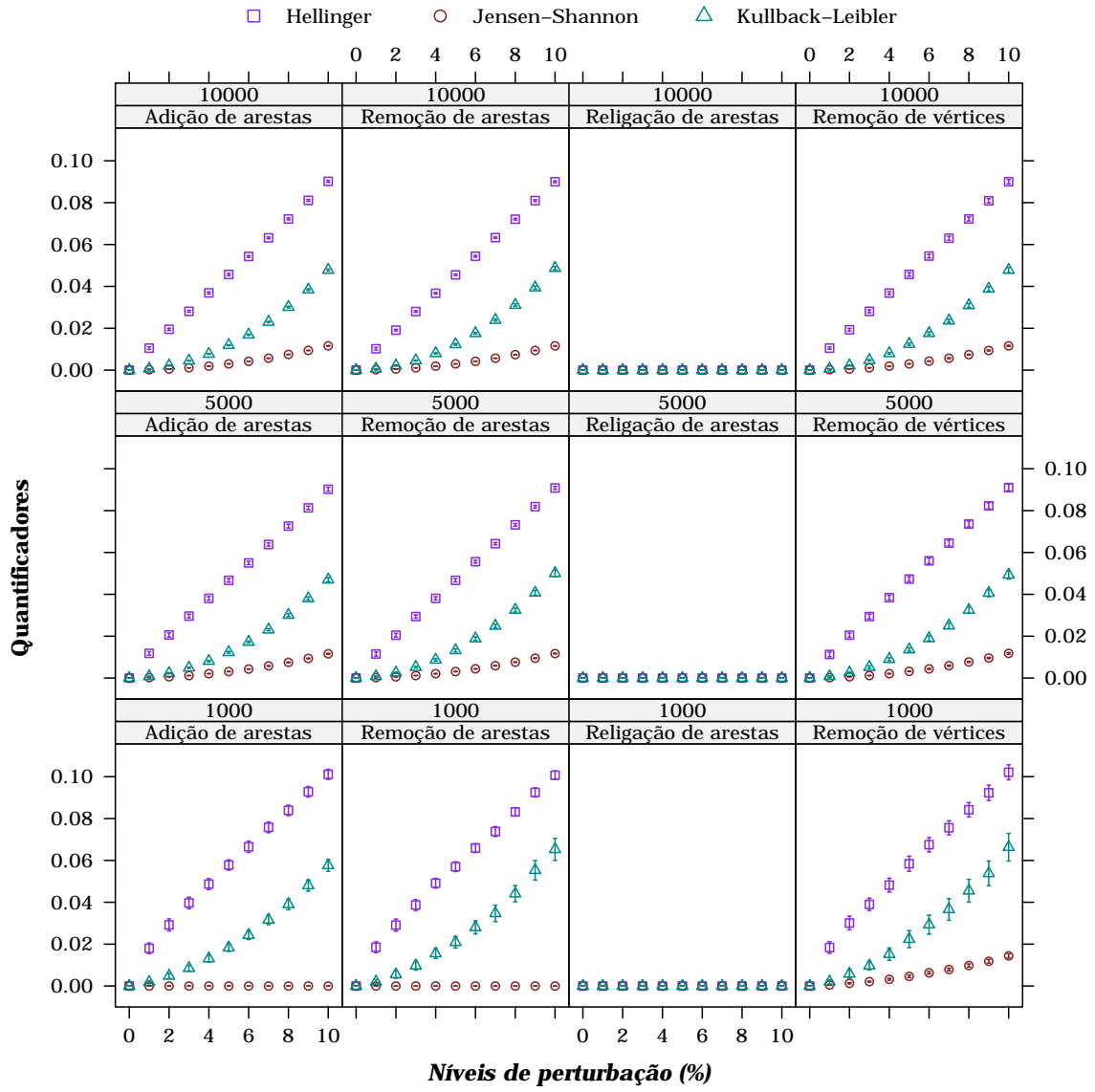
Os gráficos de todas as figuras deste capítulo estão sincronizados tanto no eixo das abscissas quanto das ordenadas, o que os torna comparáveis. Além disso apresentam o intervalo de confiança dos resultados (linhas verticais em cada ponto do gráfico).

A seguir, são apresentadas as análises sobre as diferentes perturbações e quantificadores utilizados no estudo do comportamento de cada medida de redes complexas.

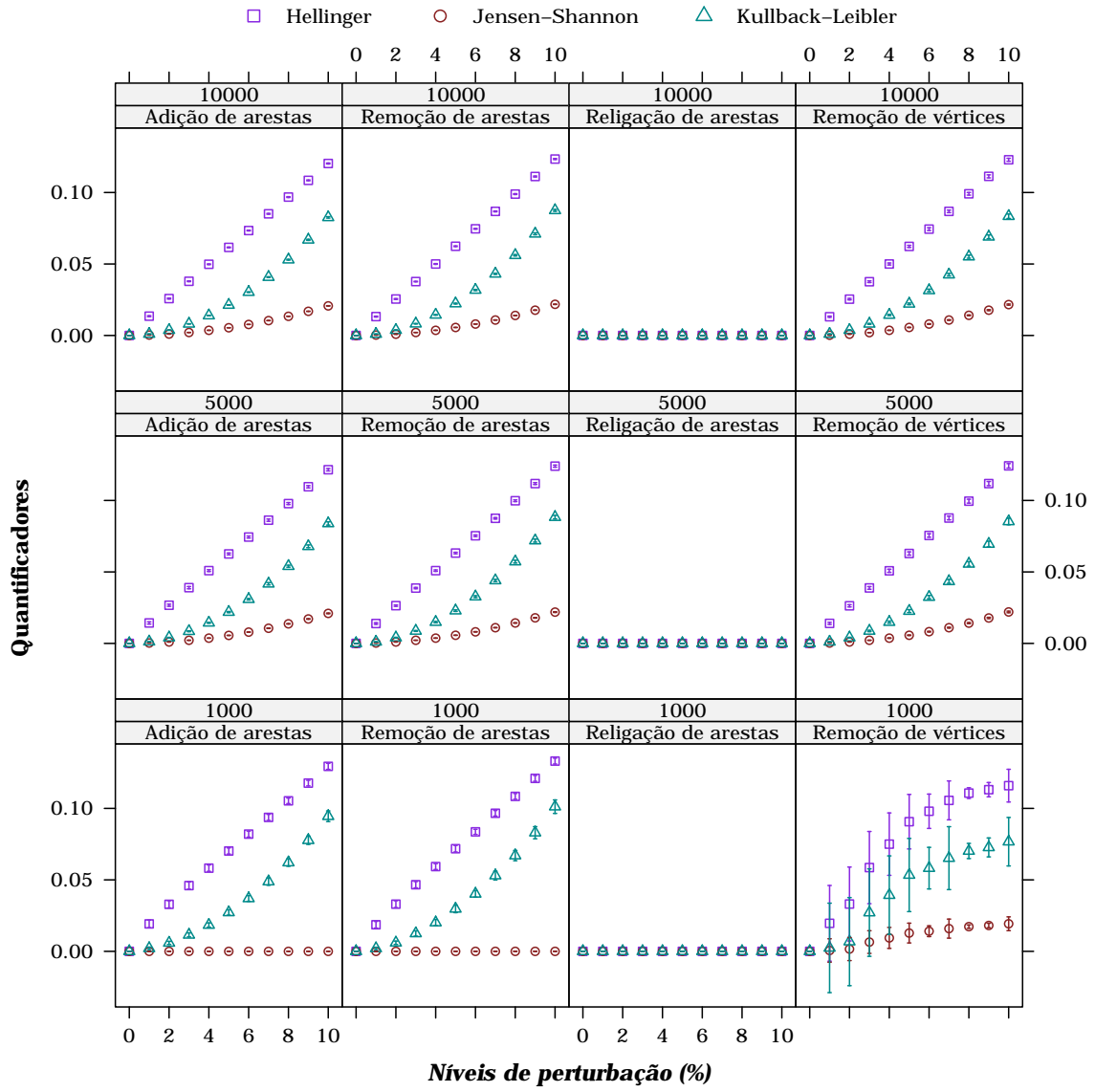
#### 4.2.2.1 Grau do vértice ( $\kappa$ )

As Figuras 4.3, 4.4 e 4.5 apresentam o comportamento dos quantificadores  $\mathcal{D}_{KL}$ ,  $\mathcal{D}_{JS}$  e  $\mathcal{D}_H$  para  $\kappa_i$  aplicado aos modelos  $\mathcal{G}_{er}$ ,  $\mathcal{G}_{ws}$  e  $\mathcal{G}_{ba}$ , respectivamente; e a Tabela 4.1 contém a média dos valores dos quantificadores no nível mais alto de perturbação (10%).

Essa medida exibe o mesmo comportamento em todos os modelos. A adição de arestas, remoção de arestas e remoção de vértices incrementam todos os quantificadores, i.e., seus valores aumentam progressivamente quando o nível de perturbação aumenta; e a religação de arestas não introduz mudanças significativas, como se vê na terceira coluna de cada grade. Observe que, os intervalos de confiança são maiores para a remoção de nós em rede de mundo pequeno (1000 vértices) e redes livres de escala (1000 e 5000 vértices), ainda assim observa-se um padrão no comportamento da curva.

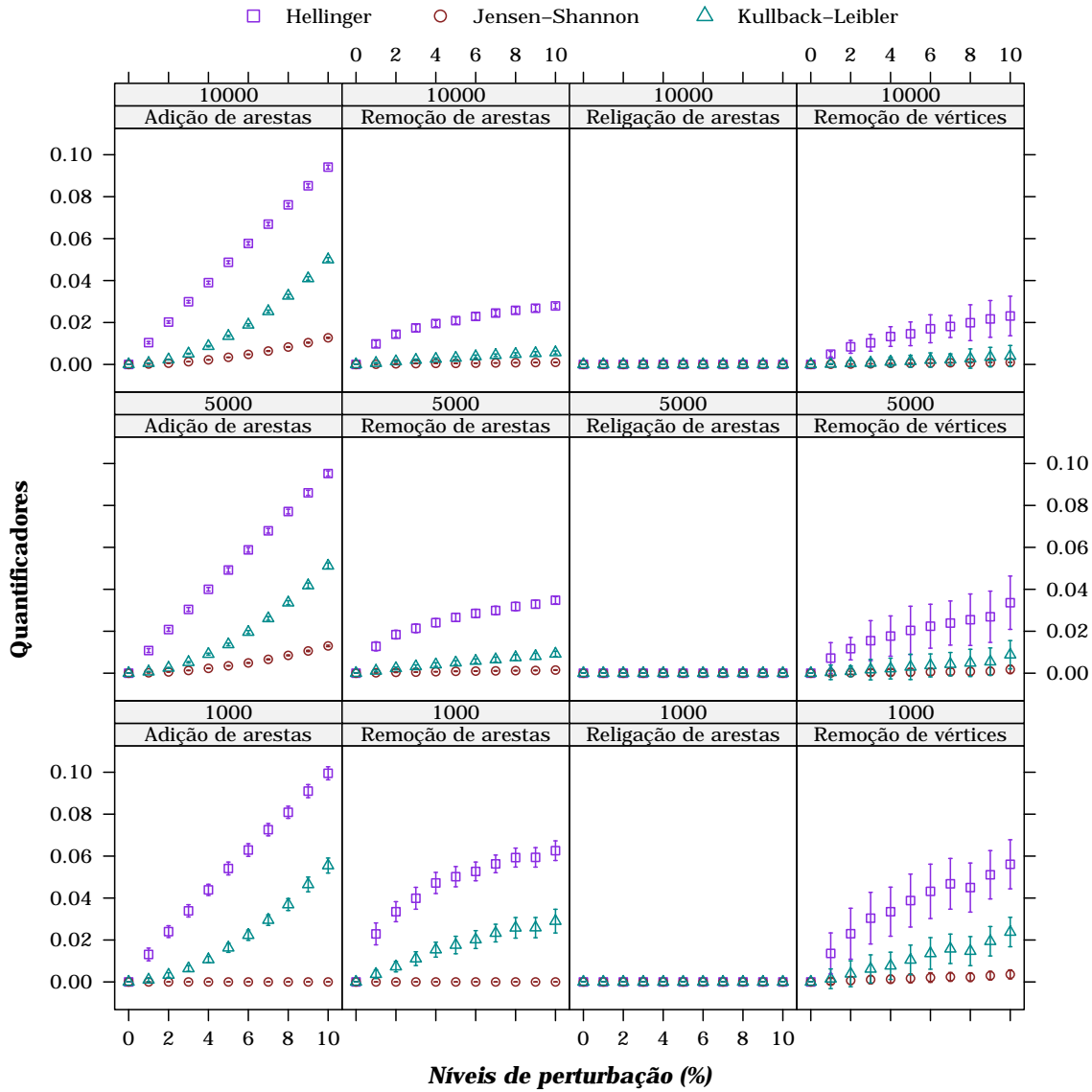


**Figura 4.3.** Variação dos quantificadores para o grau do vértice  $\kappa_i$  nas redes aleatórias.



**Figura 4.4.** Variação dos quantificadores para o grau do vértice  $\kappa_i$  nas redes de mundo pequeno.





**Figura 4.5.** Variação dos quantificadores para o grau do vértice  $\kappa_i$  nas redes livres de escala.

Como observa-se na Tabela 4.1, entre os quantificadores, a distância de Hellinger  $\mathcal{D}_H$  apresentou maior sensibilidade às perturbações que provocaram mudanças, apresentando seu maior valor (0.123) para a remoção de arestas no modelo de mundo pequeno. Em seguida vem a divergência de Kullback-Leibler  $\mathcal{D}_{KL}$  e a distância de Jensen-Shannon  $\mathcal{D}_{JS}$  (com menores variações). Considerando a adição e remoção de arestas, a distância de Jensen-Shannon não exibiu alteração perceptível para 1000 vértices e apenas um leve crescimento para 5000 e 10000

vértices.

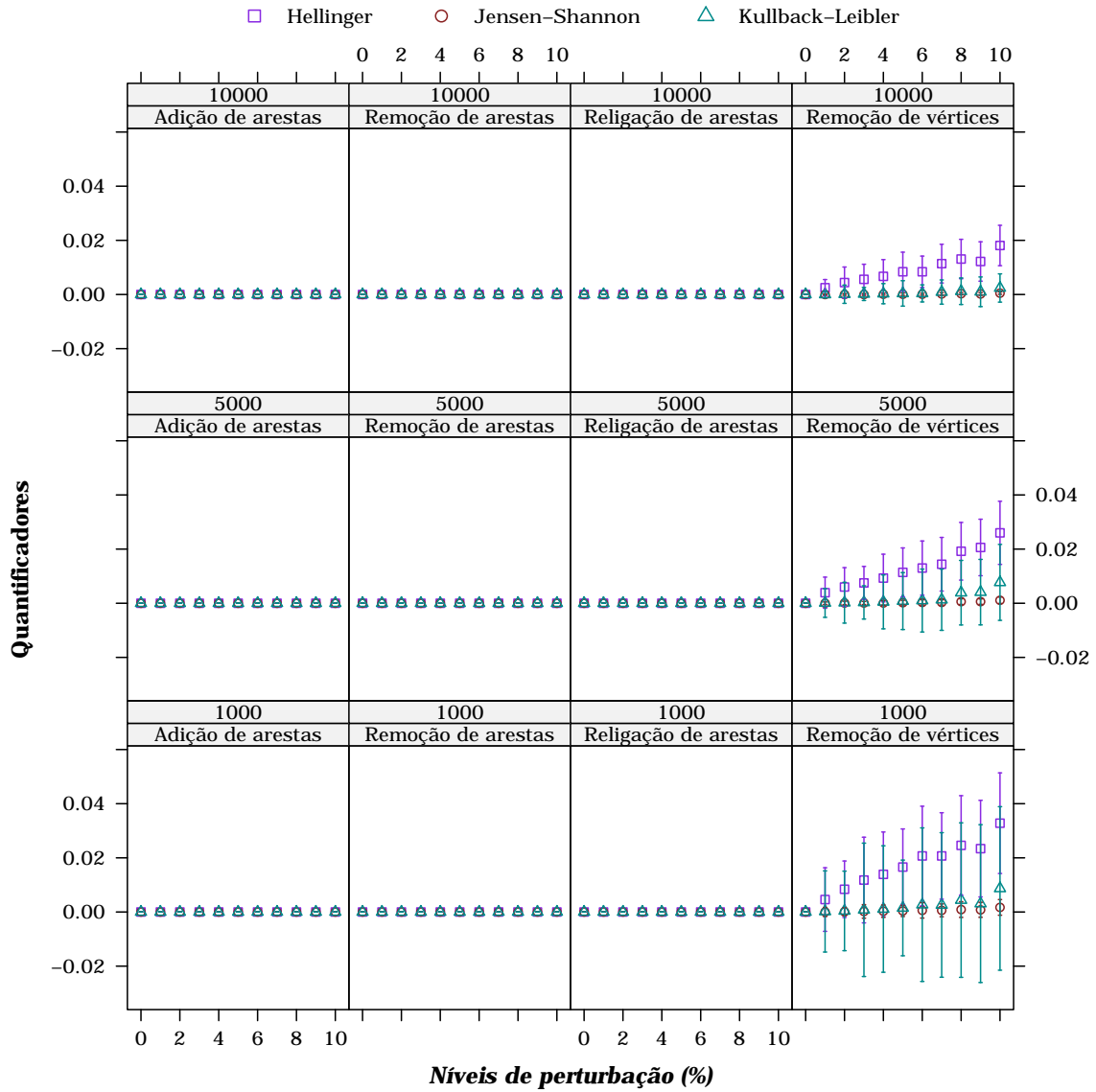
**Tabela 4.1.** Média dos valores dos quantificadores para o nível de perturbação mais intenso (10%) quando aplicado ao grau do vértice  $\kappa$ . O maior valor obtido está em negrito.

	$\mathcal{D}_{KL}$			$\mathcal{D}_{JS}$			$\mathcal{D}_H$		
	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$
$\mathcal{G}_{er}$	0.047	0.048	0	0.011	0.011	0	0.090	0.090	0
$\mathcal{G}_{ws}$	0.082	0.087	0	0.020	0.021	0	0.120	<b>0.123</b>	0
$\mathcal{G}_{ba}$	0.050	0.005	0	0.012	$10^{-3}$	0	0.094	0.027	0

#### 4.2.2.2 Coeficiente de agrupamento local ( $C_i$ )

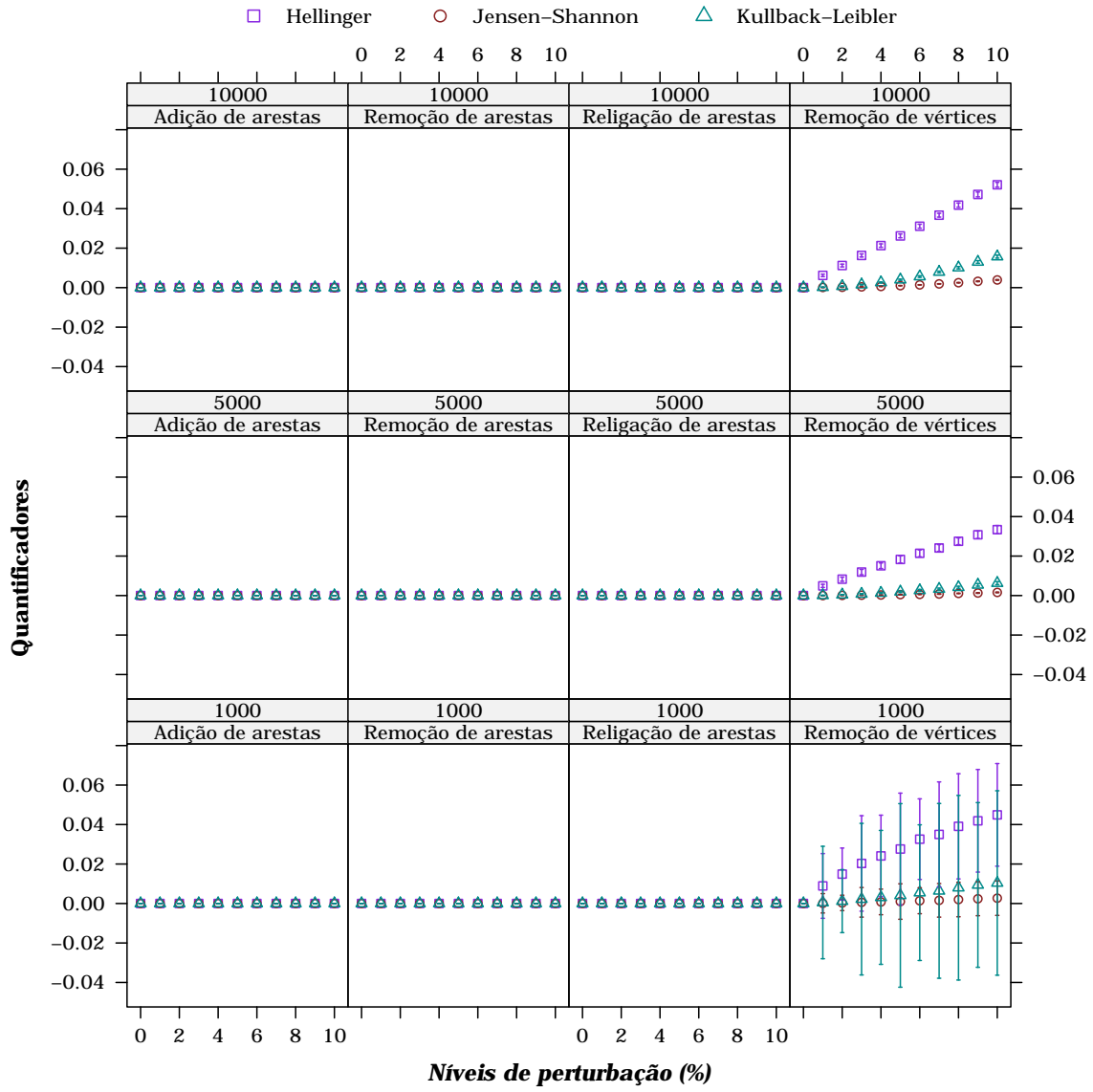
Como observa-se nas Figuras 4.6, 4.7 e 4.8, o coeficiente de agrupamento local  $C_i$  não apresentou variação para as perturbações de arestas para os modelos avaliados e para a remoção de nós nas redes livres de escala; pode-se dizer que essa medida é robusta ou insensível às perturbações aplicadas nesses casos. Como definido na seção 2.2.2,  $C_i$  é uma medida que quantifica o número de arestas que conectam os vizinhos do vértice  $v_i$  em  $\mathcal{G}$ , dessa forma a adição, remoção ou religação de arestas não alteram essa propriedade.

As redes de mundo pequeno e aleatórias são levemente sensíveis à remoção de vértices, sendo a distância de Hellinger a que apresenta maior variação: em média, 0.052 em redes de mundo pequeno e para o nível mais alto de perturbação (10%).

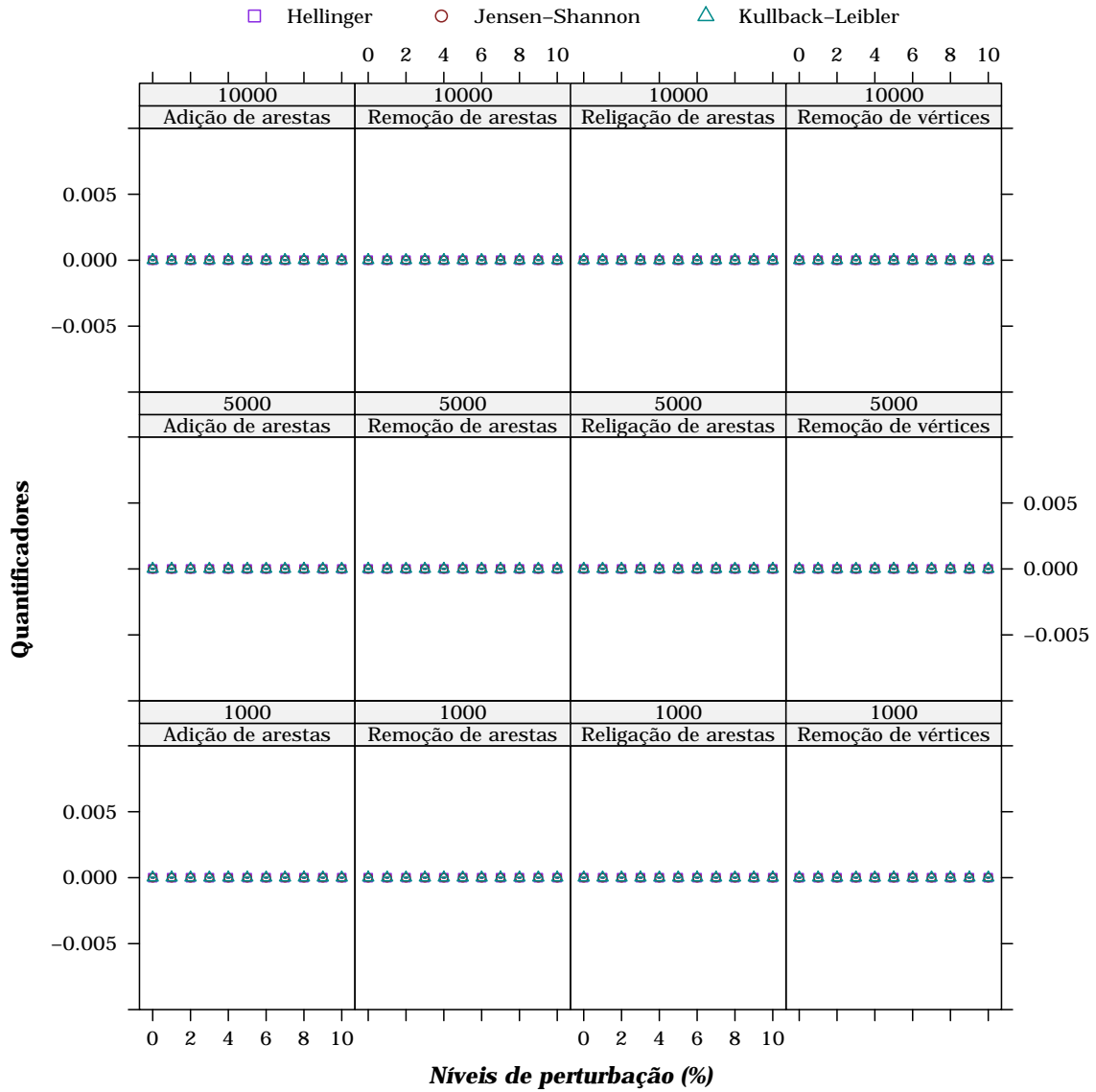


**Figura 4.6.** Variação dos quantificadores para o coeficiente de agrupamento local  $C_i$  em redes aleatórias.

Os intervalos de confiança são maiores para a remoção de vértices nas redes aleatórias (1000 e 5000 vértices) e redes de mundo pequeno (1000 vértices). Embora qualquer comparação de pares de valores subsequentes não leve à conclusão de que os valores mudam, o comportamento geral da média fornece informações suficientes para avaliar a sua dependência em relação a remoção de vértices.



**Figura 4.7.** Variação dos quantificadores para o coeficiente de agrupamento local  $C_i$  em redes de mundo pequeno.



**Figura 4.8.** Variação dos quantificadores para o coeficiente de agrupamento local  $C_i$  em redes livres de escala.

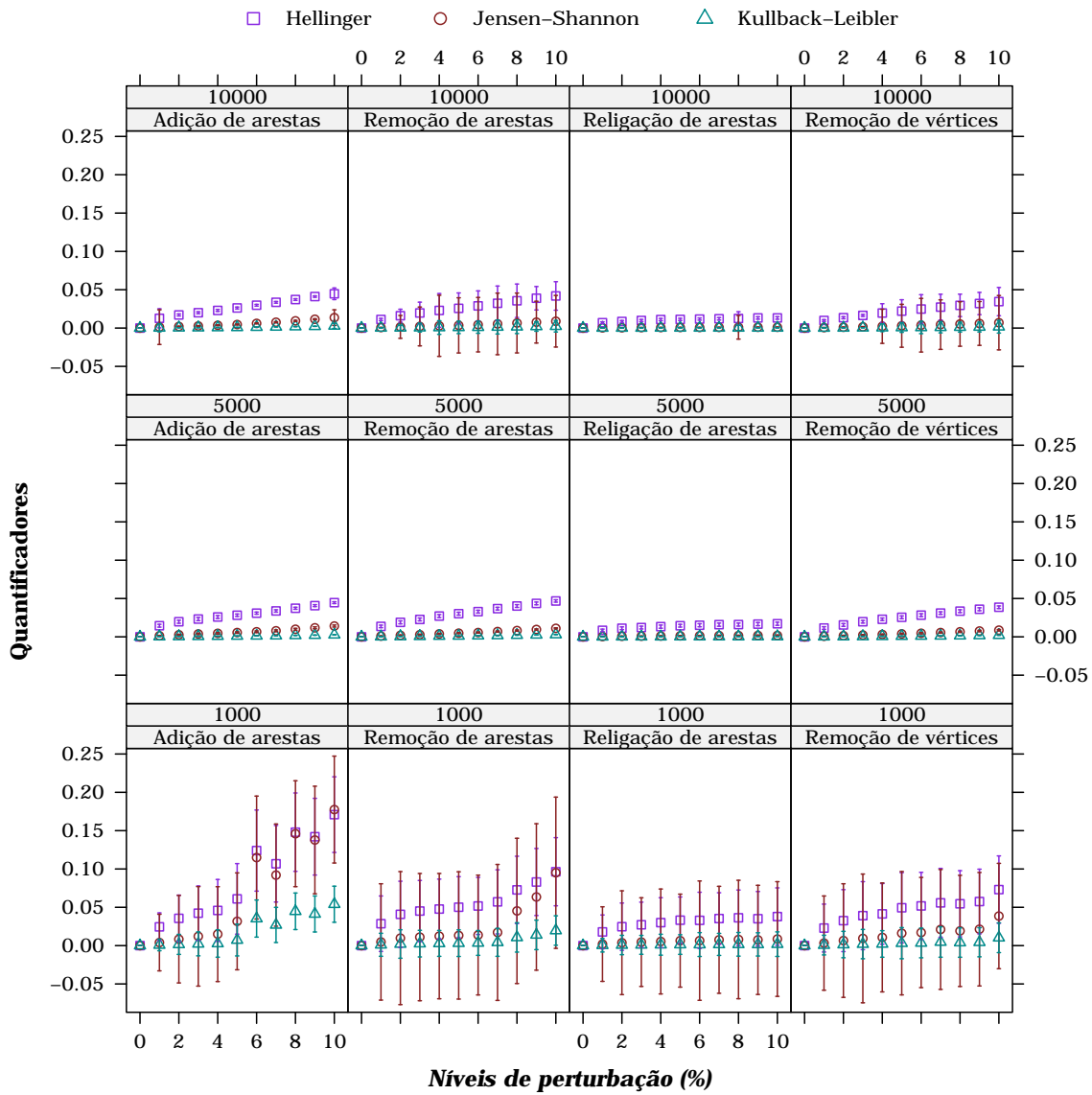
#### 4.2.2.3 Betweenness centrality ( $B_i$ )

Como apresentado nas Figuras 4.9 e 4.10, o *betweenness centrality* ( $B_i$ ) exibe o mesmo comportamento em redes aleatórias e de mundo pequeno. Todas as perturbações incrementam os quantificadores, i.e., eles aumentam a medida que o nível de perturbação aumenta. Note que o intervalo de confiança em alguns casos são maiores, como observa-se na primeira linha (de baixo para cima) das Figuras 4.9 e 4.10, mas os resultados são suficientes para perceber a dependência

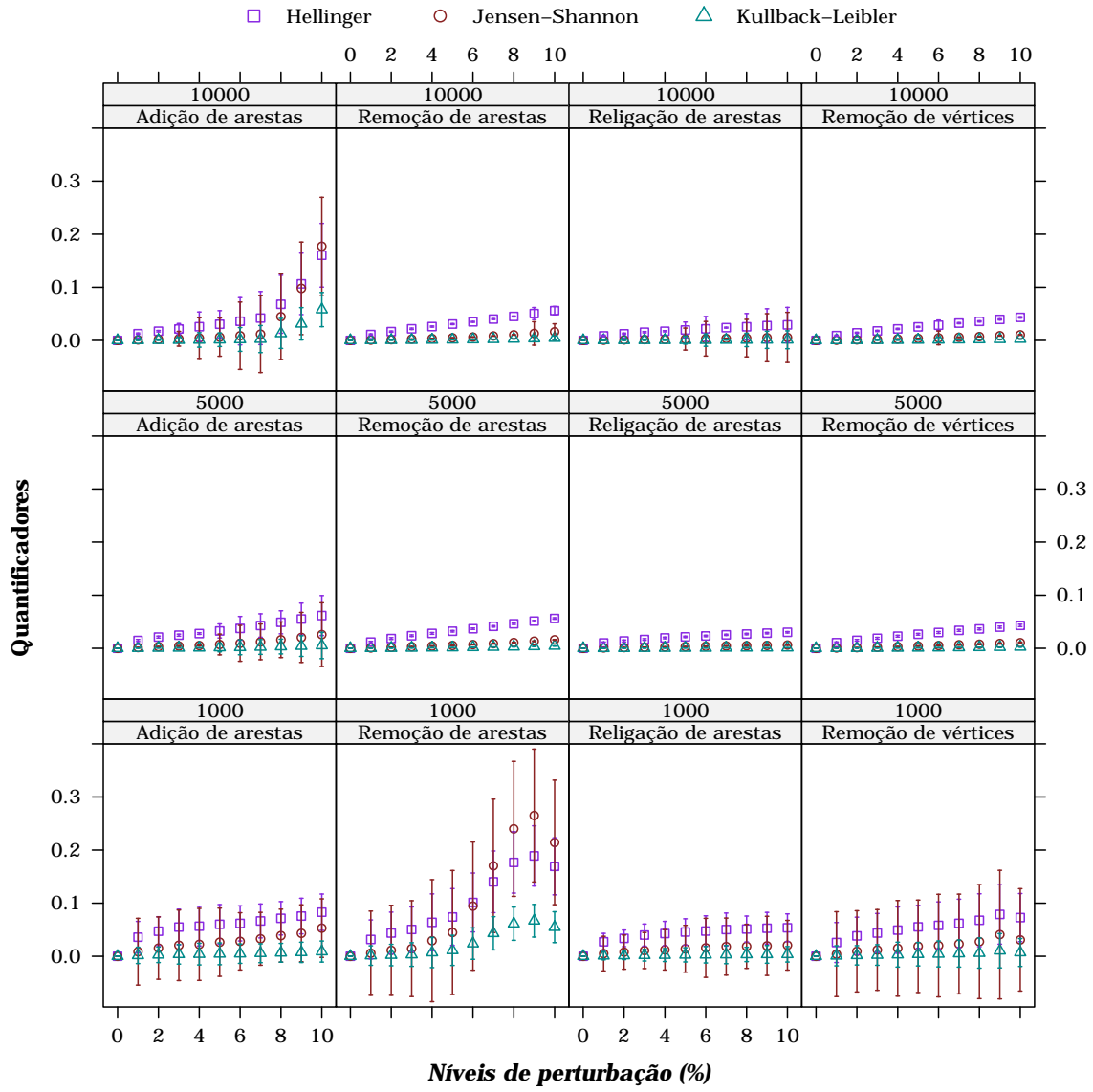
desta medida às perturbações aplicadas.

Em relação as redes livres de escala, as medidas variam fortemente quando são submetidas a pequenas perturbações, mas tendem a estabilizar, ou saturar logo depois, como observa-se na Figura 4.11.

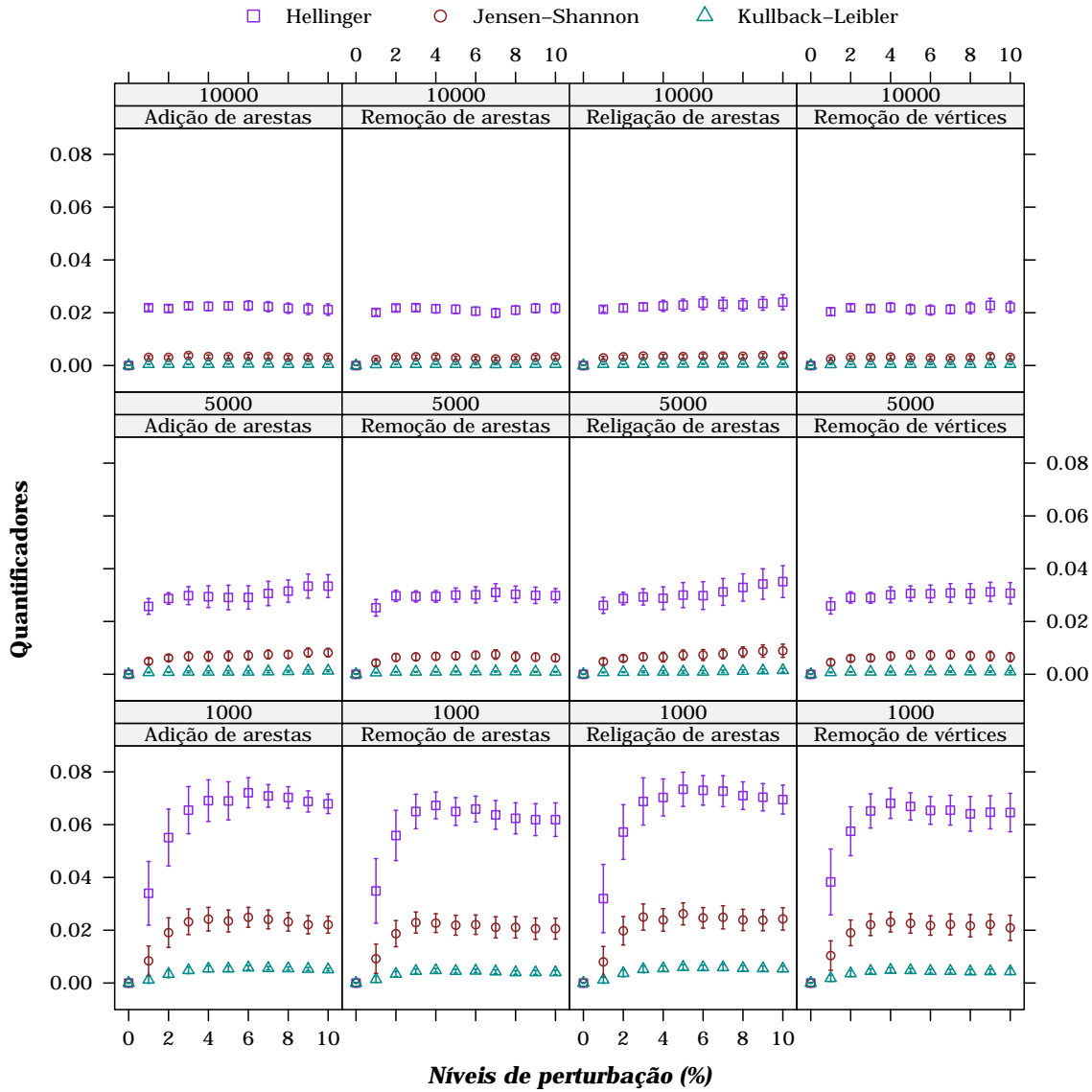
O quantificador mais sensível às perturbações é a distância de Hellinger  $\mathcal{D}_H$ , seguida pela divergência de Kullback-Leibler  $\mathcal{D}_{KL}$ . A distância de Jensen-Shannon  $\mathcal{D}_{JS}$  apresentou as menores variações.



**Figura 4.9.** Variação dos quantificadores para o *betweenness centrality*  $B_i$  em redes aleatórias.



**Figura 4.10.** Variação dos quantificadores para o *betweenness centrality*  $B_i$  em redes de mundo pequeno.



**Figura 4.11.** Variação dos quantificadores para o *betweenness centrality*  $B_i$  em redes livres de escala.

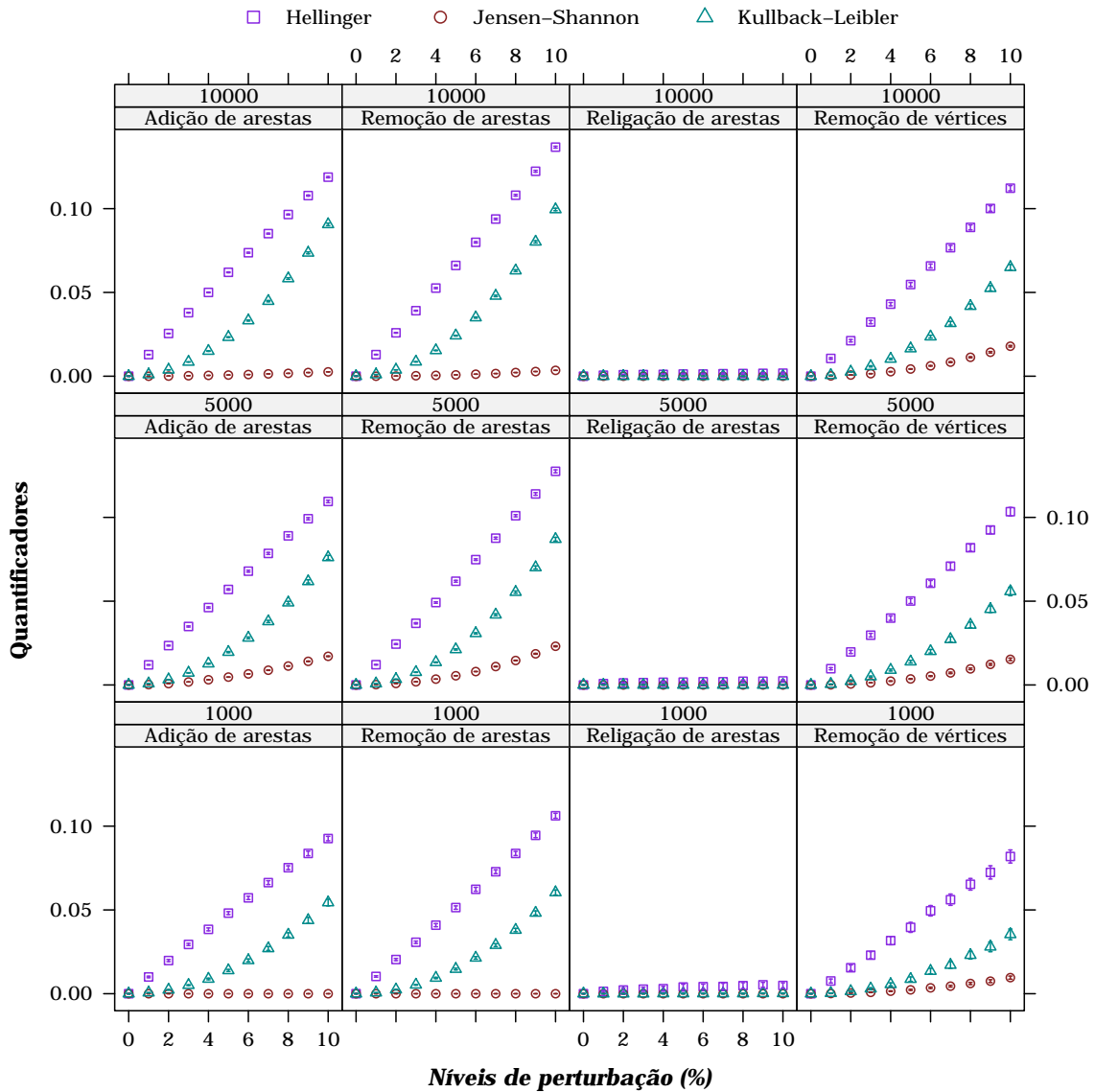
#### 4.2.2.4 Comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ )

As Figuras 4.12, 4.13 e 4.14 apresentam o comportamento dos quantificadores  $\mathcal{D}_{KL}$ ,  $\mathcal{D}_{JS}$  e  $\mathcal{D}_H$  para  $\ell_{(i,j)}$  aplicado aos modelos  $\mathcal{G}_{er}$ ,  $\mathcal{G}_{ws}$  e  $\mathcal{G}_{ba}$ , respectivamente. A Tabela 4.2 contém a média dos valores dos quantificadores no nível mais alto de perturbação (10%). O comportamento dos quantificadores para essa medida é similar em todos os modelos. À medida que o nível de perturbação aumenta, os grafos  $\eta_{ij}$  tornam-se progressivamente diferentes dos grafos iniciais  $g_j$  alterando

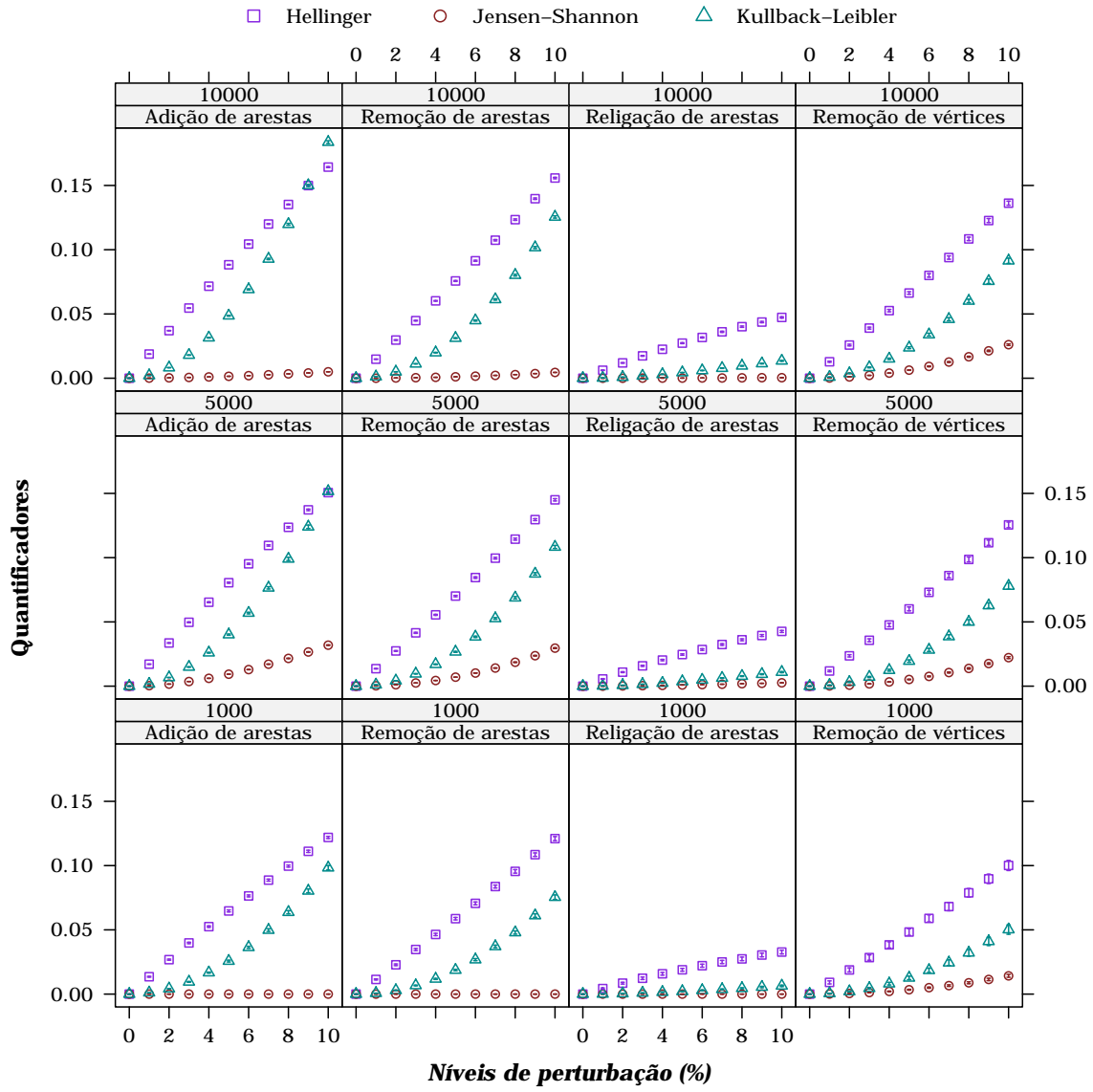


os valores de  $\ell_{(i,j)}$  e, conseqüentemente, aumentando os valores das distâncias.

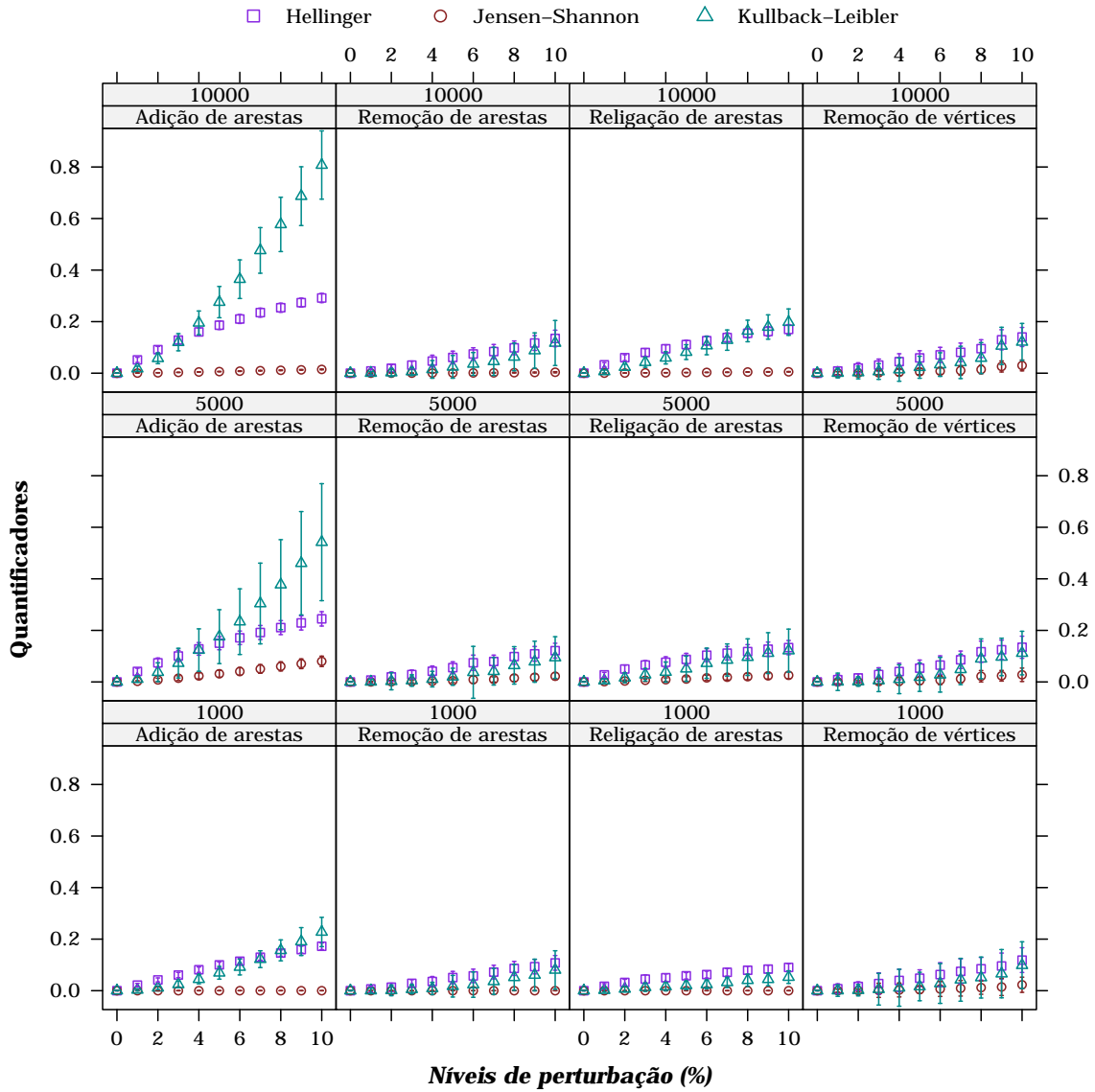
Vale ressaltar que a religação de arestas foi a perturbação que menos alterou os valores dos quantificadores, com uma variação muito pequena. Esta não apresentou diferença notável nas redes aleatórias, como observa-se na terceira coluna da Figura 4.12; e apenas um leve crescimento dos quantificadores nas redes de mundo pequeno e livres de escala, como observa-se na terceira coluna das Figuras 4.13 e 4.14, respectivamente.



**Figura 4.12.** Variação dos quantificadores para o comprimento dos caminhos mínimos  $\ell_{(i,j)}$  em redes aleatórias.



**Figura 4.13.** Variação dos quantificadores para o comprimento dos caminhos mínimos  $\ell_{(i,j)}$  em redes de mundo pequeno.

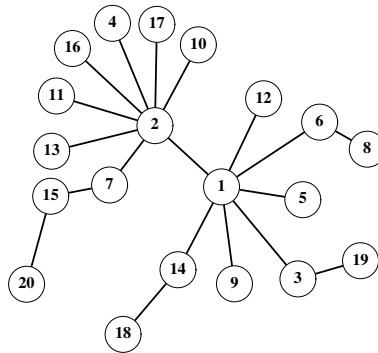


**Figura 4.14.** Variação dos quantificadores para o comprimento dos caminhos mínimos  $\ell_{(i,j)}$  em redes livres de escala.

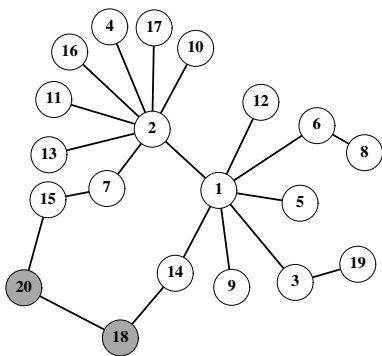
As redes livres de escala são mais sensíveis à adição de arestas em relação aos demais modelos, pois os quantificadores apresentaram valores mais altos, como observado comparando-se a Figura 4.14 com as Figuras 4.12 e 4.13. Por exemplo, a distância de Hellinger atingiu valores máximos 0.12 para redes aleatórias, 0.16 para redes de mundo pequeno e 0.81 para redes livres de escala (situação ocorrida para 10000 vértices e nível de perturbação de 10%).

A principal característica dessas redes é a presença de concentra-

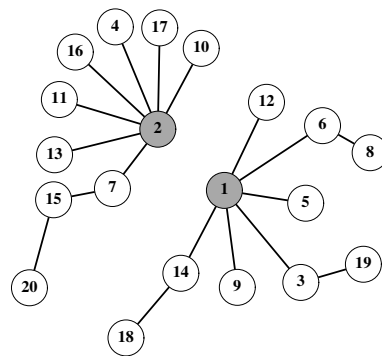
dores, dessa forma, a perturbação desses vértices pode alterar radicalmente os caminhos mínimos como ilustrado na Figura 4.15. Na Figura 4.15(a), o caminho mínimo entre os vértices  $v_{18}$  e  $v_{20}$  é  $\ell_{(18,20)} = ((v_{18}, v_{14}), (v_{14}, v_1), (v_1, v_2), (v_2, v_7), (v_7, v_{15}), (v_{15}, v_{20}))$  e seu comprimento  $|\ell_{(18,20)}| = 6$ . Supondo que a aresta  $e = (v_{18}, v_{20})$  foi adicionada a  $g_j$ , o valor de  $|\ell_{(18,20)}|$  passa a ser igual a um, como ilustrado na Figura 4.15(b). Por outro lado, como ilustrado na Figura 4.15(c), a remoção da aresta  $e = (v_1, v_2)$  de  $g_j$  torna a rede desconectada fazendo com que não exista mais caminho entre os vértices  $v_{18}$  e  $v_{20}$ , denotado por  $|\ell_{(18,20)}| = \infty$ . Porém, comparando-se a primeira e segunda coluna da Figura 4.14, observa-se que essas redes são mais sensíveis à adição de arestas do que a remoção de arestas para a medida em questão. Esse é um fato esperado, pois ao se computar os caminhos mínimos entre os vértices, aqueles que não existem são desprezados.



(a) Rede inicial  $g_j$  com  $|\ell_{(18,20)}| = 6$



(b) Rede  $g_j$  após a adição de uma aresta com  $|\ell_{(18,20)}| = 1$



(c) Rede  $g_j$  após a remoção de uma aresta com  $|\ell_{(18,20)}| = \infty$

**Figura 4.15.** Uma rede livre de escala com  $N = 20$  vértices e dois concentradores  $v_1$  e  $v_2$ .

Como observa-se na Tabela 4.2, entre os quantificadores, a distância de Hel-

linger é a mais sensível aos três tipos de perturbações, seguida da divergência de Kullback-Leibler  $\mathcal{D}_{KL}$  e Jensen-Shannon  $\mathcal{D}_{JS}$ . Esta última teve a menor variação atingindo seu valor máximo 0.014 para a adição de arestas em redes livres de escala. A divergência de Kullback-Leibler  $\mathcal{D}_{KL}$  é o quantificador mais sensível a adição de arestas em redes livres de escala, como observa-se na primeira coluna da Figura 4.14, com valor máximo igual a 0.807.

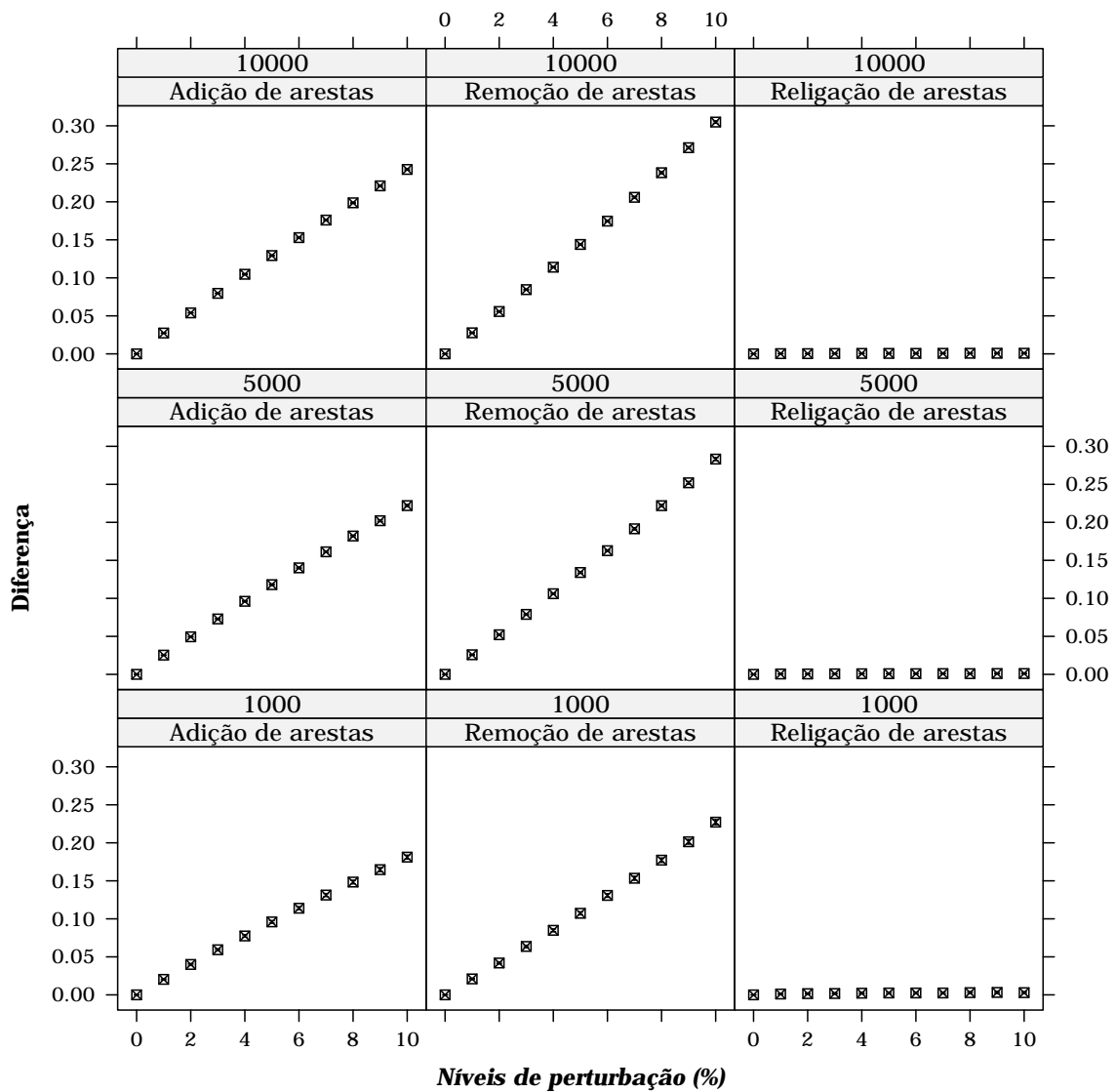
**Tabela 4.2.** Média dos valores dos quantificadores para o nível de perturbação mais intensa (10%) para o comprimento do caminho mínimo  $\ell_{(i,j)}$ . O maior valor obtido está em negrito.

	$\mathcal{D}_{KL}$			$\mathcal{D}_{JS}$			$\mathcal{D}_H$		
	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$
$\mathcal{G}_{er}$	0.090	0.099	0	0.002	0.003	0	0.118	0.136	0.0018
$\mathcal{G}_{ws}$	0.183	0.125	0.013	0.005	0.004	$10^{-4}$	0.164	0.155	0.047
$\mathcal{G}_{ba}$	<b>0.807</b>	0.117	0.198	0.014	0.003	0.005	0.291	0.134	0.169

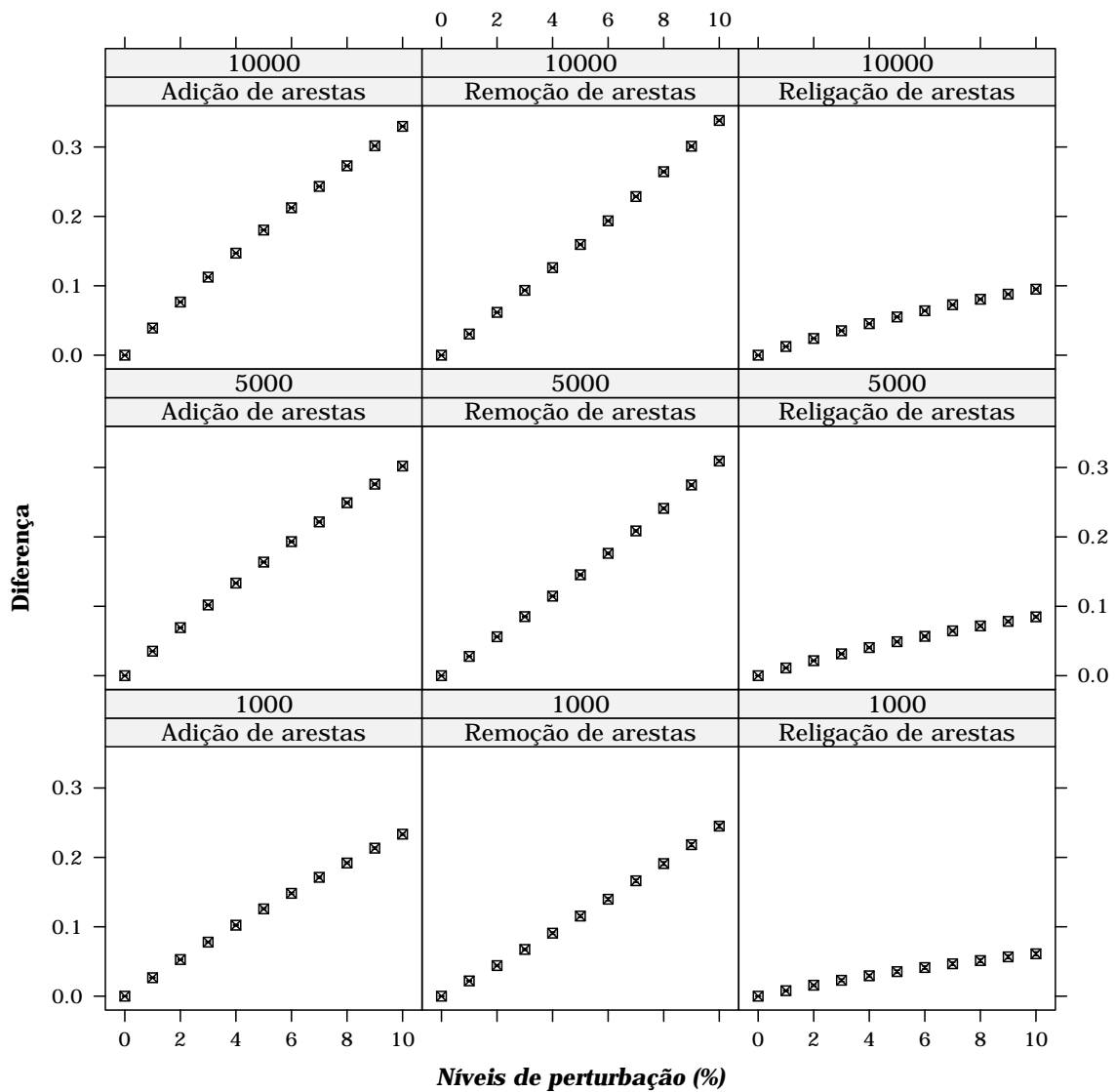
#### 4.2.2.5 Comprimento médio dos caminhos mínimos ( $L$ )

As Figuras 4.16, 4.17 e 4.18 apresentam a variação da diferença para  $L$  aplicado aos modelos  $\mathcal{G}_{er}$ ,  $\mathcal{G}_{ws}$  e  $\mathcal{G}_{ba}$ , respectivamente; e a Tabela 4.3 contém a média dos seus valores no nível mais alto de perturbação (10%).

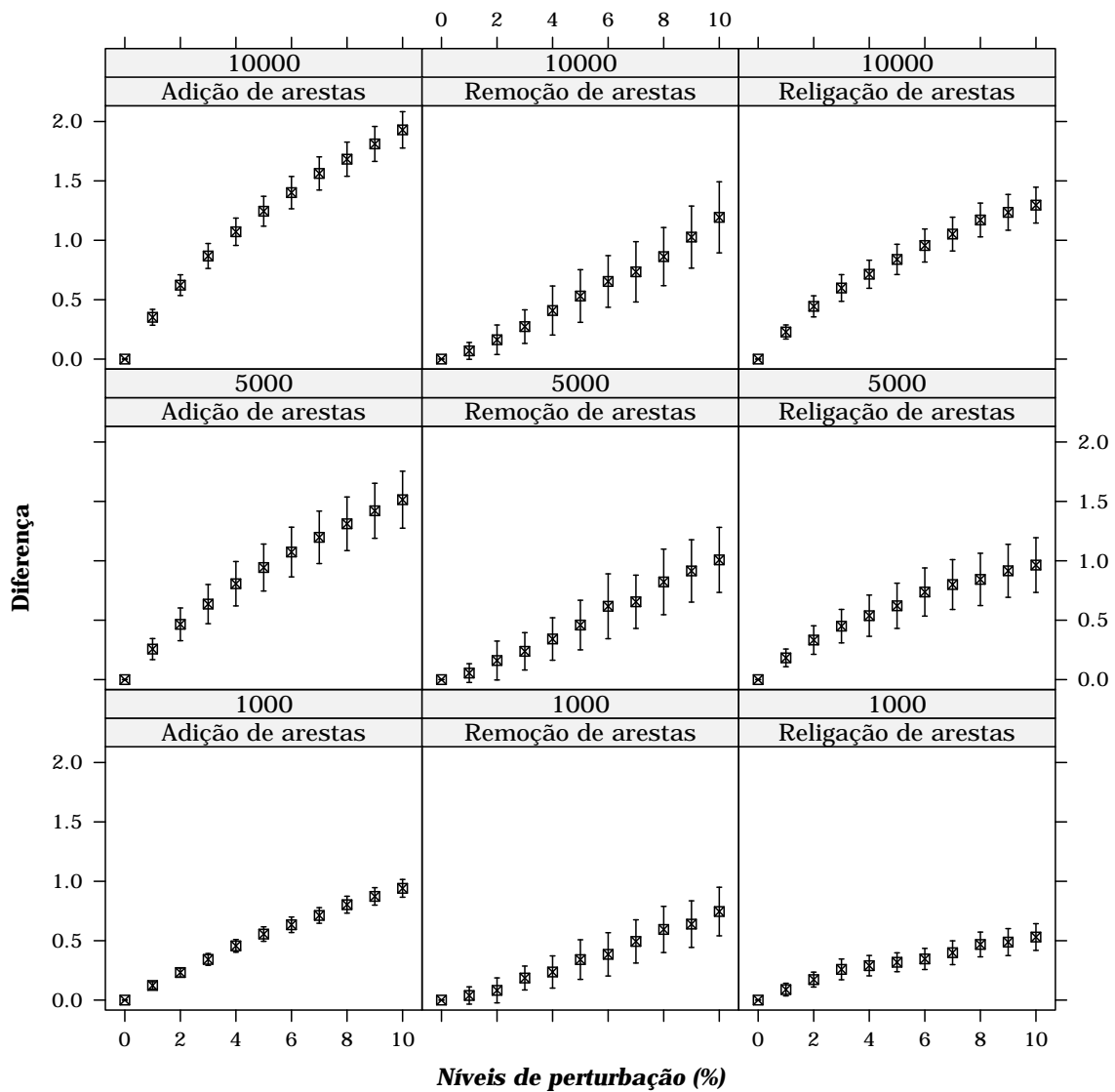
Esta medida não é sensível a religação de arestas em redes aleatórias, como observa-se na terceira coluna da Figura 4.16. As redes livres de escala são as mais sensíveis, pois apresentaram os valores mais altos para a diferença (valor máximo 1.929 para adição de arestas) como foi observado na Tabela 4.3.



**Figura 4.16.** Variação da diferença absoluta para o comprimento médio dos caminhos mínimos  $L$  em redes aleatórias.



**Figura 4.17.** Variação da diferença absoluta para o comprimento médio dos caminhos mínimos  $L$  em redes de mundo pequeno.



**Figura 4.18.** Variação da diferença absoluta para o comprimento médio dos caminhos mínimos  $L$  em redes livres de escala.

#### 4.2.3 Análise dos modelos teóricos II

Esta seção apresenta os resultados comparando-se o comportamento de cada modelo em relação as perturbações aplicadas usando-se a distância de Hellinger ( $\mathcal{D}_H$ ). O objetivo é comparar a sensibilidade dos modelos em relação às perturbações aplicadas, i.e., deseja-se saber, por exemplo, que modelo é mais sensível a adição de arestas em relação ao grau do vértice.



**Tabela 4.3.** Média dos valores da diferença para o nível de perturbação mais intensa (10%) para o comprimento médio dos caminhos mínimos  $L$ . O maior valor obtido está em negrito.

	Diferença		
	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$
$\mathcal{G}_{er}$	0.242	0.304	$10^{-4}$
$\mathcal{G}_{ws}$	0.324	0.338	0.095
$\mathcal{G}_{ba}$	<b>1.929</b>	1.193	1.296

Os resultados são apresentados nas Figuras 4.19, 4.20, 4.21 e 4.22 para as medidas: grau do vértice ( $\kappa_i$ ), coeficiente de agrupamento ( $C_i$ ), *betweenness centrality* ( $B_i$ ) e comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ), respectivamente.

Cada figura apresenta o resultado para uma medida  $t_k$  aplicada as três classes de redes complexas  $\mathcal{G}$  e é formada por um conjunto de 12 gráficos, ou uma grade de gráficos de tamanho  $3 \times 4$ , totalizando 48 gráficos. As linhas da grade correspondem aos três tamanhos de grafos estudados (de baixo para cima) 1000, 5000 e 10000 vértices; enquanto as colunas da grade descrevem os quatro tipos de perturbações aplicadas a estes modelos (da esquerda para a direita) adição, remoção e religação de arestas e remoção de vértices.

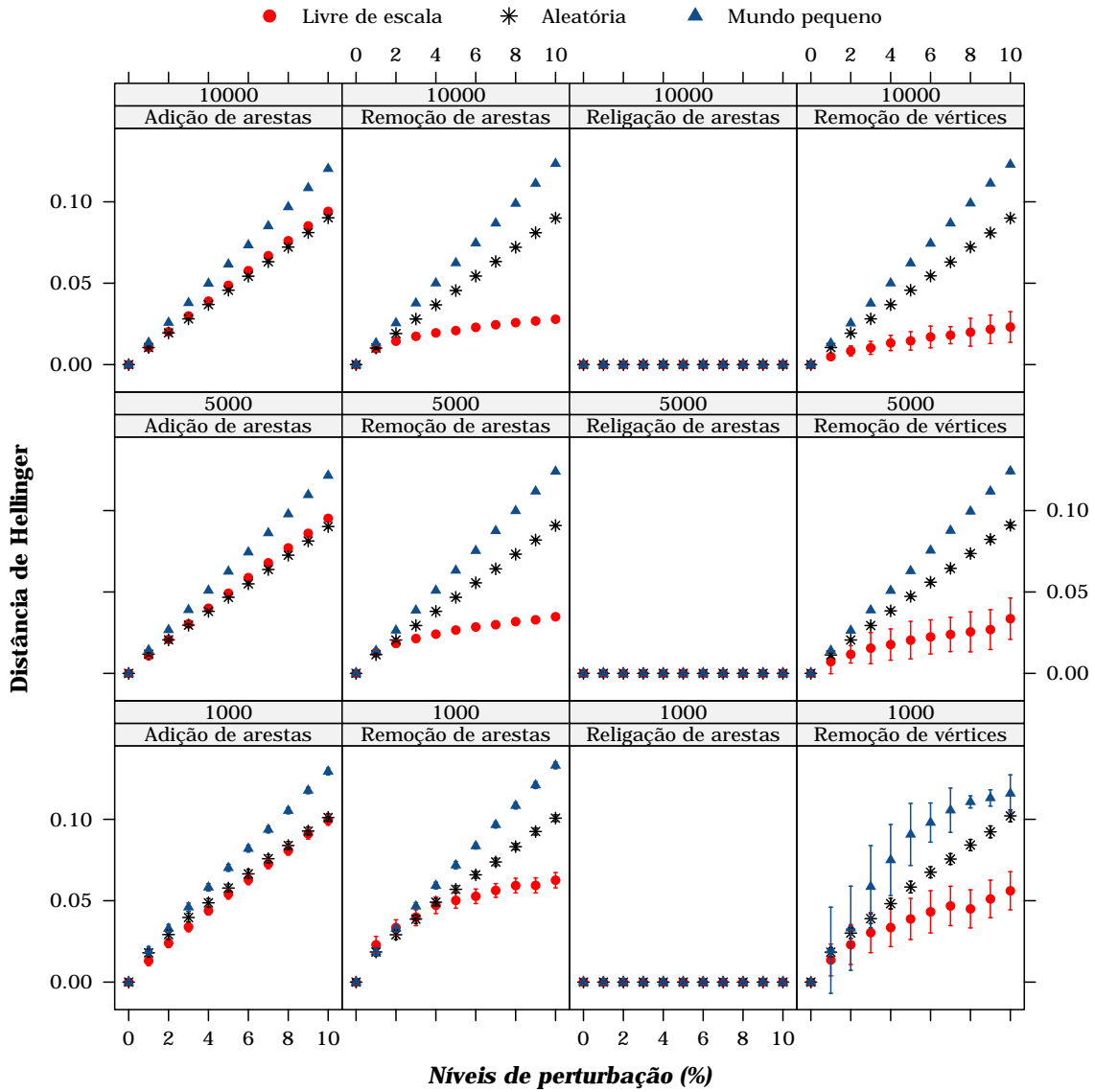
Cada gráfico do conjunto apresenta a variação da distância de Hellinger (eixo das ordenadas) em relação ao nível de perturbação aplicada à rede (eixo das abscissas). Os gráficos de cada figura estão sincronizados tanto no eixo das abscissas quanto das ordenadas, o que os tornam comparáveis.

A seguir, são apresentadas as análises sobre as diferentes perturbações e redes utilizados no estudo.

#### 4.2.3.1 Grau do vértice ( $\kappa_i$ )

A Figura 4.19 apresenta a variação da distância de Hellinger  $\mathcal{D}_H$  para a medida  $\kappa$  aplicado aos modelos  $\mathcal{G}_{er}$ ,  $\mathcal{G}_{ws}$  e  $\mathcal{G}_{ba}$ .

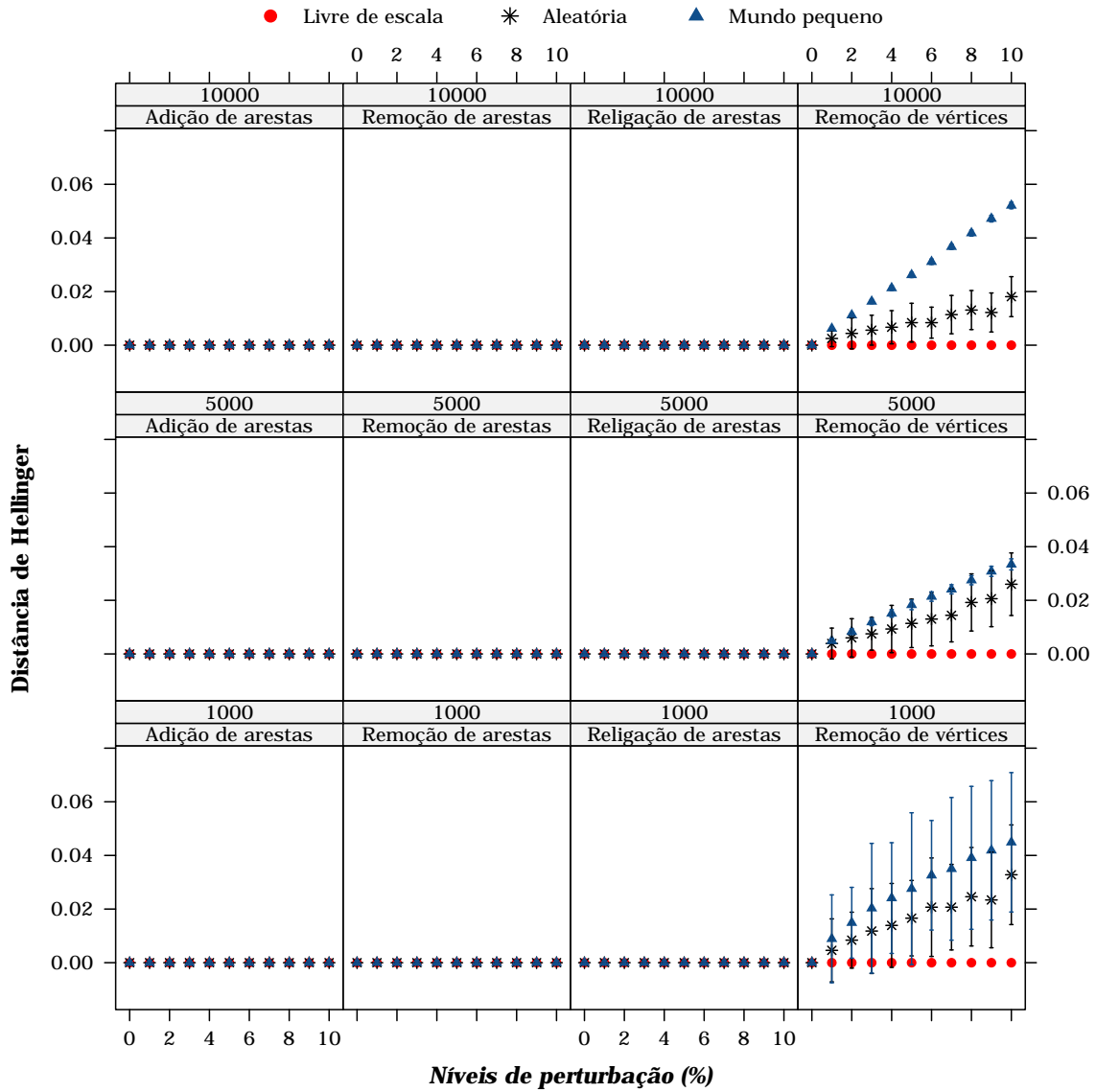
Como já citado, essa medida exhibe o mesmo comportamento em todos os modelos. A adição de arestas, remoção de arestas e remoção de vértices incrementam os valores da distância de Hellinger, i.e., seus valores aumentam progressivamente quando o nível de perturbação aumenta; e a religação de arestas não introduz mudanças perceptíveis, como se vê na terceira coluna de cada grade. Nesse caso, em todos os casos testados as redes de mundo pequeno são as mais sensíveis em relação ao grau do vértice, seguida pelas redes aleatórias e as redes livres de escala.



**Figura 4.19.** Variação da distância de Hellinger para o grau do vértice  $\kappa_i$  nos modelos teóricos.

#### 4.2.3.2 Coeficiente de agrupamento ( $C_i$ )

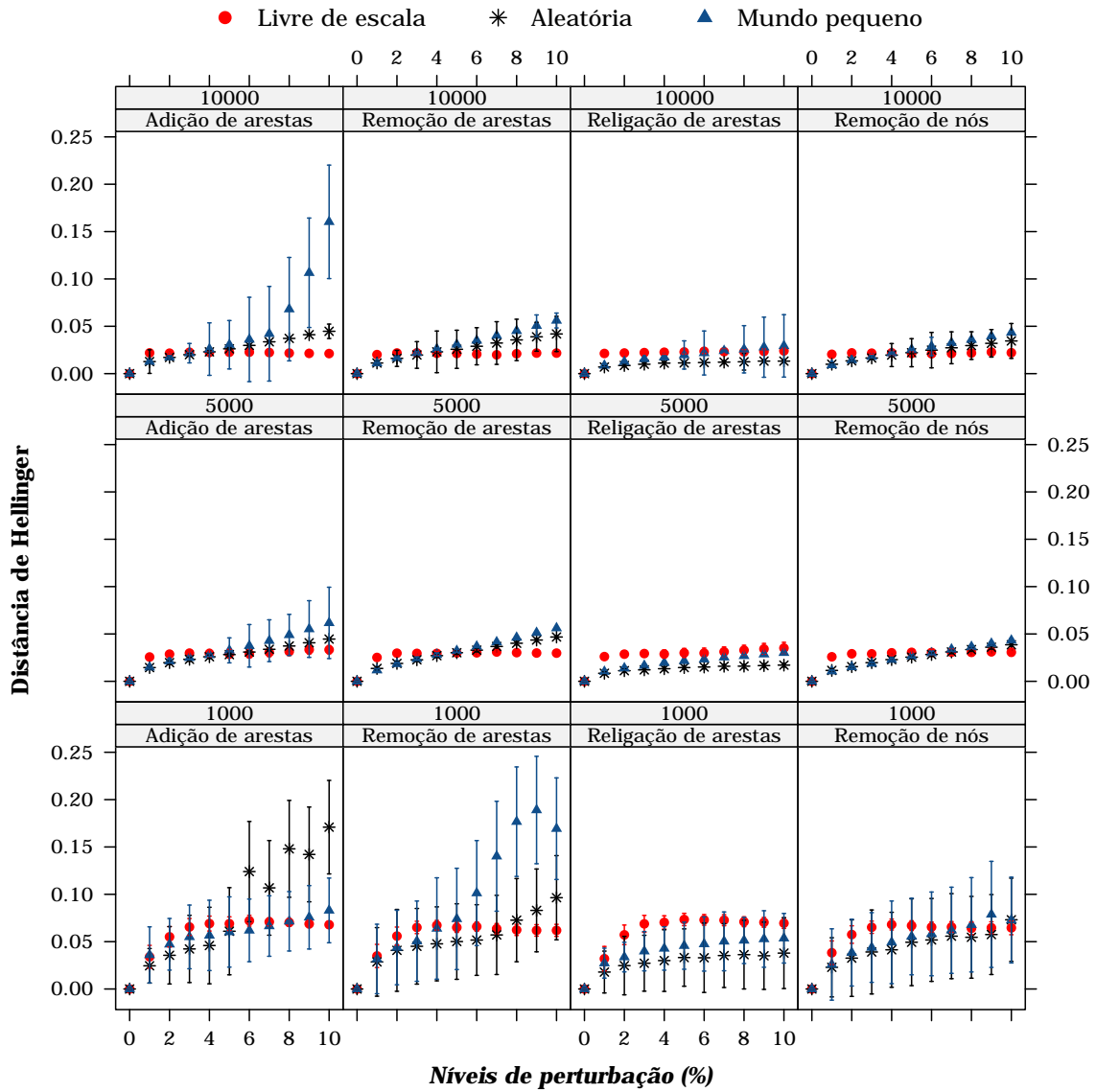
Como observa-se na Figura 4.20, o coeficiente de agrupamento local  $C_i$  não apresentou variação para as perturbações de arestas para os modelos avaliados e a remoção de nós nas redes livres de escala. Pode-se dizer que essa medida é robusta ou insensível às perturbações aplicadas nesses casos. As redes de mundo pequeno são as mais sensíveis, seguidas pelas redes aleatórias, mas com uma variação baixa da distância de Hellinger.



**Figura 4.20.** Variação da distância de Hellinger para o coeficiente de agrupamento  $C_i$  nos modelos teóricos.

#### 4.2.3.3 Betweenness Centrality ( $B_i$ )

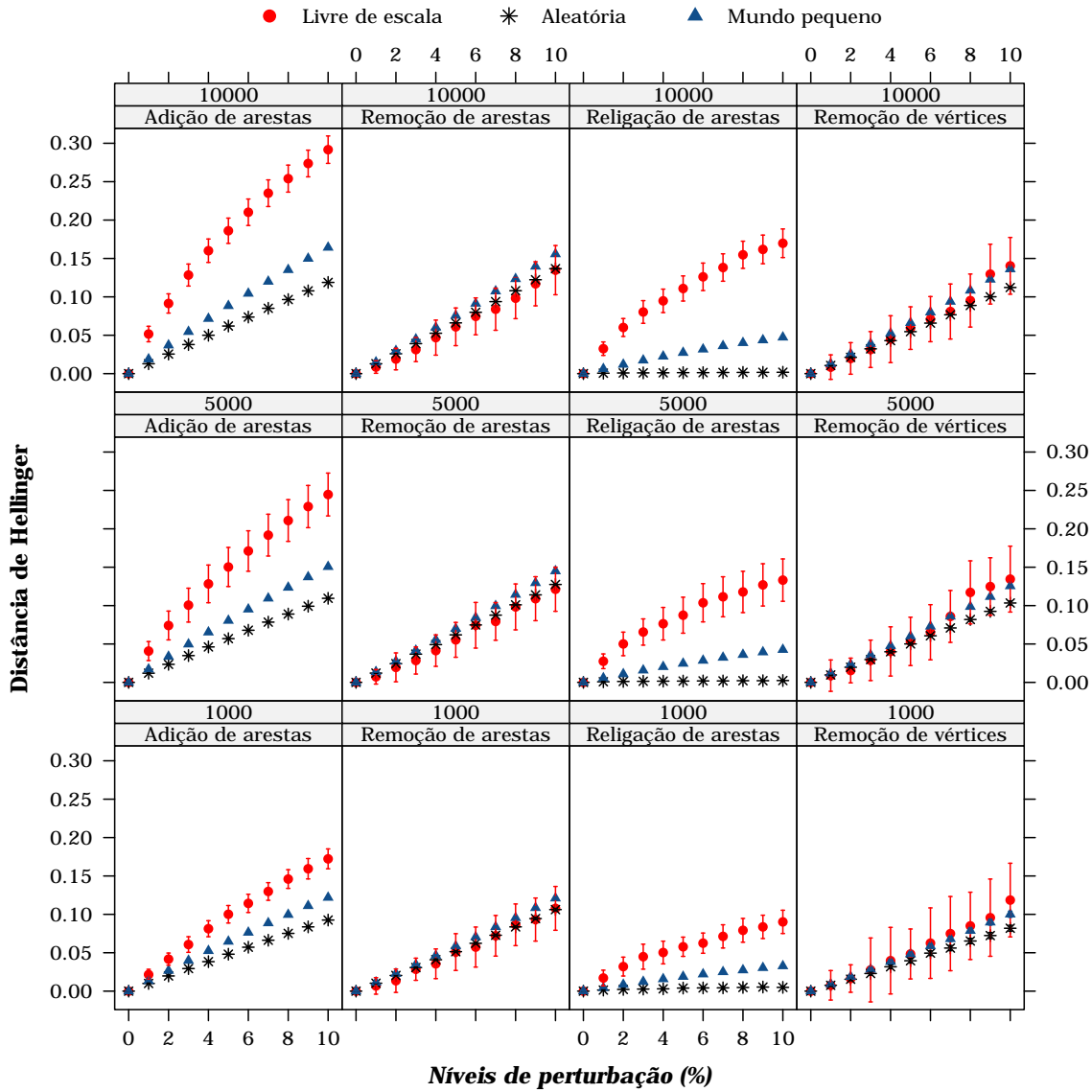
A Figura 4.21 exibe o comportamento da distância de Hellinger para o *betweenness* ( $B_i$ ). Nesse caso o comportamento de todas as redes está bem próximo, com exceção de alguns casos, em que as redes de mundo pequeno apresentam uma variação maior que as demais.



**Figura 4.21.** Variação da distância de Hellinger para o *betweenness centrality*  $B_i$  nos modelos teóricos.

#### 4.2.3.4 Comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ )

A Figuras 4.22 apresenta o comportamento do comprimento médio dos caminhos mínimos nos diferentes modelos de rede. Nesse caso, as redes livres de escala são mais sensíveis para  $\ell_{(i,j)}$  devido a presença de concentradores, em seguida vem as mundo pequeno. As redes aleatórias não apresentaram diferença notável para a religação de arestas, como observa-se na terceira coluna.



**Figura 4.22.** Variação da distância de Hellinger para o comprimento dos caminhos mínimos  $l_{(i,j)}$  nos modelos teóricos.

#### 4.2.4 Testes de hipótese para a distribuição dos graus

A distribuição dos graus  $p(\kappa)$  é uma propriedade fundamental da topologia de uma rede e determina o seu tipo de conectividade, assim uma forma de verificar o efeito das perturbações em redes complexas é estudar o comportamento dessa propriedade. A distribuição dos graus é conhecido analiticamente para os modelos aleatório ( $\mathcal{G}_{er}$ ), de mundo pequeno ( $\mathcal{G}_{ws}$ ) e livres de escala ( $\mathcal{G}_{ba}$ ), dada pelas

equações (2.8), (2.13) e (2.14), respectivamente. Assim, nesta seção são realizados testes de hipótese [Montgomery, 2005] para verificação da quebra dessa propriedade nos modelos teóricos de redes complexas.

De uma forma geral, os passos para se realizar um teste de hipótese são:

1. Identificar e formular a hipótese nula  $H_0$  e a hipótese alternativa  $H_1$ .
2. Determinar a probabilidade de rejeitar a hipótese nula quando ela é efetivamente verdadeira, i.e., o nível de significância  $\alpha$ . O nível de significância é normalmente fixado em 5%.
3. Determinar a estatística de teste que pode ser usada para avaliar a veracidade da hipótese nula.
4. Computar a estatística de teste. O  $p$ -valor é uma estatística muito utilizada para sintetizar o resultado de um teste de hipóteses. É definido como a probabilidade de, sob a hipótese nula, observarmos uma estatística de teste pelo menos tão grande quanto a observada (as estatísticas de teste são feitas para crescerem conforme nos afastamos da hipótese nula). O  $p$ -valor é usado para tomar a decisão em relação a hipótese nula.
5. Rejeitar ou não  $H_0$ , i.e., comparar o  $p$ -valor ao nível de significância  $\alpha$ . Se  $p\text{-valor} \leq \alpha$ ,  $H_0$  é rejeitada.

Duas hipóteses foram verificadas:

**Caso 1**  $H_0^1$ : a distribuição dos graus é preservada frente a perturbações, i.e.,  $H_0^1 : p(\kappa)[g_j] = p(\kappa)[\eta_{ij}]$ .

**Caso 2**  $H_0^2$ : os graus dos vértices foi preservado, i.e.,  $H_0^2 : \kappa_i[g_j] = \kappa_i[\eta_{ij}]$ .

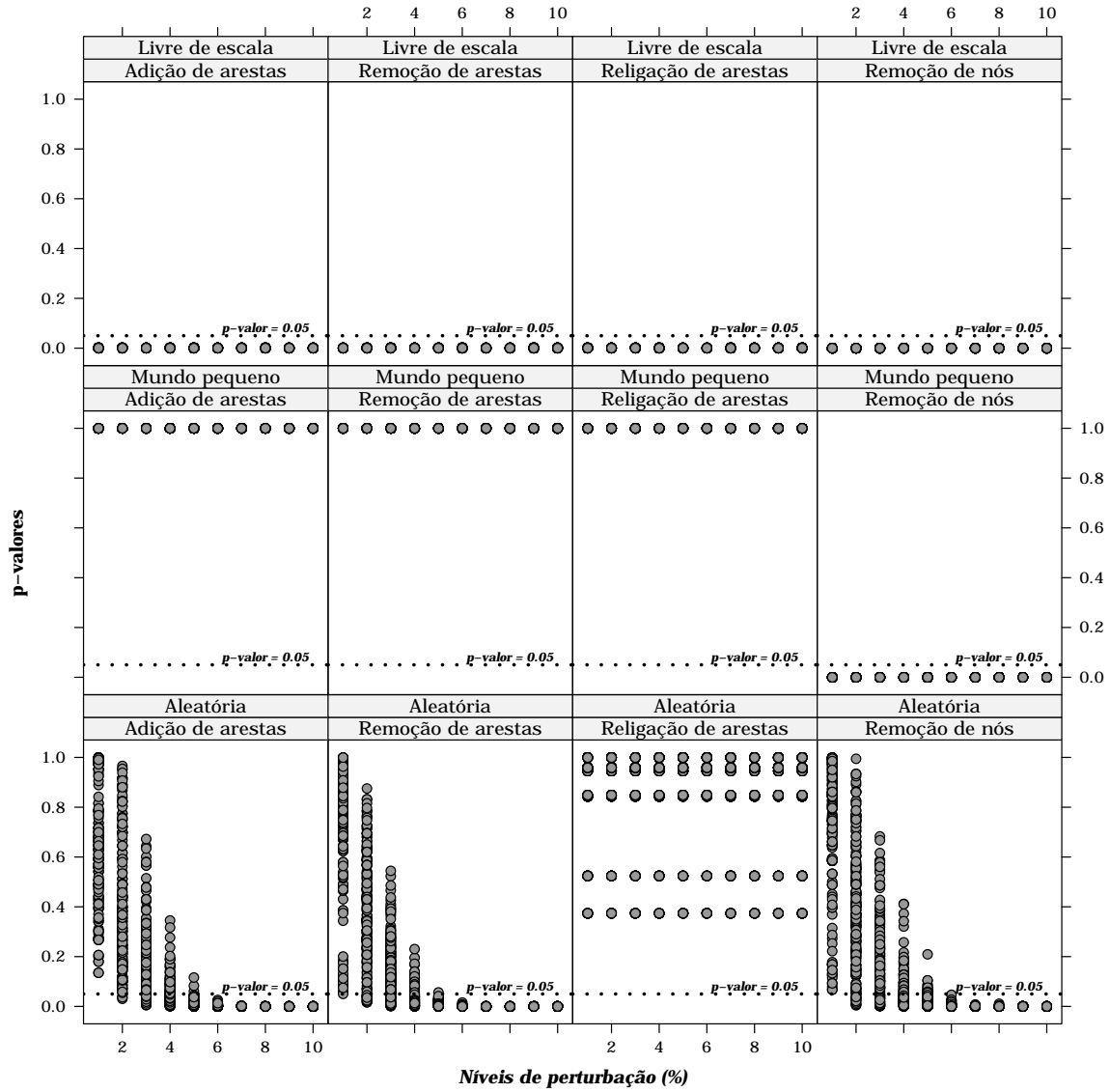
A hipótese  $H_0^1$  foi avaliada com o teste de Kolmogorov-Smirnov (teste KS) e  $H_0^2$  com o teste  $t$  de Student (teste  $t$ ) [Reschenhofer, 1997; Siegel & N. John Castellan, 1988]. Os testes foram realizados em redes com  $N = 1000$  vértices, nível de significância  $\alpha = 0.05$  e 200 amostras de teste.

No caso 1, o teste KS é usado para determinar se duas distribuições de probabilidade diferem uma da outra, tomando como base amostras finitas. Os resultados são apresentados na Figura 4.23 e na Tabela 4.4.

De acordo com os resultados, conclui-se com respeito a hipótese nula  $H_0^1$ :

- Nas redes livres de escala  $H_0^1$  foi rejeitada para todos os tipos de perturbação aplicadas, já que os  $p$ -valores são menores que 0.05 (primeira linha da Figura 4.23).
- Nas redes de mundo pequeno  $H_0^1$  não foi rejeitada para os três tipos de perturbação de arestas, já que todos os  $p$ -valores são maiores que 0.05 (segunda linha da Figura 4.23); o contrário acontece com a remoção de vértices, em que todos os  $p$ -valores são menores que 0.05.
- No modelo aleatório, a medida que o nível de adição de arestas, remoção de arestas e remoção de vértices aumenta, os  $p$ -valores aproximam-se de zero, assim  $H_0^1$  foi rejeitada para níveis de perturbação superiores 4% (terceira linha da Figura 4.23). A religação de arestas, como esperado, não leva à rejeição de  $H_0$  (terceira coluna da primeira linha da Figura 4.23).

Dizer que  $H_0^1$  é rejeitada significa que as perturbações conduziram a um colapso (ou quebra) da distribuição dos graus da rede.



**Figura 4.23.**  $P$ -valores obtidos do teste de Kolmogorov-Smirnov para 200 amostras. A linha pontilhada representa  $p$ -valores = 0.05. As linhas da grade correspondem as redes avaliadas (de baixo para cima): aleatória ( $\mathcal{G}_{er}$ ), mundo pequeno ( $\mathcal{G}_{ws}$ ) e livre de escala ( $\mathcal{G}_{ba}$ ); enquanto as colunas da grade descrevem os quatro tipos de perturbações aplicadas a estes modelos (da esquerda para a direita): adição de arestas ( $\pi_{ae}$ ), remoção de arestas ( $\pi_{re}$ ), religação ( $\pi_{rw}$ ) de arestas e remoção de vértices ( $\pi_{rv}$ ). Cada gráfico desse conjunto apresenta os  $p$ -valores (eixo das ordenadas) para as amostras testadas em relação ao nível de perturbação (eixo das abscissas).

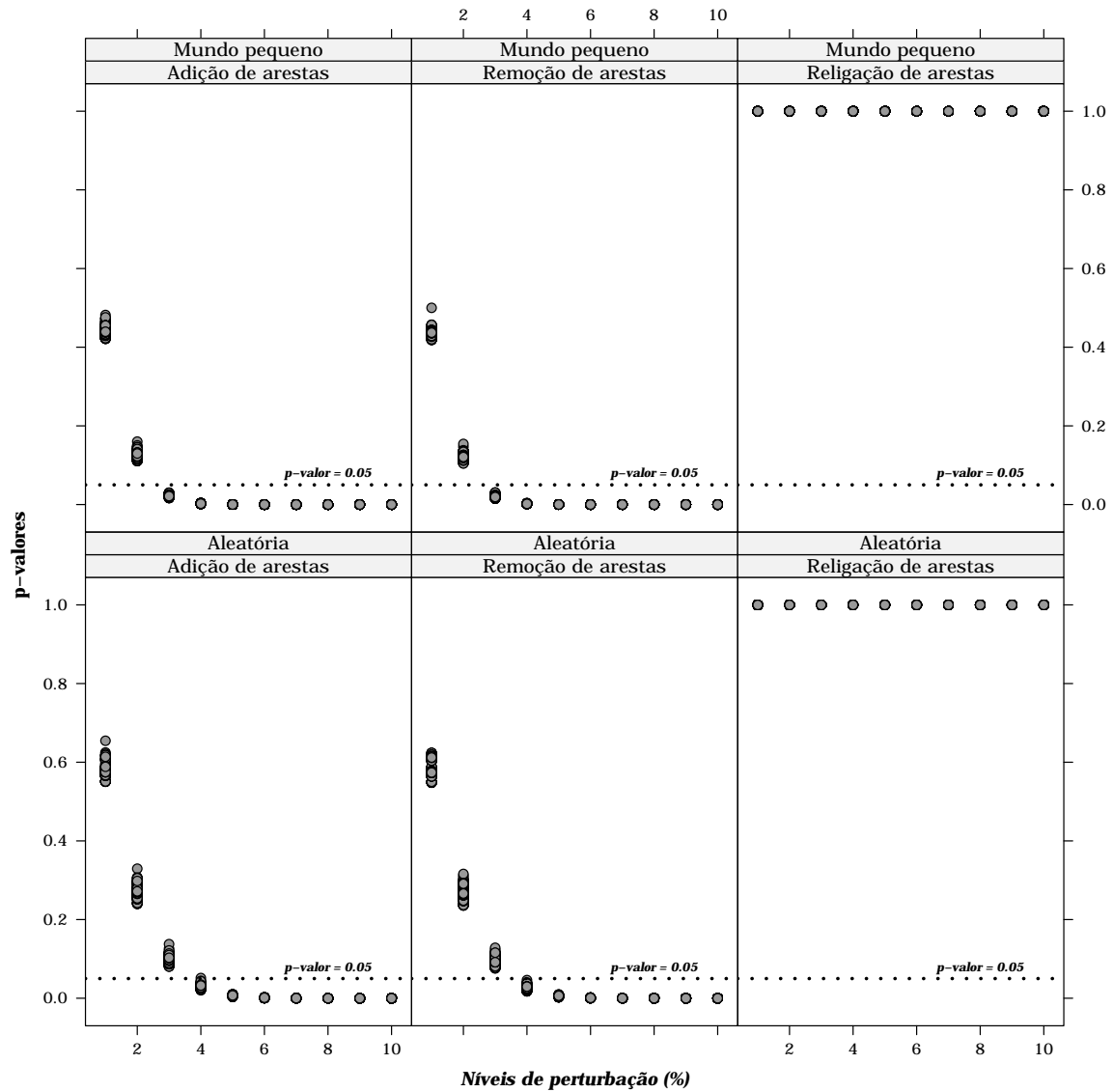


**Tabela 4.4.** Porcentagem de  $p$ -valores  $\geq 0.05$  obtidas do teste de Kolmogorov-Smirnov para 200 amostras nos níveis de perturbação 1%, 4% e 10%.

	$\pi_{ae}$			$\pi_{re}$			$\pi_{we}$		
	1%	4%	10%	1%	4%	10%	1%	4%	10%
$\mathcal{G}_{ba}$	0	0	0	0	0	0	0	0	0
$\mathcal{G}_{ws}$	100	100	100	100	100	100	100	100	100
$\mathcal{G}_{er}$	100	30	0	100	25	0	100	100	100

No caso 2, o teste- $t$  verifica se o grau dos vértices foi preservado após as perturbações aplicadas. Como no teste KS foi verificado que a distribuição dos graus é mantida apenas para as redes aleatórias e de mundo pequeno, o teste- $t$  foi aplicado somente a estes modelos. Os resultados são apresentados na Figura 4.24 e na Tabela 4.5. ] Obteve-se as seguintes conclusões a respeito deste teste:

- Nas redes de mundo pequeno, os  $p$ -valores aproximam-se de zero com o aumento do nível de perturbação aplicada. Os graus não são preservados para níveis maiores que 3%. A religação de arestas, como esperado, não modificam os graus dos vértices.
- O mesmo comportamento é observado com respeito a adição e remoção de arestas no modelo aleatório e o grau dos vértices é modificado para níveis de perturbação maiores que 4%. A religação de arestas não tem efeito nenhum sobre o grau dos vértices nestas redes, já que todos os  $p$ -valores são maiores que 0.05.



**Figura 4.24.** *P*-valores obtidos do teste *t* de Student para 200 amostras. A linha pontilhada representa *p*-valores = 0.05. As linhas da grade correspondem às redes avaliadas (de baixo para cima): aleatória ( $\mathcal{G}_{er}$ ) e mundo pequeno ( $\mathcal{G}_{ws}$ ); enquanto as colunas da grade descrevem os três tipos de perturbações aplicadas a estes modelos (da esquerda para a direita): adição ( $\pi_{ae}$ ), remoção ( $\pi_{re}$ ) e religação ( $\pi_{rw}$ ) de arestas. Cada gráfico desse conjunto apresenta os *p*-valores (eixo das ordenadas) para as amostras testadas em relação ao nível de perturbação (eixo das abscissas).

**Tabela 4.5.** Porcentagem de  $p$ -valores  $\geq 0.05$  obtidas do teste de  $t$  de Student para 200 amostras nos níveis de perturbação 1%, 4% e 10%.

	$\pi_{ae}$			$\pi_{re}$			$\pi_{we}$		
	1%	4%	10%	1%	4%	10%	1%	4%	10%
$\mathcal{G}_{ws}$	100	0	0	100	0	0	100	100	100
$\mathcal{G}_{er}$	100	0%	0	100	0	0	100	100	100

#### 4.2.5 Discussão geral dos resultados

As Tabelas 4.6 e 4.7 resumizam os resultados discutidos anteriormente. Entre todas as perturbações, a religação de arestas causou menos variação nos quantificadores estocásticos. O grau do vértice  $\kappa$  não é alterado com essa perturbação, pois a religação preserva seus valores; enquanto que o comprimento dos caminhos mínimos  $\ell_{(i,j)}$  apresentou leve variação com valores muito baixos para os quantificadores nas redes aleatórias e de mundo pequeno, atingindo valor máximo igual a 0.047 para  $\mathcal{D}_H$  em  $\mathcal{G}_{ws}$ . Já nas redes livres de escala, o valor máximo atingido foi igual a 0.198 para  $\mathcal{D}_{KL}$ .

A adição e remoção de arestas atinge diretamente o comportamento das medidas locais  $\kappa_i$  e  $\ell_{(i,j)}$ , já que estas alteram sempre seus valores causando variações nos quantificadores estocásticos. Especificamente, as redes livre de escala são as mais sensíveis a adição de arestas, já que os quantificadores atingiram os valores mais altos, 0.807 para  $\mathcal{D}_{KL}$  no nível de perturbação 10% em redes com  $N = 10000$  vértices. Vale lembrar que o uso de  $\mathcal{D}_{KL}$  exige a adição de um valor ( $\delta = 0.001$ ) às distribuições de probabilidade, e isto pode afetar o comportamento desse quantificador nesse contexto.

Em relação a sensibilidade dos quantificadores utilizados, a distância de Hellinger apresentou os maiores valores, com exceção do comprimento dos caminhos mínimos aplicado às redes livres de escala quando submetidas a adição de arestas, em que a divergência de Kullback-Leibler apresentou maiores valores. A distância de Jensen-Shannon  $\mathcal{D}_{JS}$  apresentou as menores variações para as medidas  $\kappa_i$  e  $\ell_{(i,j)}$ , i.e., esta não é sensível às perturbações aplicadas nas redes avaliadas. Este quantificador não apresentou nenhuma variação para redes com  $N = 1000$  vértices. Atingiu valor máximo de 0.0026 para  $\ell$  aplicada a redes aleatórias no nível de 10% e  $N = 10000$ .

O coeficiente de agrupamento local  $C_i$  não apresentou variação para as perturbações de arestas, assim pode-se dizer que essa medida é robusta ou insensível.

O comprimento médio dos caminhos mínimos não é sensível a religação de arestas em redes aleatórias. As redes livres de escala são as mais sensíveis, pois apresentaram os valores mais altos para a diferença.

### 4.3 Análise em redes de sensores sem fio

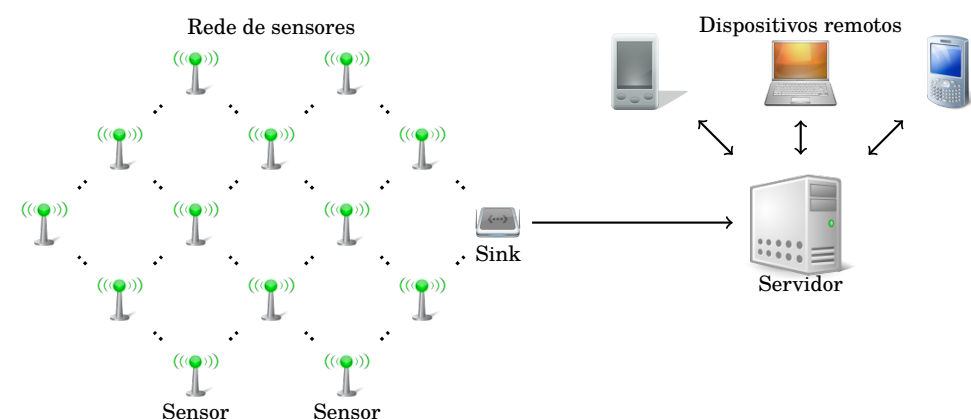
Esta seção discorre sobre a propagação de dados em redes de sensores sem fio, caracterizada como uma rede complexa, bem como sobre a avaliação do comportamento do comprimento dos caminhos mínimos. São mostrados os conceitos básicos a respeito dessas redes, e também alguns trabalhos relacionados que seguem na seção 4.3.1. A seção 4.3.2 apresenta as diferentes topologias de propagação de dados. Por fim, a seção 4.3.3 apresenta o estudo da análise de medidas.

#### 4.3.1 Definições

As Redes de Sensores sem Fio (RSSF) são uma tecnologia emergente que permite o monitoramento de variáveis físicas, tais como temperatura, som, luminosidade, vibração, pressão ou movimento. Uma RSSF consiste de um grande número de dispositivos sem fio autônomos, chamados de “nós sensores”, “sensores” ou simplesmente “nós”. Estas entidades trabalham de forma cooperativa sentindo o ambiente onde se encontram, comunicando-se e tomando decisões [Akyildiz et al., 2002]. Devido ao seu tamanho, os nós sensores possuem uma arquitetura simples e com limitações de processamento e armazenamento sendo formados por quatro componentes básicos: uma unidade perceptiva que possui alguns sensores dotados de um conversor de sinais analógicos para digitais; uma unidade de processamento com memória e processador; um transceptor; e uma fonte de energia que geralmente não é renovável. Além disso, de forma opcional, pode haver elementos que complementam a estrutura dos sensores, como sistema de localização, mecanismo de mobilidade e gerador de energia.

Tais redes são uma tecnologia promissora com uma ampla gama de aplicações em, por exemplo, biotecnologia, indústria, saúde pública e transporte [Arampatzis et al., 2005]. Destaca-se, em particular, o interesse na monitoração de ambientes hostis ou inacessíveis [Buratti et al., 2009]. Estas aplicações podem ter um caráter de *monitoramento*, em que os dados do ambiente são apenas coletados, ou de *atuação*, onde ocorre intervenção no meio monitorado [de Aquino, 2008].

As variáveis físicas são monitoradas, colhidas e armazenadas no sensor para posteriormente serem propagadas até um nó líder, chamado de nó sorvedouro (*sink*). Este é responsável pelo gerenciamento, processamento e entrega dos dados a um usuário externo [Frery et al., 2010]. O mecanismo de propagação é construído entre os nós origens e o sorvedouro, através de um mecanismo de comunicação sem fio *ad-hoc*. Essa comunicação possui restrições de energia, tempo de resposta e largura de banda. A Figura 4.25 ilustra a estrutura de uma rede de sensores sem fio, bem como os componentes externos que interagem com ela.



**Figura 4.25.** Ilustração de uma rede de sensores sem fio com seus elementos básicos e componentes externos que interagem com ela.

Por definição, a topologia de uma RSSF muda ao longo do tempo, pois alguns nós deixam de existir, por inatividade ou falhas, prejudicando a propagação dos dados. O objetivo, neste estudo, é caracterizar o mecanismo de propagação de dados em RSSFs através de medidas de redes complexas e estudar seu comportamento frente às mudanças ocorridas. As mudanças mais comuns são: adição e remoção de nós, chamadas em RSSFs de deposição e inatividade de nós, respectivamente e perturbações de arestas. Os conceitos sobre a propagação em RSSFs são apresentados na seção 4.3.2.

Alguns trabalhos da literatura desta área apresentam abordagens de propagação de dados em RSSFs modeladas como redes complexas. Helmy [2003] descreve as RSSFs como um grafo espacial mostrando que estas possuem uma tendência em formar agrupamentos. Além disso, conclui que os caminhos mínimos entre os pares de vértices são maiores que nos grafos aleatórios e mostra que é possível reduzir esse tamanho com a adição de alguns vértices, chamados de atalhos.

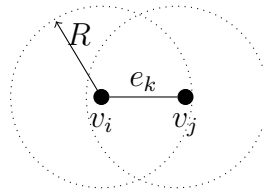
Guidoni et al. [2010] propõem um modelo *on-line* para projetar RSSFs heterogêneas com características de mundo pequeno, tendo como objetivo principal otimizar o tempo de propagação dos dados. O modelo proposto adiciona atalhos entre os nós e o sorvedouro, baseado no fluxo da propagação dos dados, obtendo assim uma rede com características de mundo pequeno. Dessa forma, permite a utilização mais eficiente da energia residual da rede e diminui a latência na propagação.

Ruela et al. [2010] apresentam uma abordagem para tornar a propagação dos dados numa RSSFs mais eficiente, a partir da inclusão de nós concentradores (*hubs*) no projeto da topologia de propagação de dados. A inclusão desses concentradores diminui o comprimento médio dos caminhos mínimos entre os nós e o sorvedouro, evitando uma sobrecarga de comunicação e reduzindo o tempo para entrega dos dados.

A seção que segue define os mecanismos de propagação de dados em redes de sensores sem fio.

#### 4.3.2 Topologias de propagação de dados em RSSFs

A conectividade física de uma RSSF pode ser descrita como um grafo geométrico não direcionado  $\mathcal{G}_{gr} = (\mathcal{V}, \mathcal{E})$ , em que  $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$  é o conjunto de nós sensores, e  $\mathcal{E} = \{e_k = (v_i, v_j) : 1 \leq i, j \leq N, 1 \leq k \leq M\}$  é o conjunto de ligações entre os nós. Cada ligação  $e_k$  é determinada pela geometria de comunicação, nesse caso uma circunferência, i.e., o conjunto de vizinhos do nó  $v_i$  é formado por todos os nós  $v_j$  que estão no máximo  $R$  unidades de distância de  $v_i$ , como ilustrado na Figura 4.26. Em RSSFs  $R$  é o raio de comunicação de um nó  $v_i$ .



**Figura 4.26.** Exemplo do estabelecimento de uma conexão  $e_k$  entre  $v_i$  e  $v_j$  em um grafo geométrico com raio  $R$ .

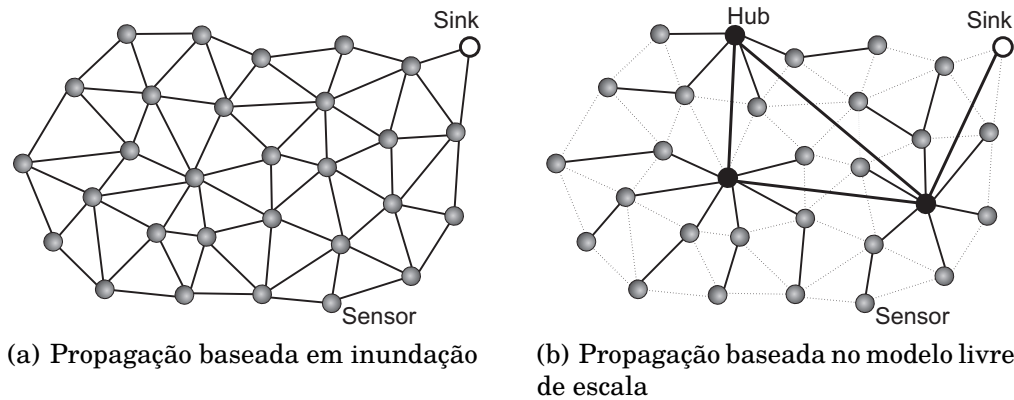
Os dados utilizam esta conectividade física para sua propagação, mas obedecendo as regras de um protocolo, que podem impor restrições sobre o uso das ligações disponíveis do conjunto  $\mathcal{E}$ , resultando em um subgrafo de comunicação  $\mathcal{G}'_{gr} = (\mathcal{V}', \mathcal{E}')$ , tal que  $\mathcal{E}' \subset \mathcal{E}$  e  $\mathcal{V}' \subset \mathcal{V}$ , portanto  $\mathcal{G}'_{gr} \subset \mathcal{G}_{gr}$ .

Duas formas de propagação de dados são avaliadas aqui:

**Propagação baseada em inundação (*flooding*):** utiliza todas as ligações disponíveis do conjunto  $\mathcal{E}$ , assim  $\mathcal{G}'_{gr} = \mathcal{G}_{gr}$ . Cada vez que um nó  $v_i$  recebe o dado, este é enviado para todos os seus vizinhos, menos aquele do qual ele recebeu. Cada nó propaga cada unidade de informação uma vez. Este processo se repete até que o dado chegue ao nó sorvedouro. A Figura 4.27(a) ilustra esta estratégia.

**Propagação baseada em redes complexas:** utiliza um subconjunto de vértices  $\mathcal{V}' \subset \mathcal{V}$  e ligações  $\mathcal{E}' \subset \mathcal{E}$ , obtendo-se  $\mathcal{G}'_{gr} = (\mathcal{V}', \mathcal{E}')$ . Cada vez que um nó  $v_i$  recebe um dado, este é enviado para o vizinho que está mais próximo do destino final, i.e., o fluxo dos dados vai sempre em direção aos nós que estão mais perto do sorvedouro. Este processo se repete até que o dado chegue ao sorvedouro.

Neste modelo de propagação, a estrutura de comunicação é determinada por propriedades topológicas da rede complexa que está sendo considerada (aleatória, de mundo pequeno ou livre de escala). A Figura 4.27(b) ilustra a propagação baseada no modelo livre de escala. Nesse caso o dado é propagado para o nó mais próximo do sorvedouro ou para um concentrador (*hub*), que possui ligações “especiais” que o tornam capaz de propagar a informação para nós mais distantes.



**Figura 4.27.** Exemplos de estratégias de propagação de dados em RSSFs.

Em alguns casos o caminho até o sorvedouro pode não ser o caminho mínimo, pois nem sempre é fácil encontrar a solução de roteamento ótima em cenários distribuídos, pois estes oferecem apenas informações locais.

Medidas específicas são utilizadas para caracterizar a propagação em RSSFs, nesse caso utilizou-se o comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ). Esta é uma medida importante, pois está diretamente relacionada ao consumo de energia da rede, pois mudanças abruptas na topologia alteram o caminho de roteamento dos dados até o sorvedouro. Vale destacar que a economia no consumo de energia é um dos principais requisitos do projeto de uma RSSF [Akyildiz et al., 2002].

#### 4.3.3 Discussão dos resultados

Esta seção apresenta a análise do comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ) em topologias de propagação de dados em RSSFs utilizando a divergência de Kullback-Leibler ( $\mathcal{D}_{KL}$ ) e a distância de Hellinger ( $\mathcal{D}_H$ ) apresentadas anteriormente<sup>†</sup>. Os resultados para distância de Jensen-Shannon ( $\mathcal{D}_{JS}$ ) foram omitidos, pois não apresentou variações significativas.

A análise das medidas, neste caso, considerou o cenário detalhado a seguir:

**Parâmetros de rede:** com o objetivo de simular RSSFs esparsas foram consideradas redes com  $N = 1000$  nós distribuídos em uma área  $A^2 = 100 \times 100$  e cada nó com raio de comunicação  $R = 5$  unidades na estratégia baseada em inundação. Dessa forma, obteve-se uma rede com densidade de aproximadamente  $d = 1.5$ , pois  $d = \pi r N / A^2$ , em que  $d$  representa o número de vizinhos de cada nó [Aquino et al., 2012].

As probabilidades de conexão  $p_c$  para geração das redes aleatórias foi 0.06, que implica um grau médio  $c = 6$ . As redes de mundo pequeno foram geradas com  $K = 3$  vizinhos e uma probabilidade de religação  $p_r = 0.3$ . As redes livres de escala foram geradas com uma aresta adicionada a cada passo da construção da rede, i.e.,  $m = 1$ .

**Perturbações:** na estratégia baseada em inundação foram aplicadas as perturbações – adição (ou deposição) e remoção (ou inatividade) de nó – com intensidades de  $\{1\%, 2\%, 3\%, \dots, 10\%\}$  do total de números de nós. Na estratégia baseada em redes complexas foram aplicadas as perturbações – adição, remoção e religação de arestas – com intensidades de  $\{1\%, 2\%, \dots, 10\%\}$  do total do número de arestas. Vale destacar que todos os tipos de perturba-

---

<sup>†</sup>A definição de  $\ell$  encontra-se na seção 2.1, pg. 7; e a definição dos quantificadores  $\mathcal{D}_{KL}$  e  $\mathcal{D}_H$  foi apresentada na seção 2.4, pq. 23.



ções aplicadas, nesse caso, preservam a geometria da rede. A Tabela 4.8 sumariza a combinação dos mecanismos de propagação vs. perturbações.

**Tabela 4.8.** Mecanismos de propagação vs. perturbações

	Inatividade	Deposição	Inserção	Remoção	Religação
Inundação	✓	✓	–	–	–
Aleatória	–	–	✓	✓	✓
Mundo pequeno	–	–	✓	✓	✓
Livre de escala	–	–	✓	✓	✓

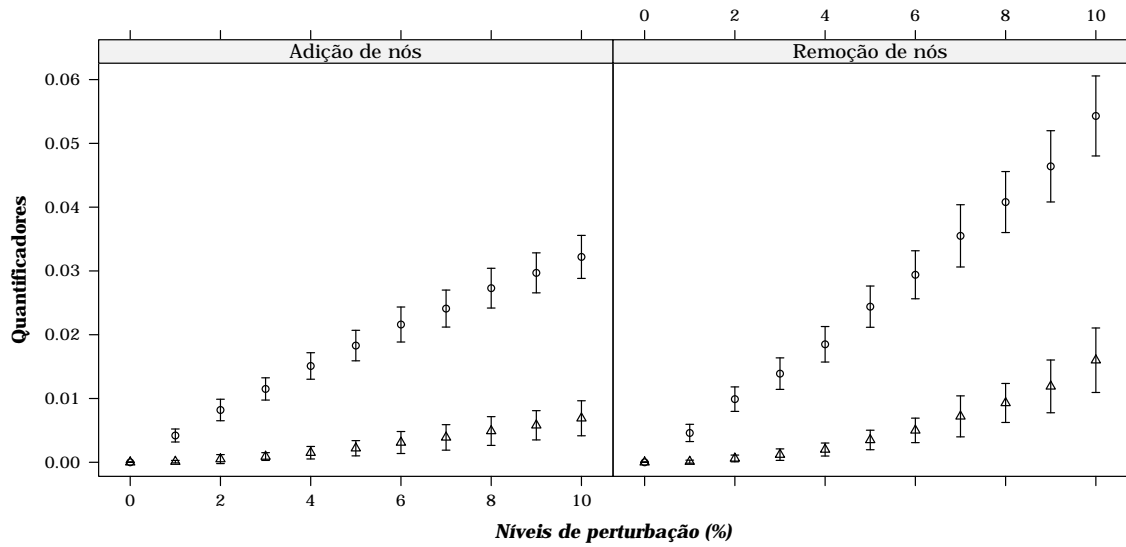
**Histogramas:** os histogramas de proporções  $\mathcal{H}(\cdot)$  foram obtidos com 200 classes de mesma largura. Como já mencionado, a divergência de Kullback-Leibler ( $\mathcal{D}_{KL}$ ) diverge para  $q(\xi) = 0$  e  $p(\xi) \neq 0$ , então deve-se adicionar um parâmetro  $\delta = 0.001$  a cada classe do histograma. Para a distância de Hellinger ( $\mathcal{D}_H$ ) utilizou-se o histograma original.

**Parâmetros de simulação:** para cada classe de rede considerada foram geradas 10 amostras diferentes e cada uma delas perturbada 100 vezes, assim tem-se  $G = 10$  e  $r_{\max} = 100$  no Algoritmo 1. Assim sendo, os resultados puderam ser apresentados com intervalos de confiança assintóticos e simétrico ao nível de significância de 95%.

**Recursos computacionais:** o algoritmo foi implementado na linguagem R [R Development Core Team, 2011] com o uso da biblioteca *igraph* para manipulação dos grafos [Csardi & Nepusz, 2006]. As simulações foram executadas em um computador Intel(R) Core(TM) i5 CPU 760 2.80 GHz com 8 GB RAM e sistema operacional Ubuntu 12.04 (64 bits).

As Figuras 4.28 e 4.29 apresentam a variação dos quantificadores em relação a  $\ell$  para as diferentes estratégias de propagação e perturbações consideradas. Cada figura é composta por vários gráficos que estão sincronizados tanto no eixo das abscissas quanto das ordenadas, o que os torna comparáveis.

**Propagação baseada em inundação** A Figura 4.28 apresenta os resultados para a estratégia baseada em inundação e é composta por dois gráficos: da esquerda para direita, adição e remoção de vértices. Cada um deles apresenta a variação dos quantificadores utilizados (eixo das ordenadas) em relação ao nível de perturbação aplicada (eixo das abscissas) à rede.



**Figura 4.28.** Variação da divergência de Kullback-Leibler ( $\Delta$ ) e distância de Hellinger ( $\circ$ ) para o comprimento dos caminhos mínimos ( $\ell$ ) na estratégia de propagação baseada na propagação por inundação.

Tanto a adição como a remoção de vértices alteram o comprimento dos caminhos mínimos  $\ell$ . A variabilidade dos quantificadores tem dependência com o nível de perturbação aplicado; eles aumentam quando o nível de perturbação aumenta, como observado na Figura 4.28. A variação dos quantificadores foi pequena, porém significativa.

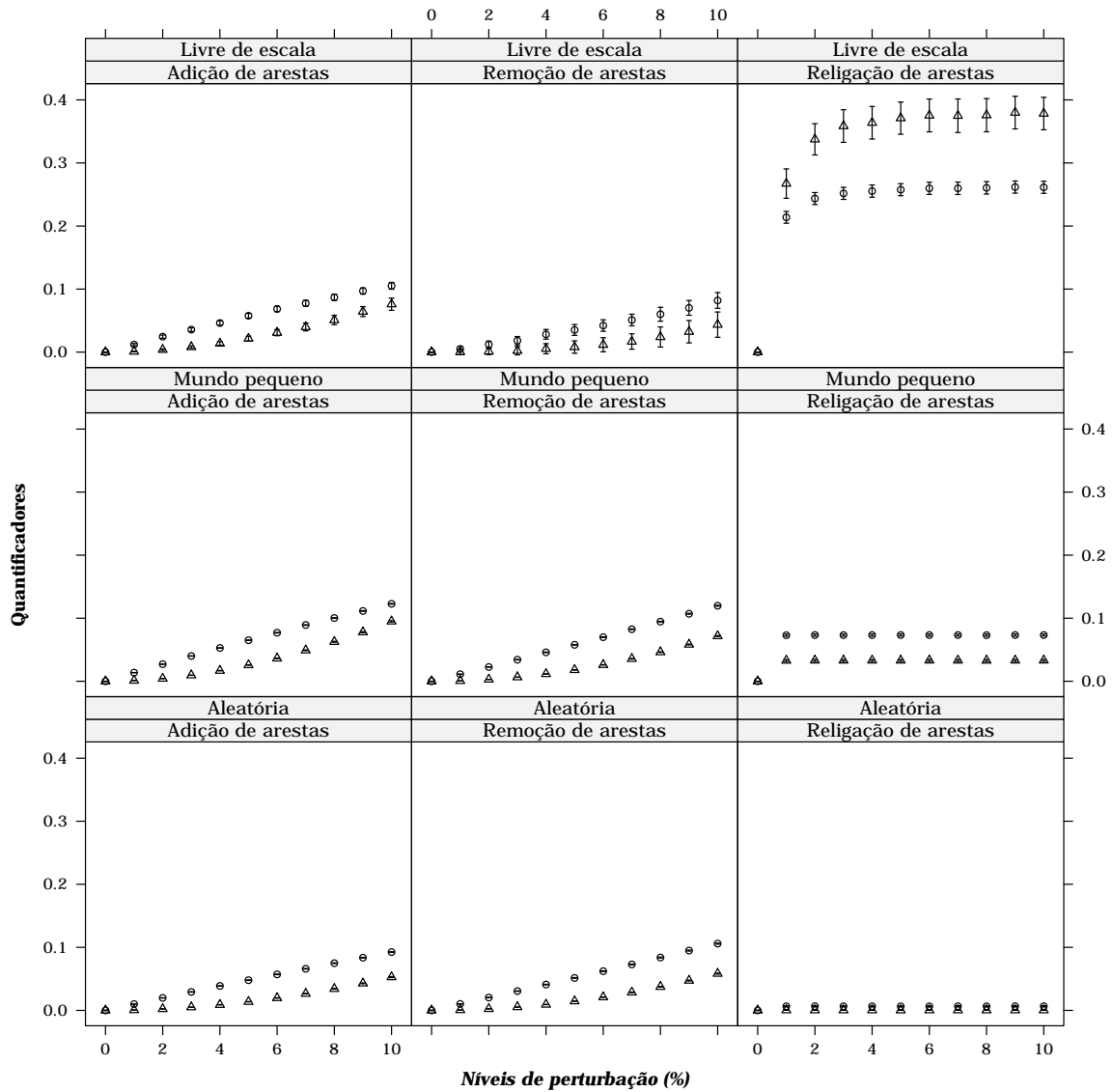
A remoção de nós teve um impacto maior sobre o comprimento dos caminhos mínimos comparada à adição de nós (sendo a variação do último mais acentuada que no primeiro); este efeito é mais visível em níveis mais altos de perturbação. Isto pode ser explicado pelo fato de que, quando novos nós forem adicionados dentro do raio de comunicação de um nó, a alteração de  $\ell$  é pequena. Porém, quando nós são removidos, muitas ligações que faziam parte dos caminhos mínimos são perdidas, alterando significativamente os valores de  $\ell$ .

A distância de Hellinger é mais sensível as perturbações que a divergência de Kullback-Leibler. Os valores máximos dos quantificadores para a adição de nós foram  $\mathcal{D}_H = 0.032$  e  $\mathcal{D}_{KL} = 0.006$  e para a remoção  $\mathcal{D}_H = 0.054$  e  $\mathcal{D}_{KL} = 0.001$ , todos no nível máximo de perturbação 10%.

Os resultados apresentados aqui são importantes no projeto de RSSFs. Baseado neste estudo pode-se ter um controle maior sobre o projeto de uma topologia. Por exemplo, quando 10% dos nós são adicionados à rede, os caminhos mínimos são alterados indicando que a adição de uma porcentagem maior de nós na rede original pode vir a ser prejudicial. Por outro lado, quando 5% dos nós tornam-se inativos (removidos) os caminhos mínimos se alteram. Esta informação pode ser utilizada pelo projetista para calibrar o ciclo de trabalho (*duty cycle*) dos nós, em que o estado de cada nó é periodicamente trocado entre os modos ativo (*awake*) e inativo (*sleeping*) [Lai & Ravindran, 2010].

**Propagação baseada em redes complexas** A Figura 4.29 apresenta os resultados para as estratégias baseadas em redes complexas sendo um conjunto de 9 gráficos ou uma grade de gráficos de tamanho  $3 \times 3$ . As linhas da grade correspondem aos três tipos de modelos de redes complexas (de baixo para cima) aleatório, mundo pequeno e livre de escala; enquanto as colunas da grade descrevem os três tipos de perturbações aplicadas a estes modelos (da esquerda para a direita) adição, remoção e religação de arestas. Cada gráfico do conjunto apresenta a variação dos quantificadores utilizados (eixo das ordenadas) em relação ao nível de perturbação aplicada (eixo das abscissas) a rede.

O comprimento do caminho mínimo é sensível à adição e remoção de ligações nos três modelos de propagação baseados em redes complexas, e os quantificadores  $\mathcal{D}_{KL}$  e  $\mathcal{D}_H$  apresentam o mesmo comportamento: quanto maior o nível de perturbação, maior são os valores dos quantificadores.



**Figura 4.29.** Variação da divergência de Kullback-Leibler ( $\Delta$ ) e distância de Hellinger ( $\circ$ ) para o comprimento dos caminhos mínimos ( $\ell_{(i,j)}$ ) nas estratégias de propagação baseadas em redes complexas.

Em relação a esses dois tipos de perturbações, a distância de Hellinger exibiu valores maiores em relação a divergência de Kullback-Leibler; note que os intervalos de confiança não se sobrepõem. A Tabela 4.9 mostra os valores médios para o nível mais intenso de perturbação (10%), que corresponde ao ponto mais a direita dos gráficos da Figura 4.29. A distância de Hellin-

ger é consistente e significativamente maior do que a divergência Kullback-Leibler por um fator de, aproximadamente, dois, em todas as situações em que a adição da ligação e remoção são aplicadas.

**Tabela 4.9.** Média dos valores dos quantificadores para o nível de perturbação mais intensa (10%) para o comprimento do caminho mínimo  $\ell_{(i,j)}$ .

	$\mathcal{D}_H$			$\mathcal{D}_{KL}$		
	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$
$\mathcal{G}_{ws}$	0.122	0.119	0.073	0.095	0.071	0.033
$\mathcal{G}_{ba}$	0.105	0.082	0.261	0.076	0.043	<b>0.378</b>
$\mathcal{G}_{er}$	0.092	0.105	0.006	0.052	0.058	$10^{-4}$

O comportamento dos quantificadores para a adição e remoção de ligações é o mesmo nos três modelos. Pode-se dizer que ambos alteram diretamente os caminhos mínimos entre os nós, já que estas perturbações alteram as distâncias entre os nós.

A religação de arestas apresenta um comportamento diferente, como é observado na terceira coluna da Figura 4.29. Nas redes aleatórias, embora os quantificadores apresentem diferenças, como visto na Tabela 4.9, a sua variação é desprezível. Já as redes de mundo pequeno são sensíveis a esta perturbação, mas o seu efeito é constante, i.e., independe da intensidade da perturbação. Com respeito as redes livres de escala, existe uma grande variação nos quantificadores, mesmo para os níveis mais baixos de perturbação, mas os valores tendem a se estabilizar. Este é o único caso em que a divergência de Kullback-Leibler apresenta valores maiores que a distância de Hellinger.

Todos as redes analisadas representam uma topologia de roteamento gerada por algum mecanismo de gerenciamento. Geralmente, essas estratégias combinam o comprimento do caminho mínimo com outros parâmetros, tais como energia, atraso e prioridade. Os resultados revelaram que, quando aproximadamente 5% de ligações são adicionadas ou removidas, a rede começa a apresentar alterações significativas em  $\ell_{(i,j)}$ . Vale destacar que essa informação pode ser utilizada para prover um melhor gerenciamento e evitar a interferência dos parâmetros de QoS, durante a geração da topologia de roteamento, uma vez que esta tarefa considera apenas as ligações que serão adicionadas ou removidas.

## 4.4 Conclusões

A análise realizada neste capítulo permitiu identificar o relacionamento entre a resistência das redes as perturbações aplicadas, concluindo-se que o comprimento dos caminhos mínimos pode ser muito sensível. Esse fato é importante, pois essa medida é base do funcionamento de muitas aplicações, como por exemplo nas redes de sensores sem fio. Além disso, é utilizada para definição de outras medidas, tais como vulnerabilidade, eficiência e medidas de centralidade.

Os modelos teóricos são caracterizadores dos mais diversos sistemas e a avaliação destes é útil na previsão do comportamento de várias aplicações reais. Como exemplos, pode-se citar a Internet, a WWW, as redes sociais e rotas ferroviárias que são consideradas redes livres de escala.

Os testes de hipótese permitiram identificar as perturbações que levaram à quebra da distribuição dos graus de uma rede, no que se refere às redes aleatórias, de mundo pequeno e livres de escala.

O uso de quantificadores que envolvem logaritmos pode não ser uma boa escolha para este tipo de caracterização, pois a ocorrência de zeros acarretam em problemas numéricos e, possivelmente, a uma interpretação incorreta do comportamento das redes em relação às perturbações.

**Tabela 4.6.** Comportamento geral dos quantificadores para todos os modelos e perturbações avaliadas. Cada célula corresponde ao resultado de uma medida aplicada a uma rede para um tipo de perturbação e um quantificador estocástico. As células marcadas com + indicam que a medida é sensível e com – que não é sensível. As células em cinza indicam o quantificador que apresentou maior valor.

		$\kappa_i$				$C_i$				$B_i$				$\ell_{(i,j)}$			
		$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{rn}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{rn}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{rn}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{we}$	$\pi_{rn}$
$\mathcal{G}_{er}$	$\mathcal{D}_H$	+	+	–	+	–	–	–	+	+	+	+	+	+	+	–	+
	$\mathcal{D}_{JS}$	+	+	–	+	–	–	–	–	+	+	+	+	+	+	–	+
	$\mathcal{D}_{KL}$	+	+	–	+	–	–	–	+	+	+	+	+	+	+	–	+
$\mathcal{G}_{ws}$	$\mathcal{D}_H$	+	+	–	+	–	–	–	+	+	+	+	+	+	+	+	+
	$\mathcal{D}_{JS}$	+	+	–	+	–	–	–	–	+	+	+	+	+	+	–	+
	$\mathcal{D}_{KL}$	+	+	–	+	–	–	–	+	+	+	+	+	+	+	+	+
$\mathcal{G}_{ba}$	$\mathcal{D}_H$	+	+	–	+	–	–	–	–	+	+	+	+	+	+	–	+
	$\mathcal{D}_{JS}$	+	+	–	+	–	–	–	–	+	+	+	+	+	+	–	+
	$\mathcal{D}_{KL}$	+	+	–	+	–	–	–	–	+	+	+	+	+	+	–	+

**Tabela 4.7.** Comportamento geral dos modelos teóricos para as perturbações avaliadas. Cada célula corresponde ao resultado de uma medida aplicada a uma rede para um tipo de perturbação para a distância de Hellinger. As células marcadas com  $x$  indicam o modelos mais sensíveis.

	$\kappa_i$				$C_i$				$B_i$				$\ell_{(i,j)}$			
	$\pi_{ae}$	$\pi_{re}$	$\pi_{rw}$	$\pi_{rn}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{rw}$	$\pi_{rn}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{rw}$	$\pi_{rn}$	$\pi_{ae}$	$\pi_{re}$	$\pi_{rw}$	$\pi_{rn}$
$\mathcal{G}_{er}$	x															
$\mathcal{G}_{ws}$		x		x				x								
$\mathcal{G}_{ba}$											x		x		x	



## Capítulo 5

### Conclusões

A análise da variabilidade de medidas em redes complexas fornece uma visão do comportamento de uma classe de redes frente a perturbações. Essa análise pode ser de grande utilidade no projeto de soluções em redes, visto que fornece uma descrição elucidativa a respeito do funcionamento de uma aplicação. Neste trabalho, foram utilizadas ferramentas da Teoria da Informação, em particular, a divergência de Kullback-Leibler, as distâncias de Jensen-Shannon e Hellinger para comparar diferentes estados dos modelos de redes complexas, sendo estas denominadas aleatória, de mundo pequeno, livre de escala; e também em redes com topologia geométrica. Foram simuladas diferentes adversidades que podem ocorrer em uma rede, chamadas de perturbações. Estas foram divididas em dois tipos: perturbações de nós (adição e remoção), inseridas em redes com topologia geométrica; e perturbações de arestas (adição, remoção e religação), aplicadas aos modelos clássicos de redes complexas.

Para entender o impacto dessas perturbações nas redes foram avaliadas quatro medidas: o grau do vértice, o coeficiente de agrupamento local, o *betweenness centrality*, o comprimento dos caminhos mínimos e o comprimento médio dos caminhos mínimos. Cada uma delas foi aplicada às redes e submetidas a diferentes níveis de perturbação.

Nos cenários avaliados, os resultados mostraram que o coeficiente de agrupamento local é uma medida robusta ou insensível, pois não sofre alteração. Por outro lado, o grau do vértice e as medidas relacionadas a distância apresentaram sensibilidade, e em muitas foram alteradas de acordo com a intensidade da perturbação aplicada.

Além disso, os testes de hipótese realizados identificaram quais perturbações levam a quebra da distribuição dos graus, uma propriedade fundamental

das redes complexas.

Constatou-se também que o uso de quantificadores que envolvem logaritmos ou divisões pode não ser uma boa escolha para este tipo de caracterização, já que a ocorrência de zeros acarreta problemas numéricos e, possivelmente, conduz a uma interpretação incorreta dos resultados.

Essa metodologia também foi utilizada na análise de medidas em diferentes mecanismos de propagação de dados, em redes de sensores sem fio, o que permitiu uma visão mais ampla do seu comportamento e funcionamento. Nesse caso, foram utilizadas a divergência de Kullback-Leibler e a distância de Hellinger para a comparação dos estados da rede lógica em relação aos comprimentos dos caminhos mínimos. As estratégias de propagação consideradas foram de dois tipos: (i) baseada em inundação, mais comum em redes de sensores sem fio; (ii) e baseada em redes complexas (modelos aleatório, de mundo pequeno e livre de escala). Para a primeira estratégia baseada em inundação aplicou-se as perturbações de nós; enquanto que para a segunda as perturbações de arestas. Com isso, identificou-se o comportamento dos caminhos mínimos em diferentes níveis de perturbações.

Essa modalidade de avaliação permite o projeto de soluções mais adequadas para cada tipo de aplicação. Ademais, considerando diferentes tipos de perturbações, identificou-se o quanto as redes são resistentes a mudanças, sendo possível projetar soluções mais robustas e eficientes para tais perturbações.

Para trabalhos futuros, sugere-se: (i) a análise de outras medidas em modelos teóricos e aplicações; (ii) identificação de níveis de perturbação que levam a quebra de suas propriedades; (iii) a utilização de outros quantificadores estocásticos; (iv) o uso de aproximações contínuas para as distribuições de probabilidades (histogramas); (v) em particular, para redes de sensores sem fio sugere-se a avaliação da sensibilidade com variante na densidade das redes, bem como a proposição de soluções mais robustas como por exemplo, algoritmos de roteamento que considerem as informações fornecidas por medidas de redes complexas.

Anexo A

Artigo aceito para publicação

Cent. Eur. J. Phys. • 1-12  
Author version

Central European Journal of Physics

## Structural Changes in Data Communication in Wireless Sensor Networks

Editorial

Raquel S. Cabral <sup>1</sup> \*, Andre L. L. Aquino <sup>2</sup> †, Alejandro C. Frery <sup>2</sup> ‡  
Osvaldo A. Rosso <sup>2,3</sup> §, Jaime A. Ramírez <sup>1</sup> ¶

<sup>1</sup> Graduate Program in Electrical Engineering - Federal University of Minas Gerais - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil

<sup>2</sup> Scientific Computing and Numerical Analysis Laboratory - Federal University of Alagoas - Av. Lourival Melo Mota, s/n, 57072-900, Maceió, AL, Brazil

<sup>3</sup> Complex Systems Laboratory, Engineering Faculty - University of Buenos Aires - Paseo Colón 850, Buenos Aires, Argentina

**Abstract:** Wireless sensor networks are an important technology for making distributed autonomous measures in hostile or inaccessible environments. Among the challenges they pose, the way data travel among them is a relevant issue since their structure is quite dynamic. The operational topology of such devices can often be described by complex networks. In this work, we assess the variation of measures commonly employed in the complex networks literature applied to wireless sensor networks. Four data communication strategies were considered: geometric, random, small-world, and scale-free models, along with the shortest path length measure. The sensitivity of this measure was analyzed with respect to the following perturbations: insertion and removal of nodes in the geometric strategy; and insertion, removal and rewiring of links in the other models. The assessment was performed using the normalized Kullback-Leibler divergence and Hellinger distance quantifiers, both deriving from the Information Theory framework. The results reveal that the shortest path length is sensitive to perturbations.

**PACS (2008):** 89.75.-k, 89.70.-a, 02.50.-r

**Keywords:** complex networks • structural measures • stochastic quantifiers • information theory quantifiers

© Versita sp. z o.o.

\* E-mail: raquelcabral@gmail.com

† E-mail: alla.lins@gmail.com

‡ E-mail: acfrery@gmail.com

§ E-mail: oarosso@gmail.com

¶ E-mail: jramirez@ufmg.br

## I. Introduction

Wireless Sensor Networks (WSNs) are an emerging technology that allows the monitoring of physical variables, such as temperature, sound, light, vibration, pressure or movement [1]. A WSN consists of a large number of wireless autonomous devices, called “sensor nodes”, “sensors” or “nodes.” These entities work in a cooperative way sensing the environment, communicating among them, and taking decisions. Such networks are a promising technology in a wide range of applications, for instance, biotechnology, industry, public health, and transportation [2]. They are of particular interest for monitoring hostile or inaccessible environments [3].

The physical variables are monitored and stored in the sensor, and propagated to a sink node. The sink is a management node responsible for processing the data and delivering it to an external user [4]. The communication uses the nodes between the sensor source and sink in an ad-hoc fashion. The two most common communication strategies are based on data flooding and on complex networks. In the *flooding based*, the data communication starts from a node to its direct neighbors, then each neighbor re-propagates to the next neighbors, and so on. Each node propagates each information once and this processes repeats until the data arrives to the sink node. Fig. 1(a) illustrates this communication strategy.

In *complex network based*, the communication depends on the model used to characterize it (random, small-world or scale-free, for example). So, specific topological properties of complex networks are used to determine this kind of communication. Fig 1(b) shows a complex network based on scale-free model. An important issue in this model is the presence of nodes with high degree known as “hubs”. The communication is towards the nodes closer to the sink or to a hub. The hubs can propagate the information to distant nodes and, consequently, the information flow is concentrated on them.

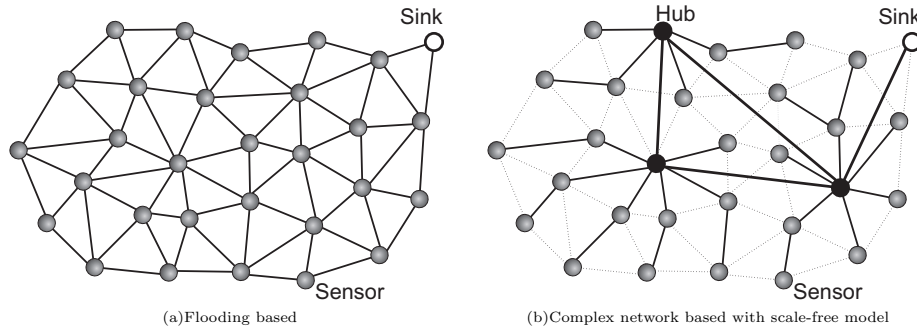


Figure 1. Examples of WSNs communications

The WSN communication strategies can be characterized by a set of measures. These measures describe different features on the network, such as connectivity, centrality, cycles or distances [5]. Example of measures are the average shortest path length, clustering coefficient, network diameter and betweenness. In this work, we use the

Raquel S. Cabral<sup>\*</sup>, Andre L. L. Aquino<sup>\*</sup>, Alejandro C. Frery<sup>\*</sup>  
 Osvaldo A. Rosso<sup>\*</sup>, Jaime A. Ramírez

shortest path length, because it is directly related to the network energy consumption and sudden changes in this measure may result in an increase or decrease in energy consumption.

Energy availability is a critical feature in WSN [1]. Moreover, classical routing solutions in WSNs employ the shortest paths [6]. If the shortest paths change, the routing solutions may become inefficient.

The communication structure in WSNs is dynamic, i.e., it changes over time. It is, therefore, important to analyze how this reflects on the measures. The most common changes are: addition and inactivity of nodes, in the flooding based communication; and insertion, removal and rewiring of links, in the complex network based communication [7]. These changes have a great impact over real WSNs design. The infrastructure mechanisms usually are designed considering topological characteristics. For instance, in a routing algorithm based on shortest path length if a node is removed the routing behaviour is affected.

In this way, the main question of this work is:

*“What is the impact of changes or perturbations on these strategies, as measured by the shortest path length?”*

This work presents the analysis of communication strategies in WSNs by means of analyzing the variation of measures. We analyze flooding, random, small-world, and scale-free networks, and the shortest path length. The variation of measures was analyzed with respect to the insertion and removal of nodes in flooding; and with respect to insertion, removal and rewiring of links in the strategy based in complex networks. Stochastic quantifiers, namely the normalized Kullback-Leibler divergence and Hellinger distance [8], were used to quantify the variation of the measure. The results reveal to which extent the measure is influenced by the perturbations considered: the shortest path length exhibits a clear dependence on the type and intensity of the perturbation.

The paper is organized as follow. Section II presents the related work. Section III analyses communication strategies in WSNs. Section IV presents the results of measures behavior and the quantifiers performance. Finally, conclusions and future directions are presented in Section V.

## II. Related work

This section presents a brief review about WSNs communication strategies that use complex networks concepts, and stochastic quantifiers applied to networks.

### II.1. WSNs communication strategies

Helmy [9] described WSNs as spatial graphs that tend to be much more clustered and with higher path lengths than random graphs. He showed that it is possible to reduce the path length of wireless networks with the addition of a few short cut links.

Guidoni et al. [10] proposed on-line models to design heterogeneous sensor networks with small world features. The proposed model takes into account the data communication flow in this kind of networks to create shortcuts

towards the sink in such a way that the communication between the sink and the sensor nodes is optimized. The network presents better small world features when the shortcuts are created, and an interesting trade off between energy and communication latency is observed.

Ruela et al. [11] improved the data communication by using hubs, an approach based on scale-free networks. The hubs introduce useful characteristics of complex networks, e.g. small average shortest path length between all sensors and the sink, and high clustering coefficient. This strategy saves resources, avoiding excessive communication and, consequently, reducing the time to data delivery.

## II.2. Information Theory quantifiers

Wang et al. [12] used entropy to provide an average measure of networks heterogeneity since it measures the diversity of the link distribution.

Boas et al. [7] analyzed the effect of perturbations in complex networks. They choosed measures based on the fact that the network characterization is made from samples rather than from the entire network. They applied three perturbations: link addition, link removal, and link rewiring.

Carpi et al. [13] proposed a new quantifier based on Information Theory for the analysis of dynamic network evolution. It is used to compute changes in topological randomness for degree distribution of the network. The quantifier, a statistical complexity measure, is obtained by the product of the normalized Shannon entropy and the normalized Jensen-Shannon distance.

In this paper, we use two stochastic quantifiers, a divergence and a distance, to quantify the changes in WSNs. We show that it is possible to identify the strength of the perturbations that leads to a breakdown of the network properties, i.e, a kind of phase transition. Moreover, we evaluate the flooding communication strategy considering node perturbations. These aspects are not addressed in the related works mentioned and constitute the main contributions.

## III. Analysis of communication strategies in WSNs

The WSN physical connectivity can be described as an undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$  is the set of sensor nodes, and  $\mathcal{E} = \{(v_i, v_j) : 1 \leq i, j \leq N\}$  is the set of links between nodes. Each link is determined by the communication geometry. In our case it is a circumference, i.e., the neighborhood of node  $v_i$  is formed by all nodes  $v_j$  at most  $R$  units away from  $v_i$ . The data employ this physical connectivity to propagate, but obeying the rules of a protocol. The protocol may impose restrictions on the use of the available links  $\mathcal{E}$ , resulting in a communication subgraph  $G = (\mathcal{G}, E)$ , such that  $E \subset \mathcal{E}$  and, therefore,  $G \subset \mathcal{G}$ .

The number of connections of a node is its degree  $\kappa$ . An important graph characteristic is its degree distribution, i.e., the collection of all its degrees.

The *flooding based communication* employs all the available links, so  $G = \mathcal{G}$ . The flooding starts propagating

Raquel S. Cabral<sup>\*</sup>, Andre L. L. Aquino<sup>\*</sup>, Alejandro C. Frery  
Oswaldo A. Rosso<sup>\*</sup>, Jaime A. Ramírez

the data from  $v_i$  to all its neighbors, then each neighbor re-propagates to its neighbors, and so on. Each node propagates each unit of information once. This process repeats until the data arrives to the sink node.

The communication based on *complex networks* employs a subset of links  $E \subset \mathcal{E}$  such that data flow towards the nodes which are closest to the sink. The source nodes always send the information to that neighbor which reduces the distance to the sink. Each node propagates the information following the same criterion. This process repeats until the data arrives to the sink node.

In some cases the path to the sink is not the optimal, because it is not always immediate to find the optimal routing solution in distributed scenarios. These scenarios offer only local information, a limitation which often prevents finding global optima. Specific measures of complex networks are used to characterize this kind of communication, for instance, the shortest path length [14, 15]. A shortest path is any path that connects two nodes and has minimal length. It is an important measure in communication strategies.

There are several complex networks models. In this work, we use the random, small-world, and scale-free models:

**Random model:** The probability to connect each pair of nodes is the same. There are two ways to build a random graph [16]:  $N$  nodes and exactly  $M$  links, and  $N$  nodes and the probability  $0 < p_c < 1$  to connect each pair of nodes. In the first description,  $M$  links are uniformly distributed among the  $N(N-1)/2$  possibilities. In the second description, which is the one we adopted here, we start with a totally disconnected graph and then connect each pair of nodes with probability  $p_c$ . In this case, the probability of observing  $0 \leq k \leq N-1$  connections in each node follows a Binomial distribution with  $N-1$  trials and probability of success  $p_c$ , leading to  $(N-1)p_c$  as mean degree.

**Small-world model:** A communication strategy is small-world when the communication network presents a high clustering coefficient and a small shortest path length [17]. There are several ways to build such strategy; in this work, we use the Watts-Strogatz model. This model starts with a circular regular topology with  $N$  nodes, each one connected to the  $k$  nearest neighbors in each direction (right or left in circular topology). Then, each link is randomly “rewired” with probability  $p_r$  [5], i.e., if the current link is  $(v_i, v_j)$  and there is no link  $(v_i, v_k)$ , rewiring consists in deleting  $(v_i, v_j)$  and creating  $(v_i, v_k)$ .

**Scale-free model:** A communication strategy is scale-free when the communication network displays a power law degree distribution  $p(\kappa) \sim \kappa^{-\lambda}$ ,  $\kappa > 0$ , with  $2 < \lambda < 3$ . The main feature of this topology is the presence of some nodes with high degree, often called “hubs”. To generate this topology we use the Barabási-Albert scale-free model that starts with a small number of nodes with  $m_0$  links, and in the next step a new node with  $m$  links is added to the network (with  $m \leq m_0$ ). The probability connection is linearly proportional to the node degree [18].

The main question stated in this work is: “*What is the impact of changes or perturbations on these strategies, as measured by the shortest path length?*” To answer this question, we need to quantify the variation of the measure when perturbations occur. The most common changes, or perturbations, are: the addition and inactivity of nodes,



in the flooding based communication; and insertion, removal and rewiring of links, in the complex network based communication.

The stochastic nature of the aforementioned quantifiers suggests the use of techniques deriving from Information Theory in order to assess their change. The normalized Kullback-Leibler divergence and Hellinger distance are two quantifiers suitable for describing the difference between distributions [8].

Consider the discrete random variables  $X$  and  $Y$  defined on the same sample space  $\Omega = \{\xi_1, \xi_2, \dots, \xi_n\}$ . The distributions are characterized by their probability functions  $p, q: \Omega \rightarrow [0, 1]$ , where  $p(\xi_i) = \Pr(X = \xi_i)$  and  $q(\xi_i) = \Pr(Y = \xi_i)$ . A metric  $\mathcal{D}$  between these two distributions is a quantifier obeying:

1.  $\mathcal{D}(p, p) = 0$ , reflexivity;
2.  $\mathcal{D}(p, q) \geq 0$ , non-negativity;
3.  $\mathcal{D}(p, q) = \mathcal{D}(q, p)$ , commutativity;
4.  $\mathcal{D}(p, q) \leq \mathcal{D}(p, r) + \mathcal{D}(r, q)$ , triangle inequality for any other probability function  $r$  defined on the same probability space.

A distance is not required to satisfy property 4, and a divergence is only required to satisfy properties 1 and 2 [19].

Assuming  $q(\xi) > 0$  for every event  $\xi \in \Omega$ , the Kullback-Leibler divergence is defined as

$$D_{KL}(p, q) = \sum_{\xi \in \Omega} p(\xi) \log \frac{p(\xi)}{q(\xi)}. \quad (1)$$

The Hellinger distance does not impose positivity on the probabilities; it is defined as

$$\mathcal{D}_H(p, q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{\xi \in \Omega} (\sqrt{p(\xi)} - \sqrt{q(\xi)})^2} = 1 - \sum_{\xi \in \Omega} \sqrt{p(\xi)q(\xi)}. \quad (2)$$

In order to make the Kullback-Leibler divergence (an unbounded positive quantity) and the Hellinger distance (which is confined to the  $[0, 1]$  interval) comparable, in the remainder of this work we will use the normalized Kullback-Leibler distance defined as  $\mathcal{D}_{KL}(p, q) = 1 - \exp\{-D_{KL}(p, q)\}$ .

## IV. Simulation results

This section presents the simulation study about the complex network measures behavior of WSNs using two quantifiers presented previously.

### IV.1. Methodology

Simulation assumptions and parameters were:

Raquel S. Cabral<sup>\*</sup>, Andre L. L. Aquino<sup>\*</sup>, Alejandro C. Frery  
Oswaldo A. Rosso<sup>\*</sup>, Jaime A. Ramírez

**Network parameters:** In order to simulate a sparse WSNs, we used  $N = 1000$  nodes deployed in an  $L^2 = 100 \times 100$  area. The communication radius of each node was  $R = 5$  units in strategies based in flooding communication. With these values, we got a network density approximated 1.5 obtained through  $d = \pi R N / L^2$ , where  $d$  represents the number of neighbours of each nodes [20]. The probability of connection in the random model was  $p_c = 0.06$ , which implies an average degree equal to 6. The nearest neighbors and probability rewiring in the small-world model were  $k = 3$  and  $p_r = 0.3$ , respectively. The number of links added in each step in the scale-free model was  $m = 1$ . The above parameters were chosen according to Boas et.al. [7].

**Perturbations:** In the strategies based in flooding communication we performed: (i) *nodes addition*, nodes were randomly added to the network; and (ii) *nodes removal*, nodes were randomly removed from the network, that represents the node inactivity. The perturbations were performed on  $\{1\%, 2\%, \dots, 10\%\}$  of the total number of nodes. In the complex networks based communication, we performed tree types of links perturbations [7]: (i) *link removal*, links were randomly removed from the network; (ii) *link addition*, two unconnected nodes were randomly selected, and a new link was established; and (iii) *link rewiring*. In this last case, the perturbations were performed in  $\{1\%, 2\%, \dots, 10\%\}$  of the total number of links.

**Normalization:** for each network, the normalized histogram  $\mathcal{H}$ , also known as histogram of proportions, was obtained with 200 bins of equal width. The  $\mathcal{D}_{KL}$  diverges for  $q(\xi) = 0$  and  $p(\xi) \neq 0$ , as defined in equation (1). In order to avoid the division by zero, a small positive constant  $\delta = 0.001$  was added to each bin, and then the histogram is normalized to add 1 [7, 21]. The original histogram is used to compute the Hellinger distance once it does not impose the positivity restriction on the probabilities.

**General parameters:** for each communication strategy, we generated 10 different networks and for each network 100 different perturbations were made. In this way, we are able to present the mean results with symmetrical asymptotic confidence intervals at the 95% significance level.

**Computational resources:** we performed the evaluation using the R platform [22], on an Intel(R) Core(TM) i5 CPU 760 2.80 GHz with 7 GB RAM, running Ubuntu 12.04 (64 bits). The **igraph** library was used to generate and modify the graphs [23].

Figures 2 and 3 present the variation of the shortest path length, for the different communication strategies and the perturbations considered. Each plot presents two quantifiers: the normalized Kullback-Leibler divergence  $\mathcal{D}_{KL}$  (denoted as “ $\Delta$ ”), and the Hellinger distance  $\mathcal{D}_H$  (denoted as “ $\circ$ ”), as functions of the level (intensity) of the perturbation.

## IV.2. Flooding communication

Figure 2 shows the shortest path variation in flooding communication; from left to right, node addition and node removal. Both addition and removal alter the shortest path length, the bigger the level of perturbation the

stronger the change. The Hellinger distance is more sensitive to changes in the shortest path length than the normalized Kullback-Leibler divergence. The variability of both measures has some dependence on the level of perturbation; it increases with the level of perturbation but then stabilizes. The removal of nodes has stronger impact on the shortest path length than node addition (the former variation is steeper than the latter); this effect is more noticeable with higher levels of perturbation. This may be explained by the fact that when new nodes are added within the communication range of a node, the shortest path is little affected. The shortest path length changes when nodes are removed because many links are lost and some of them belonged to the shortest paths.

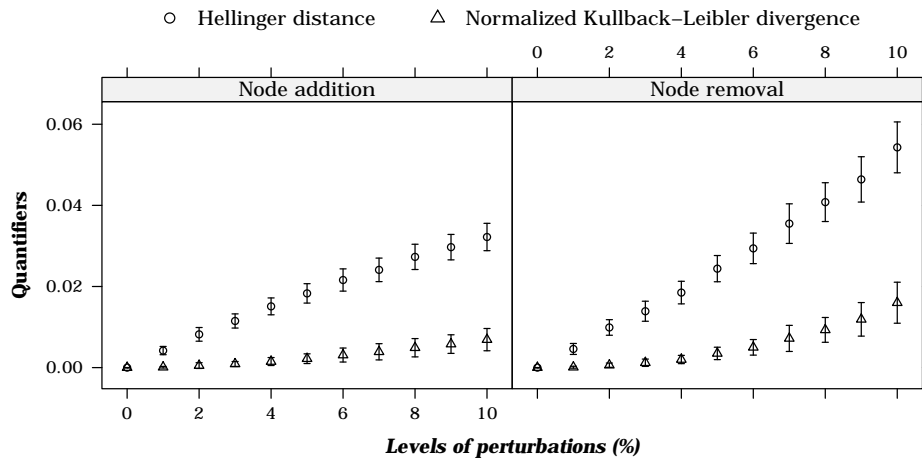


Figure 2. Hellinger distance  $\mathcal{D}_H$  and Kullback-Leibler divergence  $\mathcal{D}_{KL}$  for the shortest path length to the geometric network model (GRG).

The quantifiers are, thus, sensitive to both node addition and removal in this scenario, but with different responses. The Hellinger distance is the one which varies most, reaching, in mean, 0.032 for the addition and 0.054 for the most intense removal of nodes (10%).

The results presented are important to WSNs designers. Based on this study, a more efficient topology control can be applied. For instance, when 10% of nodes are added the shortest path length changes. This change indicates that adding more than 10% of nodes in the original network could be harmful. When 5% of nodes die (i.e. are removed) the shortest path length changes. This information can be used by the designer to calibrate the duty cycling operation, where each node periodically switches between sleeping mode and awake mode [24].

### IV.3. Complex network communication

Figure 3 presents the results in the random, small-world and scale-free communication models (top to bottom rows) and the three types of perturbations: link addition, link removal and link rewiring (left to right columns).

Raquel S. Cabral<sup>\*</sup>, Andre L. L. Aquino<sup>\*</sup>, Alejandro C. Frery  
Osvaldo A. Rosso<sup>\*</sup>, Jaime A. Ramírez

The shortest path length is sensitive to the link addition and removal applied in the three communication models, and both quantifiers ( $\mathcal{D}_{KL}$  and  $\mathcal{D}_H$ ) behave alike: the stronger the perturbation, the more the quantifier changes in direct proportion.

Regarding these two perturbations, again, the Hellinger distance exhibits more intense variations with respect to the perturbation than the normalized Kullback-Leibler divergence; notice that the confidence intervals do not overlap. Table 1 shows the mean value of the most intense perturbation (10%); this corresponds to the rightmost point of the plots of Figure 3. The Hellinger distance is consistently and significantly higher than the Kullback-Leibler divergence by a factor of, approximately, two, in all the situations where link addition and removal are applied.

The behavior of link addition and removal is the same in the tree models. Both affect the shortest path length, since these perturbations increase or decrease the distance between nodes. In particular, networks where all nodes have the same degree (regular graphs) may become small-world with just a few reconnections [17], i.e., randomly adding or removing links may result in connecting nodes which are far away, reducing the shortest path length. Link rewiring has a different behavior. Although the quantifiers exhibit differences, see Table 1, they are negligible in the random model. Link rewiring alters small-world models, but its effect seems constant, i.e., independent of the intensity of the perturbation. Regarding the scale-free model, there is a strong variation of both measures when small perturbations are applied, but the change tends to stabilize, i.e., saturates, soon after. The former is the only case where the normalized Kullback-Leibler divergence is bigger than the Hellinger distance.

The main feature of these networks is the presence of hubs, which makes them sensitive to this kind of perturbation: the shortest path length alters radically whenever a link involving hubs is added or removed. Additionally, the removal of some links makes the network disconnected.

Table 1. Mean most intense perturbation (10%) in the Hellinger distance and normalized Kullback-Leibler divergence.

	Hellinger distance			Normalized Kullback-Leibler divergence		
	Link addition	Link removal	Link rewiring	Link addition	Link removal	Link rewiring
Small-world model	0.122	0.119	0.073	0.095	0.071	0.033
Scale-free model	0.105	0.082	0.261	0.076	0.043	0.378
Random Model	0.092	0.105	0.006	0.052	0.058	$10^{-4}$

All models analysed represent a routing topology generated by some management strategies. Generally, these strategies combine the shortest path length with other QoS parameter (energy, delay or priority). The results reveal that when 5% of the links are added, or removed, the shortest path length changes. This information could be used to calibrate the management strategies to avoid the interference of QoS parameter during the routing generation, once this task considers just the links addition or removal.

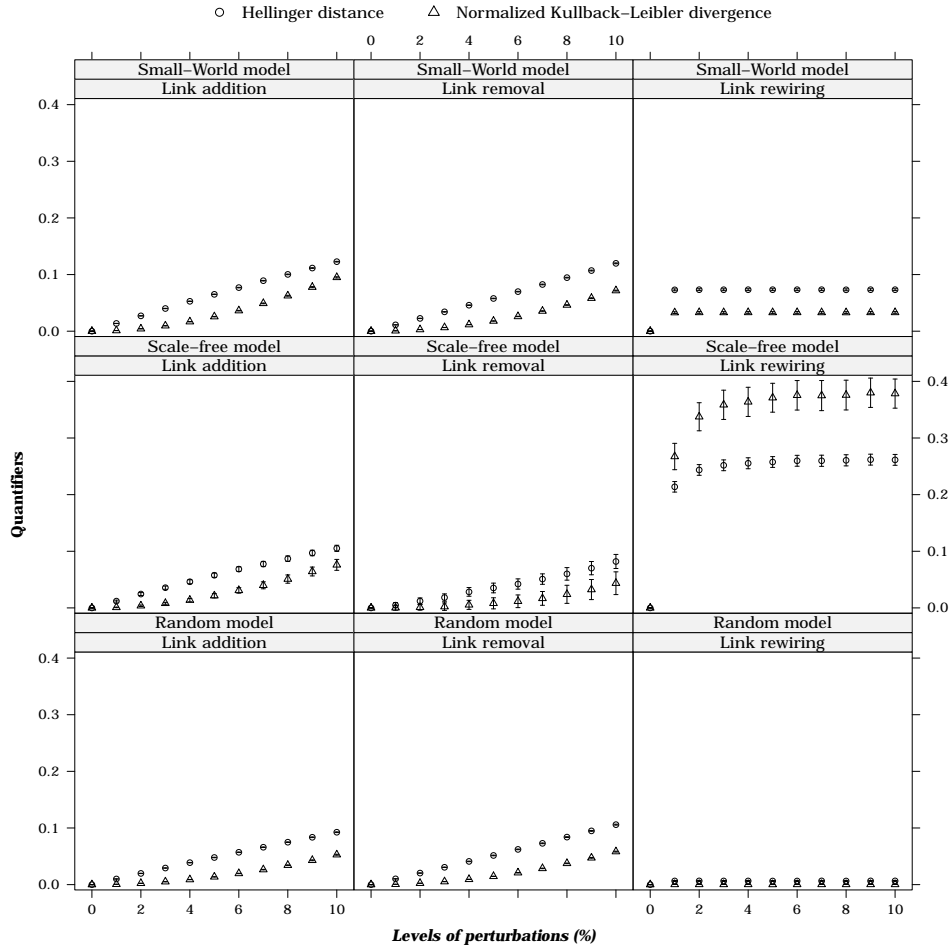


Figure 3. Hellinger distance and normalized Kullback-Leibler divergence for the shortest path length to the random network model, small-world network model and scale-free network model.

## V. Conclusion

The analysis of the variability of measures in WSNs communication strategies provides important information. It gives an insight of the behavior of the network when it is perturbed and it helps us in the design of appropriate solutions for every application.

In this paper, we used the normalized Kullback-Leibler divergence and the Hellinger distance to compare three

Raquel S. Cabral<sup>\*</sup>, Andre L. L. Aquino<sup>\*</sup>, Alejandro C. Frery<sup>\*</sup>  
 Osvaldo A. Rosso<sup>\*</sup>, Jaime A. Ramírez<sup>\*</sup>

communication models: flooding, random, small-world and scale-free (the last three are members of complex network based strategies). We performed two types of perturbations in the flooding-based strategy: addition and removal of nodes. In the complex network strategy, we used three types of perturbation: addition, removal and rewiring of links. We analyzed how the shortest path length changes with respect to different levels of each perturbation.

The shortest path length is sensitive to these changes, and in most situations it alters accordingly to the intensity of the perturbation. The analysis allows identifying the relationship between the strength of the perturbations and the change of the shortest path length. The use of quantifiers that involve logarithms or ratios may not be a good choice for this kind of characterization, because the occurrence of zeros leads to numerical problems and, possibly, to incorrect interpretation of network changes.

In addition, more efficient topology control or routing strategies in WSNs can be proposed. For instance the duty cycling or routing operations can be based on shortest path length sensibility. In the specific scenarios treated, these mechanism could be calibrated when 5% – 10% of the nodes or links are added or removed. The impact that the observed results have on the design and operation of WSNs was commented.

## ACKNOWLEDGEMENT

This work is partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) and the Research Foundation of the State of Alagoas (FAPEAL). O. A. Rosso gratefully acknowledges support from CONICET, Argentina and CNPq fellowship, Brazil.

## References

- [1] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., *IEEE Communications Magazine* **40** (2002) 102.
- [2] Arampatzis, T., Lygeros, J., and Manesis, S., A survey of applications of wireless sensors and wireless sensor networks, in *13th IEEE Mediterranean Conference on Control and Automation (MED'05)*, pages 719–724, Hawaii, USA, 2005, IEEE Computer Society.
- [3] Buratti, C., Conti, A., Dardari, D., and Verdone, R., *Sensors* **9** (2009) 6869.
- [4] Frery, A. C., Ramos, H. S., Alencar-Neto, J., Nakamura, E., and Loureiro, A. A. F., *Sensors* **10** (2010) 2150, Special Issue: Wireless Sensor Network and Its Application in Advanced Computer Science.
- [5] Costa, L. D., Rodrigues, F. A., Travieso, G., and Boas, P. R. V., *Advances in Physics* **56** (2008) 167.
- [6] Akkaya, K. and Younis, M., *Ad Hoc Networks* **3** (2005) 325.
- [7] Boas, P. R. V., Rodrigues, F. A., Travieso, G., and da F Costa, L., *Journal of Statistical Mechanics: Theory and Experiment* **2010** (2010) P03009.

- 
- [8] Nascimento, A. D. C., Cintra, R. J., and Frery, A. C., IEEE Transactions on Geoscience and Remote Sensing **48** (2010) 373.
  - [9] Helmy, A., IEEE Communications Letters **7** (2003) 490.
  - [10] Guidoni, D. L., Mini, R. A., and Loureiro, A. A., Computer Networks **54** (2010) 1266.
  - [11] Ruela, A. S., da Silva Cabral, R., Aquino, A. L. L., and Guimares, F. G., International Journal of Natural Computing Research **1** (2010) 33.
  - [12] Wang, B., Tang, H., Guo, C., and Xiu, Z., Physica A **363** (2006) 591.
  - [13] Carpi, L. C., Rosso, O. A., Saco, P. M., and Ravetti, M. G., Physics Letters A **375** (2011) 801.
  - [14] Newman, M. E. J. and Watts, D. J., Physical Review E **60** (1999) 7332.
  - [15] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U., Physics Reports **424** (2006) 175.
  - [16] Erdős, P. and Rényi, A., Publication of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960) 17.
  - [17] Watts, D. J. and Strogatz, S. H., Nature **393** (1998) 440.
  - [18] Barabási, A.-L. and Albert, R., Science **286** (1999) 509.
  - [19] Cha, S.-H. and Srihari, S. N., Pattern Recognition **35** (2002) 1355.
  - [20] Aquino, A. L., Junior, O. S., Frery, A. C., de Albuquerque, E. L., and Mini, R. A., IEEE Transactions on Computers (2013) in press (doi.ieeecomputersociety.org/10.1109/TC.2012.229).
  - [21] Cabella, B. C. T. et al., Brazilian Journal of Physics **38** (2008) 20.
  - [22] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0.
  - [23] Csárdi, G. and Nepusz, T., InterJournal **Complex Systems** (2006) 1695.
  - [24] Lai, S. and Ravindran, B., Efficient opportunistic broadcasting over duty-cycled wireless sensor networks, in *29th IEEE Conference on Computer Communications (INFOCOM'2010)*, San Diego, USA, 2010.

## Referências Bibliográficas

- Ahuja, R. K.; Magnanti, T. L. & Orlin, J. B. (1993). *Network flows: theory, algorithms and applications*. Prentice Hall.
- Akyildiz, I. F.; Su, W.; Sankarasubramaniam, Y. & Cayirci, E. (2002). A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102–114.
- Albert, R.; Jeong, H. & Barabási, A.-L. (1999). Diameter of the world wide web. *Nature*, 401:130–131.
- Albert, R.; Jeong, H. & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406:378–382.
- Almiron, M. G.; Ramos, H. S.; Oliveira, E. M. R.; de Menezes, J. G. M.; Guidoni, D. L.; Stancioli, P. O.; da Cunha, F. D.; de Aquino, A. L. L.; Mini, R. A. F.; Frery, A. C. & Loureiro, A. A. F. (2010). Redes complexas na modelagem de redes de computadores. Em *XVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pp. 1–49. Sociedade Brasileira de Computação.
- Antoniou, P. & Pitsillides, A. (2007). Understanding complex systems: a communication networks perspective. Relatório técnico, Networks Research Laboratory, Computer Science Department, University of Cyprus.
- Aquino, A. L.; Junior, O. S.; Frery, A. C.; de Albuquerque, E. L. & Mini, R. A. (2012). Musa: Multivariate sampling algorithm for wireless sensor networks. *IEEE Transactions on Computers*, 99:1.
- Arampatzis, T.; Lygeros, J. & Manesis, S. (2005). A survey of applications of wireless sensors and wireless sensor networks. Em *13th IEEE Mediterranean Conference on Control and Automation (MED'05)*, pp. 719–724, Hawaii, USA. IEEE Computer Society.



- Arianos, S.; Bompard, E.; Carbone, A. & Xue, F. (2009). Power grid vulnerability: A complex network approach. *Chaos*, 19:013119.
- Auder, B. & Iooss, B. (2009). Global sensitivity analysis based on entropy. Em *Safety, Reliability and Risk Analysis*.
- Bai, W.-J.; Zhou, T.; Fu, Z.-Q.; Chen, Y.-H.; Wu, X. & Wang, B.-H. (2006). Electric power grids and blackouts in perspective of complex networks. Em *International Conference on Communications, Circuits and Systems Proceedings*, pp. 2687–2691, Guilin.
- 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.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.
- Barabási, A.-L.; Albert, R. & Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281:69–77.
- Barabási, A.-L. & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, 288:50–59.
- Barthélemy, M. & Amaral, L. A. N. (1999). Small-world networks: Evidence for a crossover picture. *Physical Review Letters*, 82:3180.
- Basseville, M. (2013). Divergence measures for statistical data processing – an annotated bibliography. *Signal Processing*, 93:621– 633.
- Bernoulli, J. (1713). *Ars Conjectandi: Usum & Applicationem Praecedentis – Doctrinae in Civilibus, Moralibus & Oeconomicis*. Chapter 4, Traduzido para Inglês por Oscar Sheynin.
- Boas, P. R. V. (2008). *Efeito da amostragem nas propriedades topológicas de redes complexas*. Tese de doutorado, Universidade de São Paulo.
- Boas, P. R. V.; Rodrigues, F. A. & da F. Costa, L. (2009). Modeling worldwide highway networks. *Physics Letters A*, 374:22–27.
- Boas, P. R. V.; Rodrigues, F. A.; Travieso, G. & da F Costa, L. (2010). Sensitivity of complex networks measurements. *Journal of Statistical Mechanics: Theory and Experiment*, 2010:P03009.

- Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M. & Hwang, D.-U. (2006). Complex networks : structure and dynamics. *Physics Reports*, 424:175–308.
- Bollobás, B. (1998). *Modern Graph Theory*. Springer.
- Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A. & Wiener, J. (2000). Graph structure in the web. *Computer Networks*, 33:309–320.
- Buratti, C.; Conti, A.; Dardari, D. & Verdone, R. (2009). Overview on wireless sensor networks technology and evolution. *Sensors*, 9:6869–6896.
- Cabella, B. C. T.; Sturzbecher, M. J.; Tedeschi, W.; Filho, O. B.; de Araújo, D. B. & da Costa Neves, U. P. (2008). A numerical study of the kullback-leibler distance in functional magnetic resonance imaging. *Brazilian Journal of Physics*, 38(1):20–25.
- Cabral, R. S.; Aquino, A. L. L.; Frery, A. C.; Rosso, O. A. & Ramírez, J. A. (2013a). Structural changes in data communication in wireless sensor networks. *Central European Journal of Physics*.
- Cabral, R. S.; Frery, A. C. & Ramírez, J. A. (2013b). Variability analysis of complex networks measures based on stochastic distances. Submetido ao *Physica A: Statistical Mechanics and its Applications*.
- Cancho, R. F.; Janssen, C. & Solé, R. V. (2001). Topology of technology graphs: Small world patterns in electronic circuits. *Physical Review E*, 64:046119.
- Carchiolo, V.; Malgeri, M.; Mangioni, G. & Nicosia, V. (2008). Emerging structures of p2p networks induced by social relationships. *Computer Communications*, 31:620–628.
- Carpi, L. C.; Rosso, O. A.; Saco, P. M. & Ravetti, M. G. (2011). Analyzing complex networks evolution through information theory quantifiers. *Physics Letters A*, 375:801–804.
- Cha, S.-H. & Srihari, S. N. (2002). On measuring the distance between histograms. *Pattern Recognition*, 35:1355–1370.
- Chassin, D. P. & Posse, C. (2005). Evaluating north american electric grid reliability using the barabási–albert network model. *Physica A*, 355:667–677.

- Costa, L. D.; Rodrigues, F. A.; Travieso, G. & Boas, P. R. V. (2008). Characterization of complex networks: a survey of measurements. *Advances in Physics*, 56:167–242.
- Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*. Wiley-Interscience.
- Crucitti, P.; Latora, V. & Marchiori, M. (2004). A topological analysis of the italian electric power grid. *Physica A*, 338:92–97.
- Crucitti, P.; Latora, V.; Marchiori, M. & Rapisarda, A. (2003). Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications*, 320:622–642.
- Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- de Aquino, A. L. L. (2008). *Redução de dados em redes de sensores sem fio baseada em stream de dados*. Tese de doutorado, Universidade Federal de Minas Gerais.
- de Kunder, M. (2013). The size of the world wide web (the internet). Acesso: 22 de Abril de 2013.
- Demetrius, L. & Mankeentrop, T. (2005). Robustness and network evolution-an entropic principle. *Physica A*, 346:682–696.
- Diaconis, P. & Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–830.
- Ebel, H.; Mielsch, L.-I. & Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, 66:035103.
- Endres, D. & Schindelin, J. (2003). A new metric for probability distributions. *IEEE Transactions On Information Theory*, 49:1858–1860.
- Erdős, P. & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae (Debrecen)*, 6:290–297.
- Erdős, P. & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61.

- Faloutsos, M.; Faloutsos, P. & Faloutsos, C. (1999). On power-law relationships of the internet topology. *SIGCOMM Computer Communications Review*, 29:251–262.
- Flory, P. J. (1952). Molecular size distribution in three dimensional polymers. *Journal of the American Chemical Society*, 74(11):2718–2723.
- Frery, A. C.; Ramos, H. S.; Alencar-Neto, J.; Nakamura, E. & Loureiro, A. A. F. (2010). Data driven performance evaluation of wireless sensor networks. *Sensors*, 10:2150–2168. Special Issue: Wireless Sensor Network and Its Application in Advanced Computer Science.
- Ghamry, W. K. & Elsayed, K. M. F. (2012). Network design methods for mitigation of intentional attacks in scale-free networks. *Telecommunication Systems*, 49:313–327.
- Gol'dshtein, V.; Koganov, G. & Surdutovich, G. (2004). Vulnerability and hierarchy of complex networks. *Disordered Systems and Neural Networks*, cond-mat/0409298v1.
- Guidoni, D. L.; Mini, R. A. & Loureiro, A. A. (2010). On the design of resilient heterogeneous wireless sensor networks based on small world concepts. *Computer Networks*, 54:1266–1281.
- Helmy, A. (2003). Small worlds in wireless networks. *IEEE Communications Letters*, 7:490–492.
- Holmgren, A. J. (2006). Using graph models to analyze the vulnerability of electric power networks. *Risk Analysis*, 26:955–969.
- Huang, L.; Lai, Y.-C. & Chen, G. (2008). Understanding and preventing cascading breakdown in complex clustered networks. *Physical Review E*, 78:036116.
- Jeong, H. (2010). Ccnr (biological networks). Acesso:02 de Fevereiro de 2010.
- Jiang, B. & Claramunt, C. (2004). Topological analysis of urban street networks. *Environment and Planning B*, 31:151–162.
- Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *Annals of mathematical statistics*, 22:79–86.

- Lai, S. & Ravindran, B. (2010). Efficient opportunistic broadcasting over duty-cycled wireless sensor networks. Em *29th IEEE Conference on Computer Communications (INFOCOM'2010)*, San Diego, USA.
- Lee, J. (2007). Scale-free networks in the presence of constraints: An empirical investigation of the airline route network. *Seoul Journal of Business*, 13:3–19.
- Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Transactions On Information Theory*, 37:145–151.
- Liu, S.; Liao, Z.; Feng, Y. & Rong, G. (2010). Topological properties of refinery system: A complex network approach. Em *8th IEEE International Conference on Control and Automation*, pp. 9–11, Xiamen, China.
- Liu, Y. & Gu, X. P. (2007). Skeleton-network reconfiguration based on topological characteristics of scale-free networks and discrete particle swarm optimization. *IEEE Transactions on Power Systems*, 22:1267–1274.
- Lüdtke, N.; Panzeri, S.; Brown, M.; Broomhead, D. S.; Knowles, J.; Montemurro, M. A. & Kell, D. B. (2008). Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. *Journal of the Royal Society Interface*, 5:223–235.
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2:60–67.
- Mishkovski, I.; Biey, M. & Kocarev, L. (2011). Vulnerability of complex networks. *Commun Nonlinear Sci Numer Simulat*, 16:341–349.
- Montgomery, D. (2005). *Design and Analysis of Experiments*. John Wiley & Sons.
- Nascimento, A. D. C.; Cintra, R. J. & Frery, A. C. (2010). Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on geoscience and remote sensing*, 48(1):373–385.
- Newman, M. E. J. (2000). Models of the small world: A review. *Journal of Statistical Physics*, 101:819–841.
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98:404–409.
- Newman, M. E. J.; Barábasi, A.-L. & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton University Press.

- Newman, M. E. J. & Watts, D. J. (1999). Scaling and percolation in the small-world network model. *Physical Review E*, 60:7332–7342.
- Onnela, J.-P.; Saramäki, J.; Hyvönen, J.; Szabó, G.; Lazer, D.; Kaski, K.; Kertész, J. & Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104:7332–7336.
- Pardo, L.; Morales, D.; Salicrú, M. & Menéndez, M. L. (1995). Generalized divergence measures: Information matrices, amount of information, asymptotic distribution, and its applications to test statistical hypotheses. *Information Sciences*, 84(8):181–198.
- Pastor-Satorras, R.; Vázquez, A. & Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, 87:258701.
- Paul, G.; Sreenivasan, S. & Stanley, H. E. (2005). Resilience of complex networks to random breakdown. *Physical Review E*, 72:056130.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rapoport, A. (1951). Nets with distance bias. *Bulletin of Mathematical Biology*, 13(2):85–91.
- Reggiani, A.; Nijkamp, P. & Cento, A. (2010). Connectivity and concentration in airline networks: a complexity analysis of lufthansa's network. *European Journal of Information Systems*, 19:449–461. Não consegui o artigo.
- Reijneveld, J. C.; Ponten, S. C.; Berendse, H. W. & Stam, C. J. (2007). The application of graph theoretical analysis to complex networks in the brain. *Clinical Neurophysiology*, 118:2317–2331. Invited review.
- Reschenhofer, E. (1997). Generalization of the kolmogorov-smirnov test. *Computational Statistics & Data Analysis*, 24(4):422–441.
- Rodrigues, F. A. (2007). *Caracterização, classificação e análise de redes complexas*. Tese de doutorado, Universidade de São Paulo.
- Rosas-Casals, M. (2010). Power grids as complex networks: topology and fragility. Em *Proceedings of the Complexity in Engineering (COMPENG '10)*, pp. 21–26, Rome, Italy.

- Ruela, A. S.; da Silva Cabral, R.; Aquino, A. L. L. & Guimarães, F. G. (2010). Memetic and evolutionary design of wireless sensor networks based on complex network characteristics. *International Journal of Natural Computing Research*, 1(2):33–53.
- Salicrú, M.; Morales, D. & Menéndez, M. L. (1994). On the application of divergence type measures in testing statistical hypothesis. *Journal of Multivariate Analysis*, 51:372–391.
- Saltelli, A.; Ratto, M. & Andres, T. (2008). *Global Sensitivity Analysis. The Primer*. Wiley.
- Schintler, L. A.; Gorman, S. P.; Reggiani, A.; Patuelli, R.; Gillespie, A.; Nijkamp, P. & Rutherford, J. (2005). Complex network phenomena in telecommunication systems. *Networks and Spatial Economics*, 5:351–370.
- Sharma, G. & Mazumdar, R. (2005). Hybrid sensor networks: a small world. Em *Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing(MobiHoc'05)*, pp. 366–377, New York, NY, USA.
- Siegel, S. & N. John Castellan, J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Humanities/Social Sciences/Languages, Columbus, OH, USA, 2 edição. ISBN 0070573573.
- Singh, B. K. & Gupte, N. (2005). Congestion and decongestion in a communication network. *Physical Review E*, 71:055103.
- Solé, R. V. & Valverde, S. (2004). Information theory of complex networks: On evolution and architectural constraints. *Lecture Notes in Physics*, 650:189–207.
- Soramäki, K.; Bech, M. L.; Arnold, J.; Glass, R. J. & Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379:317–333.
- Steyvers, M. & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29:41–78.
- Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410:268–276.
- Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *Journal of Statistical Planning and Inference*, 49:137–162.

- Vedral, V. (2002). The role of relative entropy in quantum information theory. *Reviews of Modern Physics*, 74:197–234.
- Wang, B.; Tang, H.; Guo, C. & Xiu, Z. (2006). Entropy optimization of scale-free networks robustness to random failures. *Physica A*, 363:591–596.
- Wang, J. W.; Rong, L. L. & Zhang, L. (2009). Effect of attack on scale-free networks due to cascading failure. *Modern Physics Letters B*, 23:1577–1587.
- Wang, W.-X.; Ni, X.; Lai, Y.-C. & Grebogi, C. (2012). Optimizing controllability of complex networks by minimum structural perturbations. *Physical Review E*, 85:026115.
- Wang, X. F. & Chen, G. (2003). Complex networks: small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine*, 3:6–20.
- Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393:440–442.
- Wen, L.; Dromey, R. G. & Kirk, D. (2009). Software engineering and scale-free networks. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 39:845–854.
- Wu, J.; Gao, Z. & Sun, H. (2004). Urban transit system as a scale-free network. *Modern Physics Letters B*, 18:1043–1049.
- Wu, J. & Yang, S.-H. (2008). Small-world model-based polylogarithmic routing using mobile nodes. *Journal of Computer Science and Technology*, 23:327–342.
- Yook, S.-H.; Oltvai, Z. N. & Barabási, A.-L. (2004). Functional and topological characterization of protein interaction networks. *Proteomics*, 4:928–942.
- Zhang, J.; Xu, X.; Hong, L.; Wang, S. & Fei, Q. (2012). Attack vulnerability of self-organizing networks. *Safety Science*, 50:443–447.
- Zhou, S. & Mondragón, R. J. (2004). Accurately modeling the internet topology. *Physical Review E*, 70:066108.