

**UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS
ÁREA DE CONCENTRAÇÃO: FÍSICA**

MARILY APARECIDA BENICIO

CENTRALIDADE DA CAMINHADA ALEATÓRIA EM REDES COMPLEXAS

PONTA GROSSA

2013

MARILY APARECIDA BENICIO

CENTRALIDADE DA CAMINHADA ALEATORIA EM REDES COMPLEXAS

Dissertação submetida ao programa de Pós-Graduação em Ciências - Área de concentração: Física - da Universidade Estadual de Ponta Grossa como requisito parcial para obtenção do título de Mestre em Física.

Orientador: Prof. Dr. Sandro Ely de Souza Pinto.

PONTA GROSSA

2013

Ficha Catalográfica
Elaborada pelo Setor de Tratamento da Informação BICEN/UEPG

B467 Benicio, Marily Aparecida
Centralidade da caminhada aleatória em
redes complexas/ Marily Aparecida Benicio.
Ponta Grossa, 2013.
130f.

Dissertação (Mestrado em Ciências -
Área de Concentração: Física),
Universidade Estadual de Ponta Grossa.
Orientador: Prof. Dr. Sandro Ely de
Souza Pinto.

1. Centralidade. 2. Caminhada aleatória.
3. Redes complexas. I. Pinto, Sandro Ely de
Souza. II. Universidade Estadual de Ponta
Grossa. Mestrado em Ciências. III. T.

CDD: 519.2

TERMO DE APROVAÇÃO

Marily Aparecida Benício

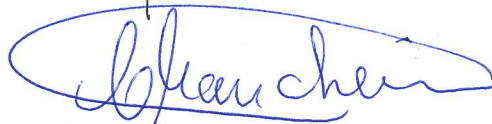
“Centralidade da Caminhada Aleatória em Redes Complexas”,

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Ciências - Física da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora.

Orientador:



Prof. Dr. Sandro Ely de Souza Pinto
Departamento de Física, UEPG/PR



Prof. César Manchein
Departamento de Física, UDESC/SC



André Maurício Brinatti
Departamento de Física – UEPG/PR

Ponta Grossa, 09 de março de 2013.

Dedico este trabalho:

À Deus.

À meu esposo, amigo e companheiro Edison, pela paciência, amor e por me ajudar a alcançar os meus sonhos.

À Pastorina, minha mãe e José Rubens, meu pai, por todo amor e por tudo que me ensinaram.

AGRADECIMENTOS

Em meio a tantos sonhos, esforços e conquistas, percebo as muitas oportunidades que surgiram e o quanto aproveitei dessas oportunidades para minha formação pessoal e profissional, percebo o quanto amadureci não somente como pessoa, mas principalmente no campo das ideias. Em meio a tudo isso, o sentimento de gratidão não apenas completa como transborda a minha alma.

Como Mestre maior e Criador de todas as coisas, agradeço primeiramente a Deus. Que me permitiu aprender sobre as ciências, pelas quais percebemos o mundo e a Matemática, pela a qual a escrevemos, e me abençoou com os desejos e anseios de pesquisadora. Por ter criado as condições necessárias para que eu estivesse hoje escrevendo estas linhas. Por ter me ajudado a ter força e determinação, além disso, ter me acompanhado no percorrer deste caminho.

Agradeço ao Prof.º Dr. Sandro Ely de Souza Pinto, meu orientador, pela confiança, dedicação e paciência. Por ter me ajudado no decorrer de toda a pesquisa, me orientado nas horas de dúvidas, e me acalmado nas horas difíceis.

Agradeço aos colegas do Grupo de Dinâmica Não Linear e Sistemas Complexos, Ms. Rodrigo Pereira e Ms. Romeu Miquéias, os quais me auxiliaram nas simulações numéricas, ao Ms. Mauricio A Ribeiro, que contribuiu para as construções gráficas, ao Ms. Roberto Vosgeral, pelo apoio teórico, ao Rafael Ribaski, que me ajudou durante toda a pesquisa em vários aspectos do desenvolvimento do trabalho e a Mônica A S Viana, por me ajudar no enfrentamento de inúmeras dificuldades. Agradeço o apoio, a orientação e paciências, estou certa de que sozinha o caminho seria imensamente mais árduo. As contribuições que trouxeram foram fundamentais para meu crescimento e aprendizado, contribuindo para o trabalho de pesquisa, minha formação profissional e pessoal.

Agradeço a meus pais, pelo apoio, incentivo, dedicação, carinho e pela confiança de que alcançaria os meus objetivos, por tudo que fizeram e batalharão para que eu tivesse a oportunidade de crescer, estudar e hoje estar percorrendo o caminho que escolhi para mim, vocês fazem parte dos meus sonhos e de minhas conquistas.

Agradeço a minha família e amigos pela compreensão, pelos diversos momentos de ausência, pelo apoio e carinho.

Agradeço ao meu esposo, você é o maior incentivador deste trabalho, tendo me ajudado de diversas formas. Agradeço pelo seu amor, carinho e confiança e por tudo que você representa a minha vida.

“Transportai um punhado de terra todos os dias e fareis uma montanha”.

Confúcio.

RESUMO

Os estudos sobre redes complexas nos auxiliam a compreender e modelar muitas situações do mundo real. O mundo é abundante em redes que podem ser encontradas em diversos contextos reais. O termo redes faz referência às relações estabelecidas entre dois conjuntos e podem ser representadas por meio da teoria de grafos. A classificação das redes complexas se dá de acordo com os modelos criados para representá-las, tais como as redes Aleatórias, redes de Pequeno Mundo, redes Sem Escala e redes Hierárquicas. Dentro da perspectiva de redes complexas, um estudo que pode trazer contribuições importantes é análise do fenômeno de difusão de informação em redes, os quais podem ser entendidos através do processo da caminhada aleatória, a qual se caracteriza por ser um processo estocástico utilizado como um mecanismo de transporte e pesquisa em redes complexas. A caminhada aleatória nas redes complexas pode ser utilizada para verificar o comportamento de cada modelo de rede frente à dissipação. Cada modelo de rede apresenta um comportamento diferente com relação ao número de caminhantes aleatórios que passam por nó da rede ao longo do tempo. Este número de caminhantes irá depender da estrutura das redes geradas por cada modelo e das medidas de Centralidade de cada nó. As medidas de centralidade dos vértices da rede são úteis para comparar a eficiência dos nós com relação ao recebimento e envio de informações sendo indicativos da rapidez com a qual, este transporte acontece. O objetivo deste trabalho é estudar o processo da caminhada aleatória e utilizá-la para analisar a eficiência das medidas de Centralidade, inferindo o número de caminhantes aleatórios que passam pelos nós nas redes complexas. As medidas de centralidade analisadas são as centralidades do Grau, Centralidade de Intermediação por Menores Caminhos, Centralidade da Caminhada Aleatória. Para comparar a eficiência das referidas medidas de Centralidade nos diferentes modelos de redes, foram realizadas simulações numéricas. Com estas, percebeu-se que o comportamento da difusão de caminhantes varia para cada modelo de rede. Para a rede Aleatória o fluxo de caminhantes se dá de maneira uniforme não sendo possível destacar algum vértice de maior importância dentro da rede. Pode-se observar que a medida de Centralidade da Caminhada Aleatória é a que mostrou maior eficiência ao apontar o um maior fluxo de caminhantes aos vértices que possuíam um maior valor para essa medida.

Palavras Chaves: Redes Complexas, Caminhada Aleatória, Centralidade.

ABSTRACT

Studies of complex networks help us to understand and model many real world situations. The world abounds in networks that can be found in many real contexts. The term network refers to relations between two sets and can be represented by means of graph theory. The classification of complex networks is given according to the models created to represent them, such as Random networks, networks of Small World, No Scaling networks and hierarchical networks. From the perspective of complex networks, a study which make significant contributions analysis is the phenomenon of diffusion of information in networks, which can be understood through the random walk process, which is characterized by a stochastic used as a mechanism transportation and research in complex networks. A random walk in complex networks can be used to check the behavior of each network model front dissipation. Each network model presents a different behavior with respect to the number of random walkers that pass through the network node over time. The number of walkers will depend on the structure of the networks generated by each model and measures of centrality of each node. Measures of centrality of the vertices of the network are useful for comparing the efficiency of the nodes with respect to receiving and sending information being indicative of the rapidity with which this transport happens. The objective of this work is to study the process of random walk and use it to analyze the efficiency of Centrality measures, inferring the number of random walkers who pass by us in complex networks. Measures of centrality are analyzed centralities Degree, Centrality Intermediation by Minor Roads, Centralization of Random Walk. To compare the efficiency of these measures of centrality in the different network models, numerical simulations were performed. With these, it was noticed that the behavior of the diffusion of walkers varies for each network model. Random network for the flow of walkers from the evenly is not possible to highlight some vertex of utmost importance within the network. It can be observed that the measure of centrality of Random Walk is the one that showed greater efficiency by pointing a greater flow of walkers to the vertices that had a higher value for this measure.

Key Words: Complex Networks, Random Walk, Centrality.

LISTA DE FIGURAS

Figura 1.1 Caminhada Aleatória Unidimensional, x é a posição atual do caminhante, $x + 1$ e $x - 1$ são os possíveis posições para o passo seguinte, p e $1 - p$, são as respectivas probabilidade para esse passo.	26
Figura 1.2 Caminhada Aleatória Simétrica Unidimensional, ilustrando a Distância percorrida por uma partícula com relação aos passos dados, onde o número de passos é 10000.....	27
Figura 1.3 Histograma referente à Caminhada Aleatória Simétrica Unidimensional, desempenhada por uma partícula, após 10000 passos.	28
Figura 1.4 Ilustração da caminhada Aleatória, no plano bidimensional, sobre uma malha quadriculada.	32
Figura 1.5 Caminhada Aleatória Bidimensional sobre uma malha quadriculada com $t = 10000$	33
Figura 1.6 Distribuição das posições percorridas pelo caminhante após t passos.....	34
Figura 2.1 Ilustração da cidade de Königsberg, onde perdurava o problema das sete pontes.....	45
Figura 2.2 Grafo representando a cidade de Königsberg, de forma de simplificada. Os nós representam os blocos de terra e as arestas representam as pontes.....	46
Figura 2.3 a) Dodecaedro, poliedro regular com 20 vértices de grau 3, 12 faces pentagonais, 30 arestas. b) Representação no plano do Dodecaedro, com os vértices indicando as principais cidades ao redor do Mundo.	47
Figura 2.4 Representação do problema da viagem à volta do mundo com um caminho destacado pelas arestas em vermelho, indicando a solução para o problema.....	48
Figura 2.5 Grafo cujos vértices são a, b, c, d, e, f, g , e as arestas $(a, c), (b, c), (c, d), (d, e), (d, f), (d, g)$	49
Figura 2.6 a) Grafo direcionado, as arestas apresentam orientações. b) Grafo não direcionado.	50
Figura 2.7 Grafo indicando caminhos: a) simples, b) trajeto, c) Hamiltoniano e d) Euleriano.	50

Figura 2.8 Grafo G_1 com dois de seus subgrafos. a) Grafo $G_1(V, E)$ com 5 vértices, b) O subgrafo G_2 extraído a partir de G_1 formado com 4 dos seus 5 vértices, c) O subgrafo G_3 formado com 3 do total de vértices de G_1	51
Figura 2.9 Grafo G com 5 vértices e 6 arestas em a), e dois de seus subgrafos b) e c), onde (c) representa um clique de (a).	52
Figura 2.10 Em a) representação de um grafo conexo, em b) representa um grafo desconexo.	52
Figura 2.11 Grafo desconexo, com dois subgrafos conexos identificados.....	53
Figura 2.12 Grafos com 7 vértices particionados em dois subconjuntos de vértices, cada partição possui 3 e 4 vértices respectivamente. Em a) Grafo Bipartido, em b) Grafo Bipartido Completo.	53
Figura 2.13 Floresta constituída por duas árvores, em que cada árvore apresenta vértices na forma de folhas e de vértices interiores.	54
Figura 2.14 Caminho simples identificado entre dois vértices de uma árvore.....	54
Figura 2.15 a) Supressão da aresta (c, d) , a árvore deixa de ser conexa, sendo dividida em duas árvores menores. b) Adição da aresta (a, b) , o grafo deixa de ser uma árvore, devido ao ciclo, constituído pelo caminho (a, b, c, a)	55
Figura 2.16 Os grafos G_1 e G_2 representam grafos isomorfos. Pois há a função $\{(a, 1), (b, 2), (c, 3), (d, 4), (e, 5)\}$ que satisfaz a condição descrita acima.	55
Figura 2.17 Matrizes de adjacências de a) Grafo G_1 não direcionado e de b) Grafo G_2 direcionado.	56
Figura 2.18 Rede com número de nós $n = 4$ e as conexões entre os vértices.....	57
Figura 2.19 Rede com $n = 4$, indicando a probabilidade de transição de 1 para 4 em diferentes números de passos. a) Probabilidade de transição de 1 para 4 em um passo. b) Probabilidade de transição de 1 para 4 em dois passos. c) Probabilidade de transição de 1 para 4 m três passos.	61
Figura 3.1 Redes com $n = 5$, o coeficiente de agrupamento Global C definido pela Eq. (3.1) é de $3/8$. Os vértices individuais têm coeficientes de agrupamento local de $C_1 = 1$, $C_2 = 1$, $C_3 = \frac{1}{6}$, $C_4 = 0$ e $C_5 = 0$ conforme a Eq.(3.2) [34].....	67

Figura 3.2 Representação de redes aleatórias com $N = 10$ e (a) $p = 0,1$; (b) $p = 0,5$ e em (c) $p = 0,8$	70
Figura 3.3 Distribuição de grau de uma rede aleatória com $N = 1000$ e $p = 0,1$	70
Figura 3.4 Redes de Pequeno Mundo com $N = 10$, $r = 2$ e a) $p = 0$; b) $p = 0,05$; c) $p = 0,1$; d) $p = 1$; construídas a partir do modelo de Watts-Strogats.....	72
Figura 3.5 Comportamento do comprimento de caminho médio $L(p)$ e do coeficiente de agrupamento $C(p)$ em função de p , segundo o modelo de Watts e Strogatz (1998). Os valores de C e L foram normalizados pelos valores $L(0), C(0)$ para uma rede regular, que são os respectivos valores máximos. Para $0,01 < p < 0,1$ ocorre o efeito Pequeno Mundo. Figura extraída de [4].....	73
Figura 3.6 Rede Barabási-Albert, com número de nós $N = 20$ e número de arestas adicionadas a partir de cada novo nó $m = 2$	75
Figura 3.7) Distribuição de Grau seguindo uma Lei de Potência em a) de uma Rede de Escala livre, pelo modelo de Barabási-Albert, com $N = 1000$ e $m = 2$ e seu respectivo histograma em b) relacionando as frequências absolutas dos graus.....	75
Figura 3.8 Primeiras etapas da construção de uma rede sem escala determinística e com estrutura hierárquica. Onde o agrupamento principal se inicia com M vértices, cujo vértice central é a raiz da rede, neste caso $M = 5$	77
Figura 4.1 Os nós correspondentes aos índices 2 e 3 possuem o mesmo grau, o qual vale 3. Porém a centralidade do nó 3 é maior que a do nó 2.	81
Figura 4.2 Assimetria da caminhada aleatória em rede. a) probabilidade de um caminhante sair de um nó 1 para um de seus vizinhos. b) A probabilidade de um caminhante estar no nó 4 e voltar para o nó 1.	86
Figura 5.1 Os gráficos mostram a relação de τ versus k das redes: a) Aleatória com $N = 500$ e $p = 0,1$. b) Pequeno Mundo com $N = 500$ e $p = 0,1$. c) Hierárquica $N = 512$ e d) Sem Escala, modelo de Barabási-Albert para $N = 500$ e os nós cresceram com o parâmetro $m = 2$	93
Figura 5.2 Relação de entre três diferentes medidas de centralidade para as redes Aleatória. Em a) k versus CIMC, b) k versus CCA, e c) CCA versus CIMC. Cada imagem traz a correlação referente ao ajuste, a rede possui $N = 500$ e $p = 0,1$	95

Figura 5.3 Relação entre três diferentes medidas de centralidade para as redes de Pequeno Mundo, seguindo o modelo de Watts e Strogatz. Em a) k versus CIMC, b) k versus CCA, e c) CIMC versus CCA. Cada imagem traz o valor da correlação referente ao ajuste, o número de nós para rede é $N = 500$ e $p = 0,1$	96
Figura 5.4 Relação entre três diferentes medidas de centralidade para as redes Hierárquicas, com o número de nós para a rede $N = 512$ e $m=5$. Em a) k versus CCA, em b) a mesma relação k versus CCA plotado com escala logarítmica permitindo a melhor visualização dos pontos. Em c) k versus CIMC e em d) CCA versus CIMC.	98
Figura 5.5 Relação entre as medidas de centralidade para as redes Sem Escala, modelo de Barabási-Albert para $N = 500$ e os nós cresceram com o parâmetro $m = 2$. Em a) k versus CIMC, escala logarítmica, b) k versus CCA, em c) CCA versus CIMC. Cada imagem apresenta o respectivo valor de correlação referente ao ajuste.	99
Figura 5.6 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede Aleatória, com $N = 500$ e $T = 2000$. O valor de n é representado na escala de cor a direita de cada gráfico.....	102
Figura 5.7 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede de Pequeno Mundo, com $N = 200$ e $T = 1000$. O valor de n é representado na escala de cores à direita de cada gráfico.....	104
Figura 5.8 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede Barabasi-Albert, com $m = 2$, $N = 500$ e $T = 2000$.O valor de n é representado na escala de cores à direita de cada gráfico.	106
Figura 5.9 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede Hierarquica, com $m = 5$, $N = 512$ e $T = 2560$. O valor de n é representado na escala de cores à direita de cada gráfico.	108

LISTA DE SÍMBOLOS

θ	Tempo de interação
h	Distância percorrida
p	Probabilidade
q	Probabilidade
σ	Variáveis aleatórias
n	Número de variáveis aleatórias
t	Tempo
x	Posição
a	Média
b	Variância
$g(k)$	Função característica
P	Probabilidade
ρ	Densidade de probabilidade
c	Velocidade média
D	Coefficiente de Difusão
\vec{r}	Vetor das coordenadas espaciais
Π	Probabilidade
$P(n, m)$	Matriz de probabilidade
λ	Autovalor da matriz de probabilidade
δ_{ij}	Delta de Kronecker
ψ	Autovetor
$G(V, E)$	Grafo
V	Conjunto de vértices
E	Conjunto de arestas

A	Matriz adjacência
k	Grau do vértice
P_i^∞	Matriz Estacionária para o vértice i
e	Aresta
v	Vértice
w	Vértice
C	Coefficiente de aglomeração
L	Comprimento de caminho médio
d	Menor distancia entre os vertices
g	Centralidade de grau
C_p	Centralidade proximidade
C_i	Centralidade de Intermediação
M	Matriz das probabilidades de transição
D	Matriz diagonal dos graus dos vértices
V	Matriz
S	Vetor
I	Vetor
N_i	Centralidade de Intermediação por caminhada Aleatória
F_{ij}	Matriz de probabilidade de Primeira Passagem
$\langle T_{ij} \rangle$	Tempo Médio de Primeira passagem
$R^{(n)}$	Momento n
C_{CA}	Centralidade da caminhada Aleatória
τ	Tempo característico de relaxação
$a_i^{(\lambda)}$	Autovetores à esquerda do operador de evolução temporal
$b_i^{(\lambda)}$	Autovetores à direita do operador de evolução temporal
Λ_λ^t	Autovalor do operador de evolução temporal.

LISTA DE ABREVIACOES

TMPP	Tempo Medio de Primeira Passagem
CIMC	Centralidade de Intermediao por Menores Caminhos
CCA	Centralidade da Caminhada Aleatoria

SUMÁRIO

INTRODUÇÃO.....	18
CAPÍTULO I – Caminhada Aleatória.....	24
1.1 Breve Histórico	24
1.2 Caminhada Aleatória Unidimensional.....	26
1.3 Caminhada Aleatória Bidimensional.....	31
1.4 Processos Markovianos.....	37
1.5 Caminhada Aleatória em termos de cadeias de Markov.....	41
CAPÍTULO 2- Teoria de Grafos	45
2.1 Origem da teoria de Grafos.....	45
2.2 Definição e Propriedades de Grafos.....	48
2.3 Caminhada Aleatória em Redes Complexas.....	57
CAPÍTULO 3 - Redes Complexas.....	63
3.1 Histórico, Desenvolvimento e Aplicações de Redes	63
3.2 Aplicações de Redes Complexas.....	63
3.3 Propriedades das Redes Complexas.....	66
3.4 Modelos de Redes.....	69
CAPÍTULO 4 - Centralidade em Redes Complexas	79
4.1 Medidas de Centralidade em Redes Complexas.....	79
4.2 Centralidade de Grau.....	80
4.3 Centralidade de Proximidade.....	80
4.4 Centralidade de Intermediação por Menores Caminhos.....	81
4.5 Centralidade de Intermediação de Percursos Aleatórios.....	83
4.6 Centralidade da Caminhada Aleatória	85
CAPÍTULO 5 - Discussão dos Resultados	91
5.1 Aspectos Gerais	91
5.2 Tempo Característico de Relaxação.....	91

5.3 Relação entre as diferentes medidas de Centralidade.....	94
5.4 Eficiência das diferentes medidas de Centralidade.....	100
CAPÍTULO 6 – Conclusões e Trabalhos Futuros	109
CAPITULO 7 –Referências Bibliográficas.....	112
ANEXOS.....	117
ANEXO A – Algoritmo e Simulações	117
ANEXO B - Leis dos Grandes Números.....	125
ANEXO C - Teorema Central do Limite.....	127
ANEXO D - Contagem dos caminhos para os nós 2 e 3, referente a Fig.(4.1).....	129
ANEXO E - Lemas utilizados para a implementação do algoritmo para CIMC.....	130

INTRODUÇÃO

Os estudos sobre redes complexas nos auxiliam a compreender e modelar muitas situações do mundo real. O mundo é abundante em redes, e estas podem ser encontradas em diversos contextos, como por exemplo, redes sociais reais ou virtuais, redes de informação, comunicação, energia, transporte, biológicas, entre outras. Essas redes do mundo real podem ser representadas mediante classificações específicas de redes complexas, de acordo com suas especificidades e comportamento. A teoria que nos ajuda a estudar e compreender as redes complexas é a teoria de grafos [1]. Por meio desta, é possível representar as redes complexas, estudar as características específicas das relações entre o conjunto de vértices e o conjunto de arestas que formam a rede, o que permite entender alguns aspectos importantes de sua estrutura, permitindo classificá-las de acordo com suas especificidades. Existem diferentes modelos de redes complexas, dentre os quais podemos citar as redes Aleatórias, redes de Pequeno Mundo, redes Sem Escala e redes Hierárquicas. Cada um desses modelos possui suas particularidades.

As redes Aleatórias são possuem o modelo mais simples e foram estudadas por Erdős e Rényi, em 1959 [2]. O modelo proposto por estes matemáticos partem de uma rede com N vértices desconectados e m arestas que são adicionadas aleatoriamente na rede. A cada inserção de arestas entre dois vértices é associada uma probabilidade p de ocorrer ou não a conexão. De maneira simplificada, as redes Aleatórias são redes onde as ligações entre os pares de nós são dadas de maneira aleatória.

O fenômeno de Pequeno Mundo é observado em muitas redes reais, e foi inicialmente proposto pelo psicólogo social Milgran, em 1967. Ele realizou estudos por meio de redes sociais reais, onde os vértices eram as pessoas e as arestas se estabeleciam entre os vértices, caso as pessoas se conhecessem. Por meio do trabalho proposto por Milgran, pode-se perceber que a distância média entre duas pessoas quaisquer pertencente à rede social é pequena. Esta distância está relacionada com o número de pessoas intermediárias entre outras duas quaisquer, de modo que a partir de uma pessoa se consiga chegar a outra. Em média são suficientes seis pessoas intermediárias para conectá-las [3]. As redes que apresentam a propriedade de pequeno mundo têm como características principais um alto coeficiente de aglomeração e um pequeno comprimento de caminho médio, ou seja, a rede de Pequeno Mundo é uma rede que ocupa uma posição intermediária entre a rede Regular e a rede Aleatória. Em 1998, Watts e Strogatz propuseram um

modelo que reproduz o efeito de Pequeno Mundo [4]. Para tal, eles partiram de uma rede anel regular e religavam as arestas com uma probabilidade p . As ligações formam atalhos na rede, fazendo com que o alto agrupamento comum nas redes Regulares seja mantido. Por outro lado, os atalhos fazem com que a distância média entre os vértices da rede diminuam, como ocorre nas redes Aleatórias.

Barabási e Albert, [5] mostraram que muitas redes reais seguiam uma distribuição de grau lei de potência. Redes com essa característica apresentam a maioria dos nós, com baixa conectividade e um pequeno número de nós altamente conectados [6]. O diferencial do modelo proposto por estes autores é representar o caráter crescente observado nas redes reais, permitindo a inserção de novos nós. Outra característica importante é que as conexões entre os vértices ocorrem por meio de ligações preferenciais entre os nós. O crescimento da rede acontece a partir de um número pequeno de nós n_0 , e em cada novo instante de tempo, um novo nó é adicionado à rede com m ligações que serão estabelecidas preferencialmente aos nós que possuem maior grau [7]. Este modelo ficou conhecido como modelo de Barabási-Albert, e por meio dele é possível gerar redes Sem Escalas que seguem uma distribuição de grau lei de potência.

Outro exemplo de rede que se caracteriza por ser uma rede Sem Escala são as redes Hierárquicas. Estas redes possuem uma organização e podem ser construídas a partir de uma regra recursiva que pode ser determinística [8]. Muitas redes reais apresentam o coeficiente de aglomeração elevado e apresentam uma distribuição de grau em Lei de Potência. O modelo capaz de unir estas duas características na rede são as redes hierárquicas, construídas deterministicamente. A rede é construída de maneira iterativa permitindo o crescimento da rede ao longo do tempo. Inicia-se a partir de um grupo pequenos de nós densamente conectados entre si. Em cada etapa do processo de geração, réplicas da iteração anterior são geradas e adicionadas na rede por meio de ligações com o nó central do antigo agrupamento denominado de raiz. Isso faz com que sejam gerados poucos nós altamente conectados e muitos nós com baixa conexão.

As redes complexas podem ser analisadas sob diversos pontos de vista e com as mais variadas finalidades. Devido a esse fato, a apresentação de algumas propriedades e características topológicas das redes complexas se torna útil. Entre estas medidas podemos destacar o Grau do vértice, a Componente Gigante, o Caminho Médio e Centralidade dos vértices.

A característica mais simples que podemos observar em uma rede complexa é o número de conexões que os vértices da rede possuem. Esta medida é denominada de grau do vértice. Em redes que são direcionadas, é possível distinguir três tipos de graus. O grau de saída, o grau de entrada e o grau total, que é a soma dos outros dois.

Componentes de uma rede são subgrafos conexos, que podem ser identificados na rede. Subgrafos conexos são grupos de nós ligados internamente, em que sempre haverá um caminho conectando dois nós quaisquer. Em geral, as redes são compostas por várias componentes, sendo que estas são desconectadas entre si. Quando a componente conexa possuir o número de nós próximo ao número de nós total da rede, esta é denominada de componente gigante.

O coeficiente de aglomeração refere-se ao agrupamento intrínseco dos vértices de uma rede. Esta característica está relacionada com o número de triângulos formados na rede [9]. A aglomeração será alta quando dois vértices que compartilham um nó vizinho possuírem uma probabilidade alta de estarem ligados uns aos outros.

A distância entre dois vértices quaisquer da rede é o número de arestas a serem percorridas para chegar de um vértice a outro. A menor distância é obtida quando consideramos o número mínimo de arestas para percorrer, sendo denominado menor caminho. A partir desta, é possível definir duas medidas: o Diâmetro da rede e o Comprimento de Caminho Médio. O Diâmetro da rede é a máxima distância entre quaisquer dois vértices da rede. O Comprimento de Caminho Médio está relacionado com os menores caminhos que ligam dois nós da rede. Assim, o comprimento de caminho médio é a média dos menores caminhos da rede.

A medida de centralidade de um vértice na rede está relacionada com o número de vértices possíveis de serem atingidos a partir dele. Há diversas formas de centralidade, em que cada definição leva em consideração diferentes especificidades relevantes da rede. Estas medidas de centralidade, a princípio diferentes, possuem uma mesma função que é avaliar a importância dos nós em uma rede. Podemos citar alguns tipos de centralidade tais como a Centralidade de Grau, Centralidade de Intermediação e a Centralidade da Caminhada Aleatória. Assim, existem diversas medidas de centralidade que são usadas para interpretar a importância desempenhada pelos vértices da rede.

O processo da caminhada aleatória em redes complexas [10] se torna importante para compreender a eficiência das diversas medidas de centralidade. A partir da análise desse processo é possível verificar a solução para a questão que norteia esta pesquisa:

Existe uma medida de centralidade mais eficiente para cada tipo de rede complexa?

Neste sentido, a caminhada aleatória é um processo dinâmico fundamental, sendo de interesse na investigação de como a heterogeneidade estrutural afeta a dinâmica de difusão [11]. O estudo da caminhada aleatória também é interessante, por ser um mecanismo de transporte e pesquisa em redes [12]. A caminhada aleatória em redes é definida por meio da equação mestra $P_{ij}(t + 1) = \sum \frac{A_{jt}}{k_{jt}} P_{ijt}(t)$, que indica a probabilidade de um caminhante aleatório situado no nó i no tempo t se encontrar em j no tempo $t + 1$, onde A_{ij} é um elemento da matriz adjacência e k_j é o grau de cada nó da rede.

As primeiras referências com relação à caminhada aleatória foram por meio das observações feitas nas experiências realizadas por Robert Brown em 1828. Brown, observou que pequenas partículas imersas em uma solução adquiriam um movimento irregular, que ficou conhecido como movimento browniano. Outras situações em que o comportamento era semelhante à dessas partículas mais tarde receberam o nome de caminhada aleatória. A teoria sobre Caminhada Aleatórias se concretizou em meados do século XX, com a publicação de trabalhos em que a caminhada aleatória era vista como um processo estocástico [13]. Atualmente, esta teoria pode ser estudada como um mecanismo de pesquisa e transporte de informação, ou ainda ser interpretada como processos difusivos que são frequentemente encontrados na natureza. Como exemplos de tais situações, podem-se citar a propagação de doenças, reações químicas e fluxo de informações.

Os caminhantes aleatórios difundidos sobre uma rede podem ser vistos como distribuidores de informações. Se considerarmos a entrega de informação entre dois nós da rede, mediante o processo da caminhada aleatória, esta seria ótima se os caminhantes seguissem sempre pelo menor caminho entre os dois nós. Mas isso, não ocorre uma vez que o caminhante aleatório não conhece o menor caminho e sua escolha a cada passo é feita de forma equiprovável. Os nós que recebem o maior número de informações ao longo do tempo são considerados com maior centralidade dentro da rede. A partir do movimento aleatório é possível calcular a Centralidade da Caminhada Aleatória (CCA) [14] dos vértices da rede. Em redes finitas é possível calcular os caminhos ótimos entre quaisquer dois vértices e, dessa forma, determinar a Centralidade de Intermediação por Menores Caminhos (CIMC) [15]. Partindo dessas considerações, realizamos a comparação entre as três diferentes medidas de centralidade, sendo elas o Grau do Vértice, a CIMC e a CCA.

Com a realização desta pesquisa, espera-se alcançar dois objetivos principais:

- Discutir sobre a importância da caminhada aleatória em redes complexas;
- Comparar a eficiência das medidas de centralidade em diferentes modelos de redes complexas.

A organização do presente trabalho é realizada em seis capítulos, a saber:

O Capítulo 1 apresenta uma revisão sobre o processo da Caminhada Aleatória, suas primeiras observações. Seguido da formalização do processo, desenvolvido para caso unidimensional, seguindo-se com a expansão para o caso bidimensional. Posteriormente é realizada a generalização em termos de redes complexas, sendo que para isso, foi usado o formalismo do processo Markoviano.

No Capítulo 2, é feita uma explanação sobre a Teoria de Grafos e como sua utilização é importante para a representação e compreensão das redes complexas. As considerações realizadas levam em consideração o surgimento dessa teoria com Euler e até seu posterior desenvolvimento. Em seguida, são abordados os principais conceitos e propriedades dos grafos, assim como sua utilização. As representações matriciais dos grafos também são importantes para o estudo da caminhada aleatória e das redes complexas.

O Capítulo 3 apresenta o histórico e o desenvolvimento das redes complexas, assim como suas aplicações nas diferentes áreas do conhecimento. Em seguida, apresenta-se uma breve descrição das diversas formas de redes complexas encontradas no mundo real. Apresentam-se também as diversas propriedades estruturais das redes complexas, bem como, a evolução dos modelos de redes complexas desenvolvidos para melhor representar as situações reais.

No Capítulo 4 são abordadas algumas medidas de centralidades estudadas nas redes complexas, as quais são encontradas na Literatura. É realizado um resgate do surgimento do conceito de centralidade no âmbito das redes sociais, e sua posterior expansão para os demais tipos de redes. Apresentam-se em seguida, as diferentes medidas de centralidade para as redes complexas, desde as centralidades mais simples como a Centralidade de Grau até as mais sofisticadas como a CCA e CIMC, assim como as especificidades de cada uma destas.

O Capítulo 5 aborda os resultados obtidos mediante simulações numéricas da caminhada aleatória nas quatro diferentes modelos de redes, como a rede Aleatória, Pequeno Mundo, Sem Escala e Hierárquica, sendo calculado numericamente o resultado para três diferentes medidas de Centralidade. Em seguida, é realizada a análise da

eficiência destas medidas de centralidades, usando para tal, a evolução temporal da caminhada aleatória nas redes complexas, sendo que esta é organizada a partir das diferentes medidas de centralidade. Por fim, apresentam-se as principais discussões dos resultados obtidos no trabalho e suas considerações.

No compromisso assumido nesta pesquisa, espera-se divulgar a importância do estudo da caminhada aleatória, sendo que esta pode ser utilizada em diversas aplicações que ajudam a compreender melhor a estrutura das redes complexas. Além disso, espera-se compreender a eficiência das medidas de centralidade para os modelos de redes distintas. Assim como, descrever algumas possíveis aplicações e o desenvolvimento para a área de sistemas complexos.

CAPÍTULO I

Caminhada Aleatória

A ordem na natureza esconde uma desordem subjacente invisível, que só pode ser compreendida pelas regras da aleatoriedade.

Leonard Mlodinow.

1.1 Breve Histórico

As primeiras referências com relação à caminhada aleatória foram decorrentes das observações empíricas a partir das experiências realizadas por Robert Brown em 1828 [13]. Este botânico inglês observou que pequenas partículas imersas em uma solução adquiriam um movimento bastante irregular, que ficou conhecido como Movimento Browniano. Outras situações em que o comportamento era semelhante à dessas partículas foram estudadas, as quais mais tarde receberiam o nome de Caminhada Aleatória.

Inúmeras tentativas foram realizadas posteriormente para desvendar a natureza do movimento browniano. Mediante aos experimentos laboratoriais verificou-se que o movimento se intensifica quando era reduzida a viscosidade do meio ou o tamanho das partículas brownianas. Outro fator que também influencia o movimento é a elevação da temperatura da solução [16]. Em 1860, o andar errático das partículas começou a ser visto por outro ângulo. Neste período consideravam a possibilidade do movimento estar relacionado com as colisões com as moléculas do fluido.

A verdadeira explicação do fenômeno foi proposta por Einstein no seu artigo de 1905. O tratamento de Einstein representou um grande avanço científico tornando a teoria atômico-molecular uma parte fundamental da estrutura da matéria. Atualmente, o movimento browniano permanece na fronteira da pesquisa como um exemplo importante de processo estocástico, sendo abordado mais tarde por meio de caminhadas aleatórias por Mark Kac [16].

Por volta do século XVI Pascal, Fermat e Bernoulli já trabalhavam com problemas envolvendo variáveis aleatórias. Entretanto, a teoria sobre Caminhada Aleatórias se concretizou apenas por volta do século XX, com a publicação de trabalhos que consideravam a caminhada aleatória como um processo estocástico [13]. Os processos

estocásticos são fenômenos que variam em algum grau, de forma imprevisível, à medida que o tempo passa. Portanto, a caminhada aleatória se caracteriza como tais processos, pois assume valores determinados aleatoriamente para seus passos, em um determinado intervalo temporal.

O termo caminhada aleatória, *random walk*, foi proposto por Karl Pearson, em 1905. Ele escreveu uma carta para a revista *Nature*, onde descrevia um problema sobre um homem que caminhava aleatoriamente:

A man starts from a point O and walks l yards in a straight line; He then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after n of these stretches He is at a distance between r and r + δr from his starting point, O.

Karl Pearson. The Gables, East Iisley, Berks.

(HUGHES, 1995, p. 53)

A carta foi respondida por Lord Rayleigh, que havia resolvido anteriormente em 1880 um problema semelhante, porém mais geral, ao proposto por Pearson [10]. Pearson possuía particular interesse na caminhada aleatória. Ele tentava descrever um modelo para explicar o fenômeno de migração, mais precisamente o caso de mosquitos que invadem alguma região. No mesmo ano em que Pearson apresentou este problema à revista *Nature*, Albert Einstein publicou um artigo sobre o movimento browniano. O tratamento de Einstein para este movimento é um dos trabalhos intelectuais mais notáveis de todos os tempos [16], significando um grande avanço científico em diversas áreas como na química e na física. O movimento browniano é um exemplo importante de processo estocástico. O tratamento de Einstein baseou-se em uma equação parabólica, que surge naturalmente por meio do tratamento de passeios aleatórios, realizado por M. Kac [16].

A caminhada aleatória também comumente aparece em alguns trabalhos, sendo referida como o passeio do bêbado, uma vez que pode ser feita uma analogia do movimento descrito pela caminhada aleatória com o passeio incerto de uma pessoa alcoolizada.

Depois de muitas horas passadas num bar, um bêbado resolve voltar para casa. Só que ele não se lembra de onde mora e começa a andar aleatoriamente pela cidade: a cada esquina, ele escolhe uma das quatro possíveis direções, sem preferência, e anda até o próximo quarteirão. No caminho, ele tenta calcular a probabilidade de passar em frente à sua casa... (Descrição do passeio aleatório bidimensional, devida a Einstein).

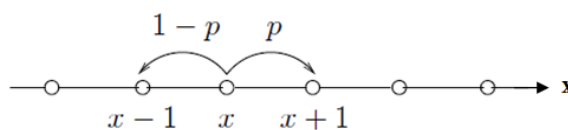
(FRIEDLI, 2011, p.15)

A teoria da caminhada Aleatória pode ser aplicada em diversas situações de estudo, como na Matemática, na Física, Química, na Economia, na Ciência da Computação, entre outros. Ela pode ser vista como um mecanismo de pesquisa e transporte de informação, ou ainda, pode ser interpretada como processos difusivos, frequentemente encontrados na natureza. Como exemplos de tais situações, pode-se citar a propagação de doenças, reações químicas e o fluxo de informações.

1.2 Caminhada Aleatória Unidimensional

A caminhada aleatória pode ser entendida a partir de uma partícula que desempenha um movimento, ou a partir de um caminhante se movendo sobre uma linha reta iniciando a sua caminhada por uma determinada origem, assumindo valores para a posição x ; levando-se em conta que a cada intervalo de tempo, designado por θ , este caminhante ou partícula percorre uma distância fixa h para direita ou para a esquerda, como ilustrado na Fig. 1.1. Se p é a probabilidade de locomoção para a direita e q probabilidade de que tome o sentido contrário, então $p + q = 1$. Quando p e q forem iguais a $\frac{1}{2}$, dizemos que a caminhada é simétrica.

Figura 1.1 Caminhada Aleatória Unidimensional, x é a posição atual do caminhante, $x + 1$ e $x - 1$ são os possíveis posições para o passo seguinte, p e $1 - p$, são as respectivas probabilidade para esse passo.



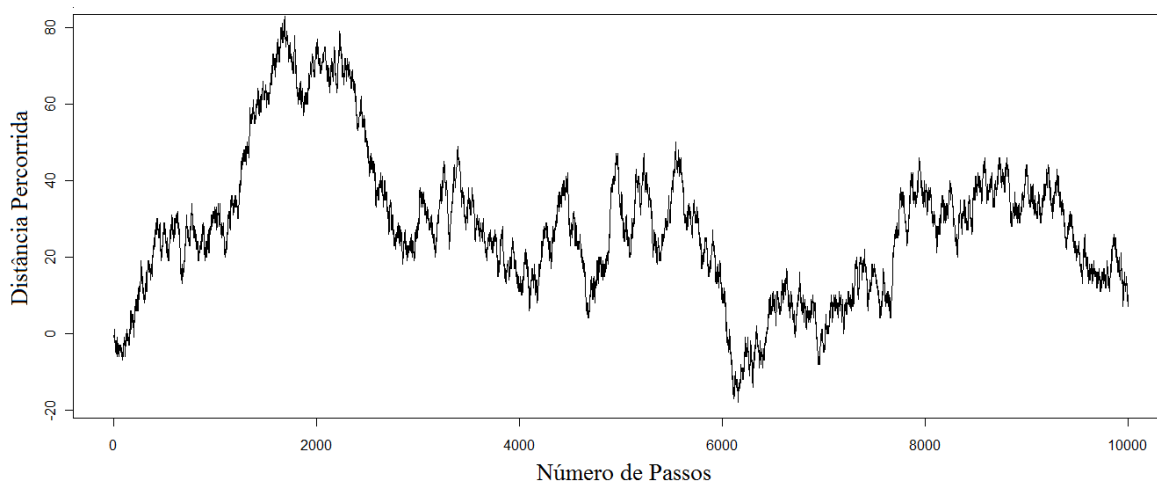
Fonte: Próprio autor.

O movimento de várias partículas pode ser descrito por meio da introdução de n variáveis aleatórias independentes $\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_n$. Estas variáveis assumem os valores $+1$ ou -1 , conforme o sentido do salto [18]. Dessa forma, se considerarmos a variável σ_j , dependendo do valor que ela possa assumir ($+1$ com probabilidade p , e -1 com probabilidade q), esta indicará se a partícula irá saltar para a direita ou para a esquerda respectivamente, no j -ésimo instante. Para determinar a posição deste caminhante em um

determinado instante $t = n\theta$, é necessário obter o valor da soma de todas as variáveis aleatórias até este instante, $m = \sigma_1 + \sigma_2 + \sigma_3 + \dots + \sigma_n$. Assim, a posição será $x = hm$.

A Fig. 1.2 ilustra a situação de uma partícula, que parte da origem de um eixo ordenado, em um movimento aleatório unidimensional. Esta partícula efetua uma caminhada aleatória seguindo um passo por instante de tempo. O passo dado pela partícula pode ser para direita ou para esquerda com igual probabilidade, assumindo os valores 1 ou -1 , respectivamente. Após uma determinada quantidade de passos ($t = 10000$), nota-se que a distância final da partícula em relação a origem é pequena, se comparada com a distância real que poderia atingir se tomasse apenas um sentido para seu passeio. As simulações foram realizada mediante a programação por meio do *software* livre R (Anexo - A).

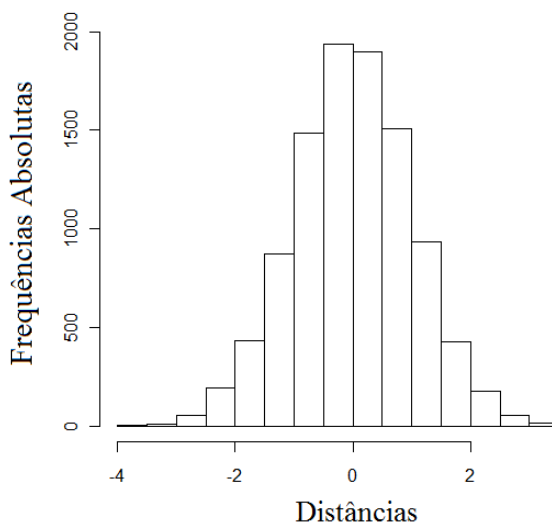
Figura 1.2 Caminhada Aleatória Simétrica Unidimensional, ilustrando a Distância percorrida por uma partícula com relação aos passos dados, onde o número de passos é 10000.



Fonte: Próprio autor.

Portanto, na caminhada aleatória simétrica unidimensional, após um número significativo de passos, o caminhante que desempenha o percurso pouco se afasta da origem, conforme pode ser observado na Fig. 1.3. A região próxima ao ponto inicial de partida da partícula é o ponto mais visitado durante o passeio sendo este o lugar mais provável de se encontrar o caminhante aleatório. O histograma mostra ainda que a distribuição de frequências tende a uma curva gaussiana, sendo esta uma característica importante da caminhada aleatória.

Figura 1.3 Histograma referente à Caminhada Aleatória Simétrica Unidimensional, desempenhada por uma partícula, após 10000 passos.



Fonte: Próprio autor.

O passeio aleatório é constituído por uma sucessão de eventos independentes. Assim, o próximo passo não depende dos resultados anteriores. Neste sentido, existem dois teoremas sobre sequência de variáveis independentes que nos ajudam a compreender melhor o comportamento destas sequências, sendo estes, A Lei dos Grandes Números e Teorema Central do Limite que podem ser vistos no Anexo – B e Anexo – C.

De acordo com as probabilidades associadas a esse evento e levando em consideração que m é uma soma de variáveis independentes é possível determinar alguns dados estatísticos [19], como os que seguem abaixo:

Média: A média de uma variável aleatória independente é definida por: $\langle X \rangle = \sum_i x_i p(x_i)$, onde x_i são os possíveis valores assumidos pela variável aleatória discreta X , e $p(x_i)$ é a função de probabilidade para esta variável [20]. Assim, para o caso do caminhante que se desloca sobre uma linha, tem-se que a média denotada por a é:

$$a = \langle \sigma_j \rangle = (+1)p + (-1)q$$

$$a = \langle \sigma_j \rangle = p - q. \quad (1.1)$$

Variância: Da mesma forma pode-se encontrar a variância de uma variável aleatória, sendo que esta é definida da seguinte maneira: $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$, em que $\langle x^2 \rangle$

representa o segundo momento da distribuição [20]. Assim, a variância de σ_j é denotada por b :

$$\begin{aligned} b &= \langle \sigma_j^2 \rangle - \langle \sigma_j \rangle^2 \\ b &= [(+1)^2 p + (-1)^2 q] - (p - q)^2 \\ b &= p + q - p^2 + 2pq - q^2. \end{aligned}$$

Como $p + q = 1$, temos:

$$\begin{aligned} b &= 1 - p^2 + 2p(1 - p) - (1 - p)^2 \\ b &= 1 - p^2 + 2p - 2p^2 - 1 + 2p - p^2 \\ b &= -4p^2 + 4p \\ b &= 4p(1 - p) \\ b &= 4pq. \end{aligned} \tag{1.2}$$

Função Característica: A função característica $g(k)$ de uma variável aleatória x é definida como a transformada de Fourier da densidade de probabilidade $\rho(x)$ associada a x , isto é:

$$g(k) = \int \rho(x) e^{ikx} dx = \langle e^{ikx} \rangle.$$

Para o caso discreto, é escrita como:

$$g(k) = \sum e^{ikx} p(x) = \langle e^{ikx} \rangle.$$

Podemos então escrever a função característica para o caso estudado como:

$$\begin{aligned} g(k) &= \sum e^{ik\sigma_j} p(\sigma_j) \\ g(k) &= pe^{ik(+1)} + qe^{ik(-1)} \\ g(k) &= pe^{ik} + qe^{-ik}. \end{aligned} \tag{1.3}$$

Mediante a função característica, é possível determinar a probabilidade $P_n(m)$ do caminhante estar distante m passos do ponto inicial, após ter decorrido n intervalos de tempo. De fato, seja:

$$G_n(k) = [g(k)]^n = (pe^{ik} + qe^{-ik})^n, \quad (1.4)$$

$G_n(k)$ é dada por:

$$G_n(k) = \sum_{m=-n}^n P_n(m) e^{ikm}, \quad (1.5)$$

em que m toma os valores $-n, -n + 2, \dots, n - 2, e n$.

A partir dessa equação, podemos escrever sua expansão binomial:

$$G_n(k) = \sum_{l=0}^n \binom{n}{l} (pe^{ik})^l (qe^{-ik})^{n-l} = \sum_{l=0}^n \binom{n}{l} p^l q^{n-l} e^{ik(2l-n)}. \quad (1.6)$$

Podemos comparar a expressão obtida na Eq.(1.5) com esta última Eq. (1.6); convenientemente, por meio da mudança de variável $m = 2l - n$, a probabilidade $P_n(m)$ pode ser assim expressa:

$$P_n(m) = \frac{n!}{\left(\frac{n+m}{2}\right)! \left(\frac{n-m}{2}\right)!} p^{(n+m)/2} q^{(n-m)/2}. \quad (1.7)$$

Trata-se de uma distribuição de probabilidade Binomial. Neste caso, como se referem à n variáveis independentes, as quais possuem a mesma distribuição de probabilidade, e de igual forma a mesma função característica $g(k)$, a média e a variância de m são dadas respectivamente, por:

$$\langle m \rangle = na = n(p - q) \quad (1.8)$$

e

$$\langle m^2 \rangle - \langle m \rangle^2 = nb = 4npq. \quad (1.9)$$

Se desejarmos obter a distribuição de probabilidade para $n \gg 1$, basta utilizar o Teorema Central do Limite [19], já que as variáveis $\sigma_1, \sigma_2, \sigma_3, \dots$ são independentes. A essa distribuição de probabilidade corresponde então uma distribuição gaussiana, confirmando o resultado obtido mediante da simulação apresentada pela Fig. 1.3:

$$P_n(m) = \frac{1}{\sqrt{2\pi nb}} \exp\left\{-\frac{(m-na)^2}{2nb}\right\}. \quad (1.10)$$

Sabendo que $t = n\theta$ e $x = mh$, a função densidade de probabilidade ρ é função de x e t , em que $\rho(x, t) = P_n(m)/h$ é dada por:

$$\rho(x, t) = \frac{1}{h\sqrt{2\pi nb}} \exp\left\{-\frac{\left(\frac{x-nah}{h}\right)^2}{2nb}\right\} = \frac{1}{\sqrt{2\pi\frac{tbh^2}{\theta}}} \exp\left\{-\frac{\left(\frac{x-\frac{tha}{\theta}}{\frac{tbh^2}{\theta}}\right)^2}{2\frac{tbh^2}{\theta}}\right\}. \quad (1.11)$$

Denotando c e D respectivamente por:

$$c = \frac{ha}{\theta} = \frac{h(p-q)}{\theta} \quad (1.12)$$

e

$$D = \frac{h^2b}{\tau} = \frac{h^24pq}{\tau}, \quad (1.13)$$

e substituindo nas equações (1.12) e (1.13) em (1.11) encontramos:

$$\rho(x, t) = \frac{1}{\sqrt{2\pi Dt}} \exp\left\{-\frac{(x-ct)^2}{2Dt}\right\}, \quad (1.14)$$

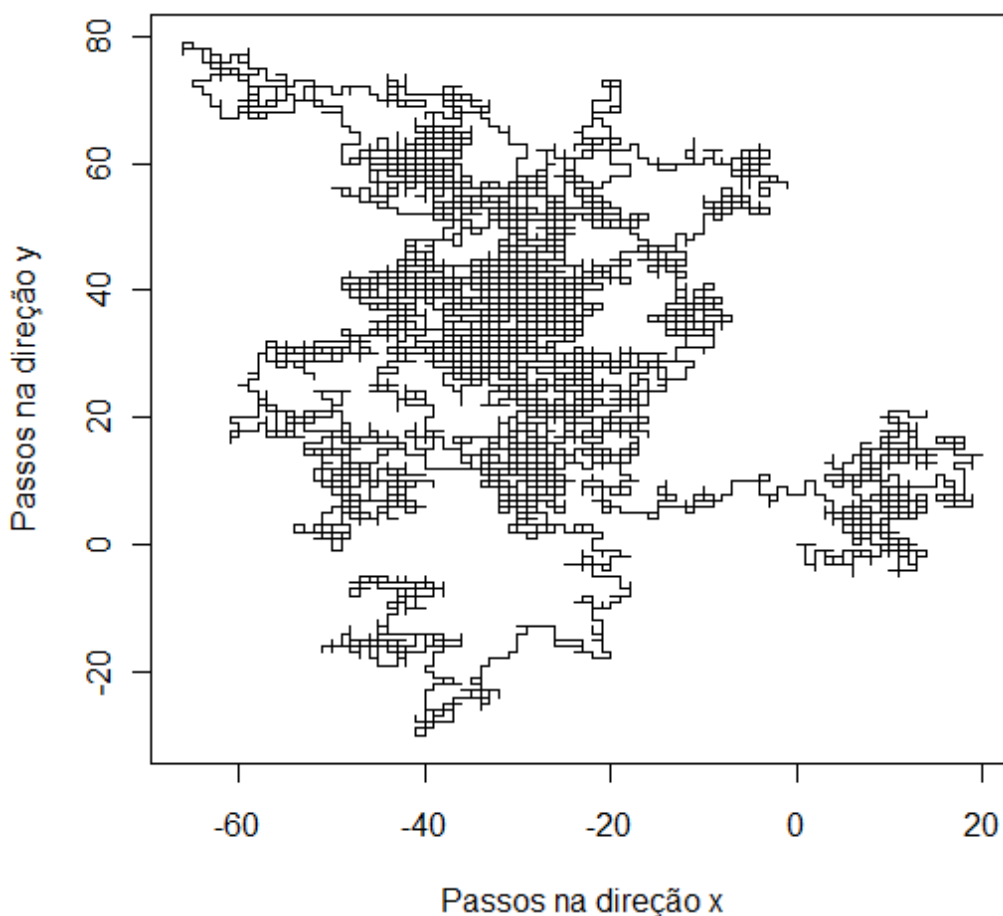
em que $\langle x \rangle = ct$ e $\langle x^2 \rangle - \langle x \rangle^2 = Dt$.

Assim, c é a velocidade média da partícula e D é o coeficiente de difusão, supondo um sistema composto por muitas partículas que desempenham passeios aleatórios independentes [19]. A densidade de partículas é proporcional a densidade de probabilidade ρ , dada pela Eq.(1.14), considerando que no instante inicial todas as partículas se encontram na origem, ocorrendo o espalhamento posterior destas partículas de acordo com a distribuição de probabilidade ρ . Para tempos longos, a densidade de partículas será uma gaussiana centrada em $x = ct$ e com largura $\Delta = \sqrt{Dt}$.

1.3 Caminhada Aleatória Bidimensional

A caminhada aleatória bidimensional pode ser modelada quando consideramos que um caminhante se movendo em um espaço de duas dimensões. Assim, a partícula se moverá para uma nova posição no plano em cada novo passo [18,19]. O intervalo de tempo

Figura 1.5 Caminhada Aleatória Bidimensional sobre uma malha quadriculada com $t = 10000$.

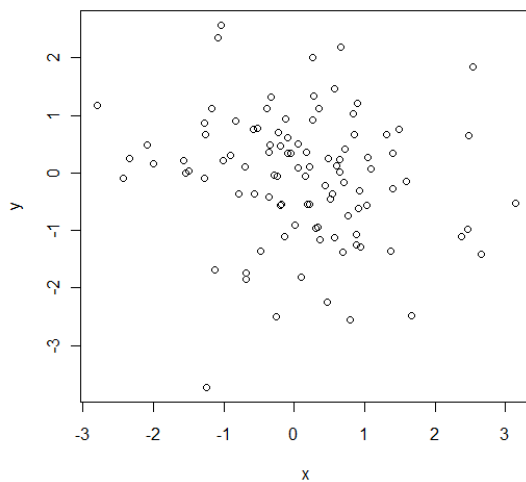


Fonte: Próprio Autor.

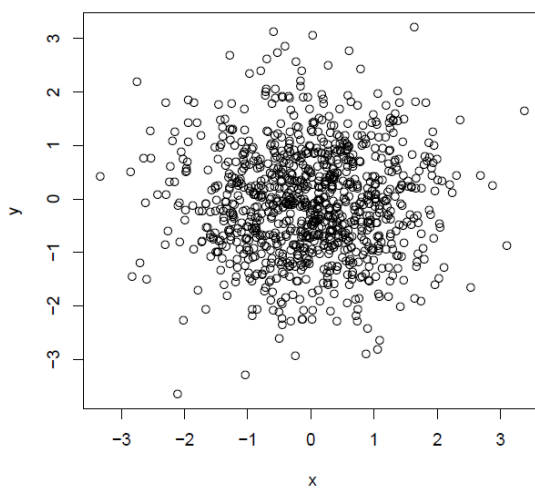
Após uma determinada quantidade de passos nota-se que a distância final da partícula em relação a origem no caso bidimensional também é relativamente pequena se considerar a distância que poderia atingir caso percorresse em apenas uma direção. Se o caminhante tomasse apenas uma das quatro possíveis direções para todos os passos do percurso, a distância que o caminhante percorreria seria $t \cdot h$, em que t é o número de iterações e h é o comprimento do passo. Porém, como ele pode escolher quaisquer das quatro direções, com igual probabilidade e ele não leva em conta a escolha passada, a distância final com relação à origem é muito menor que a distância máxima para o movimento. Este fato fica claro mediante a visualização da Fig. 1.6, em que é mostrada a distribuição das posições do caminhante aleatório após um número de passos t . Por mais que se aumente o número de passos, a partícula pouco se distancia da origem.

Figura 1.6 Distribuição das posições percorridas pelo caminhante após t passos.

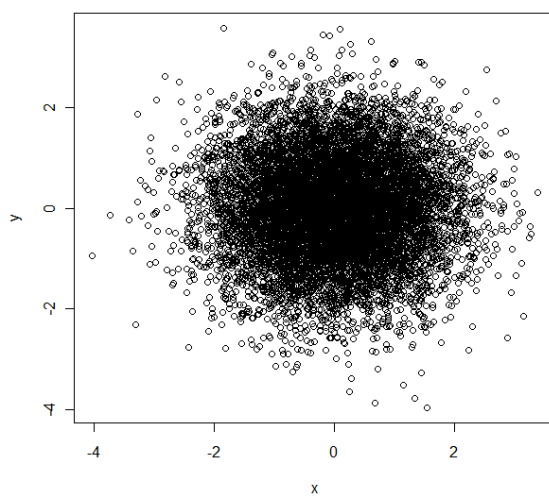
a) $t=100$



b) $t=1000$



c) $t=10000$



Fonte: Próprio Autor.

Assim como demonstramos na seção anterior, podemos identificar alguns dados estatísticos, tais como:

Função Característica: Para o caso bidimensional, a função característica correspondente ao vetor $\vec{r}_i = (x_i, y_i)$, também será composta por duas coordenadas:

$$\begin{aligned} g(\vec{k}) &= \langle e^{i\vec{k} \cdot \vec{r}_i} \rangle, \\ g(\vec{k}) &= \langle \exp\{i(k_1 x_i + k_2 y_i)\} \rangle, \end{aligned} \quad (1.17)$$

Para o caso de uma variável contínua é dado por:

$$g(\vec{k}) = \iint e^{i\vec{k} \cdot \vec{r}_i} P(\vec{r}_i) dx_i dy_i. \quad (1.18)$$

Da mesma forma, é possível escrever a função característica $G(\vec{k})$ correspondente ao vetor $\vec{r} = (x, y)$:

$$\begin{aligned} G(\vec{k}) &= \langle e^{i\vec{k} \cdot \vec{r}} \rangle, \\ G(\vec{k}) &= \langle e^{i\vec{k} \cdot (\vec{r}_1 + \vec{r}_2 + \vec{r}_3 + \dots + \vec{r}_n)} \rangle, \\ G(\vec{k}) &= \langle e^{i\vec{k} \cdot \vec{r}_1} \cdot e^{i\vec{k} \cdot \vec{r}_2} \cdot e^{i\vec{k} \cdot \vec{r}_3} \cdot \dots \cdot e^{i\vec{k} \cdot \vec{r}_n} \rangle, \\ G(\vec{k}) &= \langle e^{i\vec{k} \cdot \vec{r}_i} \rangle^n, \\ g(\vec{k}) &= [g(\vec{k})]^n. \end{aligned} \quad (1.19)$$

A densidade de probabilidade do vetor aleatório \vec{r} é calculada por meio da técnica de expansão em termos de cumulantes (faremos a expansão até a segunda ordem):

$$\begin{aligned} g(\vec{k}) &= \exp \left\{ i\vec{k} \cdot \vec{\kappa}_1 + \frac{(i\vec{k})^2}{2} \cdot \vec{\kappa}_2 \right\}, \\ g(\vec{k}) &= \exp \left\{ i(k_1 a_1 + k_2 a_2) - \frac{1}{2} (b_{11} k_1^2 + 2b_{12} k_1 k_2 + b_{22} k_2^2) \right\}, \end{aligned} \quad (1.20)$$

em que:

$$a_1 = \langle x_i \rangle \text{ e } a_2 = \langle y_i \rangle, \quad (1.21)$$

são os cumulantes de primeira ordem que coincidem com a média.

Os valores:

$$\begin{aligned} b_{11} &= \langle x_i^2 \rangle - \langle x_i \rangle^2, \\ b_{12} &= \langle x_i y_i \rangle - \langle x_i \rangle \langle y_i \rangle, \\ b_{22} &= \langle y_i^2 \rangle - \langle y_i \rangle^2, \end{aligned} \quad (1.22)$$

são os cumulantes de segunda ordem. Assim, para $t = n\tau$ grande temos:

$$G(\vec{k}) = \exp \left\{ in(k_1 a_1 + k_2 a_2) - \frac{n}{2} (b_{11} k_1^2 + 2b_{12} k_1 k_2 + b_{22} k_2^2) \right\}. \quad (1.23)$$

A densidade de probabilidade $P_n(\vec{r}) = P_n(x, y)$ do vetor $\vec{r} = (x, y)$ é obtida por meio da Antitransformada de Fourier, da equação da função Característica:

$$P_n(\vec{r}) = \frac{1}{(2\pi)^2} \iint e^{-i\vec{k} \cdot \vec{r}} G(\vec{k}) dk_1 dk_2. \quad (1.24)$$

Substituindo a expressão (1.23) em (1.24), temos:

$$P_n(\vec{r}) = \frac{1}{(2\pi)^2} \iint e^{-i\vec{k} \cdot \vec{r}} \exp \left\{ in(k_1 a_1 + k_2 a_2) - \frac{n}{2} (b_{11} k_1^2 + 2b_{12} k_1 k_2 + b_{22} k_2^2) \right\} dk_1 dk_2. \quad (1.25)$$

Resolvendo a Integral dupla obtemos a expressão que é uma gaussiana bidimensional:

$$P_n(x, y) = \frac{1}{2\pi\sqrt{n^2 D}} \cdot \exp \left\{ -\frac{1}{2nD} [b_{22}(x - na_1)^2 + 2b_{12}(x - na_1)(y - na_2) + b_{11}(y - na_2)^2] \right\}, \quad (1.26)$$

em que $D = b_{11}b_{22} - b_{12}^2$.

Da mesma forma é possível descrever a caminhada aleatória em espaços com dimensões superiores.

1.4 Processos Markovianos

Um processo estocástico é definido como sendo uma coleção de variáveis aleatórias x_t , que possuem uma relação de dependência com o parâmetro t , pertencente a um conjunto T . Este conjunto T pode ser tomado como o conjunto dos números Naturais, como ocorre frequentemente; porém é possível associar outros conjuntos para T . A variável x_t representa uma característica mensurável de interesse no parâmetro t , em que t é a variável tempo [19].

Os processos estocásticos são importantes para a descrição de sistemas que variam com o tempo [21]. Assim, a variável aleatória x_t é definida em um espaço chamado de Espaço de Estados, e representa o estado do sistema no tempo t . Podemos qualificar os processos estocásticos de duas formas: como Cadeias, denominadas Estado Discreto ou Sequências, denominadas de Estado Contínuo. No primeiro caso, o conjunto de variáveis x_t é definido sobre um conjunto finito ou infinito enumerável. Para o caso das Sequências, x_t é definido sobre um conjunto infinito não enumerável. Um processo estocástico em que o tempo t e a variável aleatória x_t podem ser escritos de forma discreta, fica completamente definido até um instante l pela distribuição de probabilidade conjunta ρ , em que x_t assume valores n_i , no instante $t = i$:

$$\rho_l(n_0, n_1, n_2, \dots, n_l). \quad (1.27)$$

Os processos Markovianos são um tipo especial de processos estocásticos, e como tal, evoluem no tempo de maneira probabilística [21]. Podemos definir um processo Markoviano, quando consideramos a probabilidade condicional de que x_t assume o valor n_{l+1} no tempo $t = l + 1$, dado que ela tenha tomado o valor n_0 em $t = 0$, o valor n_1 em $t = 1$, ... e n_l em $t = l$, e esta for igual à probabilidade condicional:

$$\rho_{l+1}(n_{l+1}|n_l), \quad (1.28)$$

o processo estocástico é classificado como um processo de Markov.

As probabilidades condicionais são denominadas probabilidade de transição:

$$\rho_{l+1}(n_{l+1}|n_0, n_1, \dots, n_l) = \rho_{l+1}(n_{l+1}|n_l), \quad (1.29)$$

A expressão acima significa que a probabilidade condicional de qualquer evento futuro ocorrer, dado que qualquer evento passado tenha ocorrido, é independente dos eventos passados, dependendo apenas do estado presente n_l . De maneira simplificada, no processo Markoviano, o estado futuro depende apenas do estado presente. Este processo também é conhecido como processo sem memória, pois o processo não leva em consideração os eventos passados. Conforme a definição de probabilidade condicional [22] tem-se que:

$$\rho_l(n_0, n_1, n_2, \dots, n_l) = \rho_l(n_l|n_{l-1}) \dots \rho_2(n_2|n_1)\rho_1(n_1|n_0)\rho_0(n_0). \quad (1.30)$$

É possível definir uma probabilidade $\Pi_l(n_l)$ de que a variável x_t assuma o valor n_l no instante $t = l$, em que este evento seja independente dos valores que a variável x_t possa ter assumido nos instantes anteriores:

$$\Pi_l(n_l) = \sum \rho_l(n_0, n_1, n_2, \dots, n_l). \quad (1.31)$$

Uma expressão de recorrência para a probabilidade $\Pi_l(n_l)$ pode ser escrita aplicando o somatório em ambos os membros da Eq. (1.30) e usando a expressão anterior Eq. (1.31). Para um dado $\Pi_0(n_0)$, é possível obter a probabilidade $\Pi_l(n_l)$ em qualquer instante:

$$\Pi_l(n_l) = \sum_{n_{l-1}} \rho_l(n_l|n_{l-1}) \Pi_{l-1}(n_{l-1}). \quad (1.32)$$

Como mencionado anteriormente as probabilidades condicionais $\rho_{l+1}(n_{l+1}|n_l)$ são denominadas Probabilidades de Transição do estado n_l para o estado n_{l+1} , e podem ser escritas como:

$$\rho_{l+1}(n_{l+1}|n_l) = P(n_{l+1}, n_l). \quad (1.33)$$

Assim, substituindo a equação (1.33) em (1.32) e fazendo uma mudança de índice $n_l = n$, e $n_{l-1} = m$, a equação fica:

$$\Pi_l(n) = \sum_m P(n, m) \Pi_{l-1}(m). \quad (1.34)$$

A probabilidade $P(n, m)$ pode ser interpretada como um elemento da matriz de probabilidade P . Desta forma, todos esses elementos são não negativos com $P(n, m) \geq 0$ e a soma dos elementos das colunas de P são iguais a 1, $\sum_n P(n, m) = 1$. Todas as matrizes que satisfazem estas duas condições são denominadas Matrizes Estocásticas [19]. A matriz $P(n, m)$ é interpretada como matriz de transição, uma vez que cada elemento desta matriz é uma probabilidade de transição do estado m para o estado n . Assim, se na Eq. (1.34) interpretarmos $\Pi_l(n)$ como sendo os elementos de uma matriz coluna Π_l , então esta equação pode ser reescrita na forma de um produto de matrizes:

$$\Pi_l = P \Pi_{l-1}. \quad (1.35)$$

Se levarmos em conta que é possível mediante a Eq.(1.35) encontrar todas as Π_i onde $i < l$, de forma que $\Pi_{l-1} = P \Pi_{l-2}$, $\Pi_{l-2} = P \Pi_{l-3}$, ... , $\Pi_2 = P \Pi_1$, $\Pi_1 = P \Pi_0$, podemos escrever Π_l em termos apenas da l -ésima potência da matriz P . Esta representa a probabilidade de transição do estado n para o estado m em l unidades de tempo da matriz coluna Π_0 inicial:

$$\begin{aligned} \Pi_l &= \underbrace{P \cdot P \cdot P \cdot \dots \cdot P \cdot P}_{l \text{ vezes}} \cdot \Pi_0, \\ \Pi_l &= P^l \cdot \Pi_0, \end{aligned} \quad (1.36)$$

ou ainda:

$$\Pi_l(n) = \sum_m P^l(n, m) \cdot \Pi_0(m). \quad (1.37)$$

A questão principal dos processos markovianos é determinar as propriedades da matriz estocástica P , para o caso em que $\lim_{l \rightarrow \infty} \Pi_l = \Pi$, em que Π é a solução estacionária, ou seja, quando o processo se estabiliza e não ocorrem mais mudanças de estado. Neste caso a matriz Π , satisfaz a equação:

$$P \Pi = \Pi. \quad (1.38)$$

As propriedades da matriz P , apresentadas a seguir, permitem entender alguns de seus aspectos relevantes [19]:

- 1) A matriz estocástica possui um autovalor igual a 1.

Isso porque a Eq.(1.38) representa um problema de autovalor $(P - \lambda I)\Pi = 0$, em que I representa a matriz identidade e $\lambda = 1$ é o maior autovalor da matriz estocástica [23]. A este autovalor é associado um autovetor com componentes não negativas, correspondente as probabilidades de estados estáveis. Como a matriz P é uma matriz homogênea, o autovalor $\lambda = 1$ pode ser degenerado, podendo haver mais de um autovetor correspondente ao mesmo. Faz-se necessário a normalização do autovetor. A matriz em questão pode possuir todos os estados comunicantes. Dois estados quaisquer i e j , são comunicantes se esses forem alcançáveis a partir um do outro, ou seja, a partir do estado i é possível chegar ao estado j após um número l de passos, $P^l(i, j) > 0$, e se o mesmo ocorrer de j para i . Não é necessário o valor de passos l ser iguais para todos os estados. Neste caso, os estados são chamados comunicantes e quando a matriz possui todos os estados comunicantes é denominada matriz irredutível.

- 2) Ao autovalor $\lambda = 1$ de uma matriz irredutível é não degenerado, isto é, existe apenas um autovetor, e a componente deste autovetor possui todas as suas componentes estritamente positivas. Essa propriedade é conhecida como Teorema de Perron-Frobenius. Este teorema permite dizer que a solução estacionária da Eq. (1.38) é única e $\Pi(n) > 0$.

A matriz estocástica é chamada de regular quando todos os elementos de alguma potência de P são estritamente positivos. Dizer que a matriz é regular significa o mesmo que dizer que ela é irredutível e aperiódica [23]. Assim, com relação a matriz regular:

- 3) Com exceção do autovalor $\lambda = 1$, todos os autovalores de uma matriz regular são em módulo, estritamente menores do que 1;
- 4) Quando $l \rightarrow \infty$, P^l converge para uma matriz cujas as colunas são todas iguais a Π , donde se concluí que $\Pi_l = P^l \Pi_0$, converge para Π qualquer que seja Π_0 .

Assim, para que o limite Π_l seja independente da probabilidade inicial, a probabilidade estacionária deve ser única. Para que isto ocorra o autovalor $\lambda = 1$ deve ser não degenerado, e isso ocorre com as matrizes regulares. Desse modo, quando as linhas da matriz P^l são aproximadamente iguais, e (quando $l \rightarrow \infty$), o processo Markoviano deixa de ser dependente da distribuição de probabilidade inicial Π_0 . A matriz probabilidade de transição P começa a se estabilizar, ou seja, seus elementos começam a se aproximar de um determinado valor ao longo do período, o que ocorre apenas se a Cadeia de Markov for Ergódica ou Irreduzível [19]. Neste caso, o $\lim_{l \rightarrow \infty} P(n, m)^l$ existe e independe de n .

1.5 Caminhada Aleatória em termos de uma cadeia de Markov

Uma caminhada aleatória pode ser entendida como um processo Markoviano, uma vez que o novo passo do caminhante depende somente da posição presente, não se levando em conta o percurso passado. Assim, para a decisão futura, existe apenas a influência do estado presente, que é a localização do caminhante. A posição do caminhante é dada por $x = nh$, onde $n = 0, 1, 2, \dots, N - 1$. Considera-se que o caminhante, quando chega ao final do percurso na posição $x = (N - 1)h$, possa caminhar para a posição $x = 0$. Da mesma forma, caso em que ele estiver em $x = 0$ pode caminhar para $x = (N - 1)h$ [19].

Como a probabilidade de escolha para um lado ou outro é a mesma, podemos escrever que a probabilidade de transição $P(n, m)$ de n para m será $P(n - 1, n) = P(n + 1, n) = \frac{1}{2}$. Para os demais casos a probabilidade é nula. Agora, se considerarmos que o caminhante possa permanecer onde está, a esta permanência associamos uma probabilidade p , e associamos a probabilidade $q = 1 - p$ de saída para outra posição. Assim, as probabilidades de transição para novas posições continuam sendo simétricas. Estas probabilidades podem ser escritas como:

$$P(n - 1, n) = P(n + 1, n) = \frac{1}{2}q, \quad (1.39)$$

$$P(n, n) = p. \quad (1.40)$$

Seja a probabilidade do caminhante estar na posição $x = nh$ no instante l , denotada por $\Pi_l(n)$. Se este caminhante iniciou o percurso a partir do ponto $x = 0$

podemos escrever a probabilidade inicial como sendo $\Pi_0(n) = \delta_{n0}$ e a evolução deste caminhante ao longo do tempo será dada pela equação:

$$\Pi_{l+1}(n) = \sum_m P(n, m) \cdot \Pi_l(m). \quad (1.41)$$

Utilizando-se da condição de periodicidade $\Pi_l(N + n) = \Pi_l(n)$ para o caso considerado e a Eq.(1.41), a probabilidade pode ser escrita como:

$$\Pi_{l+1}(n) = \frac{1}{2}q\Pi_l(n + 1) + p\Pi_l(n) + \frac{1}{2}q\Pi_l(n - 1). \quad (1.42)$$

A matriz se caracteriza como uma matriz de Toeplitz [19], que possui todos os elementos de uma determinada diagonal iguais, possuindo as seguintes propriedades:

$$1) P(n, m) = f(n - m); \quad (1.43)$$

2) Com relação aos autovetores:

$$\psi_k(n) = \frac{1}{N} e^{ikn}, \quad k = \frac{2\pi}{N} j, \quad \text{com } j = 0, 1, 2, \dots, N - 1. \quad (1.44)$$

Em que $f(n)$ é periódica $f(n + N) = f(n)$ e possui as mesmas propriedades de $P(n, m)$, ou seja $f(n) \geq 0$ e $\sum_n f(n) = 1$.

Para o caso $f(1) = f(N - 1) = \frac{q}{2}$ e $f(0) = p$ sendo zero para os demais elementos, podemos reescrever a equação de evolução, utilizando-se da propriedade 1 descrita pela Eq. (1.42):

$$\Pi_{l+1}(n) = \sum_m f(n - m) \cdot \Pi_l(m). \quad (1.45)$$

A função característica $G_l(k)$ de distribuição $\Pi_l(n)$ é :

$$G_l(k) = \sum_n e^{ikn} \Pi_l(n). \quad (1.46)$$

Para $\Pi_0(n) = \delta_{n0}$, a função característica correspondente é $G_0(k) = 1$.

Utilizando a função característica $F(k)$ de distribuição $f(n)$:

$$F(k) = \sum_n e^{ikn} f(n), \quad (1.47)$$

podemos escrever a fórmula de recorrência para a função característica:

$$G_{l+1}(k) = F(k) \cdot G_l(k), \quad (1.48)$$

ou ainda,

$$\begin{aligned} G_1(k) &= F(k)G_0(k), \\ G_2(k) &= F(k)G_1(k) = [F(k)]^2 G_0(k), \\ &\dots \\ G_l(k) &= [F(k)]^l G_0(k), \end{aligned} \quad (1.49)$$

como $G_0(k) = 1$, segue-se que:

$$G_l(k) = [F(k)]^l. \quad (1.50)$$

A função característica $F(k)$ para o passeio aleatório pode ser encontrada levando-se em conta as equações (1.39), (1.40). Mediante as quais, as probabilidades de transição para o instante inicial podem ser escritas:

$$P(n, 0) = f(n) = \begin{cases} 0 & \text{se } n \neq 1, 0, -1 \\ \frac{q}{2} & \text{se } n = 1 \text{ ou } n = -1. \\ p & \text{se } n = 0 \end{cases} \quad (1.51)$$

Juntamente com a expressão (1.43) podemos escrever:

$$\begin{aligned} F(k) &= \frac{qe^{-ik}}{2} + p + \frac{qe^{ik}}{2}, \\ F(k) &= \frac{q(e^{-ik} + e^{ik})}{2} + p, \\ F(k) &= \frac{2q\cos(k)}{2} + p, \\ F(k) &= q\cos(k) + p. \end{aligned} \quad (1.52)$$

Substituindo a expressão para $F(k)$ em (1.49), encontramos a seguinte função característica:

$$G_l(k) = [(q\cos(k) + p)]^l. \quad (1.53)$$

Considerando as propriedades de $f(n)$:

$$\begin{aligned} \lambda_k &= \sum_n e^{ink} f(n) = F(k), \\ \lambda_k &= p + q\cos(k), \end{aligned} \quad (1.54)$$

os autovetores à esquerda são $\phi_k(n) = e^{-ikn}$.

Assim, a probabilidade $\Pi_l(n)$ pode ser dada por $\Pi_l(n) = P^l(n, 0)$. De forma equivalente, P^l pode ser escrito em termos de autovalores e autovetores [19], (teorema espectral [23]) $P^l = \sum_k \lambda_k^l \psi_k \phi_k^*$:

$$\begin{aligned} \Pi_l(n) &= \sum_k \lambda_k^l \psi_k(n) \phi_k^*(0), \\ \Pi_l(n) &= \frac{1}{N} \sum_k \lambda_k^l e^{ikn}, \\ \Pi_l(n) &= \frac{1}{N} \sum_k (p + q\cos k)^l e^{ikn}. \end{aligned} \quad (1.55)$$

A Eq. (1.55) define a caminhada aleatória em termos de um processo Markoviano, para o caso unidimensional. De maneira análoga este processo pode ser realizado para o caso bidimensional e dimensões superiores. Este tratamento é interessante devido ao fato de ser uma abordagem matricial.

CAPÍTULO 2

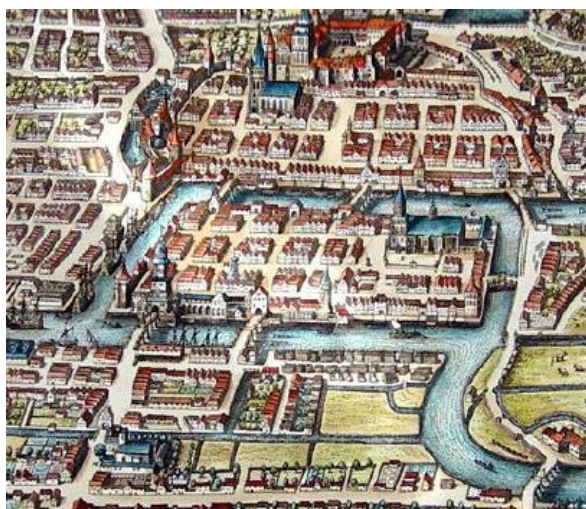
Teoria de Grafos

2.1 Origem da teoria de Grafos

A Teoria de grafos teve seu início em meados do século XVIII, quando Leonard Euler, matemático suíço, visitando a cidade de Königsberg, se deparou com um problema famoso na região. A cidade de Königsberg era uma cidade mercantil, dividida em quatro blocos de terra pelo Rio Pregel. Para facilitar o trânsito entre essas quatro regiões, seus portos e armazéns, eram unidas por sete pontes, como mostra a Fig. 2.1.

O problema que perdurava sem solução entre os habitantes dessa cidade consistia na seguinte pergunta: Seria possível atravessar cada ponte apenas uma vez e retornar ao ponto de origem? Para resolver esta questão, Leonard Euler desenvolveu um método abstrato, formulando um problema geral, que poderia ser usado para outros problemas semelhantes [24]. Este método independia da localização dos blocos de terra, da divisão realizada pelo rio e da disposição das pontes, e por meio dele se pode descobrir a possibilidade de se atravessar cada ponte exatamente uma vez.

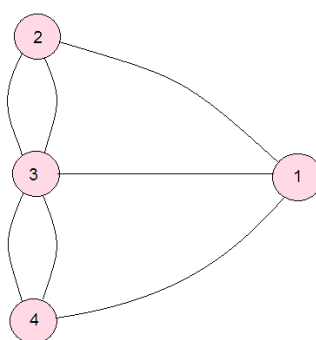
Figura 2.1 Ilustração da cidade de Königsberg, onde perdurava o problema das sete pontes.



Fonte: ARAUJO, B. Ref.[25].

Euler denotou cada região por pontos, e as pontes por linhas que uniam esses pontos. Mediante esta simplificação, ilustrada na Fig. 2.2, ele conseguiu mostrar que não era possível tal percurso. Ainda mostrou quais seriam as condições para que o percurso fosse realizado. A representação simplificada de Euler é o que chamamos de grafo. Os pontos são denominados nós ou vértices e as linhas que unem esses pontos são denominados de arestas.

Figura 2.2 Grafo representando a cidade de Königsberg, de forma de simplificada. Os nós representam os blocos de terra e as arestas representam as pontes.



Fonte: Próprio Autor.

O trabalho de Euler permitiu a formulação e demonstração do Teorema sobre as condições da existência para tal caminho, em que cada aresta é percorrida apenas uma vez, isto, para grafos não direcionados [26]. Um caminho como esse só existe se o grafo for conexo, ou seja, cada nó deve ser alcançado por todos os outros nós mediante um caminho entre eles. Na situação de Königsberg, existiam as pontes entre uma região e outra permitindo o acesso. Outra condição é que o número de arestas incidentes em cada nó deve ser par. Para o problema das sete pontes, essa condição não se cumpria.

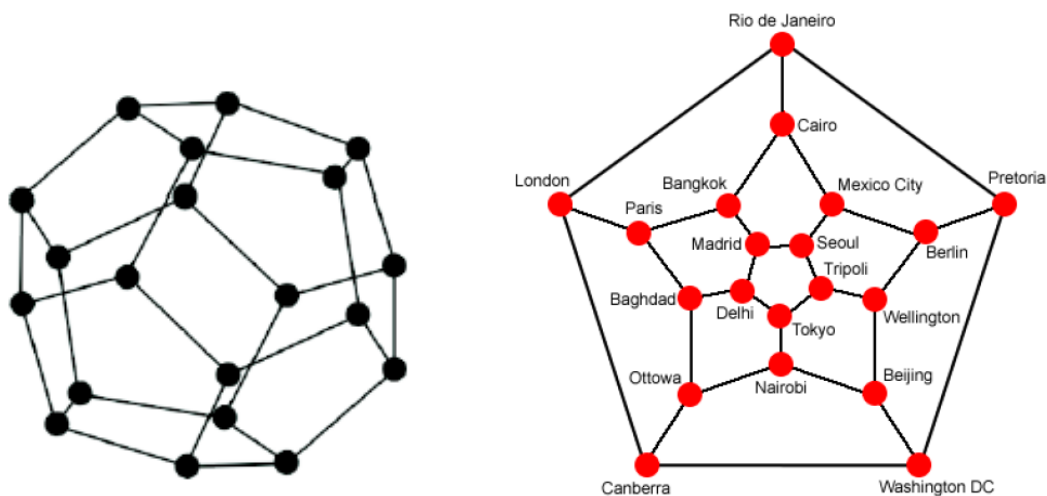
Assim, provou-se que não existia tal caminho para o problema da cidade alemã. O grafo que possui as condições expostas por Euler é conhecido como grafo Euleriano, e a solução é conhecida como circuito Euleriano [1]. O trabalho realizado por ele, para a resolução desse problema, deu início a um novo campo da matemática denominada de Teoria dos Grafos. Este trabalho foi o primeiro nesta linha, sendo publicado pela Academia de Ciências de St. Petersburg (*Academiae Scientiarum Imperialis Petropolitanae*).

Mais tarde, outros problemas semelhantes ao resolvido por Euler surgiram, permitindo explorar outros aspectos da situação, promovendo o avanço nesta área. Entre

esses problemas podemos citar o problema da Viagem à Volta do Mundo, formulado e resolvido em 1857, pelo matemático irlandês Sir William Hamilton [27, 28], o problema do Carteiro Chinês, proposto pelo matemático chinês Kwan Mei-Ko em 1962, discutido em um artigo na revista Chinese Mathematics, e também o Problema do Varredor de Rua [29].

O problema proposto por Hamilton foi elaborado com ingredientes semelhantes ao Problema das Sete Pontes de Königsberg. Este problema consistia em percorrer todos os vértices do dodecaedro, representado na Fig. 2.3 (a), passando por cada vértice uma única vez, com partida e chegada ao mesmo vértice. Este circuito foi designado por viagem à volta do mundo [27, 28]. Os vértices do dodecaedro representam 20 cidades importantes da época. Os vértices e as arestas do dodecaedro podem ser representados no plano pelo grafo da Fig. 2.3 (b).

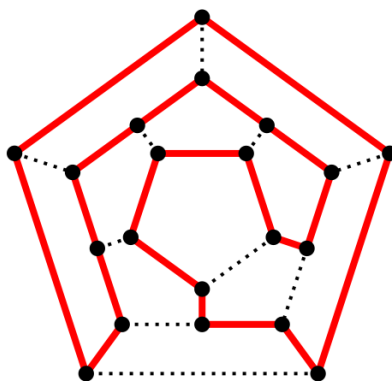
Figura 2.3 a) Dodecaedro, poliedro regular com 20 vértices de grau 3, 12 faces pentagonais, 30 arestas. b) Representação no plano do Dodecaedro, com os vértices indicando as principais cidades ao redor do Mundo.



Fonte: CARDOSO, D. M. Ref. [28]

No grafo da Fig. 2.4, as arestas em vermelho determinam um caminho fechado que, começando num vértice arbitrário, permite-nos voltar a ele depois de visitarmos cada um dos outros vértices uma única vez. Este caminho mostra como o problema proposto por Hamilton tem resposta afirmativa. Tal percurso ficou conhecido como ciclo Hamiltoniano.

Figura 2.4 Representação do problema da viagem à volta do mundo com um caminho destacado pelas arestas em vermelho, indicando a solução para o problema.



Fonte: CARDOSO, D. M. Ref. [28]

O problema formulado por Mei-Ko consistia em encontrar um caminho para o carteiro, sendo que neste caminho todas as ruas deveriam ser visitadas e o caminho deveria ser mínimo. "Dado que é impossível atravessar cada ponte exatamente uma vez e retornar para o ponto de partida, qual é o número mínimo de travessias redundantes que é preciso fazer?" Este problema possui uma forte relação com o Problema das sete pontes. Neste caso, na representação as arestas seriam as ruas e os nós seriam os cruzamentos dessas ruas, sendo que a principal diferença com o problema anterior é que o caminho admite passar por uma rua já utilizada. Assim, o carteiro deseja minimizar a sua rota passando o menor número de vezes por ruas repetidas.

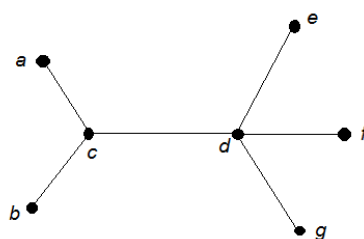
Uma variação do problema do carteiro chinês é a situação que ficou conhecida como "The New York Street Sweeper Problem" que significa O Problema dos Varredores de Rua de Nova York. É o caso em que os varredores de ruas vivenciam quando percorrem ruas de mão única. Essa situação é comum e também é vivenciada em vários outros lugares devido ao direcionamento das ruas nas cidades. Para esta situação, busca-se pelo caminho mínimo quando se tem arestas direcionadas.

2.2 Definição e Propriedades de Grafos

Os Grafos são formas abstratas utilizadas na representação de redes complexas. É uma estrutura formada por dois conjuntos. Um grafo G é definido pelo par ordenado (V, E) , em que V é um conjunto de vértices e E é um conjunto de arestas [30]. Cada aresta e pertencente ao conjunto E será denotada por $e = (v, w)$ sendo este um par de vértices.

Os vértices v e w são os extremos da aresta e são denominados vértices adjacentes ou vizinhos. A aresta e é dita incidente a ambos os vértices v e w , ou seja, a aresta parte de um e chega ao outro vértice. A Fig. 2.5 representa um grafo formado pelo conjunto de vértices $V = \{a, b, c, d, e, f, g\}$, e cada par de vértices que se conectam formam as arestas, pertencentes ao conjunto $E = \{(a, c), (b, c), (c, d), (d, e), (d, f), (d, g)\}$. O número de arestas incidentes em um vértice é chamado de grau do vértice.

Figura 2.5 Grafo cujos vértices são a, b, c, d, e, f, g , e as arestas $(a, c), (b, c), (c, d), (d, e), (d, f), (d, g)$.



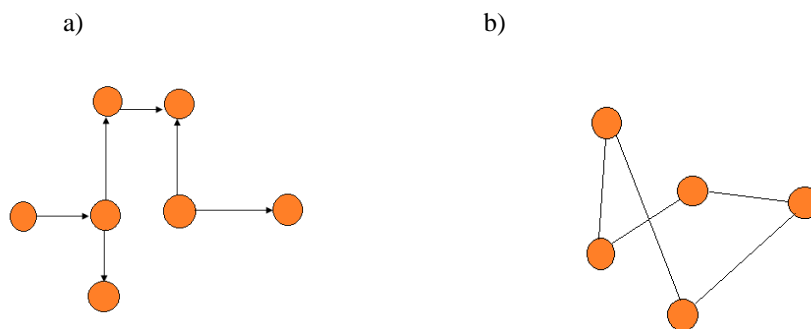
Fonte: Próprio Autor.

Um grafo pode possuir **arestas múltiplas**, o que ocorre quando um mesmo par de vértices possui mais de uma aresta [30]. Se o vértice de saída é igual ao vértice de chegada, $e = (v, v)$, a aresta é denominada de **laço**. Quando estas duas situações ocorrem em um mesmo grafo ele é denominado de **multigrafo**.

Um grafo pode ser classificado em dois tipos com relação à sua orientação, como ilustrado na Fig. 2.6. O grafo é chamado orientado quando existe na aresta uma direção indicada, identificando o vértice de origem e o vértice de chegada, denominado grafo **direcionado**, Fig. 2.6 (a). Quando não houver indicada a orientação para as arestas o grafo é dito **não direcionado** [30], Fig. 2.6 (b).

O Grafo não direcionado é aquele em que a relação (v, w) é **simétrica**. Neste caso, existe uma aresta direcionada que une v e w partindo do vértice v até alcançar o vértice w , sendo que o contrário também ocorre. Em consequência disso, podemos dizer que as arestas que ligam os vértices não possuem orientação. Se na relação definida por (v, w) existir uma aresta ligando os vértices v a w , orientada a partir de v para w e não existir uma aresta de ligação de w para v , dizemos que possui uma orientação e a relação **não é simétrica** [30].

Figura 2.6 a) Grafo direcionado, as arestas apresentam orientações. b) Grafo não direcionado.



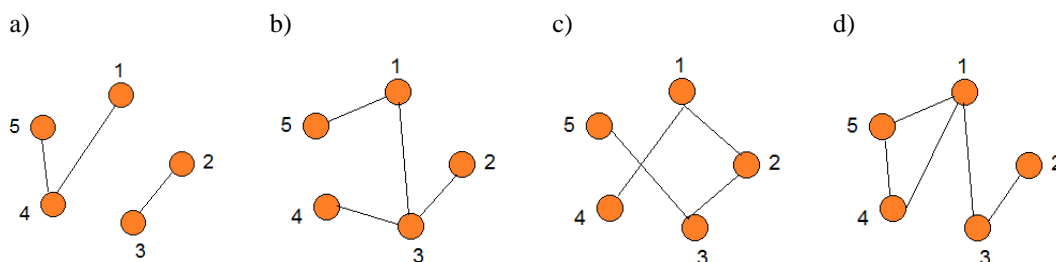
Fonte: Próprio Autor.

A sequência de arestas sucessivas de um grafo recebe o nome de **caminho** [31]. Os caminhos podem possuir características distintas, permitindo classificá-los de acordo com suas especificidades em caminhos simples, trajetos, ciclos, Hamiltonianos e Eulerianos.

O caminho é denominado **simples** (ou elementar) quando todos os vértices que formam o caminho são diferentes, não havendo repetição de vértices. Quando as arestas são todas distintas, denominamos este caminho por **trajeto**. Para o caminho em que o vértice inicial for igual o vértice final é denominado de **ciclo**. Assim, temos um ciclo quando o caminho dado pela sequência de vértices $(v_0, v_1, v_2, \dots, v_{n-1}, v_n)$ e $v_0 = v_n$, começa e termina em um mesmo nó, o ciclo deve possuir no mínimo uma aresta. Se o grafo não possuir ciclos, ele é chamado de grafo **acíclico**.

Os caminhos conhecidos como Hamiltoniano e Euleriano são dois importantes caminhos destacados ao longo da história da teoria de Grafos. Quando o caminho contém cada vértice do grafo apenas uma vez o caminho é denominado **Hamiltoniano**. Quando o caminho contém cada aresta do grafo apenas uma vez é chamado **Euleriano**. Historicamente, os caminhos Eulerianos estão associados à gênese da teoria de grafos, estando relacionada com o problema das pontes de Königsberg [31, 32].

Figura 2.7 Grafo indicando caminhos: a) simples, b) trajeto, c) Hamiltoniano e d) Euleriano.

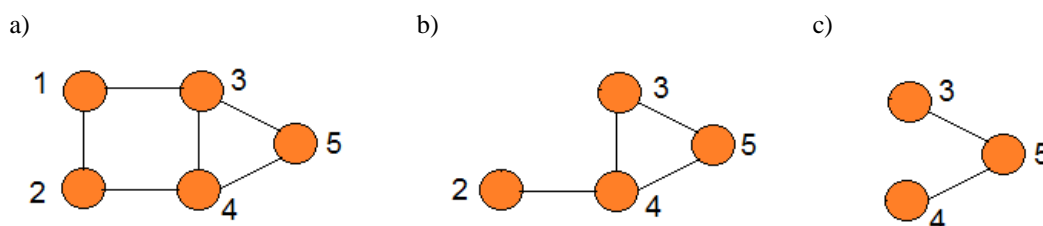


Fonte: Próprio Autor.

A Fig. 2.7 apresenta exemplos de caminhos que podemos encontrar em um grafo qualquer. No caso a) destacamos o caminho simples que começa em 1 e chega ao vértice 5, (veja que os vértices 1, 4 e 5 que compõem esse caminho aparecem apenas uma vez). No caso b) podemos identificar um caminho que possui as características de um trajeto e outros que não possuem tais características. O caminho que sai do nó 5 para o nó 2 composto pelos vértices 5, 1, 3 e 2 é um trajeto pois cada aresta desse caminho aparece uma única vez. Já o caminho que inicia pelo vértice 5 passando pelos vértices 1, 3, 4, 3 e 2, não caracteriza um trajeto, pois a aresta que une o vértice 3 e 4 é percorrida duas vezes. No caso c) é possível identificar o caminho que parte do vértice 4 e passa pelos vértices 1, 2, 3 e 5, sendo este um caminho Hamiltoniano. Já no último caso, podemos destacar um caminho Euleriano, sendo que todas as arestas do grafo devem ser percorridas, passando uma única vez por cada aresta. Assim, o caminho que inicia no vértice 2 e passa pelos vértices 3, 1, 4, 5 e 1 não repete nenhuma aresta, mesmo que o vértice 1 seja percorrido duas vezes. Neste caso, o caminho é somente Euleriano e não é Hamiltoniano.

A partir de um grafo $G(V, E)$, podemos identificar subconjuntos de vértices. Estes subconjuntos de vértices formam os chamados subgrafos. Dessa forma, um subgrafo $G_2(V_2, E_2)$ de um grafo $G_1(V_1, E_1)$ é um grafo em que o conjunto de vértice V_2 está contido em V_1 e o conjunto de arestas E_2 está contido em E_1 .

Figura 2.8 Grafo G_1 com dois de seus subgrafos. a) Grafo $G_1(V, E)$ com 5 vértices, b) O subgrafo G_2 extraído a partir de G_1 formado com 4 dos seus 5 vértices, c) O subgrafo G_3 formado com 3 do total de vértices de G_1 .



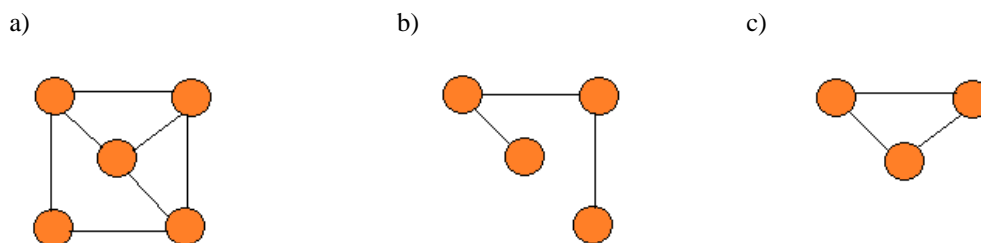
Fonte: Próprio Autor.

Quando consideramos qualquer par de vértices v, w do subgrafo G_2 de um grafo G_1 , se existe uma aresta formada por (v, w) pertence conjunto de aresta E_1 e esta aresta (v, w) pertencer também a E_2 , dizemos que o subgrafo G_2 é subgrafo induzido pelo conjunto de vértices V_2 . A Fig. 2.8 (a) traz o grafo G_1 , do qual é representado em seguida dois de seus subgrafos. O subgrafo G_2 apresentado na Fig. 2.8 (b) é um subgrafo induzido

do grafo G_1 . A Fig. 2.8 (c) apresenta o subgrafo G_3 que não é subgrafo induzido de G_1 , pois existe o par de vértices 3 e 4, que formam uma aresta em G_1 e não formam uma aresta em G_3 .

Um subgrafo denomina-se **clique** de um grafo G se ele for um subgrafo completo. Na Fig. 2.9 (a) temos a representação de um grafo qualquer G . Nas Figs. 2.9 (b) e 2.9 (c) têm-se dois de seus subgrafos, sendo que o primeiro não representa um clique, pois para ser um subgrafo completo, sempre deve existir uma aresta entre dois vértices, isso para todo par de vértices do subgrafo. Já o segundo caso representa um clique, pois, cumpre as condições de um subgrafo completo.

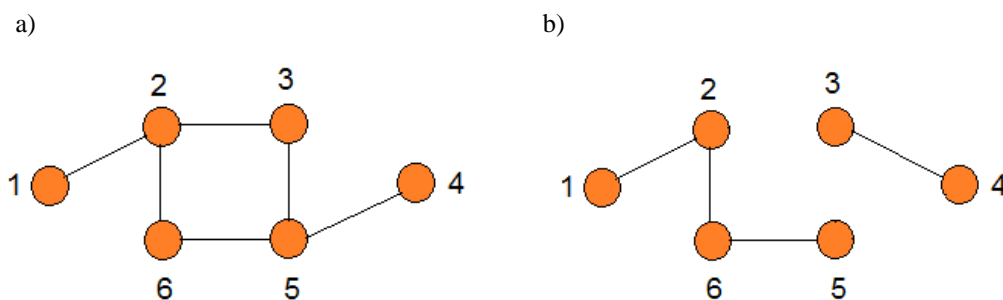
Figura 2.9 Grafo G com 5 vértices e 6 arestas em a), e dois de seus subgrafos b) e c), onde (c) representa um clique de (a).



Fonte: Próprio Autor.

Um grafo $G(V, E)$ é **conexo** quando existir um caminho entre cada par de vértices, caso contrário, é dito **desconexo**. Desse modo, se existir pelo menos um par de vértices que não esteja conectado através de um caminho o grafo é desconexo. Na Fig. 2.10 temos um grafo conexo em (a), pois sempre podemos encontrar um caminho entre quaisquer pares de vértices escolhidos. Em (b) isso não acontece, pois se escolhermos o nó 3 e o nó 1, verificamos que não existe nenhum caminho conectando esse par de nós.

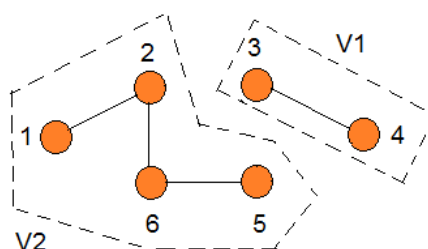
Figura 2.10 Em a) representação de um grafo conexo, em b) representa um grafo desconexo.



Fonte: Próprio Autor.

Em todo grafo desconexo é possível identificar subconjuntos de vértices independentes, os quais formam os subgrafos. Todo grafo desconexo possui pelo menos dois subgrafos conexos, também denominados de componentes conexas. Tal situação pode ser observada na Fig. 2.11, em que o grafo representado não é conexo, mas os dois conjuntos de vértices V_1 e V_2 formam subgrafos conexos. Cada subgrafo conexo corresponde a uma componente conexa do grafo em questão.

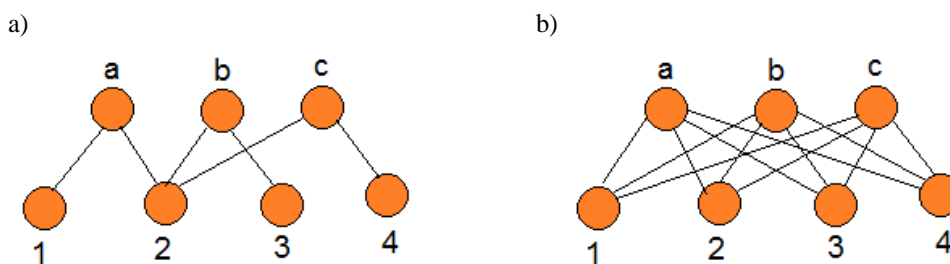
Figura 2.11 Grafo desconexo, com dois subgrafos conexos identificados.



Fonte: Próprio Autor.

Quando em um grafo $G(V, E)$ for possível distinguir partições do conjunto de vértices V em dois subconjuntos disjuntos, estas partições podem ser denotadas como V_1 e V_2 . Se estes conjuntos estiverem conectados, de tal forma que toda ligação em G conecta um vértice de V_1 a outro vértice de V_2 , sem haver arestas entre vértices de uma mesma partição [29] o grafo recebe o nome de **Bipartido**. Caso todos os vértices de uma partição estiverem conectados a todos os vértices da outra ele é denominado **Bipartido Completo**. Na Fig. 2.12 temos dois grafos com os seguintes subconjuntos de vértices $V_1 = \{a, b, c\}$ e $V_2 = \{1, 2, 3, 4\}$. Em (a) o grafo é bipartido, pois as conexões acontecem entre vértices de conjuntos distintos. Em (b) os vértices do primeiro conjunto estão ligados a todos os outros vértices do segundo conjunto, configurando um grafo bipartido completo.

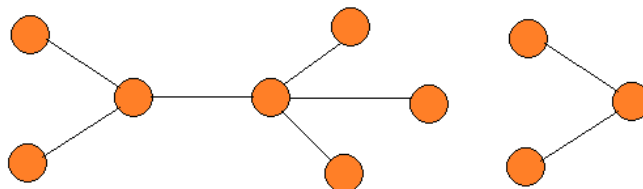
Figura 2.12 Grafos com 7 vértices particionados em dois subconjuntos de vértices, cada partição possui 3 e 4 vértices respectivamente. Em a) Grafo Bipartido, em b) Grafo Bipartido Completo.



Fonte: Próprio Autor.

O grafo denominado **floresta** se for acíclico, com $n \geq 2$ [32]. Os grafos que representam florestas podem ser constituídos de várias **árvores**, em que árvores são grafos acíclicos e conexos. Assim, as florestas não são necessariamente conexas, mas possuem componentes conexas denominadas árvores, conforme a Fig. 2.13. Uma árvore possui dois tipos de vértice. O primeiro é quando vértice possui grau igual a um, neste caso, ele é denominado de **folha**. O segundo caso é quando o grau do vértice é maior que um, sendo denominado de **vértice interior**.

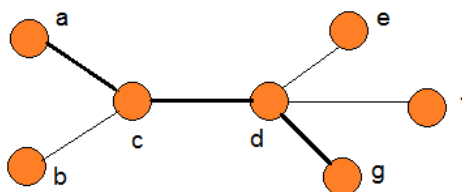
Figura 2.13 Floresta constituída por duas árvores, em que cada árvore apresenta vértices na forma de folhas e de vértices interiores.



Fonte: Próprio Autor.

Toda árvore é conexa e sem ciclos, o que significa que para todo par de vértices existe uma única sequência de ligações conectando-os. Isto pode ser observado na Fig. 2.14. Entre os vértices a e g existe apenas um caminho simples constituído pela seguinte sequência (a, c, d, g) . Isto pode ser verificado de igual forma, para todo par de vértices.

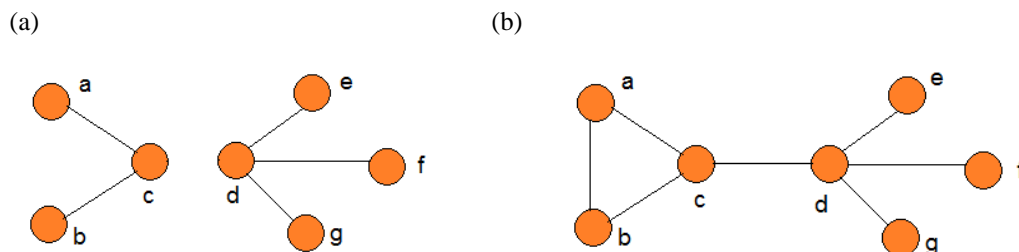
Figura 2.14 Caminho simples identificado entre dois vértices de uma árvore.



Fonte: Próprio Autor.

Outras características observadas para árvores são a supressão ou adição de uma aresta, conforme a Fig. 2.15 (a). Caso apenas uma aresta seja suprimida de uma árvore ele deixa de ser conexo. Da mesma forma, se houver uma inclusão de uma aresta, ocorre a criação de um ciclo. Isso faz com que o grafo deixe de ser acíclico, veja Fig. 2.15 (b).

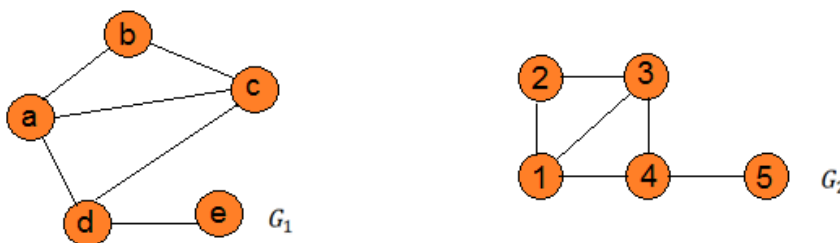
Figura 2.15 a) Supressão da aresta (c, d) , a árvore deixa de ser conexa, sendo dividida em duas árvores menores. b) Adição da aresta (a, b) , o grafo deixa de ser uma árvore, devido ao ciclo, constituído pelo caminho (a, b, c, a) .



Fonte: Próprio Autor.

Dois grafos $G_1 (V_1, E_1)$ e $G_2 (V_2, E_2)$ são iguais quando $V_1 = V_2$ e $E_1 = E_2$ e são isomorfos quando existe uma função bijetora $f: V_1 \rightarrow V_2$, tal que (v, w) pertence a E_1 se, e somente se $(f(v), f(w))$ pertence a E_2 , para todo v, w pertencente a V_1 [34]. Para representar tal situação a Fig. 2.16, exhibe dois grafos isomorfos G_1 e G_2 , cujos respectivos conjuntos de vértices $V_1 = \{a, b, c, d, e\}$ e $V_2 = \{1, 2, 3, 4, 5\}$ se relacionam através da função $\{(a, 1), (b, 2), (c, 3), (d, 4), (e, 5)\}$ satisfazendo a condição de isomorfismo.

Figura 2.16 Os grafos G_1 e G_2 representam grafos isomorfos. Pois há a função $\{(a, 1), (b, 2), (c, 3), (d, 4), (e, 5)\}$ que satisfaz a condição descrita acima.



Fonte: Próprio Autor.

Existe uma íntima relação entre a teoria de grafos e a teoria de matrizes. Há diversas representações matriciais oriundas de estruturas de grafos e o seu uso está, habitualmente, associado à necessidade da realização de cálculo envolvendo dados estruturais [32]. A matriz mais utilizada é a matriz de adjacência. Esta matriz é uma matriz quadrada, cuja ordem n corresponde ao número de vértices que um grafo possui; cada linha e cada coluna desta matriz correspondem a um vértice do grafo. Os elementos dessa matriz estão associados aos dados estruturais do grafo $G(V, E)$. Estes elementos possuem

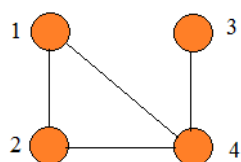
valores nulos quando não existir conexões e valores não nulos quando houver conexões entre os vértices correspondentes aos elementos. Quando se tratam de grafos valorados, nos quais as arestas possuem pesos, os valores dos elementos não nulos da matriz de adjacência recebe o peso da aresta. Para grafos não valorados, os elementos não nulos recebem o valor 1.

$$A = [a_{ij}]$$

$$a_{ij} = \begin{cases} 1, & \text{se } \exists (i,j) \in E \\ 0, & \text{se } \nexists (i,j) \in E \end{cases} \quad (2.1)$$

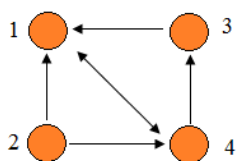
Figura 2.17 Matrizes de adjacências de a) Grafo G_1 não direcionado e de b) Grafo G_2 direcionado.

a)



$$A_{G_1} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

b)



$$A_{G_2} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Fonte: Próprio Autor.

A matriz de adjacência de um grafo não direcionado é uma matriz simétrica Fig. 2.17 (a). Para grafos direcionados, Fig. 2.17 (b), a matriz de adjacência não é simétrica, pois, quando a aresta é direcionada pode-se sair de um vértice a outro por meio dessa ligação, porém o caminho de volta pode não ser possível. Esta situação pode ser observada na relação entre o vértice 3 e 4 da Fig. 2.17 (b), em que pode-se partir do vértice 4 para o vértice 3, neste caso o elemento da matriz de adjacência, correspondente a essa posição receberá o valor 1. Para o movimento contrário, não é possível de realizá-lo saindo de 3 e para 4, pela a aresta de ligação entre elas, isso devido ao seu direcionamento, assim o elemento a_{34} receberá o valor zero.

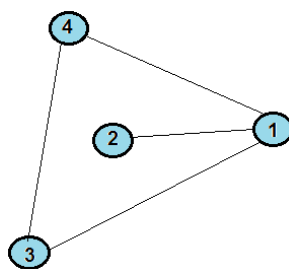
2.3) Caminhada Aleatória em Redes

As redes complexas são estruturas que podem ser observadas no mundo real de diversas formas: conjunto de pessoas que estabelecem algum tipo de relacionamento, conjunto de animais que se relacionam por meio da cadeia alimentar, conjunto de computadores que se relacionam mediante roteadores, entre outros. Enfim, existem muitas situações reais que podem ser caracterizadas por meio de redes complexas e estas redes podem ser representadas mediante a Teoria de Grafos [22,24].

Uma das aplicações da caminhada aleatória é o estudo de seu comportamento em redes complexas. Entender a caminhada aleatória nestas redes se torna importante em diversos aspectos. Por exemplo, estas redes ajudam a compreender como uma informação é disseminada em de uma rede de pessoas, ou como uma doença é transmitida em uma rede social. Estas e outras situações são estudadas por meio das redes complexas.

A rede pode ser representada de forma gráfica por meio de vértices e as conexões entre eles por arestas, ou seja, as redes podem ser representadas por meio de grafos, conforme a Fig. 2.18, por exemplo.

Figura 2.18 Rede com número de nós $n = 4$ e as conexões entre os vértices.



Fonte: Próprio Autor.

De igual forma aos grafos a Matriz de Adjacência da rede indica as conexões entre os vértices. Para o caso ilustrado na Fig. 2.18 a matriz adjacência é dada por:

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad (2.2)$$

Quando é realizada a adição dos elementos de cada linha da matriz, temos como resultado o número de arestas incidentes aos vértices da rede. Uma vez que as conexões dos vértices são representadas por meio das linhas da matriz de adjacência. Esta quantidade é denominada grau do vértice e é denotado como:

$$k_i = \sum_j A_{ij}. \quad (2.3)$$

Para o caso da Fig. 2.18 o valor do grau de cada nó é dado por:

$$\begin{aligned} k_1 &= 0 + 1 + 1 + 1 = 3 \\ k_2 &= 1 + 0 + 0 + 0 = 1 \\ k_3 &= 1 + 0 + 0 + 1 = 2 \\ k_4 &= 1 + 0 + 1 + 0 = 2 \end{aligned} \quad (2.4)$$

Supondo que um caminhante percorre um caminho aleatório, e em um determinado instante ele se encontra em vértice i , a probabilidade de que no próximo passo esse caminhante se encontre em um vértice j é dado pela probabilidade de passo 1. Essa probabilidade é denotada como o elemento de matriz definida pela razão da conexão entre i e j , que pode ser 0 e 1, pelo grau do vértice i , que é o número de saídas possíveis para diferentes vértices [14], ou seja:

$$P_{ij}(1) = \frac{A_{ij}}{k_i}. \quad (2.5)$$

Da mesma forma, é possível pensar nessa situação para dois passos, ou seja, a probabilidade de se encontrar o caminhante no vértice j , partindo do vértice i , após dois intervalos de tempo, isto é:

$$\begin{aligned} P_{ij}(2) &= \sum_{j_1 \neq j} \frac{A_{ij_1}}{k_i} \cdot \frac{A_{j_1j}}{k_{j_1}}, \\ P_{ij}(2) &= \sum_{j_1 \neq j} \frac{A_{j_1j}}{k_{j_1}} \cdot P_{ij_1}(1), \\ P(2) &= P(1)P(1) = [P(1)]^2. \end{aligned} \quad (2.6)$$

A probabilidade para o caso em que o caminhante parta de i e chegue a j após três instantes de tempo é:

$$\begin{aligned}
 P_{ij}(3) &= \sum_{j_1 \neq j; j_2 \neq j} \frac{A_{ij_1}}{k_i} \cdot \frac{A_{j_1 j_2}}{k_{j_1}} \cdot \frac{A_{j_2 j}}{k_{j_2}}, \\
 P_{ij}(3) &= \sum_{j_1 \neq j; j_2 \neq j} \frac{A_{j_2 j}}{k_{j_2}} P_{ij_2}(2), \\
 P(3) &= P(1)P(1)P(1) = [P(1)]^3.
 \end{aligned} \tag{2.7}$$

Procedendo de igual forma para os demais casos, é possível chegar à forma explícita para a probabilidade de transição P_{ij} , para ir de i para j em uma quantidade qualquer t de passos:

$$\begin{aligned}
 P_{ij}(t) &= \sum_{j_1, j_2, j_3, \dots, j_{t-1}} \frac{A_{ij_1}}{k_i} \cdot \frac{A_{j_1 j_2}}{k_{j_1}} \cdot \frac{A_{j_2 j_3}}{k_{j_2}} \dots \frac{A_{j_{t-1} j}}{k_{j_{t-1}}} = \sum_{j_1, j_2, j_3, \dots, j_{t-1}} P_{ij_1}^{(1)} P_{j_1 j_2}^{(1)} \dots P_{j_{t-1} j}^{(1)}, \\
 P_{ij}(t) &= \sum_{j_1, j_2, j_3, \dots, j_{t-1}} \frac{A_{j_{t-1} j}}{k_{j_{t-1}}} P_{ij_{t-1}}(t-1),
 \end{aligned} \tag{2.8}$$

$$\begin{aligned}
 P(t) &= \underbrace{P(1)P(1) \dots P(1)}_{t \text{ vezes}}, \\
 P(t) &= [P(1)]^t.
 \end{aligned} \tag{2.9}$$

A equação (2.8) denominada de equação mestra para a probabilidade, e através dela é possível encontrar a probabilidade de transição de i para j em qualquer tempo t . A probabilidade no tempo $t = 0$ é dada pela matriz identidade [24,33]; a probabilidade no primeiro instante de tempo é dada pela Eq. (2.5). Por meio da equação mestra, conhecendo-se o valor da probabilidade no primeiro instante de tempo é possível encontrar a probabilidade no instante $t = 2$. Uma vez sabendo o valor da Probabilidade de transição no segundo instante de tempo, é possível encontrar o valor para $t = 3$ e assim sucessivamente. De forma equivalente é possível calcular essa probabilidade de transição a partir da Eq. (2.9), mediante da multiplicação sucessiva da matriz $P(1)$. Assim, tomando a t -ésima potência da matriz de probabilidades de transição de passo 1 é possível determinar a matriz de probabilidades de transição para t passos.

Quando o tempo tende para o infinito a matriz de distribuição tende a se estabilizar, atingindo assim a distribuição estacionária. Para o exemplo ilustrado na Fig. 2.18, teremos

as matrizes de probabilidade de transição para os respectivos tempos $t = 0$, $t = 1$, $t = 2$, $t = 3$ e $t = 20$. Para $t = 0$, a probabilidade é definida como a matriz identidade, $P(0) = I$ e para a última matriz calculada, já é possível perceber a estabilização.

$$\text{Para } t = 1: \quad P(1) = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

$$\text{Para } t = 2: \quad P(2) = \begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} & \frac{5}{12} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{6} & \frac{1}{6} & \frac{5}{12} \end{bmatrix}.$$

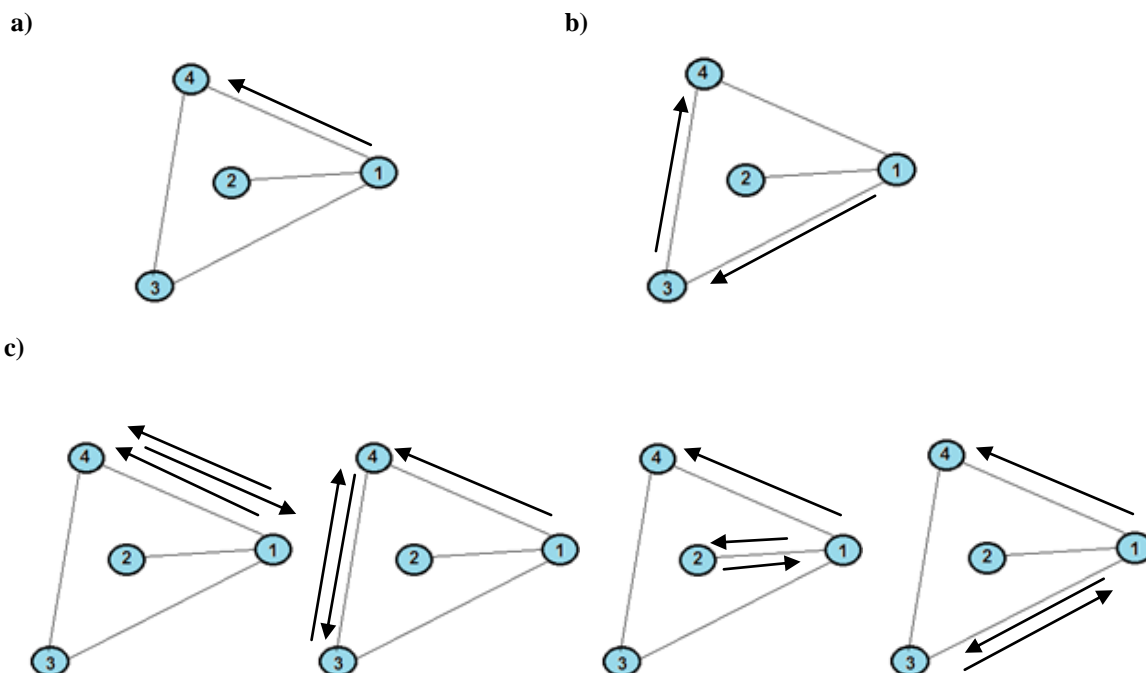
$$\text{Para } t = 3: \quad P(3) = \begin{bmatrix} \frac{1}{6} & \frac{2}{9} & \frac{11}{36} & \frac{11}{36} \\ \frac{2}{3} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{3}{11} & \frac{1}{12} & \frac{1}{6} & \frac{7}{24} \\ \frac{11}{24} & \frac{1}{12} & \frac{1}{6} & \frac{7}{24} \\ \frac{11}{24} & \frac{1}{12} & \frac{7}{24} & \frac{1}{6} \end{bmatrix}.$$

$$\text{Para } t = 20: \quad P(20) \approx \begin{bmatrix} \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Para melhor verificar como se comportam essas probabilidades, tomaremos apenas um elemento específico das matrizes de probabilidade de transição. Por exemplo, considerando apenas a probabilidade de transição do caminhante situado no vértice 1 para o vértice 4, como ilustrado na Fig. 2.19 (a). Para o tempo $t = 1$, vemos que a probabilidade de partir de 1 e chegar em 4 em um único passo é equivalente ao elemento $P_{14}(1)$, que é a transição para o primeiro passo, ou transição de passo 1.

A Fig. 2.19 (b) mostra a probabilidade de transição em dois passos. Se acaso o caminhante vai para 4 a partir de 1 no primeiro passo, para o segundo passo obrigatoriamente ele deve sair do vértice 4. Ainda, se no primeiro passo ele vai para o vértice 2, como 2 não está conectado diretamente com 4 ele sairá de 2 mas não chegará a 4 no segundo passo. Veja que se o caminhante parte de 1 e vai para 4 em dois passos, o único elemento intermediário possível é o vértice 3. Assim, para o tempo $t = 2$, a probabilidade será o produto das possibilidades de saída que são respectivamente $\frac{1}{3}$ para as probabilidades de saída do vértice 1 e $\frac{1}{2}$ para a probabilidade de saída do vértice 3. Dessa forma a probabilidade de que, a partir de 1, chegue-se a 4 passando por 3 em dois passos é $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$ que é igual ao elemento $P_{14}(2)$.

Figura 2.19 Rede com $n = 4$, indicando a probabilidade de transição de 1 para 4 em diferentes números de passos. a) Probabilidade de transição de 1 para 4 em um passo. b) Probabilidade de transição de 1 para 4 em dois passos. c) Probabilidade de transição de 1 para 4 em três passos.



Fonte: Próprio Autor.

Para o tempo $t = 3$, há várias possibilidades, conforme indicado na Fig. 2.19 (c). Como existem diversas possibilidades de percursos, a probabilidade de ir de 1 para 4 em

três passos será a soma dos diferentes percursos tomados, sendo que cada caminho pode ser encontrado mediante a multiplicação das probabilidades de saída:

$$P_{14}(3) = P_{14}P_{41}P_{14} + P_{14}P_{43}P_{34} + P_{12}P_{21}P_{14} + P_{13}P_{31}P_{14}$$

$$P_{14}(3) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{11}{36}$$

Assim, a probabilidade de que a partir de 1 chegue-se a 4 em três passos é $\frac{11}{36}$, que é igual ao valor de $P_{14}(3)$ da matriz de probabilidade de transição em 3 passos. E dessa forma se podem suceder inúmeras vezes, por meio da equação (2.8) tornando possível encontrar a probabilidade de transição de i para j para qualquer tempo t .

CAPÍTULO 3

Redes Complexas

3.1 Histórico, Desenvolvimento e Aplicações das Redes Complexas

O estudo de redes complexas é abrangente, envolvendo diversas áreas de conhecimento. O termo redes faz referência às relações estabelecidas entre dois conjuntos. O primeiro deles é o conjunto de vértices que são interligados por meio de um segundo conjunto, as arestas. Suas representações podem ser feitas de forma algébrica ou a partir de diagramas, sendo, para isso, utilizada a teoria de Grafos [5].

Pode-se dizer que as redes complexas são estruturas que não seguem um padrão regular e apresentam propriedades topológicas que não são encontradas em grafos mais simples. Na literatura não há um consenso sobre as definições de redes complexas. Sabe-se que estas apresentam características próprias, que revelam informações sobre o modo de formação, crescimento e outras informações estruturais que podem ser levadas em consideração na análise de problemas específicos.

Os primeiros estudos envolvendo redes complexas foram realizados por volta de 1930. Neste período, sociólogos utilizavam tais redes com a finalidade de estudar o comportamento da sociedade e a relação entre os indivíduos [34]. No princípio os estudos eram feitos de forma bastante restrita, devido à dificuldade de se representar e analisar sistemas que possuíam um grande número de nós. Ao longo dos anos, com o desenvolvimento da tecnologia computacional, grandes avanços se deram nesta área de estudo, permitindo análises de redes, que antes não passavam de algumas dezenas de nós, para redes com milhares de nós. Devido a esses avanços, hoje diversos aspectos do mundo real podem ser representados por meio de redes complexas.

3.2 Aplicações de Redes Complexas

O mundo real é repleto de diferentes tipos de redes. Inúmeras situações reais podem ser modeladas pelas redes complexas, desde os relacionamentos existentes na natureza, na sociedade, até relações envolvendo conjuntos de objetos. Entre os exemplos mais

conhecidos estão a *World Wide Web* (WWW), Internet, redes sociais, redes tecnológicas, redes biológicas entre outras.

Redes Tecnológicas

As redes tecnológicas têm como objeto de estudo estruturas desenvolvidas pelo homem. As redes tecnológicas incluem a rede de energia elétrica, a rede de computadores, a rede de telefonia fixa e móvel as redes de transportes aéreos e rodoviários [35] entre outras. A internet e as variadas forma de seu estudo, também se caracterizam como sendo redes tecnológicas.

Redes de computadores e a Internet

As redes de computadores são redes em que os nós são os computadores e as ligações são físicas, existindo uma ligação entre dois nós na rede se estes computadores estão fisicamente conectados por meio de um cabo. Este tipo de rede pode se apresentar em vários tamanhos, que são classificados como *local area network* (LAN) caso o número de computadores conectados entre si seja pequeno. Para o caso em que a rede possua um número grande de computadores, podendo consistir de milhares de computadores, ela é chamada de *Wide area network* (WAN) [35].

A internet também é uma rede, composta de sistemas autônomos (SA) que mantém o intercâmbio de dados entre si. Estes pacotes de dados são transferidos entre computadores designados para tal, chamados de "roteadores". A Internet, como uma rede complexa, é normalmente estudada por meio de duas formas distintas. A primeira por meio de roteadores, em que cada roteador é um nó, e o segundo por meio de SA, em que cada SA é tratado como um nó.

Redes Virtuais

Redes virtuais são redes cujas ligações, por vezes os nós, não são entes físicos. Estas redes podem ser compostas de computadores e outras tecnologias, sendo vistas como um reflexo da rede social real. As conexões existentes entre os vértices da rede são chamadas de ligações lógicas. A maior rede virtual e mais conhecida é a WWW, na qual

os vértices são as páginas HTML. As ligações entre as páginas, são indicados por meio de *hiperlink*. Se existir um *hiperlink* de uma página para outra, significa que há uma aresta dirigida entre tais páginas. Em 2009, estimou-se que o número de páginas era de centenas de bilhões [35]. Outro exemplo de redes virtuais são as redes de *e-mail*, em que cada indivíduo é um nó que está ligado a todos os indivíduos em sua lista de endereços eletrônicos.

Redes Sociais

Uma classe importante de redes complexas é a de interações sociais entre atores; os quais podem assumir o papel de indivíduos, grupos, organizações, comunidades ou sociedades, os quais estão ligados uns aos outros através de alguma relação social. Estas relações podem ser vínculos de amizade e relacionamento [36], relações de trabalho, relações políticas, relações acadêmicas, relações profissionais, relação de colaboração em filmes [37], entre outros. Pode-se destacar como um tipo de rede que se comporta de forma semelhante as redes sociais, as redes econômicas, cujos nós representam não apenas indivíduos isolados, mas sim cidades, países, empresas, indústrias. Neste caso, as arestas representam as relações comerciais existentes entre tais entes econômicos [3].

O estudo destas redes é importante por diversos aspectos, sendo fundamental para o entendimento do comportamento social e para estudos epidemiológicos por exemplo. Um estudo de suma importância nesta área foi o de Milgram [38] que levou ao conceito de pequeno mundo.

Redes Biológicas

As redes biológicas são uma das classificações mais importantes e estudadas na atualidade [39, 40], devido à grande variedade de sistemas biológicos que podem ser modelados por meio destas. Redes biológicas [35] podem ser estruturas com ligações lógicas que representam ou não interações físicas, como as interações entre proteínas e as redes ecológicas que modelam a relação entre presa e predador. Há também as redes biológicas com ligações físicas, tais como as redes de neurais, redes de proteínas e as redes de vasos sanguíneos em um organismo.

3.3 Propriedades das Redes Complexas

As redes complexas podem ser analisadas por diversos pontos de vista e com as mais variadas finalidades. Devido a esse fato, a apresentação de algumas propriedades das redes complexas é útil para esta análise.

Coefficiente de Aglomeração

O coeficiente de aglomeração refere-se ao agrupamento intrínseco dos vértices de uma rede, que também é conhecido como transitividade. Esta característica está relacionada com o número de triângulos formados pelos vértices [9]. A aglomeração é alta quando dois vértices que compartilham um nó vizinho possuem uma probabilidade alta de estarem ligados uns aos outros.

Há duas definições comuns para o coeficiente de aglomeração. A primeira é a aglomeração global, descrita pela Eq. (3.1), que está relacionada com o número de triângulos total existente na rede e o número de triplas de vértices conectados:

$$C = \frac{3 \times \text{n}^\circ \text{ de triângulos na rede}}{\text{n}^\circ \text{ de triplas de vértices conectados da rede}} \quad (3.1)$$

O fator 3 no numerador da Eq. (3.1) provém do fato de que cada triângulo, contém três triplas de vértices. A segunda forma de coeficiente de aglomeração é dada por meio da média dos agrupamentos de nós individuais. O coeficiente de aglomeração para apenas um vértice é dado pela razão abaixo:

$$C_i = \frac{\text{n}^\circ \text{ de triângulos na rede ligados ao vértice } i}{\text{n}^\circ \text{ de triplas de vértices conectados da rede centradas em } i} \quad (3.2)$$

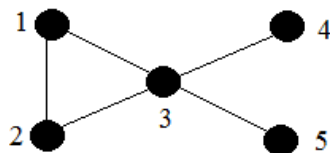
Para os vértices que não formam triplas ou triângulos, o coeficiente de aglomeração é definido como zero. O coeficiente de aglomeração para toda a rede é dado por:

$$C = \frac{1}{n} \sum_i C_i, \quad (3.3)$$

em que n representa o número de nós na rede.

Figura 3.1 Redes com $n = 5$, o coeficiente de agrupamento Global C definido pela Eq. (3.1) é de $3/8$.

Os vértices individuais têm coeficientes de agrupamento local de $C_1 = 1$, $C_2 = 1$, $C_3 = \frac{1}{6}$, $C_4 = 0$ e $C_5 = 0$ conforme a Eq.(3.2) [34].



Fonte: Próprio Autor.

A rede ilustrada na Fig. 3.1 apresenta um único triângulo formado pelos vértices 1, 2 e 3. Assim, o valor para as duas formas apresentadas de Coeficiente de agrupamento C , pode ser calculado mediante as Eqs. (3.1) e (3.2). Esta rede tem um triângulo e oito triplas de vértices conectados entre si, sendo elas $(2,1,3)$; $(1,2,3)$; $(1,3,2)$; $(1,3,4)$; $(1,3,5)$; $(2,3,4)$; $(2,3,5)$ e $(4,3,5)$; portanto, tem um coeficiente de agrupamento global de $3/8$. Os vértices individuais têm coeficientes de agrupamento local de $C_1 = 1$, $C_2 = 1$, isso porque ambos os vértices fazem parte de um triângulo e existe apenas uma tripla centrada em 1 e uma tripla centrada em 2. Para o vértice 3, temos $C_3 = \frac{1}{6}$, pois este faz parte de um triângulo e existem seis triplas centradas em 3. Como os vértices 4 e 5 não formam triângulos, os coeficientes de agrupamento local são $C_4 = 0$ e $C_5 = 0$; o valor médio da aglomeração é $C = \frac{13}{30}$ [39].

Comprimento de Caminho Médio

O comprimento de Caminho Médio está relacionado com os menores caminhos que ligam dois nós da rede. O menor caminho que conecta dois vértices é determinado pelo menor número de arestas a ser percorrida em suas conexões. Assim, o comprimento de caminho médio nada mais é do que a média dos menores caminhos para sair de um vértice e chegar a outro na rede, podendo ser calculado pela definição abaixo:

$$L = \frac{1}{N(N-1)} \sum_{i,j} d(v_i, v_j), \quad (3.4)$$

Em que $d(v_i, v_j)$ denota a menor distância percorrida entre os vértices v_i e v_j e N é a quantidade de vértices na rede.

Distribuição de Graus

A distribuição de grau p_k é a fração de vértices com grau k , o que equivale dizer que p_k é a probabilidade de que um vértice escolhido ao acaso tenha grau k , informando o modo como se distribuem as ligações aos vértices. Deste modo, a distribuição de graus é uma função de distribuição probabilística que indica a probabilidade de um determinado vértice ter k ligações. Uma maneira de quantificar essa distribuição é por meio de uma função de distribuição cumulativa [9], em que P_k é a função cumulativa de distribuição de probabilidades:

$$P_k = \sum_{k'=k}^{\infty} p_{k'} \quad (3.5)$$

O gráfico da distribuição de grau de qualquer rede pode ser feito a partir de um histograma dos graus dos vértices. Para grafos aleatórios como os estudados por Erdős e Rényi [2] a distribuição de grau é binomial, visto que cada aresta pode estar presente ou ausente na rede de maneira equiprovável. Já as redes reais possuem distribuição de grau bastante diferente das redes aleatórias. Os graus dos vértices na maioria das redes descrevem uma distribuição em lei de potência, $p_k \sim k^{-\alpha}$, em que o expoente varia entre $2 < \alpha < 3$.

Resistência

A resistência da rede está diretamente relacionada com a distribuição de grau. Esta propriedade indica a capacidade de resistência da rede quanto à remoção de alguns de seus vértices, sem que haja perda de sua funcionalidade [9]. Assim, se houver uma supressão de alguns vértices essa conectividade pode ser afetada, aumentando o comprimento desses caminhos entre os nós, ou até mesmo desligando os pares de vértices impossibilitando, a comunicação entre estes.

A resistência da rede varia com a remoção específica de determinados vértices e também com a maneira com que essa remoção é feita. Para exemplificar, a remoção pode

ser feita de maneira aleatória, ou pode focar em algum tipo de vértice específico, como os que possuem alto grau.

3.4 Modelos de Redes Complexas

Para modelar as diferentes situações de redes reais, pesquisadores buscaram propor diferentes modelos de redes que pudessem atender as principais características de cada situação. Entre os principais modelos de redes destacam-se as redes Aleatórias, as redes de Pequeno Mundo, as redes Sem Escala e as redes Hierárquicas. Estas redes foram geradas conforme o Anexo - A, e cada uma delas será apresentada a seguir:

Rede Aleatória

As redes aleatórias constituem o modelo mais simples de rede complexa. Essas redes foram estudadas em 1959, pelos matemáticos Paul Erdős e Alfréd Rényi, os quais propuseram um modelo para a obtenção dessa rede, que ficou conhecido como modelo de Erdős-Rényi [2]. Neste modelo m arestas são adicionadas aleatoriamente entre N vértices; cada inserção de uma aresta está associada a uma probabilidade independente p .

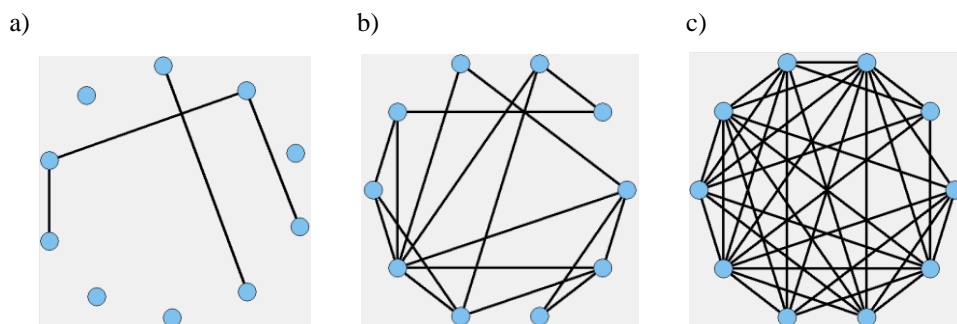
A criação de uma rede aleatória com N vértices e m arestas inicia-se a partir de um conjunto de N vértices desconectados. As possíveis combinações desses vértices são $\frac{N(N-1)}{2}$, considerando que não hajam arestas múltiplas e nem laços. Este número representa o número máximo possível de arestas na rede. Assim, as m arestas são distribuídas entre esses nós inicialmente desconectados; para cada par de vértices está associado à probabilidade p de haver uma conexão mediante a inserção de uma aresta.

Podemos assim perceber que quando $p = 0$ a rede continuará desconectada, não havendo nenhuma ligação entre os vértices e o grau médio será $\langle k \rangle = 0$. Quando $p = 1$ a rede se torna um grafo regular totalmente conectado, pois todas as ligações possíveis são estabelecidas, obtendo o grau médio em $\langle k \rangle = N - 1$. O grau médio para o caso em que é associada uma probabilidade p qualquer é $\langle k \rangle = p(N - 1)$.

Para os casos em que a probabilidade assume valores intermediários entre 0 e 1, como apresentado na Fig. 3.2, pode-se perceber que quanto menor for a probabilidade menor é o número de conexões estabelecidas. À medida que a probabilidade aumenta o

número de conexões também cresce e a rede torna-se cada vez mais próxima de uma rede regular quanto mais próximo de 1 for esta probabilidade.

Figura 3.2 Representação de redes aleatórias com $N = 10$ e (a) $p = 0,1$; (b) $p = 0,5$ e em (c) $p = 0,8$.



Fonte: Próprio Autor.

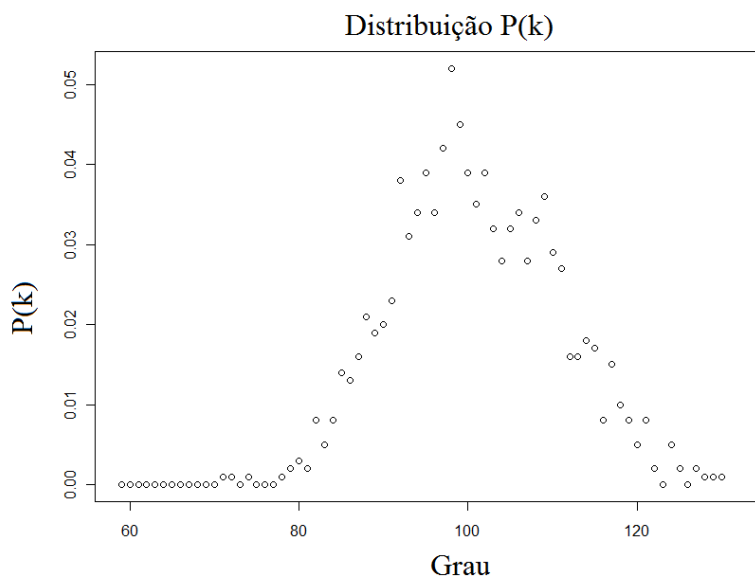
A distribuição de grau para as redes aleatórias será então uma distribuição binomial, levando-se em conta os eventos de inserção ou não de arestas, e é dado por:

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad (3.6)$$

No limite, para $N \rightarrow \infty$, as redes aleatórias obedecem à distribuição de Poisson, dada por:

$$P(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}. \quad (3.7)$$

Figura 3.3 Distribuição de grau de uma rede aleatória com $N = 1000$ e $p = 0,1$.



Fonte: Próprio Autor.

A Fig. 3.3 apresenta a distribuição de grau de uma rede aleatória formada por 1000 vértices com probabilidade de ligação $p = 0,1$. A imagem mostra que a distribuição se caracteriza por ser uma distribuição Binomial, cujo pico se localiza em torno do grau médio da rede, no presente caso $\langle k \rangle = N \cdot p = 1000 \cdot (0,1) = 100$.

Rede de Pequeno Mundo

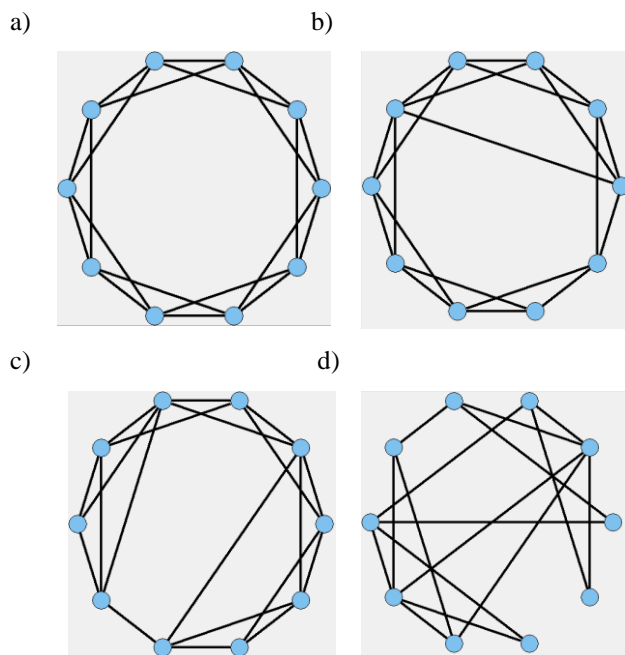
O fenômeno de Pequeno Mundo é observado em muitas redes reais. O primeiro estudo referente a esse assunto foi proposto pelo psicólogo social Milgran em 1967 [4]. Seu experimento consistia na distribuição de cartas para pessoas escolhidas de maneira aleatória da cidade de Omaha, as quais deveriam entregá-las para outras pessoas até chegar ao seu destino final. Os moradores de Omaha não conheciam os destinatários, os quais moravam em Boston. Como informações, receberam apenas o nome, endereço e ocupação. Assim, mesmo sem conhecer o destinatário, a primeira pessoa deveria entregar a carta para quem ela julgava conhecer ou estar mais próximo do destinatário. Com esse experimento Milgran concluiu que a distância média entre duas pessoas quaisquer em uma rede social é pequena, sendo que, em média, seis passos são suficientes para conectá-las, o que ficou conhecido como “seis graus de separação” [4].

Redes que apresentam a propriedade de pequeno mundo têm como características principais um alto coeficiente de aglomeração e um pequeno comprimento de caminho médio. Watts e Strogatz [4] perceberam que muitas redes apresentam padrões altamente conectados, tendendo a formar pequenas quantidades de conexões em cada vértice. Desse modo, em 1998 eles propuseram um modelo que reproduzia o efeito de pequeno mundo. Para tal, eles partiram de uma rede anel regular e religavam as arestas com uma probabilidade p . Com isso, eles obtiveram uma rede que ocupa uma posição intermediária entre a rede regular e a rede aleatória. Com as religações, atalhos se formaram na rede permitindo manter o alto agrupamento da rede regular e fazendo a distância média diminuir como acontece nas redes aleatórias.

O modelo proposto consistia em partir de uma rede anel regular, onde os N nós estavam conectados a seus $2r$ primeiros vizinhos. As ligações entre os nós estavam dispostas de forma simétrica, formando um total de $M = Nr$ arestas. A cada ligação é associada uma probabilidade p de religação e $1 - p$ para sua preservação. Quando a probabilidade p de religação é 0 a rede continua sendo a rede regular inicial, conforme a

Fig. 3.4 (a), pois não haverá possibilidade de religação. Aumentando o valor da probabilidade de religação esta passa a gerar redes com propriedades de Pequeno Mundo, Fig. 3.4 (b) e Fig. 3.4 (c), e quando a probabilidade é igual a $p = 1$, a rede assume o modo puramente aleatório Fig. 3.4 (d).

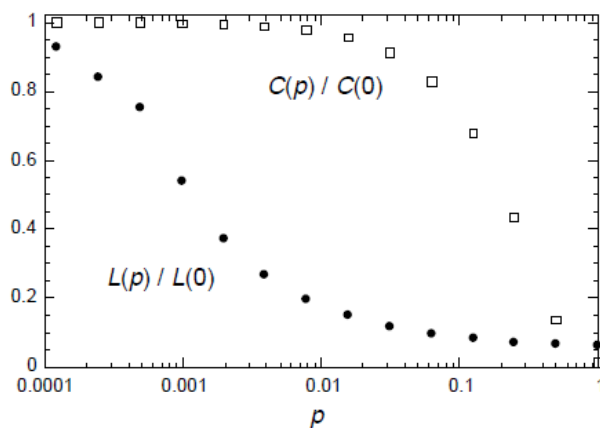
Figura 3.4 Redes de Pequeno Mundo com $N = 10, r = 2$ e a) $p = 0$; b) $p = 0,05$; c) $p = 0,1$; d) $p = 1$; construídas a partir do modelo de Watts-Strogats.



Fonte: Próprio Autor.

As propriedades estruturais dessas redes podem ser quantificadas por suas características de comprimento de caminho $L(p)$ e coeficiente de agrupamento $C(p)$, onde $L(p)$ mede a separação típica entre dois vértices do grafo, e $C(p)$ mede o coeficiente de agrupamento de uma vizinhança típica. Segundo Watts e Strogatz [4], se o grafo é conectado, então $L \sim n/2k \gg 1$ e $C \sim 3/4$ quando $p \rightarrow 0$, enquanto $L \approx L_{\text{aleatório}} \sim \ln(n)/\ln(k)$ e $C \approx C_{\text{aleatório}} \sim k/n \ll 1$ quando $p \rightarrow 1$. Assim, a rede regular em $p = 0$ possui um alto coeficiente de aglomeração, em que L cresce linearmente com o número de nós n , enquanto que a rede aleatória com $p = 1$ possui coeficiente de aglomeração baixo, em que L cresce logaritmicamente com n . Assim, para as redes regulares e aleatórias que são os casos limitantes, percebe-se que quando C é alto L também é alto. Para a rede regular acontece o oposto, esta possui baixos valores para L e C .

Figura 3.5 Comportamento do comprimento de caminho médio $L(p)$ e do coeficiente de agrupamento $C(p)$ em função de p , segundo o modelo de Watts e Strogatz (1998). Os valores de C e L foram normalizados pelos valores $L(0)$, $C(0)$ para uma rede regular, que são os respectivos valores máximos. Para $0,01 < p < 0,1$ ocorre o efeito Pequeno Mundo. Figura extraída de [4].



Fonte: Watts e Strogatz, Ref.[4].

A Fig. 3.5 revela que há um intervalo amplo de p sobre os quais $L(p)$ é quase tão pequeno quanto $L_{aleatório}$, e ainda, $C(p) \gg C_{aleatório} \sim \ln(n)/\ln(k)$. As redes de pequeno mundo resultam da queda imediata na $L(p)$, causada pela introdução de algumas arestas de longo alcance, como se fossem atalhos entre vértices distantes inicialmente. A cada aresta religada na rede, verifica-se que a distância não apenas entre o par de vértices que ele se conecta, mas entre suas vizinhanças imediatas diminua. Por outro lado, uma aresta removida de um agrupamento para criar um atalho causa um efeito linear sobre C , donde $C(p)$ permanece praticamente inalterada para p pequeno, embora $L(p)$ caia rapidamente.

Rede Sem Escala

Com os estudos das redes reais por meio das bases de dados disponíveis, verificou-se que a maioria das redes não possuía uma distribuição de grau de Poisson, como apontavam as redes aleatórias e Pequeno Mundo. Barabási e Albert [5] mostraram que muitas dessas redes seguiam uma distribuição de grau em lei de potência, onde existiam poucos nós altamente conectados e muitos nós com baixa conectividade [6]. Assim, em busca de criar um modelo mais preciso para representar algumas características das redes reais, Barabási e Albert em 1999 propuseram um novo modelo [7]. Este modelo permitia a geração de redes que seguissem uma distribuição em lei de potência, com $2 \leq \gamma \leq 3$,

conforme a Eq. (3.8). Tais redes recebem o nome de redes Sem Escala ou redes de Escala Livre, em que:

$$P(k) \sim k^{-\gamma} \quad (3.8)$$

Barabási e Albert foram os precursores no estudo de redes com tais características. Eles argumentavam que a origem da rede de Escala Livre estava fundamentada em dois princípios, que são comuns a muitas redes reais. Os modelos de redes propostos até então partiam de uma rede com número determinado de nós para as redes, que são aleatoriamente ligados ou religados. Assim o número de nós na rede permanece sempre inalterado. Porém, as redes reais não são estáticas, estas crescem ao longo do tempo, permitindo que novos nós sejam incorporados às redes. Os novos nós são inseridos mediante novas conexões estabelecidas com vértices antigos na rede, sendo que estas conexões se dão de forma preferencial.

Por exemplo, a WWW cresce exponencialmente no tempo por meio da inclusão de novas páginas web. Da mesma forma, a pesquisa de documentos cresce constantemente por meio de publicação de novos artigos. A ligação preferencial pode ser percebida nestas redes quando um novo artigo publicado cita preferencialmente um artigo anterior mais conhecido do que um artigo pouco conhecido, da mesma forma que páginas *webs* novas possuem *hiperlinks* para páginas anteriores mais conhecidas.

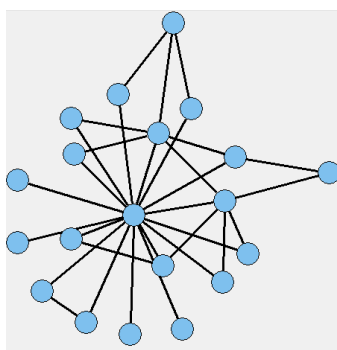
A partir dessas duas principais características, crescimento e ligação preferencial, desenvolveu-se o modelo de Barabási e Albert [7]. O crescimento se dá a partir de um número pequeno de nós n_0 , e em cada novo instante de tempo, um novo nó é adicionado à rede com m novas ligações, que serão estabelecidas entre o novo nó e os m nós já existentes na rede. As escolhas destes m nós da rede são realizadas preferencialmente aos nós que possuem maior grau. Dessa forma, um novo nó i será conectado a um nó preexistente j com grau k_j , com probabilidade:

$$P_{ij}(k) = \frac{k_j}{\sum_n k_n} . \quad (3.9)$$

Assim, os nós com maior número de conexão na rede tendem a ser escolhidos para as novas ligações. Devido ao favorecimento dos nós com maior conectividade, os nós ricos ficam mais ricos, e os pobres, tendem a ficarem mais pobres [5]. Esse resultado é

conhecido pelo efeito Mateus, por alusão a uma parábola deste Evangelho: “Pois a quem tem, mais será dado, e terá em grande quantidade. Mas a quem não tem, até o que tem lhe será tirado. (Mateus 25:29)”.

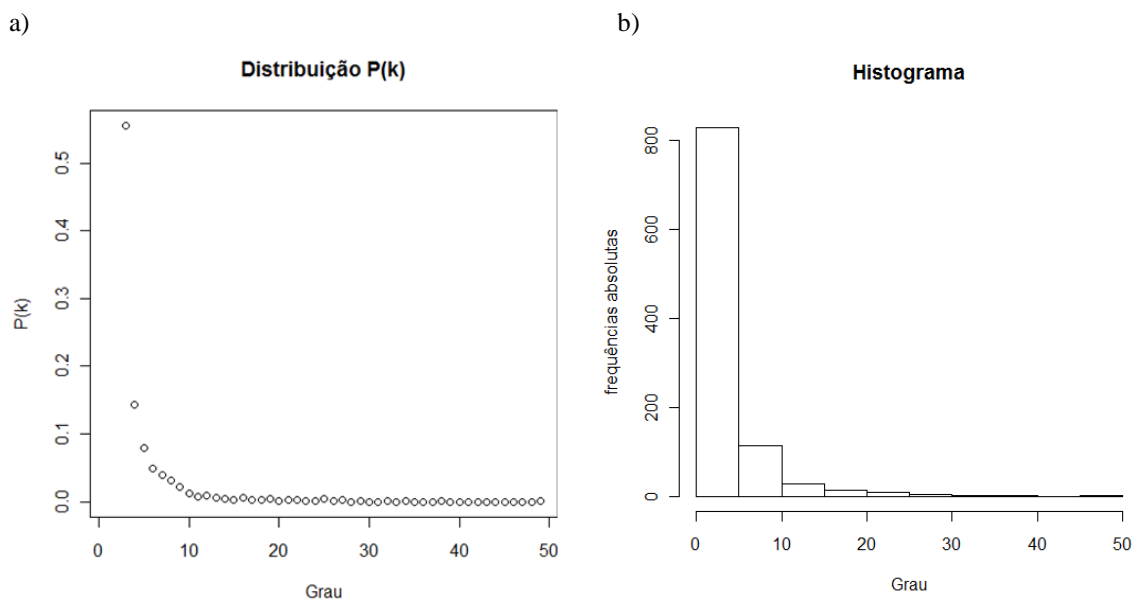
Figura 3.6 Rede Barabási-Albert, com número de nós $N = 20$ e número de arestas adicionadas a partir de cada novo nó $m = 2$.



Fonte: Próprio Autor.

A Fig. 3.6 apresenta uma rede de Barabási-Albert com $N = 20$, a qual cresce a cada instante de tempo com $m = 2$, ou seja, a cada iteração duas novas arestas são adicionadas, formando nós mais conectados e os nós mais recentes na rede menos conectados.

Figura 3.7) Distribuição de Grau seguindo uma Lei de Potência em a) de uma Rede de Escala livre, pelo modelo de Barabási-Albert, com $N = 1000$ e $m = 2$ e seu respectivo histograma em b) relacionando as frequências absolutas dos graus.



Fonte: Próprio Autor.

Após t instantes de tempo, a rede terá $N = n_0 + t$ nós e mt novas arestas. O modelo gera redes com distribuição de grau $P(k) \sim k^{-3}$ lei de potência. A Fig. 3.7 apresenta a distribuição de grau de uma rede de Barabási-Albert, a qual possui uma grande diversidade de valores para o grau dos nós, incluindo valores elevados para poucos nós e muitos nós com baixa conexão. A partir do modelo de Barabási e Albert, que geram redes com $\gamma = 3$ especificadamente, foram desenvolvidas variações deste modelo para tentar explorar outros valores para γ [7,41,42].

Rede Hierárquica

Um comportamento que desperta interesse em muitas redes reais é a organização hierárquica. Redes com este tipo de organização são redes sem escala que são construídas a partir de uma regra recursiva, sendo que esta regra pode ser determinística. Uma vez que se aplica a regra repetidas vezes produz-se uma rede com estrutura semelhante em várias ordens de magnitude [8, 43].

A maioria das redes que aparecem na natureza parecem seguir algum tipo de organização. Nestas redes, duas características podem ser destacadas: Elevado coeficiente de aglomeração e distribuição de grau em Lei de Potência. Estas características coexistem nas maioria das redes reais [44] e o modelo capaz de reuní-las são as redes hierárquicas construídas deterministicamente.

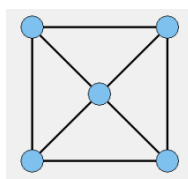
Seguindo um modelo determinístico é possível gerar redes sem escala e com organização hierárquica [44]. A rede é construída de maneira iterativa, iniciando a partir de um grupo pequeno de nós densamente conectados entre si. A cada etapa do processo de geração, réplicas da iteração anterior são geradas e adicionadas na rede. Estas cópias são então conectadas através de ligações com o agrupamento antigo. Nas iterações os elementos gerados nas etapas anteriores se repetem, mantendo um alto coeficiente de agrupamento.

O nó com maior grau é denominado “raiz”, que é o nó onde se inicia o processo de crescimento iterativo. A cada nova iteração, cópias da raiz no instante anterior vão sendo introduzidas na rede, e a raiz original vai aumentando seu grau à medida que a rede cresce, assim como as cópias que vão sendo introduzidas apresentam grau maior ou igual que as anteriores, gerando assim uma hierarquia de nós bastante conectados, que são uma característica comum de redes de escala livre. A distribuição de grau da rede hierárquica é

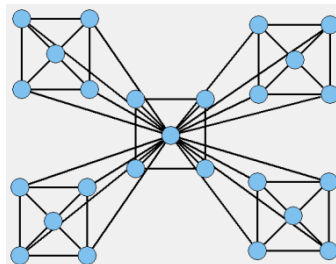
determinada por uma Lei de Potência, por haver nós na rede com poucas conexões e alguns nós muito conectados.

Figura 3.8 Primeiras etapas da construção de uma rede sem escala determinística e com estrutura hierárquica. Onde o agrupamento principal se inicia com M vértices, cujo vértice central é a raiz da rede, neste caso $M = 5$.

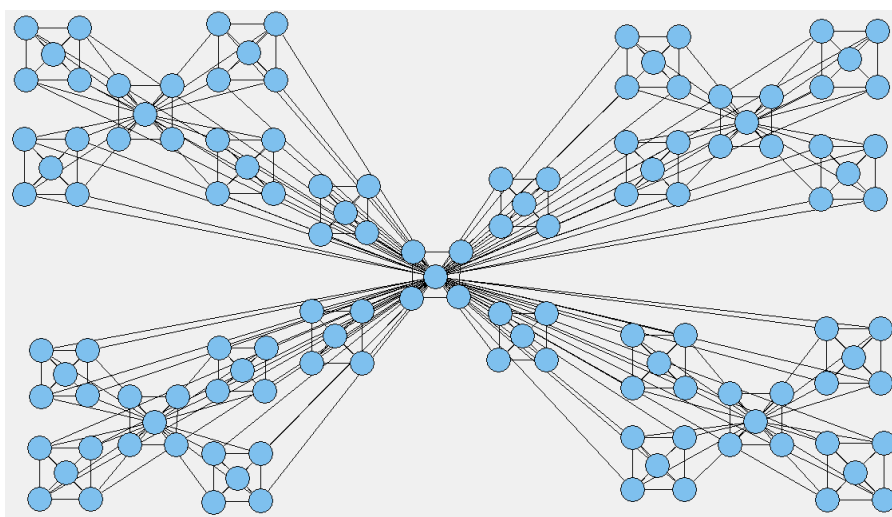
a) $t = 1$



b) $t = 2$



c) $t = 3$



Fonte: Próprio Autor.

A Fig. 3.8 mostra uma construção iterativa; inicialmente há um agrupamento totalmente conectado de cinco nós mostrados em (a). Cria-se quatro réplicas idênticas, ligando os nós periféricos de cada grupo ao nó central do agrupamento original, obtendo-se

uma rede com $N = 25$ na segunda iteração (b). No passo seguinte, cria-se quatro réplicas do conjunto obtido na etapa anterior, e conecta-se os nós periféricos novamente ao nó central do agrupamento inicial (c), obtendo-se uma rede com $N = 125$ [43]. Esse processo pode ser continuado indefinidamente levando-se a uma rede hierárquica.

CAPÍTULO 4

Centralidade em Redes Complexas

4.1 Medidas de Centralidade em Redes Complexas

O conceito de centralidade surgiu no contexto das redes sociais. As análises focavam o relacionamento interpessoal e a importância desempenhada por determinadas pessoas em uma rede, dada a função desenvolvida ou dado o número de conexões que possuíam. As conexões representavam os relacionamentos compreendidos em sua estrutura e organização. A partir dessas observações em redes sociais o conceito de centralidade pode ser expandido para outros tipos de redes complexas [45]. De forma intuitiva, podemos dizer que centralidade de um vértice na rede está relacionada ao número de vértices possíveis de serem atingidos a partir dele. Portanto, quanto mais central for o vértice, quando partindo dele, maior for o número de vértices possíveis de se atingir, com maior rapidez ou facilidade.

Há diversas formas de centralidade, cada definição leva em consideração diferentes especificidades relevantes da rede. De igual maneira, existem diferentes formas de mensurá-la. Estas medidas de centralidade, em princípio diferentes, possuem uma mesma função que é avaliar a importância dos nós em uma rede, considerando atuações distintas.

Podemos citar alguns tipos de centralidades comumente encontradas na literatura.

- Centralidade de Grau [46, 47]: recebe esta denominação por ser uma função do número de arestas incidentes a um vértice da rede.
- Centralidade de Proximidade [48,49]: está relacionada com a distância total de um vértice a todos os demais.
- Centralidade de Eficiência [50]: minimiza as distâncias de um vértice para alcançar qualquer outro vértice no grafo.
- Centralidade de Intermediação [51, 45]: mede quantos caminhos geodésicos entre todos os pares de vértices do grafo passam através de um vértice determinado.
- Centralidade de Intermediação de Percursos Aleatórios [15], o qual considera todos os caminhos, geodésicos e não geodésicos, passando através de um

vértice possíveis de serem usados por um percurso aleatório. Assim, existem diversas medidas de centralidade que são utilizadas para interpretar diferentes aspectos da rede.

4.2 Centralidade de Grau

A medida mais simples de centralidade é o grau, e deve ser também a concepção mais intuitiva. O grau corresponde ao número de ligações incidentes em um vértice da rede. Sendo assim, quanto maior o grau, maior é o número de conexões, o que garante ao vértice uma certa popularidade devido ao grande número de relações existente entre ele e os demais vértices da rede. Um nó que se encontra com um alto grau pode ser visto como um canal maior de informações. Shaw [47], foi o primeiro a usar o grau do vértice como uma medida de centralidade, sendo esta considerada a mais adequada até por volta de 1979, quando novas concepções surgiram.

Para se determinar a centralidade de grau é necessário computar o número de adjacências de um vértice v_k . Assim, considerando-se um grafo G com n vértices e v_k como sendo um vértice de G , a centralidade de grau de v_k , denotada por g_k , é o número de arestas incidentes a v_k , isto é,

$$g_k = \sum_{j=1}^n a_{kj}, \quad (4.1)$$

Em que, a_{kj} são elementos da matriz de adjacência de G .

4.3 Centralidade de Proximidade

O conceito de Proximidade está relacionado com a distância geodésica média entre um vértice e os outros vértices alcançáveis a partir deste. A Proximidade pode ser considerada como uma medida de tempo para levar informações de um dado vértice para outros na rede. Existem diversas medidas de centralidade usando o conceito de proximidade [46, 48, 49]. A mais simples destas medidas foi proposta por Sabidussi [49], denominada Centralidade de Proximidade. É baseada na soma das distâncias de um vértice em relação aos demais vértices da rede.

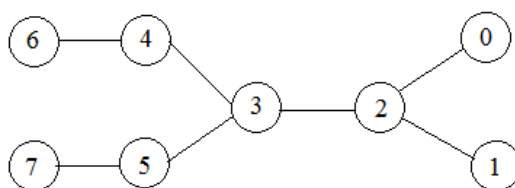
Se G um grafo conexo que representa uma rede complexa, com n vértices e v_k um vértice de G , a centralidade de proximidade de v_k é dada pelo inverso da soma das distâncias $d(v_j, v_k)$ de v_k a todos os demais vértices do grafo, ou seja:

$$C_p(v_k) = \frac{1}{\sum_{j=1}^n d(v_j, v_k)}. \quad (4.2)$$

4.4 Centralidade de Intermediação por Menores Caminhos

A centralidade de intermediação de um nó está relacionada com a importância estrutural que este nó possui dentro de uma rede, sendo medido a partir da fração de menores caminhos entre pares de nós que passam por um vértice específico. Esta medida pode ser vista como a influência de um nó, permitindo uma maior propagação de informação através da rede [35]. A centralidade do nó pode ser relacionada com o grau que possui. Porém, apenas o grau do nó não é suficiente para determinar se ele possui ou não maior centralidade (Anexo-D), como podemos observar na figura abaixo:

Figura 4.1 Os nós correspondentes aos índices 2 e 3 possuem o mesmo grau, o qual vale 3. Porém a centralidade do nó 3 é maior que a do nó 2.



Fonte: Próprio Autor.

A importância do nó também pode ser observada verificando-se as consequências sobre os outros nós quando acontece sua remoção da rede. Assim, quando as informações estiverem sendo distribuídas sobre a rede, pode-se verificar a centralidade do vértice, considerando o quanto influenciará no tráfego de informações a remoção de um desses nós.

No caso da Fig. 4.1, se suprimirmos o nó 3, a rede se partirá em três redes menores sem comunicação entre elas. Já se suprimirmos o nó 2, teremos que apenas dois nós da rede ficarão isolados, verificando assim, que a importância, ou seja, a centralidade de

intermediação do nó 3 na rede é maior que a do nó 2. Leva-se em conta que a rede abordada neste exemplo é uma rede genérica onde a função do nó não está sendo considerada (a verificação de remoção de nós pode variar com as funções assumidas pelos nós em diferentes redes reais).

Uma das definições mais aceitas de centralidade é baseada na contagem de caminhos geodésicos que passam por um nó [51] denominada de Centralidade de Intermediação por Menores Caminhos (CIMC). Assim, tomando-se um nó i qualquer da rede, deve-se contar o número de caminhos mínimos para cada par de nós da rede passando por i .

A CIMC é essencial na análise de várias redes do mundo real, tais como, as redes sociais, redes de transportes, redes de informação, entre outras. Em todas essas redes é importante a determinação do caminho mais curto entre os nós, pois, se conhecemos tal caminho, a eficiência no transporte seja de informação ou de qualquer outro material ocorre de maneira mais rápida e eficaz. Altas centralidades significam que um nó pode alcançar os demais nós da rede, de maneira rápida, por meio de caminhos curtos.

A centralidade de Intermediação pode ser descrita por um grafo G . Supondo que as redes estejam conectadas, sem laços, não direcionadas e sem arestas múltiplas, pode-se definir um caminho entre dois nós $s \in V$ e $t \in V$, como sendo uma sequência de nós conectados que se inicia com s e finaliza com t [45]. Se o grafo for ponderado, ou seja, se possuir pesos para suas arestas, o comprimento desse caminho será a soma dos pesos de suas arestas. Caso o grafo não seja ponderado, a cada aresta é associado o peso 1.

Denota-se por $d_G(s, t)$ a distância entre os nós s e t em um grafo G . Assumindo que $d_G(s, s) = 0$ para cada $s \in V$, e como a rede não possui direcionamento, as distâncias entre dois vértices s e t são iguais independente de qual seja o nó de partida, ou seja $d_G(s, t) = d_G(t, s)$ para $s, t \in V$.

Outra medida importante para a análise da importância de um nó na rede é o número de caminhos mínimos que ligam dois pares de vértices quaisquer, s e t da rede. Esse número é denotados por σ_{st} [45]. Também devido ao não direcionamento da rede, temos que $\sigma_{st} = \sigma_{ts}$, e por convenção, a medida de $\sigma_{ss} = 1$. Quando tratamos do caminho mínimo que conecta dois vértices quaisquer s e t da rede, que passa por um terceiro nó i , denotamos por $\sigma_{st}(i)$. A centralidade de intermediação de um nó i , é o número de caminhos mais curtos entre todos os pares s e t de nós da rede passando por i , ou seja:

$$C_i = \sum \sigma_{st}(i). \quad (4.3)$$

Caso exista mais de um caminho mais curto conectando dois pares de nós, a centralidade de intermediação será dado por:

$$C_i = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}. \quad (4.4)$$

Na Fig. 4.1, vemos que pelo nó 2 passam 11 caminhos, pelo 3 passam 16 caminhos, confirmando assim a maior centralidade do vértice 3, conforme o anexo D.

4.5 Centralidade de Intermediação de Percursos Aleatórios

A centralidade de Intermediação de Percursos Aleatórios foi proposta por Newman [15], e assim como a Intermediação por Menores Caminhos de Freeman [51] é calculado por meio da contagem do número de caminhos aleatórios que passam por um vértice i conectando dois pares de vértices quaisquer s e t da rede. A principal diferença entre estas duas propostas é que mesmo os caminhos que não são geodésicos são considerados para a medida da centralidade, o que torna a medida mais condizente com as redes reais. Nestas redes a informação ou fluxo, não conhece qual é o caminho mínimo a ser percorrido, e assim a informação percorre ao acaso até encontrar o seu destino, incluindo dessa forma não apenas os percursos ideais como também leva em consideração outros caminhos.

Para tal medida, considera-se uma rede formada por vários vértices e as ligações entre os mesmos. Um mensageiro deve partir de um vértice de origem s até um vértice de destino t ; porém o mensageiro não conhece o caminho que o leva até o seu alvo. Dessa maneira, o mensageiro a cada etapa do percurso escolhe aleatoriamente e uniformemente o próximo passo a partir das ligações adjacentes ao vértice atual, perfazendo um percurso aleatório. A probabilidade de transição de um caminhante que se encontra em um nó j e que poderá ser encontrado no próximo passo no nó i é dada por:

$$P_{ij} = \frac{a_{ij}}{k_j}. \quad (4.5)$$

Denotando por M a matriz das probabilidades de transição entre os vértices, podemos escrever $M = AD^{-1}$, onde D é a matriz diagonal dos graus dos vértices e A é matriz de adjacência da rede.

Como o percurso aleatório é absorvente, para $j = t$ tem-se que $M_{ij} = 0$. Isso significa que quando o destino é alcançado o caminho chega ao fim. A maneira conveniente, segundo Newman [15], para representar esta situação é remover a linha e a coluna referente à t na matriz M , sendo que as demais transições não são afetadas por esse procedimento. A nova matriz com os elementos removidos é representada por $M_t = A_t D_t^{-1}$. A matriz $(A^r)_{ij}$ fornece o número de caminhos que unem o vértice i ao vértice j , com comprimento r (Anexo - E). A probabilidade de que um percurso aleatório inicie em s e após r etapas, esteja em j é dado por $[M_t^r]_{js}$ e a probabilidade que na etapa posterior $r + 1$ esteja em um vértice adjacente i é dada por $k_j^{-1}[M_t^r]_{js}$.

Tomando o somatório sobre todos os possíveis percursos de comprimento r de 0 até ∞ , que vão de j para i , temos que:

$$\sum_{r=0}^{\infty} k_j^{-1}[M_t^r]_{js} = k_j^{-1}[(I - M_t)^{-1}]_{js}, \quad (4.6)$$

matricialmente, obtemos:

$$V = D_t^{-1}(I - M_t)^{-1}.S = (D_t - A_t)^{-1}.S. \quad (4.7)$$

Em que S é um vetor fonte definido como:

$$S = \begin{cases} +1, & \text{para } i = s \\ -1, & \text{para } i = t \\ 0, & \text{caso contrário} \end{cases} . \quad (4.8)$$

Portanto, a probabilidade de um percurso aleatório passe ao longo da aresta entre i e j é dado pela diferença absoluta entre as probabilidades $|V_{st}(i) - V_{st}(j)|$ e a probabilidade de passar através do vértice i é a metade da soma das probabilidades nas arestas incidentes a este:

$$I_{st}(i) = \frac{1}{2} \sum_j A_{ij} |V_{st}(i) - V_{st}(j)| = \frac{1}{2} \sum_j A_{ij} |T_{is} - T_{js} - T_{it} + T_{jt}| \text{ para } j \neq s, t. \quad (4.9)$$

A expressão não é válida para os vértices de origem e destino; nestes casos, pode-se tomá-los como a unidade, $I_{st}(s) = I_{st}(t) = 1$. Assim, a centralidade de intermediação do vértice i é a média sobre todos os caminhos aleatórios entre s e t e pode ser definida da seguinte forma:

$$N_i = \frac{2 \sum_{s < t} I_{st}(i)}{n(n-1)}. \quad (4.10)$$

4.6 Centralidade da Caminhada Aleatória

A Centralidade da Caminhada Aleatória (CCA) proposta por Noh e Rieger [14] é uma medida da velocidade com que as mensagens caminham na rede. Neste caso o processo de transporte em redes seria o mais rápido se percorresse o menor caminho entre dois vértices. Porém, o caminho ótimo só é encontrado se for possível conhecer a conectividade global da rede. A importância desta medida é que não depende da conectividade global; apenas com o conhecimento da conectividade local é possível determiná-la.

No contexto de transporte e de procura, o Tempo de Primeira Passagem é uma característica importante da caminhada aleatória, e pode ser entendido como o tempo exigido para o processo atingir o estado j a partir do estado i . Quando $j = i$ temos o tempo de recorrência, que significa simplesmente o número de transições para o estado retornar ao estado i .

O tempo de primeira passagem é uma variável aleatória, em que a distribuição de probabilidade a ele associada, depende das probabilidades e transição. A probabilidade associada ao tempo de primeira passagem é $F_{ij}(t)$ denominada de probabilidade de primeira passagem de sair de i para j em t passos.

Primeiramente, para a transição em um único passo as duas probabilidades consideradas são iguais. As probabilidades seguintes podem ser encontradas de forma recursiva, como descrita abaixo:

$$F_{ij}(1) = P_{ij}(1) = P_{ij}$$

$$F_{ij}(2) = \sum_{k \neq j} P_{ik} \cdot F_{kj}(1)$$

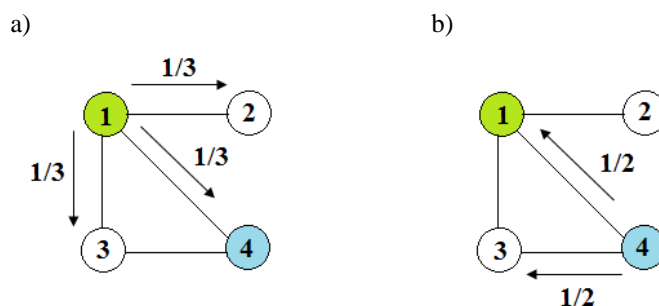
$$\begin{aligned}
 F_{ij}(3) &= \sum_{k \neq j} P_{ik} \cdot F_{kj}(2) \\
 &\vdots \\
 F_{ij}(t) &= \sum_{k \neq j} P_{ik} \cdot F_{kj}(t-1).
 \end{aligned}
 \tag{4.11}$$

Se $\sum_t^\infty F_{ij}(t) < 1$ significa que o processo inicialmente em i jamais pode encontrar-se no estado j . Quando $\sum_t^\infty F_{ij}(t) = 1$, $F_{ij}(t)$ pode ser considerada como uma distribuição de probabilidade para o tempo de primeira passagem. Ainda, se considerarmos $i = j$ e $\sum_t^\infty F_{ii}(t) = 1$ o estado é recorrente. Assim, pode-se determinar o Tempo Médio de Primeira Passagem (TMPP) de um estado para outro, da seguinte forma:

$$\langle T_{ij} \rangle = \begin{cases} \infty, & \text{se } \sum_t^\infty F_{ij}(t) < 1, \\ \sum_t^\infty t F_{ij}(t), & \text{se } \sum_t^\infty F_{ij}(t) = 1. \end{cases}
 \tag{4.12}$$

O movimento da caminhada aleatória possui uma assimetria intrínseca em sua definição, com relação ao movimento em uma direção e sua oposta, como pode ser visto na Fig. 4.2, o que leva também a uma assimetria no TMPP. Noh e Rieger [14] realizaram a derivação de uma fórmula exata para o TMPP de um caminhante aleatório que vai de um nó i para outro nó j , denotado por $\langle T_{ij} \rangle$ em uma rede arbitrária. A diferença entre $\langle T_{ij} \rangle$ e $\langle T_{ji} \rangle$ determinará a centralidade da caminhada aleatória. Ela descreve a centralização de informações vagando sobre as redes, e o potencial em receber e dissipar mais rapidamente informações.

Figura 4.2 Assimetria da caminhada aleatória em rede. a) probabilidade de um caminhante sair de um nó 1 para um de seus vizinhos. b) A probabilidade de um caminhante estar nó 4 e voltar para o nó 1.



Fonte: Próprio Autor.

Considere uma rede conectada e não direcionada com N nós ($i = 1, \dots, N$) e ligações entre estes nós. O processo estocástico em tempo discreto da caminhada aleatória nesta rede que é descrita pela Eq. (2.8). Quando o tempo tende ao infinito a distribuição atinge o estado estacionário, denotado por $P_j^\infty = \lim_{t \rightarrow \infty} P_{ij}(t)$.

Ao considerar as probabilidades de transição do vértice i para j e comparando as expressões para P_{ij} e P_{ji} definidos pela Eq. (2.8) segue-se que:

$$P_{ij}(t) = \sum_{j_1, j_2, j_3, \dots, j_{t-1}} \frac{A_{ij_1}}{k_i} \cdot \frac{A_{j_1 j_2}}{k_{j_1}} \cdot \frac{A_{j_2 j_3}}{k_{j_2}} \dots \frac{A_{j_{t-1} j}}{k_{j_{t-1}}}.$$

Assim, multiplicando a probabilidade de transição pelo grau do vértice de saída temos:

$$\begin{aligned} k_i P_{ij}(t) &= \sum_{j_1, j_2, \dots, j_{t-1}} k_i \cdot \frac{A_{ij_1}}{k_i} \cdot \frac{A_{j_1 j_2}}{k_{j_1}} \cdot \frac{A_{j_2 j_3}}{k_{j_2}} \dots \frac{A_{j_{t-1} j}}{k_{j_{t-1}}} \\ &= \sum_{j_1, j_2, \dots, j_{t-1}} k_j \cdot \frac{A_{j j_{t-1}}}{k_j} \cdot \frac{A_{j_{t-1} j_{t-2}}}{k_{j_{t-1}}} \dots \frac{A_{j_2 j_1}}{k_{j_2}} \cdot \frac{A_{j_1 i}}{k_{j_1}} = k_j P_{ji}(t). \end{aligned}$$

Este fato ocorre devido ao não direcionamento da rede; basta considerar que $A_{j_k j_{k+1}} = A_{j_{k+1} j_k}$. Para o caso estacionário a equação acima pode ser reescrita como $k_i P_j^\infty = k_j P_i^\infty$ e, portanto, uma solução que satisfaz esta igualdade é dada por:

$$P_i^\infty = \frac{k_i}{N}. \quad (4.14)$$

Em que $N = \sum_i k_i$ [14]. A distribuição estacionária está assim relacionada com o grau do vértice em questão e sua normalização. Dessa forma, quanto mais ligações um nó possui, maior será a probabilidade de ser visitado por um caminhante aleatório.

Para descrever a rapidez com que acontece o movimento aleatório, é necessário verificar o TMPP. Para tal, é necessário considerar a probabilidade da primeira passagem $F_{ij}(t)$ que satisfaz a equação:

$$P_{ij}(t) = \delta_{t0} \delta_{ij} + \sum_{t'=0}^t P_{jj}(t-t') F_{ij}(t'). \quad (4.15)$$

O símbolo delta de Kronecker assegura a condição inicial $P_{ij}(0) = \delta_{ij}$ e $F_{ij}(0)$ é definido como zero. Introduzindo a Transformada de Laplace $\tilde{f}(s) \equiv \sum_{t=0}^{\infty} e^{-st} f(t)$, na Eq. (4.15) a expressão torna-se:

$$\sum_{t=0}^{\infty} e^{-st} P_{ij}(t) = \sum_{t=0}^{\infty} e^{-st} \left(\delta_{t0} \delta_{ij} + \sum_{t'=0}^t P_{jj}(t-t') F_{ij}(t') \right).$$

O segundo membro pode ser resolvido mediante a aplicação do Teorema da Convolução, resultando em:

$$\begin{aligned} \tilde{P}_{ij}(s) &= \delta_{ij} + \tilde{F}_{ij}(s) \tilde{P}_{jj}(s), \text{ ou ainda,} \\ \tilde{F}_{ij}(s) &= [\tilde{P}_{ij}(s) - \delta_{ij}] / \tilde{P}_{jj}(s). \end{aligned} \quad (4.16)$$

Como as redes consideradas são finitas, o caminho aleatório é recorrente [14,18], de modo que o TMPP é dado conforme a Eq. (4.12):

$$\langle T_{ij} \rangle = \sum_{t=0}^{\infty} t F_{ij}(t) = -\tilde{F}'_{ij}(0). \quad (4.17)$$

Uma vez que todos os momentos $R_{ij}^{(n)} \equiv \sum_{t=0}^{\infty} t^n \{P_{ij}(t) - P_j^{\infty}\}$ da relaxação decaem exponencialmente $P_{ij}(t)$ são finitos (em particular, temos o tempo característico de relaxação do nó i como sendo $\tau_i = R_{ii}^{(0)} = \sum_{t=0}^{\infty} \{P_{ii}(t) - P_i^{\infty}\}$) pode-se expandir \tilde{P}_{ij} por uma série em s :

$$\tilde{P}_{ij}(s) = \frac{k_j}{N(1-e^{-s})} + \sum_{n=0}^{\infty} (-1)^n R_{ij}^{(n)} \frac{s^n}{n!}. \quad (4.18)$$

Inserindo esta série na Eq. (4.16) e expandindo-o como uma série de potência em s , temos que:

$$\tilde{F}_{ij}(s) = \frac{\frac{k_j}{N(1-e^{-s})} + \sum_{n=0}^{\infty} (-1)^n R_{ij}^{(n)} \frac{s^n}{n!} - \delta_{ij}}{\frac{k_j}{N(1-e^{-s})} + \sum_{n=0}^{\infty} (-1)^n R_{ij}^{(n)} \frac{s^n}{n!}},$$

$$\tilde{F}_{ij}(s) = \frac{k_j + N(1 - e^{-s}) \cdot \sum_{n=0}^{\infty} (-1)^n R_{ij}^{(n)} \frac{s^n}{n!} - N(1 - e^{-s}) \cdot \delta_{ij}}{k_j + N(1 - e^{-s}) \sum_{n=0}^{\infty} (-1)^n R_{ij}^{(n)} \frac{s^n}{n!}}.$$

Derivando a expressão acima e calculando para o caso $n = 0$, obtemos

$$\tilde{F}'_{ij}(0) = \frac{-Nk_j R_{ij}^{(0)} + Nk_j \delta_{ij} + Nk_j R_{jj}^{(0)}}{k_j^2},$$

Aplicando-se a Eq.(4.17) obtemos:

$$\langle T_{ij} \rangle = \begin{cases} \frac{N}{K_j}, & \text{para } j=i \\ \frac{N}{K_j} [R_{jj}^{(0)} - R_{ij}^{(0)}], & \text{para } j \neq i. \end{cases} \quad (4.19)$$

Tomando a diferença entre $\langle T_{ij} \rangle$ e $\langle T_{ji} \rangle$ para $i \neq j$ resulta em

$$\langle T_{ij} \rangle - \langle T_{ji} \rangle = N \left(\frac{R_{jj}^{(0)}}{K_j} - \frac{R_{ii}^{(0)}}{K_i} \right) - N \left(\frac{R_{ij}^{(0)}}{K_j} - \frac{R_{ji}^{(0)}}{K_i} \right).$$

O último termo desaparece devido à Eq. (4.13). Portanto, obtemos,

$$\langle T_{ij} \rangle - \langle T_{ji} \rangle = N \frac{R_{jj}^{(0)}}{K_j} - N \frac{R_{ii}^{(0)}}{K_i}.$$

Reescrevendo em termos da Eq.(4.14) a qual define a probabilidade estacionária:

$$\langle T_{ij} \rangle - \langle T_{ji} \rangle = \frac{R_{jj}^{(0)}}{P_j^{\infty}} - \frac{R_{ii}^{(0)}}{P_i^{\infty}},$$

$$\langle T_{ij} \rangle - \langle T_{ji} \rangle = C_{CA}^{-1}(j) - C_{CA}^{-1}(i). \quad (4.20)$$

Em que $C_{CA}(i)$ é definido como:

$$C_{CA}(i) = \frac{P_i^\infty}{\tau_i}. \quad (4.21)$$

Esta quantidade $C_{CA}(i)$ é denominada Centralidade da Caminhada Aleatória [14], uma vez que quantifica o quanto um nó i está localizado sobre o seu potencial para receber informações aleatórias difundindo sobre a rede. Assim, se considerados dois nós i e j com $C_{CA}(i) > C_{CA}(j)$ e supondo que cada um destes nós lança um sinal simultaneamente na rede espera-se que o nó i que possui maior CCA receba o sinal emitido mais rapidamente que j . Assim, a CCA pode ser considerada como uma medida para a eficácia na comunicação entre nós.

CAPÍTULO 5

Discussão dos Resultados

5.1 Aspectos Gerais

Foram analisados diferentes tipos de redes complexas e o comportamento das diversas formas de centralidade para cada tipo de rede. As redes analisadas foram às redes Aleatórias, as redes de Mundo Pequeno, as redes Sem Escala proposta pelo modelo de Barabási-Albert [5] e as redes determinísticas Hierárquicas [43]. Das redes acima citadas, as duas primeiras são modelos de redes que possuem um número fixo de nós, não havendo inserção de novos nós na rede. No entanto, as duas últimas redes seguem modelos de crescimento não estático, ao contrário, são modelos que permitem o crescimento a cada nova iteração até um número pré-estabelecido. Para estes modelos dinâmicos de rede em crescimento, a rede de Barabási-Albert possui como característica possibilidades de novas conexões não equiprováveis, seguindo uma probabilidade de ligação preferencial que varia de acordo com o número de conexões atuais dos nós da rede. Para a rede Hierárquica, o processo de construção e crescimento da rede não é probabilístico e sim determinístico, onde a cada iteração os novos nós se conectam com os nós antigos formando cópias da estrutura anterior.

5.2 Tempo Característico de Relaxação

Para cada modelo de rede, foram realizadas simulações de caminhada aleatória com as redes já evoluídas, encontrando-se em um estado estático, onde não há mais o crescimento da rede. Para realizar estas simulações, em cada rede foi resolvida numericamente a equação mestra da caminhada aleatória com condição inicial $P_i(t = 0) = \delta_{ik}$, pois é a partir dessa equação que é possível calcular os valores da CCA (Anexo - A). Como a medida de CCA depende explicitamente das conexões locais e do tempo característico de relaxação, foi calculado também, para cada modelo de rede, o tempo de relaxação τ_k para cada k .

Como visto no Cap. (4) a CCA é determinada por duas características da rede, o grau k e o tempo característico de relaxação τ . A ordem de grandeza do tempo característico de relaxação τ está relacionada com o segundo maior autovalor do operador de evolução temporal em $P_{ii}(t) = P_i^\infty + \sum_{\lambda=2}^N a_i^{(\lambda)} b_i^{(\lambda)} \Lambda_\lambda^t$, em que $a_i^{(\lambda)}$ e $b_i^{(\lambda)}$ são os autovetores à esquerda e à direita, respectivamente, do operador de evolução temporal, relativos ao autovalor Λ_λ^t . Se a ordem dos autovalores de acordo com os módulos é $|\Lambda_2| \geq |\Lambda_3| \geq \dots \geq |\Lambda_N|$, então o comportamento é assintótico, $P_{ii}(t) - P_i^\infty \sim a_i^{(2)} b_i^{(2)} \Lambda_2^t$ e $\tau_i \approx a_i^{(2)} b_i^{(2)} / |\ln|\Lambda_2^t||$. Desse modo, o tempo de relaxação τ_i tem uma dependência para o nó somente através de um fator peso, que é presumivelmente fraco. Por outro lado, a dependência do grau é explícita. Estas considerações também podem ser observadas nas simulações de acordo com a Fig. 5.1, onde cada gráfico corresponde a um tipo específico de rede analisada.

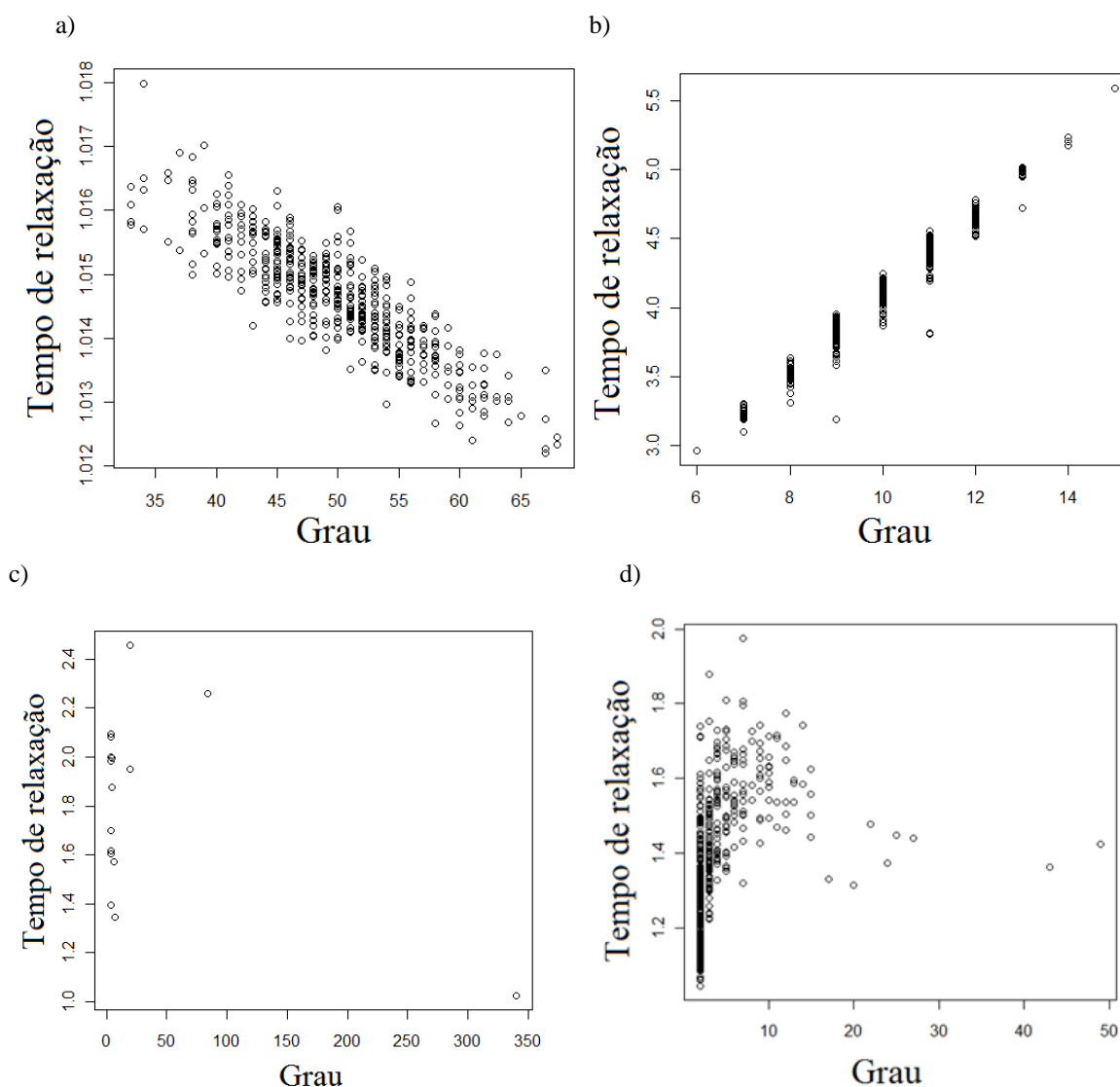
O foco principal da análise realizada na simulação ilustrada na Fig. 5.1 não é encontrar um comportamento específico para a relação “grau x tempo de relaxação”, e por esta razão não foi realizado o ajuste dos pontos obtidos, ainda mais que, para modelos diferentes de rede a distribuição do grau em relação ao tempo de relaxação também se comporta de forma diferente, devido a características próprias de cada modelo rede e suas especificidades estruturais. Assim, a análise importante para essa relação é verificar o intervalo numérico que se limitam as distribuições em ambas as grandezas, verificando dessa forma, a influência causada por elas no cálculo da CCA.

Por meio de simulações ilustradas na Fig. 5.1 é possível perceber que o grau é distribuído amplamente em comparação com o tempo de relaxação em todos os casos. O grau se distribui em um intervalo grande de valores, enquanto o tempo de relaxação se estabelece em um intervalo muito menor, confirmando que a CCA é influenciada de forma significativa pela dependência com relação ao grau, enquanto que a influência do tempo de relaxação é dada de forma restrita, causando pequena influência na medida de CCA. A maior variação para o grau ocorre nas redes Sem Escala. Para τ , a maior variação ocorre na rede de Pequeno Mundo.

A rede analisada na Fig. 5.1 (a) é uma rede Aleatória que possui 500 nós e probabilidade $p = 0,1$. Para esta rede o grau varia entre $30 \leq k \leq 70$ e o tempo de relaxação, é distribuído de maneira muito restrita no intervalo $1 \lesssim \tau \lesssim 1.02$. As medidas para CCA, para o modelo de rede em questão, não devem apresentar flutuações

significativas devido ao pequeno intervalo de valores assumido por τ . Para a rede de Pequeno Mundo com $N = 500$ e $p = 0,1$, podemos observar na Fig. 5.1 (b) que o grau varia entre $5 \leq k \leq 15$. Este modelo de rede possui uma menor variação para os valores do grau, sendo estes distribuídos em torno do grau médio. Por outro lado, a rede de Pequeno Mundo apresenta a maior variação em relação ao tempo de relaxação, sendo este distribuído no intervalo $3 \lesssim \tau \lesssim 5,5$, e por isso, a medida de CCA para esta rede possui uma maior interferência de τ , acarretando em maiores flutuações em sua comparação com o grau.

Figura 5.1 Os gráficos mostram a relação de τ versus k das redes: a) Aleatória com $N = 500$ e $p = 0,1$. b) Pequeno Mundo com $N = 500$ e $p = 0,1$. c) Hierárquica $N = 512$ e d) Sem Escala, modelo de Barabási-Albert para $N = 500$ e os nós cresceram com o parâmetro $m = 2$.



Fonte: Próprio Autor.

Na Fig. 5.1 (c) é ilustrado o resultado da simulação para a rede Hierárquica, com número de nós igual 512, e agrupamento inicial composto por cinco nós, ($m = 5$). Para esta rede o grau varia entre $2 \leq k \leq 340$ e o tempo de relaxação é distribuído de maneira muito restrita no intervalo $1 \lesssim \tau \lesssim 2,4$. Assim, como na rede Aleatória a flutuação não é significativa para a CCA. E por fim, o gráfico da Fig. 5.1 (d) apresenta a simulação para a rede de Barabási-Albert com $N = 500$ e quantidade de nós adicionados em cada iteração é $m = 2$; o grau varia entre $2 \leq k \leq 50$ e por ser uma rede Sem Escala, assim como a rede Hierárquica, o comportamento é semelhante para o tempo de relaxação, $1 \lesssim \tau \lesssim 2$ (portanto a flutuação no comportamento da CCA não é significativa).

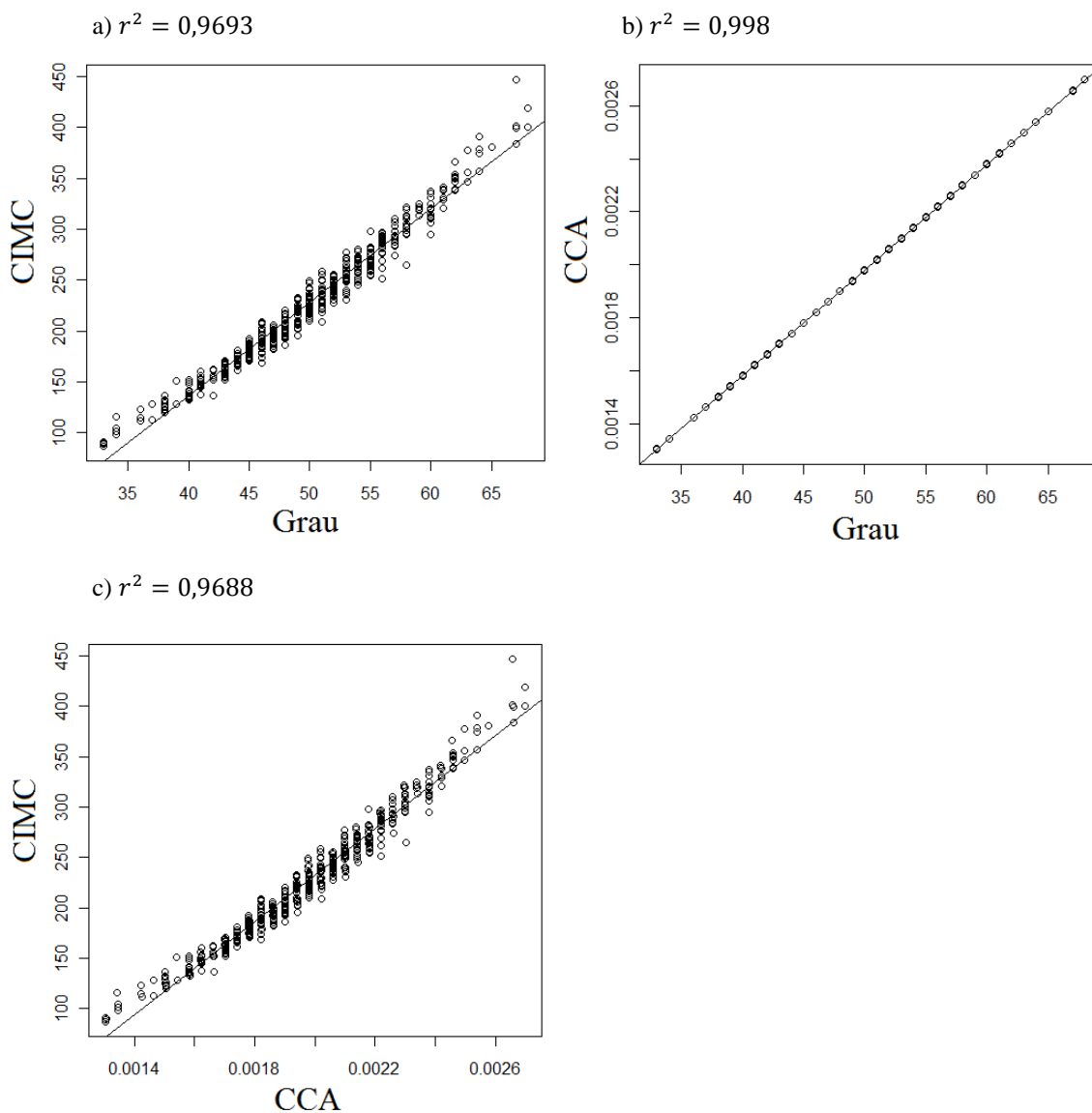
Diversas simulações foram realizadas, com os mesmos parâmetros, para cada rede, com as quais se pode extrair a média para chegar aos dados apresentados. Em todos os casos, foram realizadas também simulações para diferentes tamanhos de redes, não sendo encontrada qualquer ampliação significativa da distribuição de τ , para nenhum dos modelos de rede analisada, indicando que a CCA é determinada principalmente pelo grau da distribuição estando pouco relacionada com seu tamanho.

5.3 Relação entre as diferentes medidas de Centralidade

Pode-se inferir diferentes medidas de centralidade para uma rede complexa, como já visto no Capítulo 4. Entre estas medidas escolheu-se para as comparações a Centralidade de Grau, a CIMC e a CCA. Assim, para verificar a relação entre as diferentes medidas de centralidade de redes complexas foram construídos os gráficos das Fig. 5.2, Fig. 5.3, Fig. 5.4 e Fig. 5.5. Estas imagens estão relacionadas às comparações entre as medidas de centralidade realizadas para as quatro classificações de redes complexas, sendo elas a rede Aleatória, Pequeno Mundo, Hierárquica e Escala Livre, respectivamente.

A primeira imagem, Fig. 5.2 (a), se refere as comparações em relação às medidas de centralidades para a rede Aleatória. Por meio desta imagem é possível perceber que as medidas de centralidades apresentam um caráter crescente, proporcional ao grau. A Fig. 5.2 (b) apresenta o gráfico de CCA *versus* grau. Este mostra que a CCA é aproximadamente proporcional ao grau, não havendo muitas discrepâncias na relação entre estas duas medidas, em razão da flutuação exercida pelo tempo de relaxação τ ser pequena. A reta de regressão linear para esta rede possui correlação positiva.

Figura 5.2 Relação de entre três diferentes medidas de centralidade para as redes Aleatória. Em a) k versus CIMC, b) k versus CCA, e c) CCA versus CIMC. Cada imagem traz a correlação referente ao ajuste, a rede possui $N = 500$ e $p = 0,1$.



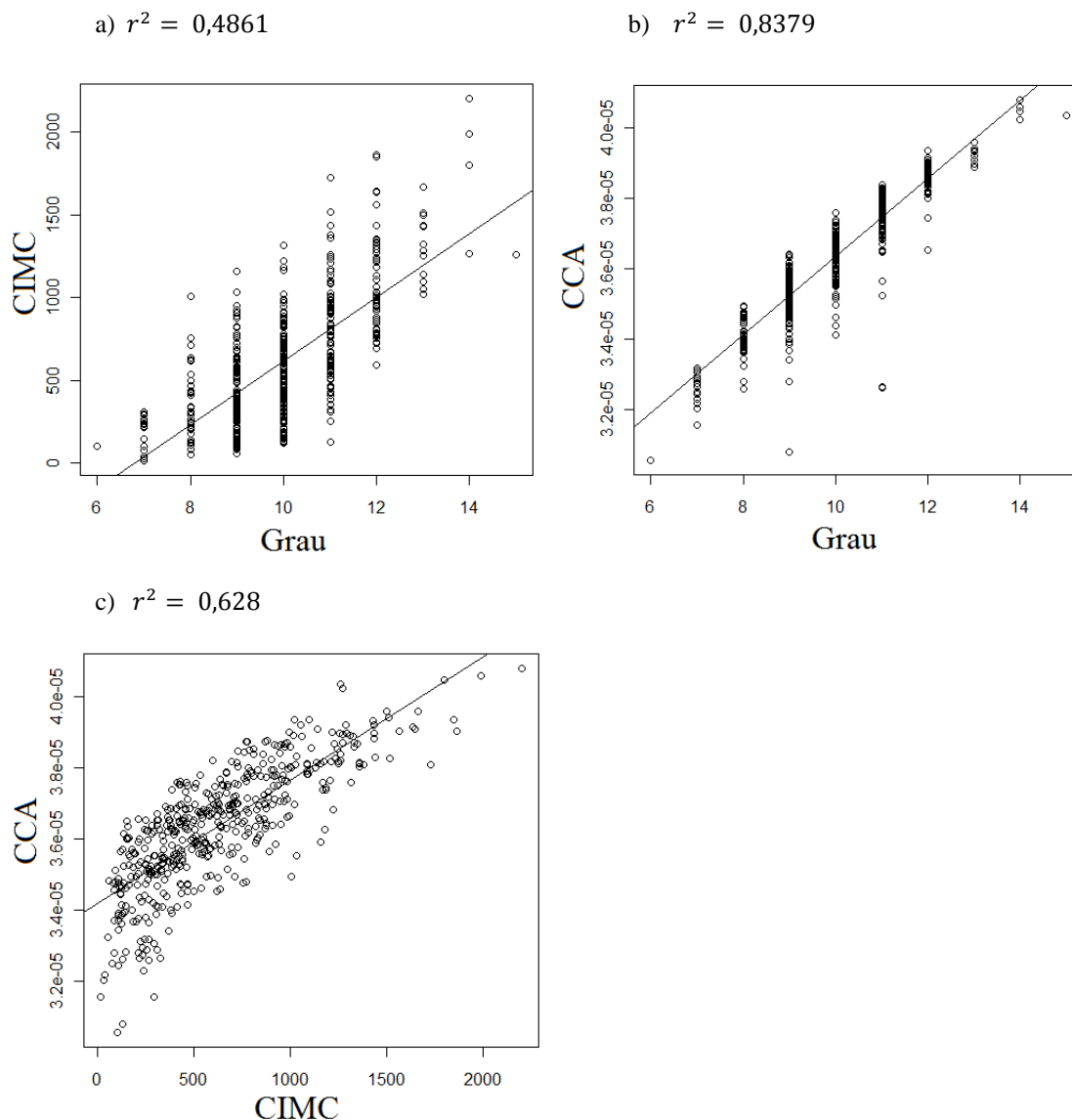
Fonte: Próprio Autor.

Nas Fig. 5.2 (a) e Fig. 5.2 (c) é possível verificar a relação existente entre Grau x CIMC e CCA x CIMC, respectivamente. Ambos os casos, possuem um comportamento crescente, porém este crescimento não é uniforme, pois o grau não influencia de maneira determinante a medida de CIMC, sendo que nós que possuem o mesmo valor para o grau podem possuir números diferentes de menores caminhos passando por eles. É claro que

quanto maior o valor do grau de um vértice, também é maior a probabilidade de possuir um maior número de caminhos mínimos.

As redes de Pequeno Mundo mostram que o comportamento das medidas de centralidade continua ter um caráter crescente em suas relações, fato este que pode ser observado na Fig. 5.3. Porém, estas redes apresentam uma maior variação para os valores de suas medidas em relação aos valores assumidos para os graus dos vértices da rede.

Figura 5.3 Relação entre três diferentes medidas de centralidade para as redes de Pequeno Mundo, seguindo o modelo de Watts e Strogatz. Em a) k versus CIMC, b) k versus CCA, e c) CIMC versus CCA. Cada imagem traz o valor da correlação referente ao ajuste, o número de nós para rede é $N = 500$ e $p = 0,1$.



Fonte: Próprio Autor.

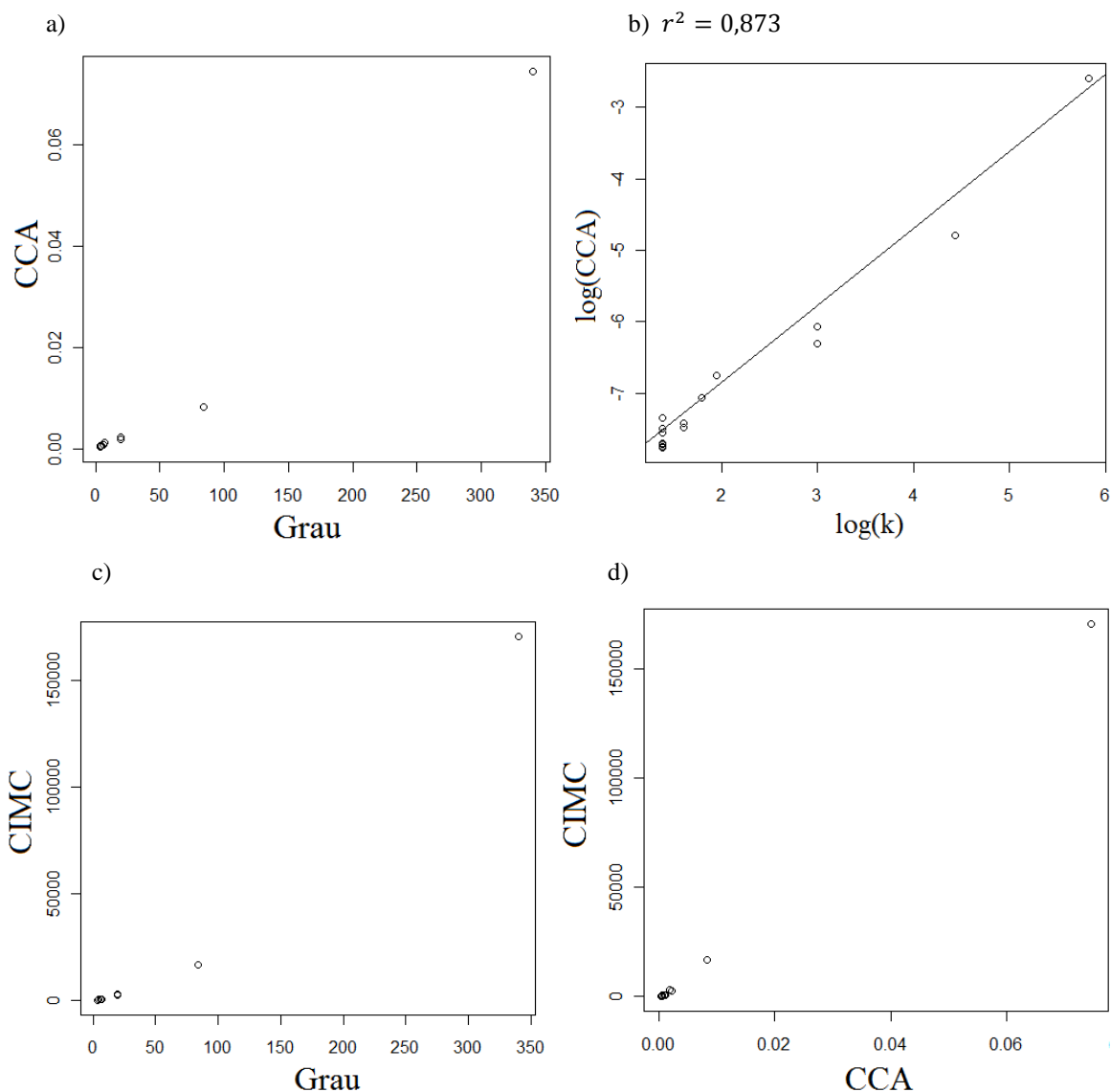
Na Fig. 5.3 (a) verifica-se que existe um grande número de nós que possuem o mesmo grau e diferentes valores para a CIMC. Na Fig. 5.3 (b) ocorre a mesma situação para os valores da CCA. Estes fatos podem ser entendidos, devido as redes de Pequeno Mundo possuírem uma pequena variação entre os valores dos graus, onde estes se concentram, para rede aqui analisada, entre 6 para o grau mínimo e 15 para o grau máximo. Assim muitos vértices com o mesmo grau apresentam valores diversos tanto para a CIMC quanto para CCA, respectivamente. Ainda referente a Fig. 5.3 (b) vemos que as discrepâncias com relação ao grau são maiores pois, como visto anteriormente na Fig. 5.1 (b), tal rede possui maior variação para τ , influenciando de forma mais significativa as medidas de CCA, (quanto maior for o tempo de relaxação, mais impacto causa no cálculo da CCA, permitindo menor uniformidade da medida com relação ao grau).

A comparação entre as medidas de CIMC em relação a CCA, apresentadas na Fig. 5.3 (c), mostra também que o comportamento é crescente. No entanto, os pontos referentes às medidas se apresentam de forma bastante dispersa no diagrama, apresentando um comportamento não uniforme. Os pontos com relação à reta de ajuste apresentam certa discrepância que pode ser assim percebida pelos dados da regressão linear que apontam o crescimento positivo, mas distante de 1.

As redes Hierárquicas mostram que o comportamento das medidas de centralidade continua com um caráter crescente em suas relações, como ilustradas na Fig. 5.4. Esta rede tem como característica possuir muitos nós com o mesmo valor para o grau, devido à sua forma de geração, por ser uma rede cujo processo de crescimento se faz mediante a repetição de cópias idênticas das iterações anteriores, revelando uma característica fractal.

A Fig.5.4 (a) mostra as relações entre o grau e a CCA; podemos perceber o acúmulo dos pontos na região de baixo valor. Assim, de forma a facilitar a visualização do comportamento da relação entre a centralidade de Grau e a CCA, o gráfico da Fig. 5.4 (b) apresenta essa mesma relação em escala logarítmica. O ajuste linear confirma que a medida de CCA apresenta um comportamento linearmente crescente em relação ao grau para este modelo de rede. Dessa forma, se confirma a estimativa realizada anteriormente, devido ao valor de τ ser restrito a um intervalo pequeno de valores entre 1 e 2, causando uma flutuação pequena com relação a estas duas grandezas.

Figura 5.4 Relação entre três diferentes medidas de centralidade para as redes Hierárquicas, com o número de nós para a rede $N = 512$ e $m=5$. Em a) k versus CCA, em b) a mesma relação k versus CCA plotado com escala logarítmica permitindo a melhor visualização dos pontos. Em c) k versus CIMC e em d) CCA versus CIMC.



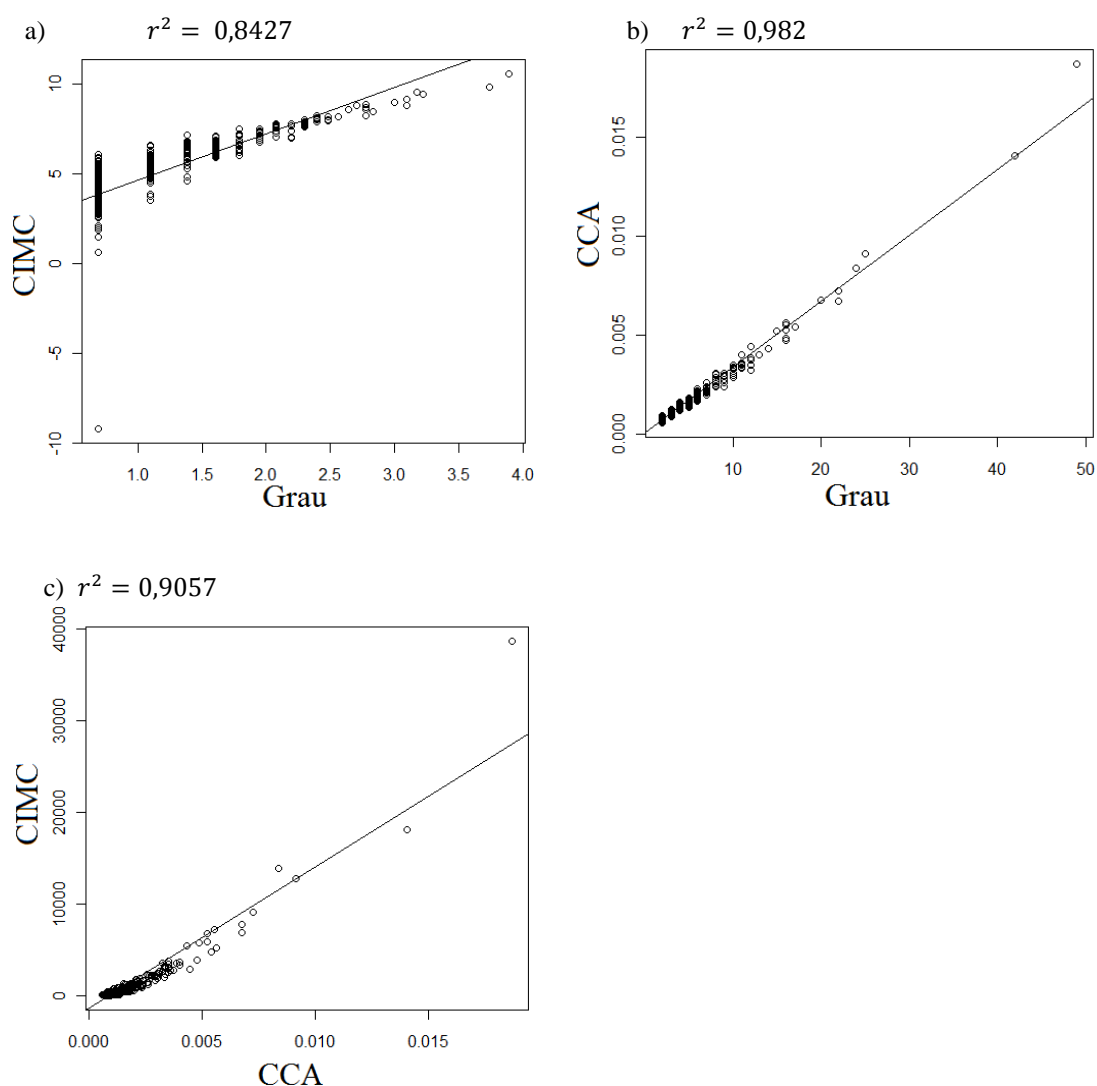
Fonte: Próprio Autor.

As Fig. 5.4 (c) e Fig. 5.4 (d) apresentam as outras duas comparações, sendo estas a relação entre o grau e a CIMC e a relação entre CIMC e a CCA, respectivamente. Nos dois casos observam-se comportamentos semelhantes. O crescimento linear entre as grandezas relacionadas também é verificado para a rede Hierárquica.

As redes de Barabási-Albert mostram que o comportamento das medidas de centralidade continua a ter um caráter crescente em suas relações, conforme Fig. 5.5.

Como esta rede e a rede Hierárquica são dois modelos de rede Sem Escala, espera-se que o comportamento entre estas seja semelhante. Porém, como cada uma delas utilizam princípios particulares para a geração, os resultados, apesar de serem semelhantes, também possuem algumas especificidades próprias.

Figura 5.5 Relação entre as medidas de centralidade para as redes Sem Escala, modelo de Barabási-Albert para $N = 500$ e os nós cresceram com o parâmetro $m = 2$. Em a) k versus CIMC, escala logarítmica, b) k versus CCA, em c) CCA versus CIMC. Cada imagem apresenta o respectivo valor de correlação referente ao ajuste.



Fonte: Próprio Autor.

A rede Barabási-Albert é uma rede com caráter probabilístico para suas conexões, e isso acarretam uma gama muito maior de valores para os graus dos vértices,

diferentemente da rede Hierárquica. Isso faz com que existam também diferentes valores para CCA que está diretamente relacionada com o grau dos vértices, e também para CIMC, visto que os menores caminhos entre os vértices são também influenciados pelas medidas do grau. É importante observar que para este modelo de rede o índice do nó também pode ser interpretado como uma medida de centralidade para a rede, pois, os nós mais antigos possuem maior probabilidade de receberem novos nós, tornando-os mais conectados e consequentemente, mais centrais dentro da rede.

O gráfico apresentado pela Fig. 5.5 (a) exhibe a relação entre o grau e a CIMC; é apresentado na escala logarítmica para facilitar a visualização do comportamento crescente, que é linear conforme o ajuste. A rede de Barabási-Albert apresenta um grande número de vértices com baixa conectividade e muitos desses vértices podem assumir o mesmo valor para o grau. Percebe-se então que há um acúmulo de nós possuindo diferentes valores de CIMC para um mesmo valor de k na região que abrange os valores menores de k .

Para a relação CCA x grau a Fig. 5.5 (b) mostra que o comportamento entre essas duas grandezas é aproximadamente proporcional. No entanto, a CCA não é monotonicamente crescente com o grau, devido à flutuação de τ , ilustrado anteriormente na Fig. 5.1 (d). O valor de τ para esta rede varia entre 1 e 2, apresentando uma flutuação no comportamento desta relação. Entretanto, como os valores assumidos por τ estão dentro de uma faixa de valores pequenos e relativamente próximos de 1, esta flutuação também é pequena.

Mediante as Fig. 5.5 (a) e Fig. 5.5 (b) é possível verificar a grande influência causada pelo grau com relação à CCA e CIMC, (note que em ambos os gráficos há um aglomerado de pontos na região com baixo grau e para poucos nós altamente conectados há também uma alta CCA e alta CIMC). A Fig. 5.5 (c) apresenta a relação entre as grandezas CCA x CIMC e como nas demais redes apresentadas, esta relação possui um caráter crescente.

5.4 Eficiência das diferentes medidas de Centralidade

As medidas de Centralidade são úteis quando se compara os movimentos do caminho aleatório entre dois nós. Por exemplo, sejam os nós de índices i e j com a centralidade de i maior que a centralidade de j . Em média, um caminhante aleatório

começando em j chega até i mais rapidamente que o segundo caminhante, que inicia o movimento em i até chegar a j . Em outras palavras, o nó que possui maior centralidade possui uma probabilidade maior de ser visitado pelo caminhante aleatório mais rapidamente que o que possui menor valor para a centralidade [14]. Neste sentido o caminhante aleatório pode ser interpretado como um mensageiro de informações, em que nós com maiores Centralidades são mais eficientes em receber informações do que nós com valores menores para a Centralidade.

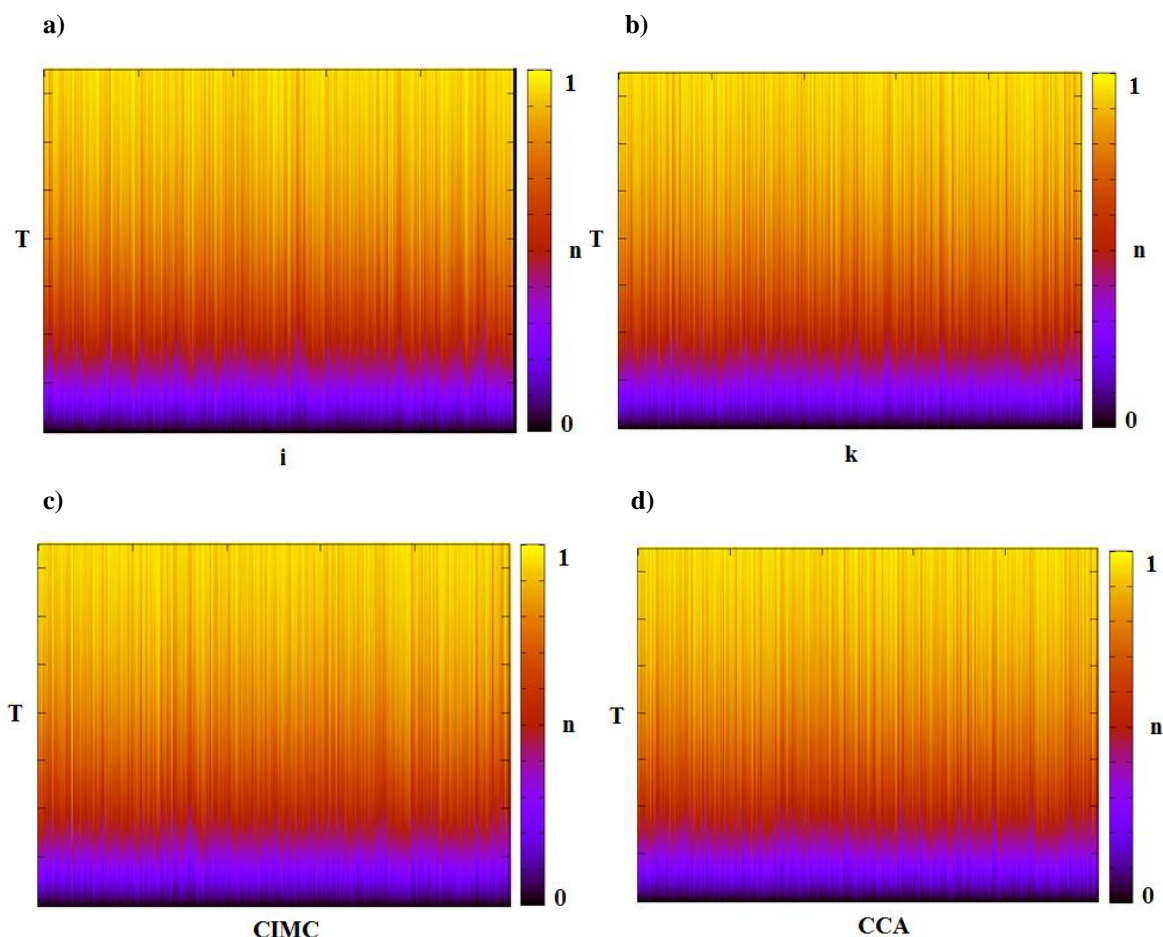
Dessa forma, foram realizadas simulações numéricas para estudar a relação entre a eficiência das medidas de Centralidade para redes complexas distintas. Para realizar esta quantificação, foi considerada uma situação onde inicialmente todos os nós em uma rede estavam ocupados por diferentes caminhantes aleatórios (Anexo - A). Eles iniciam o movimento da caminhada no tempo $t = 0$, e é medido o valor da fração de caminhantes, n_i , que passaram pelo nó i , em função do tempo t .

Assumimos que não houve interações entre os caminhantes. Consideramos que estes caminhantes são mensageiros de informações e que entregam informação a cada nó que visitam. Assim, quanto maior o número de caminhantes que passam por um nó, maior é o número de informações recebidas por este nó. Para quantificar o número de caminhantes que atravessam o vértice i ao longo do tempo, foi utilizada a medida n_i . A informação é proporcional ao número de caminhantes que passam por um determinado nó e a cada instante de tempo a informação é acumulada. De igual forma, a medida de n_i também é cumulativa, sendo que a cada iteração é considerada a fração de caminhantes que passam pelo vértice i no instante atual, juntamente com as frações de caminhantes que já passaram por este nó nos instantes anteriores. Ao final do processo da caminhada aleatória a fração de caminhantes assume valores entre $(0,1]$. Quando $n_i = 1$, significa que todos os caminhantes distribuídos inicialmente passaram por i . Inicialmente os caminhantes aleatórios são distribuídos uniformemente pela rede e n_i é proporcional à quantidade de caminhantes que passam por cada nó. Os nós com maiores valores de Centralidade terão maior probabilidade de uma fração n_i maior para qualquer tempo.

Para realizar as simulações foram considerados novamente, os quatro modelos de redes: rede Aleatória, rede de Pequeno Mundo, rede Hierárquica e rede Sem Escala de Barabási-Albert, onde foram medidos os valores de n_i a cada nova iteração. A Fig. 5.6 apresenta o resultado da caminhada aleatória nas redes complexas ao longo do tempo. A organização desta caminhada está disposta de quatro formas diferentes para cada rede. Na

primeira imagem, os vértices da rede são organizados em ordem crescente com relação ao índice i de cada nó e, como o índice não se caracteriza por ser uma medida de centralidade (com exceção da Rede Sem Escala em que pode ser indicado como um parâmetro de centralidade devido à forma de construção de rede), as imagens da evolução temporal organizada a partir do índice geralmente não apresentam um padrão de organização. Na segunda imagem os nós são organizados em ordem decrescente do grau k ; na terceira imagem a fração de caminhantes para cada nó está disposta em ordem decrescente da CIMC, e por fim, a última imagem está organizada na forma decrescente de CCA.

Figura 5.6 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede Aleatória, com $N = 500$ e $T = 2000$. O valor de n é representado na escala de cor a direita de cada gráfico.



Fonte: Próprio Autor.

A escala de cor localizada à direita de cada imagem mostra a fração de caminhantes para cada nó. O valor de n_i está indicado nas extremidades da escala de cores, seu valor se restringe ao intervalo $(0,1]$. Inicialmente todos os nós estão ocupados por um caminhante,

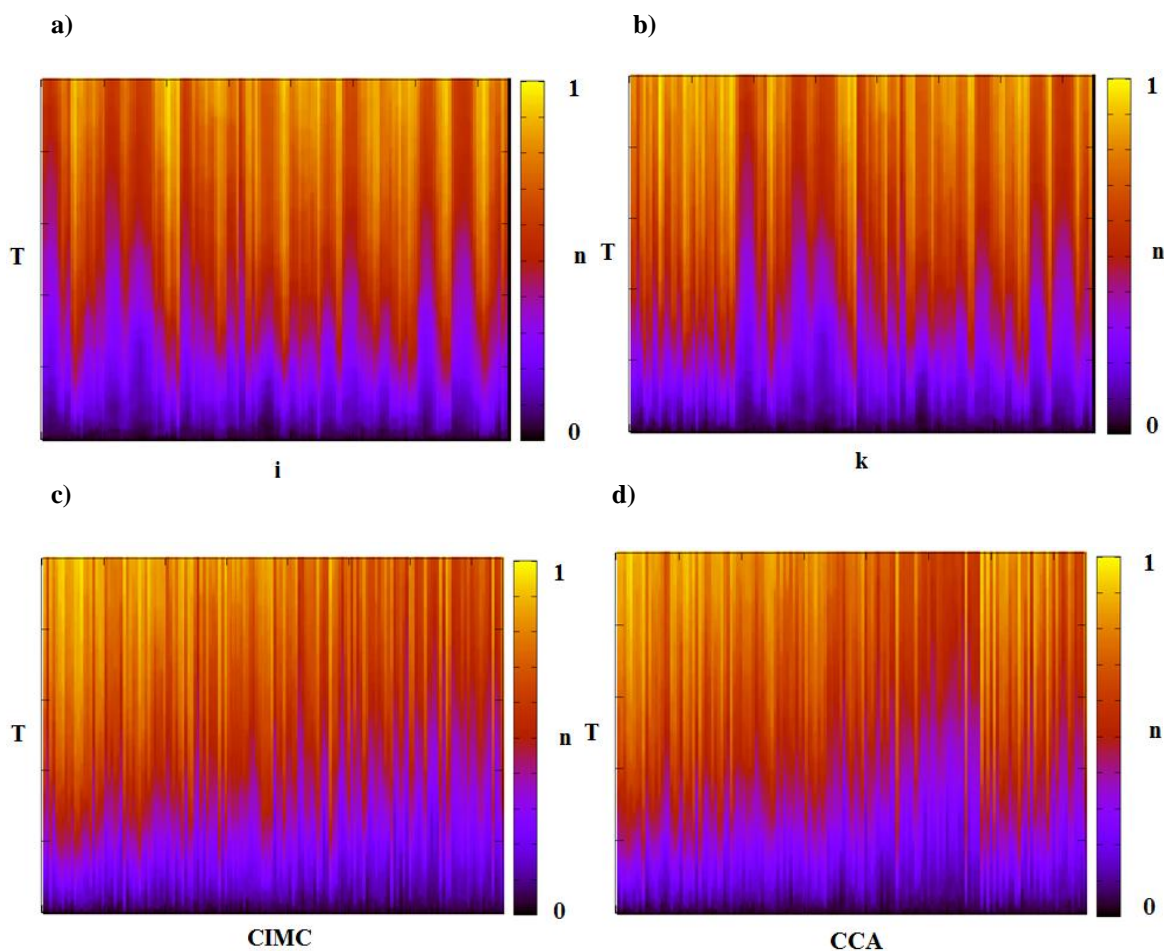
se a rede em questão possuir 100 nós, por exemplo, a fração n_i será de 1/100. Ou seja, a fração de caminhantes no início da evolução temporal é muito pequena, próxima de zero. A cada iteração o caminhante deixa o nó onde está e vai para outros nós. Intuitivamente os nós com maior centralidade receberão mais caminhantes ao mesmo tempo, enquanto outros podem não receber caminhantes em alguma iteração.

As simulações para a rede Aleatória, Fig. 5.6 (a), indicam que os vértices, de maneira geral, são visitados por todos os caminhantes de maneira uniforme, e que a fração de caminhantes em todos os vértices após o tempo $T = 2000$ é 1 ou muito próximo desse valor, indicando que por esses nós passaram quase que a totalidade de caminhantes dispostos na rede. Observa-se ainda na Fig. 5.6 que não é possível destacar algum vértice de maior centralidade. Em razão disso, também não é possível destacar qual medida de centralidade possui maior eficiência na difusão de informação realizada através da caminhada aleatória. Para todas as formas de disposição feitas pelas diferentes medidas de centralidade Fig. 5.6 (b, c, d), se observa o mesmo comportamento para a caminhada aleatória, não havendo diferenças significativas se o vértice possui o maior ou menor grau para a dispersão dos caminhantes, sendo da mesma forma com relação às demais medidas de centralidade.

Assim, para redes aleatórias a difusão de caminhantes possui um caráter praticamente monótono. As razões para esse comportamento decorrem dos valores do grau dos vértices da rede pouco se distanciarem do valor do grau médio. Isso faz com que as medidas de CCA para essa rede, as quais são proporcionais ao grau, também não se distanciem da média, assumindo valores muito próximos. O mesmo ocorre para a CIMC, visto que a rede aleatória tem o comprimento de caminho médio baixo e a CIMC é definida a partir da contagem dos caminhos mais curtos que conectam dois nós quaisquer.

Analisando as simulações para a rede de Pequeno Mundo, Fig. 5.7, é possível perceber que o comportamento com relação à rede aleatória é diferente em alguns aspectos e semelhante em outros. Para a rede de Pequeno Mundo a distribuição de grau possui caráter semelhante a da rede Aleatória, a maioria dos vértices permanece com seus valores de grau próximos do valor do grau médio da rede. Isto acarreta que a medida de Centralidade de Grau, Fig. 5.7 (b), não exerça uma influência na disseminação dos caminhantes, pois a diferença entre os mesmos é muito pequena devido a seus valores serem próximos da média.

Figura 5.7 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede de Pequeno Mundo, com $N = 200$ e $T = 1000$. O valor de n é representado na escala de cores à direita de cada gráfico.



Fonte: Próprio Autor.

A evolução temporal dos caminhantes aleatórios para esta rede possui características semelhantes, quanto ao índice e ao grau, apresentada pela rede Aleatória. Apesar de que, para rede de Pequeno Mundo seja possível destacar alguns nós da rede, os quais apresentam uma maior fração de caminhantes, formando regiões mais claras no gráfico. Porém, estes nós que formando estas regiões quando organizados em ordem decrescente do grau, não apresentam um comportamento uniforme. Uma vez que, existem alguns nós com valores baixos para o grau e com fluxo maior de caminhantes que outros que possuem valores maiores para o grau apresentando fluxo baixo de caminhantes. Assim, o grau não se caracteriza em uma medida eficiente para classificar os nós com maior centralidade.

Nas Fig. 5.3 (a) e Fig. 5.3 (b) é possível notar uma grande variação entre os valores de CCA e CIMC com relação ao grau k . Estes gráficos mostram que estas duas últimas

imagens, Fig. 5.7 (c, d) seriam diferentes da organização feita a partir do grau. A medida de CCA depende do grau dos nós e do tempo de relaxação, e para a rede de Pequeno Mundo, τ possui o maior peso assumindo valores mais altos e em um intervalo grande em comparação com as demais redes, causando uma maior diferença no comportamento da evolução dos caminhantes aleatórios, Fig. 5.7, organizada a partir destas duas medidas (CCA e grau). A medida de CIMC não está relacionada com τ , mas a diferença apresentada entre esta e o grau está relacionada com os atalhos criados na rede no processo de geração de Watts-Strogatz para a rede de Pequeno Mundo. Assim, um nó pode possuir um grau baixo, mas se este servir de atalho para outros vértices terá um alto valor para CIMC. Isso ocorre porque, os vértices que estabelecem atalhos fazem com que o comprimento de caminho médio diminua, e conseqüentemente, diminuindo a distância entre os nós através de caminhos mais curtos que passam por eles.

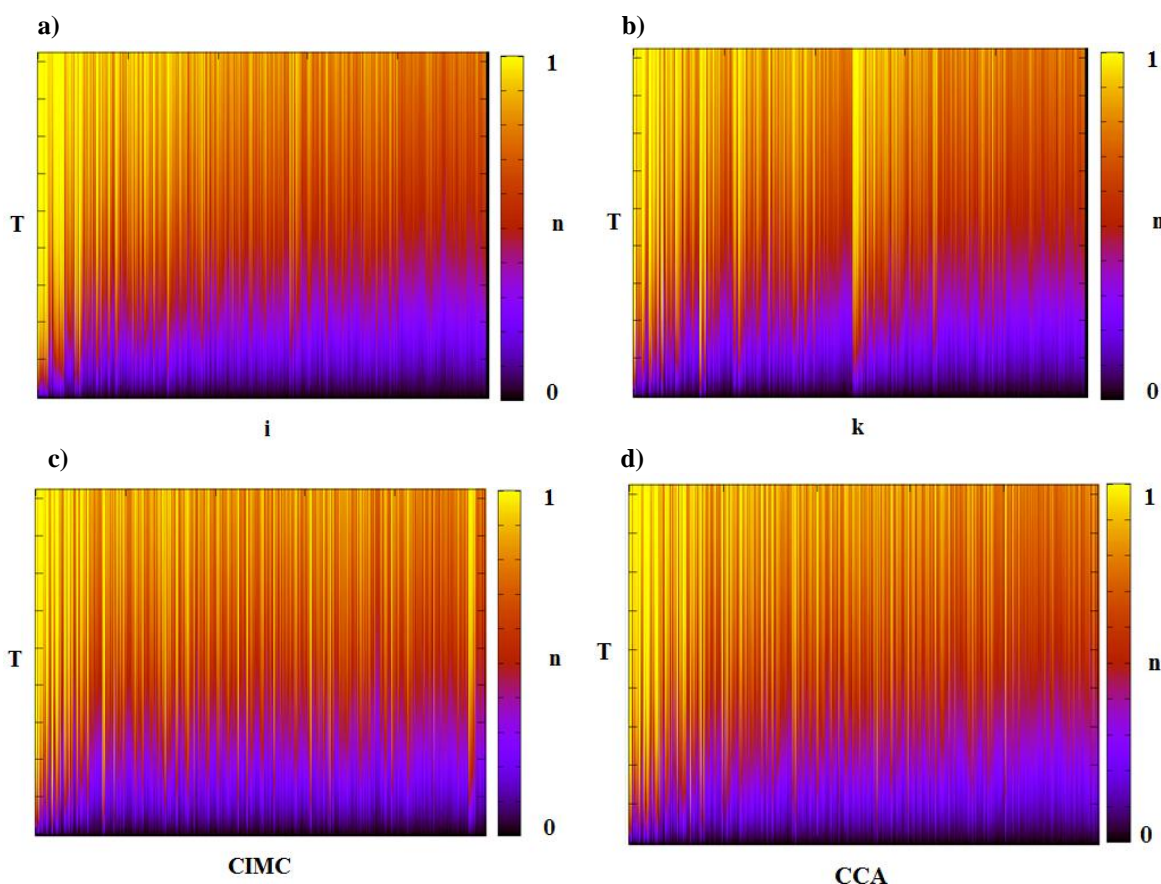
Portanto, para as imagens da Fig. 5.7 (c, d), que estão organizados em relação à CIMC e CCA, é possível destacar um comportamento crescente para a evolução temporal dos caminhantes aleatórios. É importante notar ainda na Fig. 5.7 que, para os nós com valores maiores de centralidade, há também um maior número de caminhantes aleatórios percorrendo-os, como é mostrado por meio das regiões mais claras que ocorrem no canto superior esquerdo do gráfico. Isto significa, que quando organizamos a fração de caminhantes por nó a partir dos valores decrescentes de CCA e de CIMC podemos visualizar a eficiência destas medidas, pois quanto maior é a CCA de um nó, mais rapidamente os caminhantes passam por esse nó. O mesmo é válido para a CIMC.

Analisando as simulações para rede Sem Escala de Barabási-Albert na Fig. 5.8, é possível perceber que todas as medidas de Centralidade apresentam um comportamento crescente. Para os nós com valores maiores de centralidades, (com exceção do índice i que está organizada de maneira contrária as demais, sendo disposta de forma crescente), há também um maior número de caminhantes aleatórios percorrendo-os. Estes formam na imagem regiões mais claras, em que os vértices são visitados rapidamente por um maior número de caminhantes.

A Fig. 5.8 (a) refere-se à simulação para a rede Sem Escala. Como visto anteriormente, para o modelo de rede de Barabási-Albert o valor i do índice também pode ser visto como uma medida de centralidade. Pois, o índice do nó corresponde ao tempo de iteração, durante os quais novos nós são adicionados à rede. Quanto menor é o índice do vértice mais antigo ele é dentro da rede e maior é a probabilidade de este receber ligações

com novos vértices. Assim, os nós mais velhos possuem maior número de conexões e por isso, possuem maior possibilidade de serem mais centrais da rede.

Figura 5.8 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede Barabasi-Albert, com $m = 2$, $N = 500$ e $T = 2000$. O valor de n é representado na escala de cores à direita de cada gráfico.



Fonte: Próprio Autor.

A Fig. 5.8 (b) referente ao grau k , mostra que a fração n é disposta com algumas irregularidades, apresentando um comportamento crescente, porém não monótono [9]. Para esta rede a centralidade de grau se mostra mais eficiente que para as redes de Pequeno Mundo e Aleatória, isso porque diferentemente das anteriores, a distribuição de grau da rede Sem Escala, se caracteriza por uma Lei de Potência. Assim, os valores para os graus apresentam uma grande diferença com relação à média, havendo poucos vértices que possuam alto valor para o grau e muitos outros vértices com valores baixos para o grau.

Este fato influencia diretamente também nas demais medidas de centralidade, pois a CCA depende principalmente do grau, sendo proporcional a este, sobretudo porque o τ para a rede de Barabasi-Albert é pequeno, como pode ser visto na Fig. 5.1 (d), não

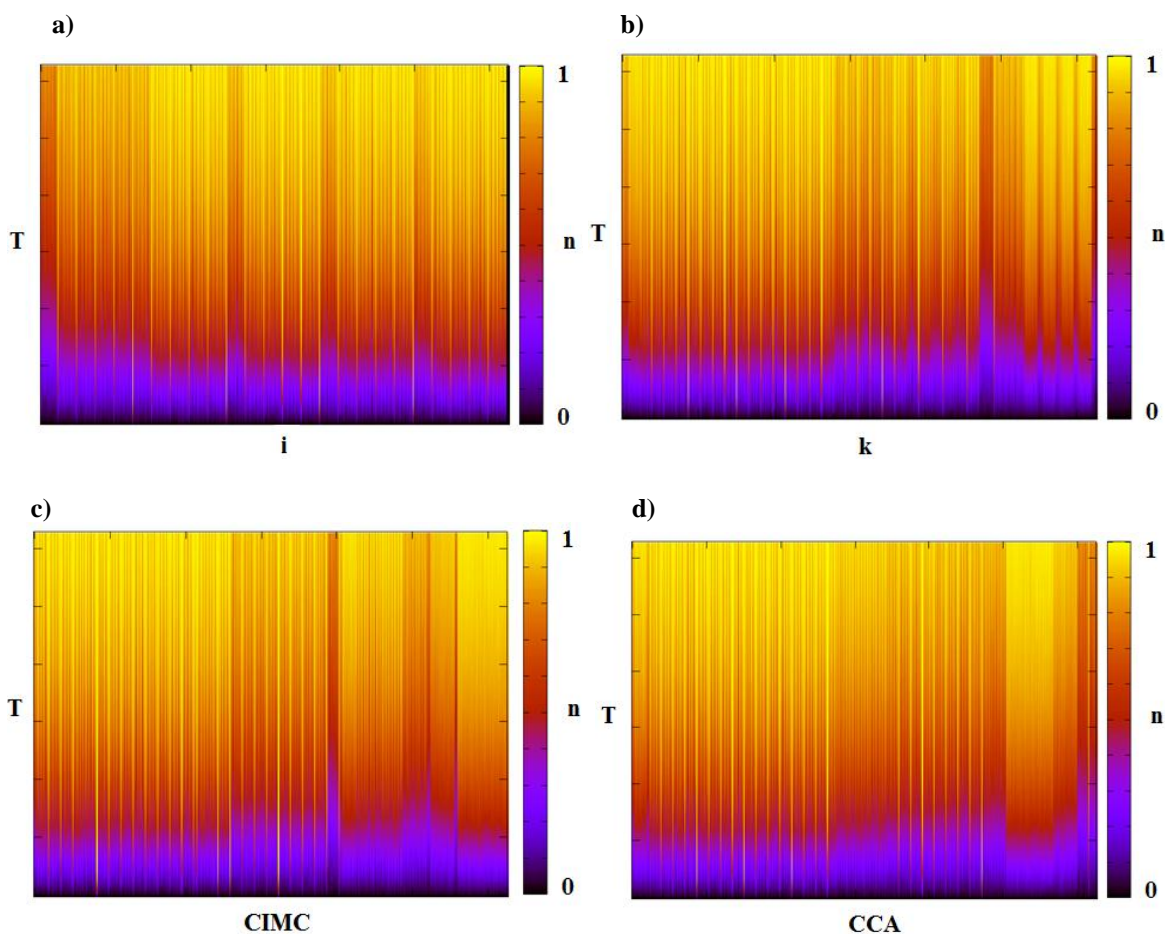
causando interferências significativas nesta medida. A CIMC também sofre influências da forma de distribuição de grau da rede de Barabási-Albert. Os nós com muitas conexões são às vezes a única ponte entre os nós que estão à margem da rede com poucas conexões para demais nós da rede, fazendo com que muitos caminhos mais curtos passem pelos nós fortemente conectados.

As imagens das Fig. 5.8 (c) e Fig. 5.8 (d) apresentam o comportamento da evolução temporal, comprovando as semelhanças entre as formas de organização desta evolução para a rede de Barabási-Albert, principalmente o caráter crescente, cujos nós com maiores centralidades possuem maior fração de caminhantes, mostrando assim a eficiência destas duas últimas medidas. A imagem com relação à CIMC torna-se mais suave se comparada com a organização feita através do grau, mas ainda as tendências não monotônica permanecem. No entanto, em função do CCA, a imagem torna-se muito mais suave e quase monótona.

A rede Hierárquica, por ser também uma rede Sem Escala, possui um comportamento semelhante à rede de Barabási-Albert, exceto com relação ao índice, justamente porque a regra da construção da rede hierárquica possui características diferentes, conforme abordado no Capítulo 3. Assim, a sua evolução temporal possui as mesmas características abordadas anteriormente para a rede de Barabási-Albert, principalmente porque a distribuição de grau continua sendo em Lei de Potência. A primeira imagem referente à Fig. 5.9 (a) mostra a existência de franjas mais claras e mais escuras que se repetem periodicamente, devido às características em sua geração que é realizada por cópias das iterações anteriores.

A Fig. 5.9 (b) apresenta a disposição dos nós da rede de acordo com o grau. É possível perceber o comportamento crescente indicando as regiões mais claras com maior número de caminhantes aleatórios para os nós com grau alto. Análogo à rede de Barabási-Albert, existem nós com muitas conexões e outros com valores baixos para o grau. Dessa forma, a centralidade de grau é uma medida que se mostra eficiente, apesar de haver nós com baixo número de ligações com a fração de caminhantes alta. Isso ocorre por haver nós, com baixa conexão, ligados diretamente com nós com alto grau (principalmente com o nó raiz). Dessa forma, estes nós de baixa conexão possuem uma centralidade que não é captada pelas medidas de centralidade; é como se houvesse uma centralidade por associação, já que estão diretamente ligados a nós importantes. Esta característica persiste para as demais centralidades, devido à relação existente com o grau.

Figura 5.9 Evolução temporal da fração de caminhantes n que passam por um nó como uma função do índice do nó i , do grau k , da CIMC e da CCA na rede Hierarquica, com $m = 5$, $N = 512$ e $T = 2560$. O valor de n é representado na escala de cores à direita de cada gráfico.



Fonte: Próprio Autor.

As medidas de CIMC e CCA são influenciadas pelo grau de maneira análoga à rede anterior, e possuem as mesmas características com relação ao tempo de relaxação e em relação ao maior número de menores caminhos que passam pelos nós com maior grau. Assim, as duas últimas imagens, Fig. 5.9 (c) e Fig. 5.9 (d) exibem a fração de caminhantes em função da CIMC e da CCA, respectivamente. Percebe-se um refinamento melhor na terceira imagem em comparação a segunda organizada a partir do grau. A última imagem apresenta a organização da fração de caminhantes dada por meio da CCA, se comparada com as imagens anteriores, verifica-se que a distribuição apresenta um aspecto crescente e mais suave que as demais. Portanto, a fração de caminhantes é maior para os nós que possuem valores maiores para a CCA. A evolução temporal de n_i organizada em função de CCA é distribuída de forma suave, praticamente monótona para a rede Hierárquica, indicando que a CCA é a medida de centralidade mais eficiente para esta rede.

CAPÍTULO 6

Conclusões e Trabalhos Futuros

A fundamentação teórica apresentada neste trabalho tratou (Capítulo 1), sobre processo estocástico da Caminhada Aleatória em Redes Complexas, sua importância para a determinação da medida de Centralidade da Caminhada Aleatória, além de sua utilidade para a estimativa do fluxo de caminhantes por vértice da rede, sendo este, um indicador de maior ou menor centralidade. A caminhada aleatória desempenha papel fundamental por suas aplicações no campo da percolação e difusão de informação. A pesquisa abordou a teoria de grafos (Capítulo 2), sua importância para a representação das redes complexas e para o desenvolvimento de estudos nesta área. As redes complexas, suas características específicas e propriedades também foram abordadas (Capítulo 3). Estas podem ser classificadas de acordo com suas características estruturais e modo de geração em modelos distintos de rede. Nas redes complexas é possível calcular algumas medidas importantes, como grau dos vértices, coeficiente de agrupamento, comprimento de caminho médio, centralidade dos vértices, entre outras. A centralidade dos vértices (Capítulo 4) pode ser calculada de formas variadas e sua função é verificar a importância desempenhada por cada nó. Quatro modelos de redes foram escolhidos para se realizar as simulações (Capítulo 5), as redes Aleatória, Pequeno Mundo, Sem Escala e Hierárquica. Nestas redes foram calculadas três diferentes medidas de Centralidade: o grau, a CIMC e a CCA.

No intuito de responder a problemática levantada na pesquisa, foram realizadas simulações numéricas para a caminhada aleatória em redes, sendo este processo fundamental para inferir a eficiência das medidas centralidades calculadas nas mesmas. Podemos perceber que para os diferentes modelos de redes analisados, a rede aleatória é a única rede em que as medidas de centralidade não destacam nenhuma característica relevante. Isso porque os valores do grau dos vértices da rede pouco se afastam do valor do grau médio, fazendo com que as medidas de CCA para essa rede também não se distanciem da média, uma vez que são proporcionais ao grau. Da mesma forma, ocorre para a CIMC, visto que a rede aleatória tem o comprimento de caminho médio baixo e a CIMC é definida a partir da contagem dos caminhos mais curtos que conectam dois nós quaisquer da rede. Assim nenhuma medida de centralidade inferida para esta rede se

mostrou eficiente em quantificar a verdadeira importância de cada nó presente na rede. Para as demais redes, as medidas de Centralidade abordadas, mostraram-se eficientes para detectar os nós com maior fluxo de caminhantes, sendo que, a que se mostrou com maior eficiência foi a CCA.

Para a medida de CCA, a evolução temporal de caminhantes aleatórios se apresentou de forma suave e sem muitas irregularidades, mostrando que para nós com maior valor de CCA exibem também maior fluxo de caminhantes. A medida de CCA consegue apontar a importância dos nós com uma maior precisão que as demais medidas de centralidade, como verificado mediante as simulações numéricas. A informação quando distribuída sobre uma rede real, de forma a sair de um vértice i até um vértice j , na maioria das situações, não conhece qual caminho percorrer e tão pouco sabe qual o menor caminho a ser tomado para alcançar o seu destino, assim ela percorre aleatoriamente sobre as redes até atingir seu destino final, o que torna a CCA mais próxima das situações reais, pois ela é determinada a partir da equação mestra que descreve o processo de percursos aleatórios.

Os objetivos do presente trabalho foram atingidos, tendo em vista que foi possível realizar a discussão das diferentes abordagens apresentadas para o processo da caminhada aleatória. Além de verificar que este processo interfere diretamente para o cálculo da CCA. Ainda foi possível adotar caminhada aleatória como um mecanismo de distribuição de informação, em que este pode ser utilizado como parâmetro para interpretação de dados e construções de imagens. Neste sentido, o processo da caminhada aleatória permitiu comparar as diferentes de medidas de centralidade verificando a eficiência de cada uma dela para as diferentes modelos de rede.

Dentre os movimentos do percurso aleatório entre dois nós, a caminhada para o nó com valor maior para as medidas de Centralidade, acontece mais rapidamente que para outros nós com menores valores para essas medidas. Além disso, argumentamos que em um dado intervalo de tempo, nós com valores maiores de Centralidade são visitados por mais caminhantes aleatórios que foram distribuídos inicialmente de maneira uniforme, confirmado pelas simulações numéricas nas redes de Pequeno mundo, rede Sem Escala e rede Hierárquica.

Podemos considerar os caminhantes aleatórios como mensageiros de informações através da rede. Nossos resultados implicam que a informação não se distribui uniformemente em redes heterogêneas; a informação é centralizada para nós com valores maiores de Centralidade. Estes nós possuem a vantagem de conhecer as novas informações

antes que os outros nós com menor centralidade, a desvantagem apresentada em redes reais e com capacidade finita é que os com maior centralidade são muito carregados, devido ao acúmulo de caminhantes, isto pode ocasionar congestionamentos [12]. Portanto, deve-se tomar cuidado no gerenciamento da rede, de forma, a prevenir congestionamento nos nós com fluxo muito grande de caminhantes.

Um exemplo de rede Pequeno Mundo em que é importante conhecer os nós com maior centralidade são as redes sociais. Em uma situação, que se fizesse necessário transmitir uma mensagem para uma rede de pessoas, na qual todos os integrantes dessa rede devem conhecer a informação, sendo está espalhada por meio de alguns indivíduos escolhidos nesta rede. Neste caso, as medidas de Centralidade se tornam importantes, pois a escolha das pessoas, as quais começarão a disseminação de informação pode ser feita pelas que apresentam maiores centralidade. Uma vez que, por estas passam maior número de informação e permite que a difusão se estabeleça em um tempo menor.

A partir das medidas de centralidade dos nós é possível compreender aspectos relevantes das redes reais, tais como as redes de transportes, em que os nós podem ser visto como estradas, sendo que as maiores centralidade podem ser sobrecarregados e gerando congestionamentos. As redes sociais, em que pessoas com maior centralidade podem contribuir para casos epidêmicos. No campo comercial e no *marketing*, a centralidade também apresenta relevância, colaborando para o levantamento de locais com maior fluxo de pessoas, sendo estes favoráveis para divulgação.

Como trabalhos futuros, pretende-se modelar redes reais como as mencionadas, aplicar o processo de caminhada aleatória e calcular os valores de centralidade para os vértices da rede, identificando os vértices que apresentam maior centralidade. O presente trabalho considera apenas, a caminhada aleatória para redes não direcionadas, ficando outra oportunidade a generalização para redes dirigidas, sendo que tais redes modelam inúmeras situações reais. As medidas de centralidade abordadas na presente pesquisa abrangem apenas os vértices da rede, o estudo de medidas de centralidade para as ligações entre os nós é uma questão relevante, sendo que a CIMC para as arestas já foram estudadas em [52]. A generalização para a centralidade por meio da caminhada aleatória para as arestas, de forma a encontrar uma expressão, como foi realizada no caso da CCA, seria também interessante.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BONDY, A.; MURTY, U. **Graph Theory**, Nova Iorque, Springer, 2007.
- [2] ERDOS, P.; RÉNYI, A.; On random graphs, **Publicationes Mathematicae** 6, 290–297, 1959.
- [3] BONANNO, G.; CALDARELLI, F.; LILLO, S.; MICCICHÈ, N.; VANDEWALLE, R. N. MANTEGNA; Networks of equities in financial markets, **European Physical Journal B**, 38, 363–371, 2004.
- [4] WATTS, D. J.; STROGATZ, I.M. Collective dynamics of 'small-world' networks, **Nature**, 393, 440-442, 1998.
- [5] BARABÁSI, A. L. **Linked: How everything is connected to everything else and what it means for business, science and everyday life**, Cambridge, Plume, 2003.
- [6] ALBERT, R.; JEONG, H.; BARABÁSI, A.-L.; Attack and error tolerance of complex networks, **Nature**, 406, 378–382, 2000.
- [7] BARABÁSI, A.-L.; ALBERT, R.; Emergence of scaling in random networks. **Science**, 286 (5439), 509-512, 1999.
- [8] DOROGOVTSSEV, S. N.; GOLTSET, A. V.; MENDES, J. F.; Pseudofractal scale-free web, **Physical Review E**, 65, 066122, 2002.
- [9] NEWMAN, M. E. J. Structure and function of complex networks, **SIAM Review**, 45, 167-256, 2002.
- [10] HUGHES, R. D. **Random Walks: Random Walks and Random Environments**, Oxford, Clarendon, 1995.
- [11] JESPERSEN, S.; SOKOLOV, I.M. e BLUMEN, A. Blumen, Phys. Relaxation properties of small-world networks. **Physical Review E**, 62, 4405, 2000.
- [12] HOLME, P. Congestion and centrality in traffic flow on complex networks, **cond-mat**, 0301013, 2003.

- [13] SANTOS, M.C. **Busca Markoviana Super Difusiva: Origens e Relevância das Leis de Potência**. Dissertação de Mestrado, Universidade Federal do Paraná, 2008.
- [14] NOH, J. D.; RIEGER, H.; Random Walks on Complex Networks, **Phys. Rev. Lett.** 92, 118701, 2004.
- [15] NEWMAN, M. E. J.; A measure of Betweenness centrality based on random walks, **Social Networks**, 39-54, 2005.
- [16] SILVA, J.M.; LIMA, J.A.S. Quatro abordagens para o movimento browniano. **Revista Brasileira Ensino Física**, 25-35, 2007.
- [17] FRIEDLI, S. **Dinâmica Estocástica: três exemplos**. Universidade Federal de Minas Gerais, 2011.
- [18] FELLER, W. **An Introduction to Probability Theory and Its Applications**. Princeton, John Wiley Computer, 1971.
- [19] TOMÉ, T.; OLIVEIRA, M. J. **Dinâmica Estocástica e Irreversibilidade**. São Paulo, Edusp, 2001.
- [20] SPIEGEL, M. R.; SCHILLER, J.; SRINIVASAN, R. A. **Probabilidade e Estatística**. São Paulo, Makron Books do Brasil, 1978.
- [21] SANTOS, R. J. **Cadeia de Markov**, Imprensa Universitária, Universidade Federal de Minas Gerais, 2006.
- [22] BURIONI, R.; CASSI, D. Random walks on graphs: ideas, techniques and results. **Journal of Physics A: Mathematical and General**, 45-78, 2005.
- [23] ANTON, H.; RORRES, C.; **Elementary Linear Algebra**, Applications Version, John Wiley & Sons, 2010.
- [24] BURIONI, R.; CASSI, D.; VEZZANI, A. Random walks and physical models on infinite graphs: an introduction **Random Walks and Geometry**. V Kaimanovich, K Schmidt and W. Woess, Berlin, 2004

- [25] ARAUJO, B.M., **Identificação de Outlier em Redes Complexas baseadas em Caminhada Aleatória**. Dissertação de Mestrado, Universidade de São Paulo, 2010.
- [26] NEWMAN, J. Leonhard Euler and the Königsberg Bridges. **Scientific American**.66-70, 1953.
- [27] CARDOSO, D. M., **Teoria de Grafos e Aplicações**. Dissertação de Mestrado, 2005.
- [28] PICADO, J. **Teoria de Grafos**. Universidade de Coimbra, 2010.
- [29] GOLDBARG, M. C.; LUNA, H. P. L. **Otimização Combinatória e Programação Linear**. Rio de Janeiro, Editora Campus, 2005.
- [30] FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**. Universidade de São Paulo, 2011.
- [31] BALAKRISHNAN, V. K. **Graph Theory**. Schaum's Outlines. 1997.
- [32] NETTO, P. O. B. **Grafos**. Edgard Blücher. 2001.
- [33] WOESS, W. **Random Walks on Infinite Graphs and Groups**. Cambridge University Press, Cambridge, 2000.
- [34] BOLLOBÁS, B.; RIORDAN, O.M. Mathematical results on scale-free random graphs, in **Handbook of Graphs and Networks: From the Genome to the Internet**, Wiley-VCH, Berlin, 2002.
- [35] COHEN, H.; HAVLIN, S. **Complex Networks Structure, Robustness and Function**. Cambridge, Cambridge University Press, 2010.
- [36] LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E.; A° BERG, Y. The web of human sexual contacts, **Nature**, 907–908, 2001.
- [37] NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of National Academy Sciences**, 404-409, 2001.

- [38] MILGRAM, S.; The small world problem, *Psychology Today*, **Sociometry**, 60-67, 1967.
- [39] HE, Y.; CHEN, Z. J.; EVANS, A. C.; Small-world anatomical networks in the human brain revealed by cortical thickness from MRI, **Cerebral Cortex**, 2407–2419, 2007.
- [40] MANCARI, D.; VARETTO, G.; BUCCI, E.; MANIERO, F.; GUIOT, C.; Fractal parameters and vascular networks: facts and artifacts, **Theoretical Biology and Medical Modelling**, 1-12, 2004.
- [41] JEONG, H., NÉDA, Z.; BARABÁSI, A.-L.; Measuring preferential attachment for evolving networks, **cond-mat/0104131**, 2001.
- [42] KRAPIVSKY, P. L.; REDNER, S.; LEYVRAZ, F.; Connectivity of growing random networks, **Physical Review Letters**, 85, 4629, 2000.
- [43] RAVAZ, E.; BARABÁSI, A.-L.; Hierarchical organization in complex networks, **Physical Review E**, 67, 026112, 2003.
- [44] RAVASZ, E.; SOMERA, A. L.; MONGRU, D. A.; OLTVAI, Z. N.; BARABÁSI, A.-L. Hierarchical organization of modularity in metabolic networks, **Science**, 1551, 2002;
- [45] BRANDES, U.; A faster algorithm for betweenness centrality. **Journal of Mathematical Sociology**, 163–177, 2001.
- [46] BEAUCHAMP, M. A.; An improved index of centrality, **Behavioral Science**, 161-163, 1965.
- [47] SHAW, M. E.; Advances in Experimental Social Psychology, **Communication networks**, In L. Berkowitz, 111-147, 1964.
- [48] BAVELAS, A., Communication patterns in task oriented groups, **Journal of the Acoustical Society of America**, 725-30, 1950.
- [49] SABIDUSSI, G.; The centrality index of a graph, **Psychometrika**, 581-603, 1966.

- [50] HAGE, P.; HARARY, F.; Eccentricity and centrality in networks, **Social Networks**, 57-63, 1995.
- [51] FREEMAN, L. C.; A Set of Measures of Centrality Based on Betweenness, **Sociometry**, 35-41, 1977.
- [52] BARRAT, A.; BARTHELEMY, M.; PASTOR-SARROTAS, R.; VESPIGNANI, A. The architecture of complex weighted networks. **Proceedings of the National Academy of Sciences of the USA**, 3747–3752, 2004.
- [53] NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical Review E**, 026113, 2004.
- [54] REDNER, S. How popular is your paper? An empirical study of the citation distribution, **European Physical Journal B**, 131–134, 1998.
- [55] NEWMAN, M. E. J.; Clustering and preferential attachment in growing networks, **Physical Review E**, 025102, 2001.
- [56] ADAMIC, L. A.; LUKOSE, R.M.; PUNIYANI, A. R.; HUBERMAN, B. A. Search in power-law networks, **Physical Review E**, 046135, 2001.
- [57] TORGO, L. **Introdução à Programação em R**, Universidade do Porto, 2006.

ANEXOS

A - Algoritmos e Simulações

As simulações e gráficos apresentados foram criados a partir do software R. O R é ao mesmo tempo uma linguagem de programação e um ambiente de desenvolvimento integrado, para cálculos estatísticos e gráficos [57]. Possui como característica ser um software livre, em que possui disponibilidade para vários sistemas operacionais. A vantagem na utilização do R é que este apresenta uma série de pacotes que são incluídos em sua instalação. Além, dos muito outros disponíveis na rede de distribuição do R, em que estes pacotes permitem a utilização de funções específicas para determinadas áreas de estudo. Para o estudo das redes complexas e teoria de grafos, o R disponibiliza uma biblioteca própria para criação e análise das mesmas. Assim os gráficos e as simulações apresentadas neste trabalho foram construídos mediante a utilização deste software.

Primeiramente foi realizada a simulação para a caminhada aleatória, a qual pode ser desenvolvida mediante um algoritmo simples, que gera números aleatórios, os quais são utilizados na determinação do próximo passo. Segue abaixo o algoritmo para a caminhada aleatória em uma dimensão sendo que as dimensões superiores podem ser realizadas de maneira análoga.

Caminhada em 1 dimensão:

Números de passos: N ;

Declarando as variáveis:

Passo do caminhante:

$$x_{\text{passos}} \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{"numeric"});$$

Distância percorrida:

$$D_x \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{"numeric"});$$

Vetor com números aleatórios:

$$x \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{numeric});$$

$soma \leftarrow 0$;

```

for (t in 1:N)
{
  x[t] ← runif(N);
  if (x[t] ≤ 0,5)
  {
    xpassos[t] ← 1;
  }
  else
  {
    xpassos[t] ← -1;
  }
Dx[t] ← somax + xpassos[t];
somax ← Dx[t];
}

```

Criando as Redes:

Para a geração das redes complexas foi utilizada a biblioteca *igraph*, este é um dos pacotes incluídos no R, é útil para criação e manipulação de grafos direcionados e não direcionados. Além disso ele permite a geração de modelos de redes conhecidos na literatura, incluindo implementações para problemas clássicos da teoria dos grafos.

```
library(igraph);
```

Rede Anel regular com números de nós iguais a N e ligações com seus k primeiros vizinhos:

```
g <- graph.lattice(N, nei = k, dim = 1, circular = T).
```

Rede de Pequeno Mundo segundo o modelo de Watts e Strogatz, com números de nós iguais a N , ligados inicialmente com seus k primeiros vizinhos, em que as arestas entre os nós são religadas com probabilidade p , para que aconteça o efeito pequeno mundo a probabilidade $0,01 \leq p \leq 0,1$, devido a este fato, escolheu-se para a simulação o valor mais utilizado na literatura, em que a probabilidade assume o valor $p = 0,1$;

$g < -watts.strogatz.game(size = N, nei = k, p = 0.1, dim = 1).$

Rede Aleatória, segundo o modelo proposto por Erdos e Renyi segundo, com números de nós iguais a N inicialmente desconectados, em que as arestas são inseridas na rede com probabilidade p ;

$g < -erdos.renyi.game(N, p).$

Rede Sem Escala, segundo o modelo de Barabasi e Albert, com números de nós iguais a N e m é o numero de novas arestas que são adicionadas na rede a cada iteração;

$g1 < -barabasi.game(N, power = 1, m).$

Cálculo das Medidas de Centralidade

As medidas de Centralidade de Grau e de Intermediação por menores caminhos de uma rede podem ser determinadas a partir do pacote *igraph*, o qual já traz algumas quantidades importantes para o estudo e análise das redes complexas. Estas medidas podem ser obtidas a partir dos comandos *degree(g)* para a centralidade de Grau e *betwennees(g)*, para a CIMC. Para verificar a medida de centralidade da Caminhada Aleatória proposto por Noh e Rieger [14], foi desenvolvido o algoritmo abaixo:

Algoritmo CCA:

library(igraph);

N (números de nós da rede);

T (Tempo de iteração);

g (Rede gerada a partir de um dos modelos, com N nós).

Vetor com o número de grau para cada nó da rede:

$grau \leftarrow degree(g);$

Conectividade Global da rede:

$$sum(grau);$$

Probabilidade Estacionária segundo Noh e Rieger:

$$PE \leftarrow grau/sum(grau);$$

Matriz de Probabilidade estacionária:

$$PE \leftarrow matrix(c(PE), N, N);$$

$$PE \leftarrow t(PE);$$

Matriz de Acoplamento:

$$A \leftarrow get.adjacency(g)/grau;$$

Construção da matriz Identidade:

Declarando as variáveis:

$$i \leftarrow integer()$$

$$j \leftarrow integer()$$

$$I \leftarrow matrix(ncol = N, nrow = N)$$

$$\text{for } (i \text{ in } 1:N)$$

$$\{$$

$$\text{for } (j \text{ in } 1:N)$$

$$\{$$

$$I[i, j] \leftarrow 0;$$

$$I[i, i] \leftarrow 1;$$

$$\}$$

$$\}$$

A seguir a iteração da matriz de probabilidade, segundo a Equação mestra da caminhada Aleatória Eq. (2.8), em que a matriz P é a matriz de probabilidade de transição. Inicialmente esta matriz é igual a matriz Identidade. A matriz S é obtida do somatório de todas as matrizes de transição em cada iteração, esta é utilizada para o cálculo do Tempo de Relaxação;

$P \leftarrow I;$

$S \leftarrow I.$

Vetor C representará a CCA, em que cada elemento do vetor corresponde a um nó da rede.

$C \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{"numeric"});$

Vetor R representará o Tempo de Relaxação, em que cada elemento corresponde a um nó da rede.

$R \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{"numeric"});$

Multiplicação das Matrizes:

```

for (j in 1:T)
{
    Pv ← P
    P ← crossprod(t(Pv), A)
    S ← P + S
    for (i in 1:N)
    {
        R[i] ← S[i, i] - (j + 1) * PE[i, i]
        C[i] ← PE[i, i]/R[i]
    }
}

```

Simulação da Caminhada Aleatória em Redes

Na caminhada Aleatória em redes complexas, é considerado que cada nó da rede, para o tempo inicial, possua um caminhante aleatório. A cada iteração este caminhante escolherá um de seus vizinhos para realizar o próximo passo. O algoritmo que descreve tal movimento segue abaixo:

library (igraph);

N (números de nós da rede);

W (números de caminhantes aleatórios);

T (Tempo de iteração).

Vetor n representa o vetor que guarda a informação dos nós da rede:

```
 $n \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{"numeric"});$ 
```

Vetor n_a representa o vetor com os nós antigo da rede:

```
 $na \leftarrow \text{vector}(\text{length} = W, \text{mode} = \text{"numeric"});$ 
```

Vetor n_f representa o vetor com a fração de caminhantes da rede:

```
 $nf \leftarrow \text{vector}(\text{length} = N, \text{mode} = \text{"numeric"});$ 
```

Cada nó da rede recebe um caminhante no instante inicial, este deixara o nó de origem, chamado nó antigo (na), e escolherá um nó para visitar:

```
 $\text{for } (i \text{ in } 1:W)$ 
 $\{$ 
 $na[i] \leftarrow i;$ 
 $\}$ 
```

O vetor n atualiza o valor da fração de caminhantes, mediante a função *which*, que apresentam quais caminhantes estão em cada nó em uma iteração e a função *length*, a qual indica o tamanho do vetor, que no caso representa a quantidade de coordenadas que o vetor *which* possui. Ou seja, indica o número de caminhantes por nó para a iteração, permitindo encontrar a nova fração de caminhantes.

```
 $\text{for } (i \text{ in } 1:N)$ 
 $\{$ 
 $n[i] \leftarrow \text{length}(\text{which}(na == (i)))$ 
 $n[i] \leftarrow n[i]/W$ 
 $\}$ 
```

A matriz M representa a rede, em que a cada coluna refere-se a um nó. Quando se adiciona os elementos de cada coluna dessa matriz, obtêm-se um vetor, cuja coordenada traz a quantidade de caminhantes aleatórios por nó da rede, que inicialmente é 1. A fração

de caminhantes aleatórios por nó, representado pelo vetor nf , é obtido dividindo estas coordenadas pelo total de caminhantes na rede.

```

M ← matrix(ncol = N, nrow = N);
M ← I;
nf ← colSums(M);
nf ← nf/W;

```

A escolha para o próximo passo é feita a partir da função *neighbors*. Esta função faz o levantamento dos nós vizinhos de cada nó, sendo estes, as novas possibilidades para o passo seguinte. Caso o vértice da rede estiver conectado com apenas um vizinho obrigatoriamente o caminhante que se encontrar nesse nó deverá retornar para seu vizinho. Quando o nó possuir mais de um vizinho o caminhante sorteará de forma aleatória, por meio da função *sample*, qual de seus vizinhos será seu novo destino.

Para a evolução temporal dos caminhantes aleatórios é necessário guardar as informações passadas, quanto a frações de caminhantes que já percorreram pelos nós da rede. Isso por que a fração de caminhantes pode ser vista como informação, em que esta é cumulativa. Assim, quanto maior o número de caminhantes que percorreram um nó da rede ao longo das iterações, maior será também a informação obtida por esse nó. Para tal situação se faz necessário atualizar a matriz M com o número de caminhantes que passam pelo nó.

```

for(t in 1:T)
{
  for (i in 1:N)
  {
    viz ← neighbors(g, na[i]);
    if(length(viz) == 1)
    {
      na[i] ← viz[1];
    }
  }
}

```

```
        else
        {
            na[i] ← sample(viz, 1);
        }
    }

    for (k in 1:N)
    {
        M[k, na[k]] ← 1;
    }
    nf ← colSums(M)
    nf ← nf/W
}
```

As imagens das redes assim como os gráficos de comparação entre as medidas de centralidade foram desenvolvidos a partir das ferramentas apresentadas pelo próprio software R. Já as imagens em escala de cores foram construídas por meio do software livre Gnuplot, sendo este próprio para construção de gráficos, assim como análise de gráficos e imagens.

B - Leis dos Grandes Números

Considere uma sequência de N variáveis aleatórias independentes e com a mesma distribuição de probabilidade para cada uma delas, sendo estas variáveis denotadas por $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_N$.

A lei dos grandes números diz que $\frac{1}{N} \sum_{j=1}^N \varepsilon_j \rightarrow a$ quando $N \rightarrow \infty$; $a = \langle \varepsilon_j \rangle$

A condição para que o teorema seja válido é que a média exista.

Esse teorema permite a interpretação frequencial da probabilidade de um evento. Suponha que estejamos interessados em saber a frequência de ocorrência, de um evento A, em uma sequência de N ensaios. Se denotarmos por ε_j a variável que toma o valor 1 se A ocorre e o caso contrário, então $m = (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_N)$ será o número de vezes que A acontece e a frequência $f = \frac{m}{N}$. Por outro lado, $a = \langle \varepsilon_j \rangle = p$ onde p é a probabilidade de ocorrer A. Logo a Lei de Grandes Números garante que $f = p$; $N \rightarrow \infty$.

No limite $N \rightarrow \infty$ o único resultado possível para variável y é a , o que é equivalente a dizer que y assume o valor a com probabilidade 1 ou ainda que a densidade de probabilidade associada a y é:

$$\rho(y) = \delta(y - a).$$

Dem: Seja $G_y(k)$ a função característica correspondente a variável aleatória y definindo por:

$$y = \frac{1}{N} \sum_{j=1}^N \varepsilon_j,$$

$g(k)$ é a função característica de cada uma das variáveis ε_j . Então

$$G_y(k) = \langle e^{iky} \rangle = \langle e^{ik \frac{1}{N} \sum_{j=1}^N \varepsilon_j} \rangle,$$

$$G_y(k) = \langle e^{i \frac{k}{N} \varepsilon_1 + i \frac{k}{N} \varepsilon_2 + i \frac{k}{N} \varepsilon_3 + \dots + i \frac{k}{N} \varepsilon_N} \rangle,$$

$$G_y(k) = \prod_{j=1}^N \langle e^{i \frac{k}{N} \varepsilon_j} \rangle,$$

$$G_y(k) = [g(k/N)]^N.$$

Como a média existe, podemos escrever:

$$g(k) = 1 + ika + O(k),$$

$$\left[g\left(\frac{k}{N}\right) \right]^N = \left[1 + \frac{ika}{N} + O\left(\frac{k}{N}\right) \right]^N = 1 + \frac{ika}{N} \rightarrow e^{ika}.$$

Quando $N \rightarrow \infty$

$$G_y(k) = e^{ika}.$$

Aplicando a antitransformada de Fourier:

$$\rho(y) = \frac{1}{2\pi} \int g(k) e^{-ikx} dk,$$

$$\rho(y) = \frac{1}{2\pi} \int e^{ika} e^{-iky} dk,$$

$$\rho(y) = \frac{1}{2\pi} \int e^{-ik(y-a)} dk,$$

$$\delta(x) = \frac{1}{2\pi} \int e^{-ikx} dk,$$

$$\rho(y) = \delta(y - a).$$

C- Teorema Central do Limite (TCL).

O teorema central do Limite afirma que a variável aleatória z definida por:

$$z = \frac{1}{\sqrt{Nb}} \left\{ \sum_{j=1}^N \varepsilon_j - Na \right\},$$

e possui distribuição gaussiana:

$$\rho(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Para que o TCL seja válido, basta existir a média a e a variância b .

Dem: Seja $g(k)$ a função característica de cada uma das variáveis aleatórias ε_j e considere a expansão em cumulantes. Devemos ter:

$$g(k) = \exp \left\{ iak - \frac{1}{2}bk^2 + O(k^2) \right\}.$$

Pois o primeiro cumulantes é a média a e o segundo é a variância b .

A função característica $G_z(k)$ correspondente a z é:

$$G_z(k) = \langle e^{ikz} \rangle = \langle \exp \left\{ i \frac{k}{\sqrt{Nb}} \sum_{j=1}^N (\varepsilon_j - a) \right\} \rangle,$$

ou

$$G_z(k) = \prod_{j=1}^N \langle e^{i\tilde{k}\varepsilon_j} \rangle e^{-i\tilde{k}a} = \{g(\tilde{k})e^{-i\tilde{k}a}\}^N,$$

com $\tilde{k} = \frac{k}{\sqrt{Nb}}$.

Como ainda:

$$G_z(k) = \exp \left\{ -\frac{1}{2}Nb\tilde{k}^2 + NO(\tilde{k}^2) \right\},$$

agora $Nb\tilde{k}^2 = k^2$ e sendo $O(\tilde{k}^2) = O(N^{-1})$ então $NO(N^{-1}) \rightarrow 0$; quando $N \rightarrow \infty$.

Portanto corresponde a distribuição gaussiana.

$$\rho(z) = \frac{1}{2\pi} \int g(k) e^{-ikz} dk,$$

$$\rho(z) = \frac{1}{2\pi} \int e^{-\frac{k^2}{2}} e^{-ikz} dk,$$

$$\rho(z) = \frac{1}{2\pi} \int e^{-ikz - \frac{k^2}{2}} dk,$$

$$\rho(z) = \frac{1}{2\pi} \sqrt{2\pi} \exp \left[\frac{-z^2 - 0}{4 \cdot \frac{1}{2}} \right],$$

$$\rho(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}}.$$

Para N suficientemente grande esse resultado constitui uma boa aproximação de modo que, em termos da variável $m = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_N = \sqrt{Nb}z + Na$ ele se escreve:

$$P_N(m) = \frac{1}{\sqrt{2\pi Nb}} \exp \left\{ -\frac{(m - Na)^2}{2Nb} \right\},$$

pois $\rho(z)dz = P_N(m)dm$ e $dm = \sqrt{Nb}dz$.

D- Contagem dos caminhos para os nós 2 e 3, referente a Fig. 4.1.

Menores caminhos passando pelo nó 2.

0-2-1	}	Total = 11 caminhos
0-2-3		
0-2-3-4		
0-2-3-5		
0-2-3-6		
0-2-3-7		
1-2-3		
1-2-3-4		
1-2-3-5		
1-2-3-6		
1-2-3-7		

Menores caminhos passando pelo nó 3.

0-2-3-4	}	Total=16 caminhos.
0-2-3-5		
0-2-3-4-6		
0-2-3-5-7		
1-2-3-4		
1-2-3-5		
1-2-3-4-6		
1-2-3-4-7		
2-3-4		
2-3-5		
2-3-4-6		
2-3-5-7		
4-3-5		
5-3-4-6		
4-3-5-7		
6-4-3-5-7		

E- Lemas utilizados para a implementação do algoritmo para CIMC.

A implementação do Algoritmo foi realizado a partir da utilização de alguns Lemas e Teoremas, já conhecido na Literatura, como BRANDES [45].

Lema 4.1 (critério Bellman):

Um vértice $i \in V$ encontra-se em um caminho mais curto entre os vértices $s \in V$ e $t \in V$, se e somente se, $d_G(s, t) = d_G(s, i) + d_G(i, t)$.

Chamando “par-dependência” $\delta_{st}(i) = \frac{\sigma_{st}(i)}{\sigma_{st}}$, sendo a razão entre os menores caminhos que ligam s e t passando por um dado nó intermediário i e o de total de menores caminho entre s e t .

De acordo como Lema 1, o caminho mais curto entre s e t só passa por i , se $d_G(s, t) = d_G(s, i) + d_G(i, t)$. Assim se, $d_G(s, t) < d_G(s, i) + d_G(i, t)$ significa que o caminho mais curto não passa por i . Desse modo, podemos escrever que o número de caminhos mais curtos que passam por i é:

$$\sigma_{st}(i) = \begin{cases} 0, & \text{se } d_G(s, t) < d_G(s, i) + d_G(i, t) \\ \sigma_{si} \cdot \sigma_{it}(i), & \text{caso contrário} \end{cases}$$

Determinando o número de caminhos mais curtos que passam por um dado nó i , podemos encontrar a centralidade de intermediação para esse nó, fazendo:

$$C_B(i) = \sum_{s \neq t \neq i \in V} \delta_{st}(i).$$

Para realizar a contagem dos caminhos, fez-se uso do produto da matriz adjacência, mais especificado no lema 2.

Lema 2 (contagem caminho algébrico)

Seja $A^k = (a_{u,v}^{(k)})$ a k -ésima potência da matriz de adjacência de um grafo não ponderado. Então, $a_{u,v}^{(k)}$ é igual o número de caminhos de u para v de comprimento exatamente k .