



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA
MESTRADO EM FÍSICA

ANTONIO MAURICIO ROCHA ALENCAR JÚNIOR

ANÁLISE DOS EFEITOS DA DIMENSIONALIDADE EM REDES COM ESTATÍSTICA
DE TSALLIS

FORTALEZA

2023

ANTONIO MAURICIO ROCHA ALENCAR JÚNIOR

ANÁLISE DOS EFEITOS DA DIMENSIONALIDADE EM REDES COM ESTATÍSTICA DE
TSALLIS

Dissertação apresentada ao Curso de Mestrado em Física do Programa de Pós-Graduação em física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Física. Área de Concentração: Física da matéria condensada.

Orientador: Prof. Dr. Humberto de Andrade Carmona.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A353a Alencar Júnior, Antonio Mauricio Rocha.

Análise dos efeitos da dimensionalidade em redes com estatística de Tsallis / Antonio Mauricio Rocha Alencar Júnior. – 2023.

76 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Física, Fortaleza, 2023.

Orientação: Prof. Dr. Humberto de Andrade Carmona.

1. Modelo de crescimento da conexão preferencial. 2. Modelo de Barabási-Albert. 3. Estatística de Tsallis. 4. q-Exponencial. 5. Assortatividade. I. Título.

CDD 530

ANTONIO MAURICIO ROCHA ALENCAR JÚNIOR

ANÁLISE DOS EFEITOS DA DIMENSIONALIDADE EM REDES COM ESTATÍSTICA DE
TSALLIS

Dissertação apresentada ao Curso de Mestrado em Física do Programa de Pós-Graduação em física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Física. Área de Concentração: Física da matéria condensada.

Aprovada em: 16/08/2023.

BANCA EXAMINADORA

Prof. Dr. Humberto de Andrade
Carmona (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Soares de Andrade Júnior
Universidade Federal do Ceará (UFC)

Prof. Dr. Erneron Alves de Oliveira
Universidade de Fortaleza (UNIFOR)

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

A minha família, por todo apoio e suporte que eles têm me dado sempre. Sem eles eu jamais conseguiria chegar onde eu estou.

Aos meus amigos, pelas conversas frutíferas e banais ao longo dos semestres.

Aos professores do departamento, que ao longo da minha formação sanaram minhas dúvidas e me ajudaram em meus desenvolvimentos em minha pesquisa, em especial aos professores Saulo, Carmona e Sampaio.

Ao meu grupo, Confraria: Mateus, Barbara, Rute, Laura, Adriel e Thiago. Que me acompanharam e me acompanham desde minha entrada na graduação, nos momentos bons e ruins.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

"É a diferença de possibilidade e probabilidade. As pessoas acham que é a mesma coisa, mas é totalmente diverso. A entropia é um conceito mais sutil que a energia: a energia funciona no mundo das possibilidades. A entropia respeita o que é improvável." (TSALLIS, 2014, entrevista para Globo.)

RESUMO

Nesse trabalho foi realizado um estudo acerca dos efeitos da dimensionalidade de redes com estatística de Tsallis, tendo essas redes distribuição de graus na forma $P(k) \sim e_q^{-k/\lambda}$, sendo a q-exponencial na forma $e_q^z \equiv [1 + (1 - q)z]^{1/(1-q)}$ e no limite termodinâmico possui a forma $P(k) \sim k^\gamma$. Redes com distribuição de graus em lei de potência são conhecidas como redes livre de escala. Dentre os modelos capazes de gerar essas redes, estão os modelos de configuração e o de crescimento da conexão preferencial, proposto por Samuráí *et. al*, sendo o modelo de foco do trabalho. A estatística de Tsallis, foi desenvolvida por Constantino Tsallis, sendo oriunda da generalização da entropia de Boltzmann-Gibbs. No modelo de crescimento da conexão preferencial é realizado uma alteração na conexão preferencial proposta por Barabási, onde surge um termo de distância euclidiana d-dimensional, obtendo uma distribuição de graus que otimiza a entropia que define a estatística de Tsallis. Focamos em estudar as propriedades do comprimento médio do caminho mais curto, o diâmetro e a assortatividade dessa rede, analisando os efeitos da dimensionalidade nessas propriedades da rede.

Palavras-chave: Modelo de crescimento da conexão preferencial; Modelo de Barabási-Albert; q-Exponencial; Estatística de Tsallis; Assortatividade.

ABSTRACT

In this work, a study was carried out on the effects of the dimensionality of networks with Tsallis statistics, having these networks degree distribution in the form $P(k) \sim e_q^{-k/\lambda}$, being the q-exponential in the form $e_q^z \equiv [1 + (1 - q)z]^{1/(1-q)}$ and in the thermodynamic limit has the form $P(k) \sim k^\gamma$. Networks with power-law degree distribution are known as scale-free networks. Among the models capable of generating these networks, are the configuration models and the preferential attachment growth model, proposed by Samuraj et al., being the model of focus of the work. The Tsallis statistic was developed by Constantino Tsallis and comes from the generalization of the Boltzmann-Gibbs entropy. In the growth model of the preferential connection, a change is made in the preferential connection proposed by Barabási, where a d-dimensional Euclidean distance term appears, obtaining a degree distribution that optimizes the entropy that defines the Tsallis statistic. We focus on studying the properties of the average length of the shortest path, the diameter and the assortativity of this network, analyzing the effects of dimensionality on these network properties.

Keywords: Preferential attachment growth model; Barabási–Albert model; q-Exponential; Tsallis statistics; Assortativity.

LISTA DE FIGURAS

- Figura 1 – O quadro (a) mostra uma rede com oito sítios ($N = 8$) e oito ligações ($L = 8$). Em (b) é mostrado a distribuição de graus da rede em (a). 20
- Figura 2 – Os círculos enumerados de 0 até 4 representam sítios, com as conexões entre eles sendo as ligações. Tendo os conjuntos de caminhos dos sítios 0 e 3, destacados de verde e azul. Nós temos a sequência de ligações $E = \{(0, 1), (1, 2), (2, 4), (4, 3)\}$, com comprimento do caminho $\ell_{03} = 4$, para o caminho verde, enquanto o azul possui a sequência de ligações $E = \{(0, 1), (1, 3)\}$, com comprimento do caminho $\ell_{03} = 2$ 22
- Figura 3 – Exemplificação do processo de interpolação para o caso de $N = 20$ sítios e $\langle k \rangle = 4$. A rede inicia-se no formato regular regular, com os sítios conectados aos seus $\langle k \rangle = 4$ vizinhos mais próximos. É escolhido um sítio na rede, selecionando a ligação com seu sítios mais próximo no sentido horário. Com uma probabilidade p , essa ligação é realocada para outro sítio aleatório da rede. Esse processo é repetido até percorrer todos os sítios da rede. Podemos observar que para $p = 0$ a rede permanece a mesma, enquanto para o caso $p = 1$ todos os sítios são realocados, tornando a rede aleatória. Tendo as redes com fenômeno de mundo pequeno ocorrem num ponto de transição em que $p \in (0, 1)$ 25
- Figura 4 – k_{nn} representa o grau médio dos sítio da vizinhança e k o grau dos sítios. Na imagem (A) nós temos o comportamento da curva para uma rede assortativa ($0 < R \leq 1$), representando a rede de afiliação na categoria Astro-ph do arXiv (LESKOVEC *et al.*, 2007). Na imagem (B) para uma rede desortativa ($-1 \leq R < 0$), representando a rede de interação proteína-proteína (PPI, do inglês *protein-protein interactions*) (JEONG *et al.*, 2001). E na imagem (C) de uma rede não-assortativa ($R = 0$), representando o modelo de Barabási (BARABASI; ALBERT, 1999). 32
- Figura 5 – Do lado esquerdo nós temos um possível microestado do sistema, enquanto do lado direito temos outro microestado do mesmo sistema. 34

- Figura 6 – Os círculos enumerados cinza representam os sítios, com m representando o número de ligações que o novo sítio adicionado fará, $m_0 = 2$, \mathcal{N} a constante de normalização e k_i o grau do i -ésimo sítio. Na figura 6.(a) nós temos a rede inicial, com sítio 1 e 2 conectados. Na figura 6.(b) é realizado um sorteio do sítio que fará ligação com o novo sítio 3, tendo a probabilidade dada pela Eq (4.1), tendo sítio sorteado 1, realizamos uma ligação entre esses dois sítios. Na figura 6.(c) o sorteio é repetido, pois é necessário realizar $m = 2$ ligações. Na figura 6.(d) nós temos o novo sítio fazendo todas as ligações. O processo é repetido mais $N - 3$ vezes para obtermos uma rede com N sítios. 41
- Figura 7 – Na figura abaixo nós temos dois exemplos de rede de Barabási. Os sítios são representados pelos círculos, a intensidade da cor e o tamanho do sítio representam seu grau. Na Figura 7.(a) a rede possui $m_0 = 2$ com $m = 1$, enquanto na Figura 7.(b) nós temos $m_0 = 4$ e $m = 3$. Como para representada em (b) possui valor de m maior, a rede possuirá mais ligações. Ambas as redes possuem $N = 500$ sítios. 41
- Figura 8 – Os círculos coloridos de azul são os valores calculados a partir da distribuição de graus oriundo do modelo de Barabási. Foram utilizadas $M = 50$ amostras para o cálculo da distribuição, com valor de $m_0 = 1$ e $m = 1$. Enquanto a seta preta representa a regressão linear dos dados, dos quais podemos inferir o valor de γ a partir do coeficiente linear. As redes geradas possuem $N = 10^6$ sítios. 45
- Figura 9 – Os círculos pretos, cinza, carmesim, verdes, azuis e rosas representam os valores dos diâmetros para $A = 1, 3, 5, 7, 9, 10$, respectivamente. Enquanto Δd representam a variação no diâmetro da rede. Podemos observar que para o prefator $A = 1$, uma variação de N na ordem de 10^6 representa uma variação de $\Delta d = 2.54$, mostrando que o diâmetro dessa rede cresce muito lentamente com o número de sítios. 46
- Figura 10 – Nós temos r como sendo a distância e $p(r)$ a distribuição dada pela Eq. (5.3). Na figura (a) nós podemos observar os efeitos do parâmetro α_g sobre a distribuição dada pela Eq. (5.3), sendo $d = 1$ a dimensão. Enquanto na Figura (b) podemos observar os efeitos da dimensão, fixado $\alpha_g = 2.0$ 49

- Figura 11 – Nós temos $\log r$ como sendo a distância e $\log p(r)$ a distribuição dada pela Eq. (5.3). Na figura (a) nós podemos observar os efeitos do parâmetro α_g sobre a distribuição dada pela Eq. (5.3), sendo $d = 1$ a dimensão. Enquanto na Figura (b) podemos observar os efeitos da dimensão, fixado $\alpha_g = 2.0$. . . 50
- Figura 12 – Os círculos representam os sítios, enquanto k_i representa o grau do i -ésimo sítio, sendo \mathcal{N} a constante de normalização e r_{CM} o centro de massa. Na Figura 12.(a) nós temos a rede com os sítios 1 e 2 conectados entre si, a distância r_* foi gerada seguindo a distribuição dada pela Eq. (5.2), na Figura 12.(b) nós temos a região onde o novo sítio pode ser alocado, um círculo de raio r_* . Na Figura 12.(c) foi realizado um sorteio, sendo o sítio 2 selecionado para conectar ao sítio 3. Na Figura 12.(d) o sítio 2 e 3 são conectados, atualizando o centro de massa e o grau dos sítios. O processo é repetido $N - 3$ até que a rede possua N sítios. 50
- Figura 13 – No gráfico expresso, nós temos que d representa a dimensão, enquanto α_g é o parâmetro relacionado com a distância dos novos sítios para o centro de massa e α_a o parâmetro relacionado as distâncias na conexão preferencial. Para todos os conjuntos fixamos $\alpha_a = 2.0$. Foram utilizadas $M = 82$ amostras com $N = 10^5$ sítios. 52
- Figura 14 – No gráfico expresso, nós temos que d representa a dimensão, enquanto α_g é o parâmetro relacionado com a distância dos novos sítios para o centro de massa e α_a o parâmetro relacionado as distâncias na conexão preferencial. Para todos os valores fixamos $\alpha_g = 2.0$. Foram utilizadas $M = 82$ amostras com $N=10^5$ sítios. 53
- Figura 15 – No gráfico expresso, nós temos que d representa a dimensão, fixamos o valor de $\alpha_g = 2.0$ para todas as curvas, enquanto variamos α_a . Foram utilizadas $M = 82$ amostras para redes com $N = 10^5$ sítios. As retas pretas representam a curva que melhor se ajusta ao conjunto de dados, com os valores de q e λ para cada uma das curvas expressa na Tabela 2. O ajuste foi realizado numa região dos dados, devido ao platô nas curvas oriundo da grande quantidade de sítios com mesmo grau e poucos sítios com grande valor de grau. Esse platô desaparece no limite termodinâmico, onde $k \gg 1$ 54

- Figura 16 – No gráfico expresso, nós temos que $\langle \ell \rangle$ representa o comprimento médio dos menores caminhos da rede e α_g é o parâmetro relacionado as distâncias dos novos sítios para o centro de massa, foi utilizado $\alpha_a = 1.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de $\langle \ell \rangle$ para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios. 56
- Figura 17 – No gráfico expresso, nós temos que $\langle \ell \rangle$ representa o comprimento médio dos menores caminhos da rede e α_a é o parâmetro relacionado a conexão preferencial modificada, foi utilizado $\alpha_g = 1.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de $\langle \ell \rangle$ para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios. 56
- Figura 18 – No gráfico expresso, nós temos que d_{max} representa o diâmetro da rede e α_g é o parâmetro relacionado as distâncias dos novos sítios para o centro de massa, foi utilizado $\alpha_a = 1.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de d_{max} para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios. 57
- Figura 19 – No gráfico expresso, nós temos que d_{max} é o diâmetro da rede e α_a é o parâmetro relacionado as distâncias dos novos sítios para o centro de massa, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de d_{max} para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios. 57
- Figura 20 – No gráfico expresso, nós temos que $\langle \ell \rangle$ representa o comprimento médio do menor caminho da rede, N é o tamanho da rede, d a dimensão e α_a o parâmetro relacionado as distâncias na conexão preferencial, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 402, 252, 202, 92, 92, 118, 75, 12$ amostras, para as redes com $N = 5.000, 10.000, 20.000, 40.000, 80.000, 160.000, 320.000$ sítios, respectivamente. 58

- Figura 21 – No gráfico expresso, nós temos que d_{max} representa o diâmetro da rede, N é o tamanho da rede, d a dimensão e α_a o parâmetro relacionado as distâncias na conexão preferencial, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 402, 252, 202, 92, 92, 118, 75, 12$ amostras, para as redes com $N = 5.000, 10.000, 20.000, 40.000, 80.000, 160.000, 320.000$ sítios, respectivamente. 58
- Figura 22 – No gráfico expresso, nós temos que R representa o coeficiente de assortatividade e α_g o parâmetro as distâncias para o centro de massa do sistema, foi utilizado $\alpha_a = 1.0$ para todos os conjuntos de amostras. Os círculos pretos é o valor de R para dimensão $d = 1$, estrelas verdes $d = 2$, triângulos azuis $d = 3$ e quadrado vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios. 60
- Figura 23 – No gráfico expresso, nós temos que R representa o coeficiente de assortatividade e α_a o parâmetro relacionado as distâncias do centro de massa da rede, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras. Os círculos pretos é o valor de R para dimensão $d = 1$, estrelas verdes $d = 2$, triângulos azuis $d = 3$ e quadrado vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios. 60
- Figura 24 – No gráfico expresso, nós temos que k_{nn} representa o grau da vizinhança, k o grau do sítio, α_g o parâmetro relacionado as distâncias geradas com relação ao centro de massa do sistema e d a dimensão. Para todos os casos foram utilizados $\alpha_a = 2.0$. Foram utilizados $M = 41$ amostras para redes com $N = 10^5$ sítios. Foi realizado *Logarithmic binning* nos dados. 61
- Figura 25 – No gráfico expresso, nós temos que k_{nn} representa o grau da vizinhança, k o grau do sítio, α_a o parâmetro relacionado as distâncias na conexão preferencial e d a dimensão. Para todos os casos foram utilizados $\alpha_g = 2.0$. Foram utilizados $M = 41$ amostras para redes com $N = 10^5$ sítios. Foi realizado *Logarithmic binning* nos dados. 61

- Figura 26 – No gráfico expresso, nós temos que R é o coeficiente de assortatividade, N o número de sítios na rede, α_a o coeficiente relacionado as distâncias na conexão preferencial, d a dimensão, λ e q os parâmetros da distribuição de graus da rede. Os diferentes marcadores representam diferentes valores de α_a , foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 402, 252, 202, 92, 92, 118, 75, 12$ amostras, para as redes com $N = 5.000, 10.000, 20.000, 40.000, 80.000, 160.000, 320.000$ sítios, respectivamente. 62
- Figura 27 – No gráfico expresso, nós temos que R é o coeficiente de assortatividade, α_a o coeficiente relacionado as distâncias na conexão preferencial, d a dimensão. Os diferentes marcadores representam diferentes valores de α_a , foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 81$ amostras para redes com $N = 10^5$ sítios. As regiões avermelhadas representam a área onde $q = 1.33, \lambda = 0.30$, fora dessa região os valores de q decresce e λ aumenta, segundo as Eqs. (6.2), (6.1). 62
- Figura 28 – Fluxograma de um algoritmo para implementação do modelo de crescimento da conexão preferencial. Os símbolos destacados em azul serão os parâmetros que iremos variar ao longo do trabalho. 72

LISTA DE TABELAS

- Tabela 1 – Conjunto de comprimentos dos menores caminhos entre os sítios da rede representado na Figura 2, com valor do diâmetro, $d_{max} = 3$, destacado em vermelho. 22
- Tabela 2 – Os parâmetros q e λ foram calculados utilizando o algoritmo LM (do inglês *Levenberg-Marquardt algorithm*) (GAVIN, 2019), enquanto R^2 é o coeficiente de determinação, que representa precisão dos dados mensurados com relação a função hipótese (S.A. SLINKER B.K., 2016), no nosso caso, a função hipótese é a distribuição expressa na Eq. (5.5). $R^2 \in [0, 1]$, quanto mais próximo de 1, maior a tendência da função hipótese descrever aqueles dados. 55

SUMÁRIO

1	INTRODUÇÃO	16
2	CONCEITOS BÁSICOS DE CIÊNCIAS DE REDES	19
2.1	Propriedades das redes	19
2.1.1	<i>Coefficiente de agrupamento</i>	20
2.1.2	<i>Comprimento médio do caminho mais curto e diâmetro</i>	21
2.1.3	<i>Algoritmo BFS</i>	23
2.1.3.1	<i>Fenômeno de mundo pequeno</i>	24
2.1.4	<i>Assortatividade</i>	26
2.1.4.1	<i>Caracterização discreta</i>	27
2.1.4.2	<i>Assortatividade por propriedades escalares</i>	30
3	CONCEITOS DE FÍSICA ESTATÍSTICA	33
3.1	Temperatura, entropia e ensambles	34
3.2	Estatística de Boltzmann-Gibbs	36
3.3	Estatística de Tsallis	38
4	O MODELO DE BARABÁSI-ALBERT	40
4.1	Grau médio e distribuição de graus	42
4.2	Caminho médio, menor caminho e diâmetro	46
5	O MODELO DE CONEXÃO PREFERENCIAL ENVOLVENDO DISTÂNCIAS EUCLIDIANAS	48
5.1	Distribuição de graus da rede	51
6	RESULTADOS	52
6.1	Distribuição de graus da rede	52
6.1.1	<i>Comprimento médio do menor caminho e diâmetro da rede</i>	55
6.1.2	<i>Assortatividade</i>	59
7	CONCLUSÃO	64
	REFERÊNCIAS	65
	APÊNDICE A –ALGORITMO: MODELO BARABÁSI-ALBERT	68
	APÊNDICE B –ALGORITMO: MODELO DO CRESCIMENTO DA CONEXÃO PREFERENCIAL	70
	APÊNDICE C –OTIMIZAÇÃO DA ENTROPIA DE TSALLIS	73

1 INTRODUÇÃO

O primeiro trabalho envolvendo a ideia de rede — também conhecido como grafo na matemática — surgiu com Euler em 1736, com o famoso problema das pontes de Königsberg (BOAVENTURA, 2017). A forma como Euler propôs para resolver esse problema deu origem ao que hoje é conhecido como grafo. Em geral, os termos “grafo” e “rede” são usados de forma intercambiável, sendo o primeiro utilizado na perspectiva Matemática e o último na perspectiva de Ciência de Sistemas Complexos. Ambos se referem a uma coleção de nós ou sítios conectados por arestas ou ligações. Nessa dissertação usaremos o termo “rede”, Além dos termos sítios e ligações para os elementos da rede. Uma rede é representada por $G(V, E)$, onde V representa o conjunto de sítios e E o conjunto de ligações (do inglês *edges*). Os sítios e ligações podem possuir diferentes significados, dependendo do problema que estamos lidando. Podemos, de forma abstrata, relacionar os sítios apenas como pontos, e as ligações como as conexões entre esses sítios. Quando um sítio i está conectado ao sítio j , a ligação correspondente é representada por (i, j) . O número de ligações que cada sítio possui é conhecido como grau do sítio, sendo representado comumente na literatura por k_i , onde o subscrito i é para indicar o grau do sítio i .

Diversas propriedades mostram-se recorrentes ao longo do estudo de uma ampla variedade de redes, dentre os quais estão a distribuição de graus, comprimento médio do caminho mais curto da rede, diâmetro e a assortatividade. A distribuição de graus da rede representa a forma como estão distribuídos os graus de todos os sítios da rede, retratando a probabilidade de selecionarmos aleatoriamente um sítio tendo grau k em toda rede. A assortatividade nesse trabalho, será utilizada para checar uma correlação entre os graus dos sítios da rede, ou seja, checar se sítios muitos conectados tendem a se conectar-se a sítios muito conectados, com poucas conexões ou se não há uma tendência na conexão com relação ao grau dos sítios. O comprimento médio do caminho mais curto da rede representa a média sobre todos conjuntos de menores caminhos da rede, enquanto o diâmetro da rede é o maior desse conjunto de menores caminhos. O cálculo do comprimento médio do menor caminho da rede pode ser feito por meio de dois tipos de navegações: navegação com conhecimento local e global. Na navegação com conhecimento local nós possuímos informações apenas do sítio alvo (sítio onde queremos chegar) e dos vizinhos dos sítios fonte (sítio de saída), enquanto na navegação conhecimento global, nós possuímos conhecimento sobre todos os pares de ligações da rede (LI *et al.*, 2013).

Um das redes mais fascinantes é a *World Wide Web (WWW)*. Essa rede é formada por documentos em hipermídia (reunião de várias mídias num ambiente computacional, suportado

por sistemas eletrônicos de comunicação (BAIRON, 2012)), cada documento fica alocado numa página, sendo a representação do sítio da rede, tendo a ligação entre eles sendo dada por um *hyperlink* que leva a outro documento. Atualmente a WWW é a maior rede criada pelo ser humano, com uma estimativa de um trilhão de documentos (BARABÁSI, 2013). Diversos estudos foram realizados com o foco de obter as propriedades dessas largas redes, dentre os trabalhos primordiais, onde foi realizado o cálculo do diâmetro da WWW (ALBERT *et al.*, 1999).

Redes reais como a WWW são dinâmicas, ou seja, crescem com o tempo. Esse crescimento normalmente se dá de forma preferencial: os sítios mais conectados se tornam cada vez mais conectados. Baseando-se nessa característica Albert-Laszlo Barabási, desenvolveu um modelo mínimo chamado de modelo Barabási-Albert, onde um sítio adicionado a rede possui uma probabilidade maior de se conectar a um sítio com um valor de grau maior, esse processo é chamado de conexão preferencial e gera redes com distribuição de graus dadas através de uma lei de potência, chamadas de redes livre de escala (BARABASI; ALBERT, 1999). Esse modelo é usado em diversos trabalhos, dentre os quais está o estudo de processos aleatórios com alta variação (JOHNSTON; ANDERSEN, 2022) e o efeito da aprendizagem social observacional na vacinação e doenças (ORABY; BALOGH, 2022).

A estatística de Tsallis foi inicialmente proposta por Constantino Tsallis por meio de uma entropia conhecida como entropia de Tsallis, que generaliza a entropia de Boltzmann-Gibbs. Essa generalização ocorre por meio do parâmetro q que está associado com a quantidade normalmente escalada em sistemas multifractais (TSALLIS, 1988). Observando que é possível realizar modificações na conexão preferencial para que possamos representar sistemas físico-estatísticos (BARABASI, 2001a), Soares *et. al* propuseram um modelo que pudesse retratar sistemas termostáticos com a estatística de Tsallis (SOARES *et al.*, 2005).

Soares *et. al* propuseram um modelo cuja conexão preferencial depende de distâncias euclidianas bidimensionais. Foi observado através de análises numéricas que a distribuição de graus da rede possui não somente a mesma forma funcional que otimiza a entropia de Tsallis – q -exponencial – mas uma relação direta entre as propriedades dos sistemas com mesma entropia, sujeitas aos vínculos de normalização da distribuição e grau médio constante (SOARES *et al.*, 2005).

Posteriormente, realizou-se uma generalização d -dimensional do modelo proposto por Soares, chamado de modelo de crescimento da conexão preferencial (BRITO *et al.*, 2016).

Nesse trabalho foi realizado uma análise numérica acerca dos efeitos da dimensionalidade na distribuição de graus, observando que a distribuição de graus possuía uma dependência da dimensão sobre a qual a rede é construída.

Nosso trabalho foca-se em dar continuidade ao trabalho de Samuráí *et. al*, focando-se em analisar os efeitos da dimensionalidade nas seguintes propriedades da rede: assortatividade dos graus da rede, diâmetro e comprimento médio do menor caminho da rede. O cálculo do comprimento médio do menor caminho da rede e diâmetro foram realizados através de um algoritmo *Breadth-first search (BFS)*.

2 CONCEITOS BÁSICOS DE CIÊNCIAS DE REDES

2.1 Propriedades das redes

Uma rede é um conjunto de elementos $G(V, E)$, cujo V representa os sítios e E as ligações. O grau de um sítio representa o número de ligações que ele possui, tendo essa definição em mente, podemos também introduzir a ideia de número de ligações, comumente denotado pela letra L , com esse valor expressando o número total de interações entre os sítios (BARABÁSI, 2013). Tendo uma rede não-direcionada, isto é, uma rede onde as ligações entre os sítios não possuem uma direção, podemos relacionar L e k_i :

$$L = \frac{1}{2} \sum_{i=1}^N k_i, \quad (2.1)$$

com o fator $1/2$ aparecendo devido ao fato do sítio i estar conectado ao sítio j , assim como j está conectado ao sítio i , isto é, somando o grau de todos os sítios obtemos o dobro do número de ligações. Tendo esse conceito de grau de um sítio bem definido, podemos definir também a distribuição de graus da rede dada por $P(k)$, que representa a probabilidade de selecionarmos aleatoriamente um sítio possuindo grau k (BARABÁSI, 2013). Essa distribuição é dada por:

$$P(k) = \frac{N_k}{N}, \quad (2.2)$$

onde N_k é o número de sítios com grau k , N é o número total de sítios da rede. A distribuição $P(k)$ deve ser normalizada, isto é

$$\sum_{k=1}^{k_{max}} P(k) = 1, \quad (2.3)$$

tendo que

$$\sum_{k=1}^{k_{max}} N_k = N, \quad (2.4)$$

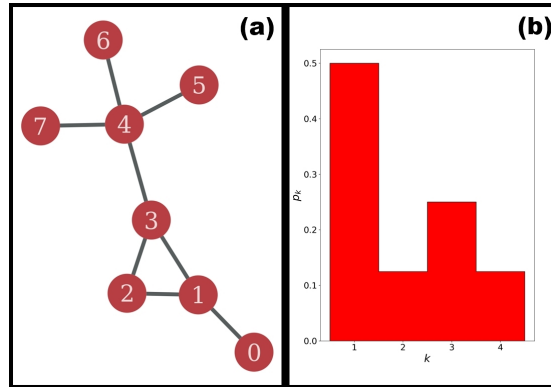
onde k_{max} é o máximo grau da rede.

Com a distribuição de graus dada pela Eq. (2.2), podemos calcular o grau médio da rede através da equação:

$$\langle k \rangle = \sum_{i=1}^N P(k_i) k_i. \quad (2.5)$$

Na Figura 1 nós mostramos uma rede com 8 sítios e 8 ligações distribuídas entre os sítios de forma desuniforme, ou seja, os graus estão distribuídos de forma que há sítios com mais ligações do que outros.

Figura 1 – O quadro (a) mostra uma rede com oito sítios ($N = 8$) e oito ligações ($L = 8$). Em (b) é mostrado a distribuição de graus da rede em (a).



Fonte: elaborado pelo autor

A distribuição cumulativa representa a probabilidade dos valores do nosso conjunto de amostras K (que está relacionado ao conjunto de todos graus possíveis da rede), assuma um valor menor do que o grau k (DEISENROTH *et al.*, 2020). Como nosso caso, nós estamos lidando com distribuições discretas, essa distribuição de graus cumulativa da rede será dada simplesmente por

$$P_c(k) = \sum_{k'=k}^{k_{max}} P(k'). \quad (2.6)$$

2.1.1 Coeficiente de agrupamento

O coeficiente de agrupamento é uma medida de como os sítios da rede tendem a agrupar-se (BARABÁSI, 2013). Esse coeficiente pode ser calculado de forma local ou global. O coeficiente de agrupamento local é dado por:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}, \quad (2.7)$$

onde L_i representa o número de ligações entre os vizinhos do sítio i , k_i é o grau desse sítio. O valor de C_i está entre 0 e 1, tendo os casos extremos:

- $C_i = 0$ os sítios vizinhos do i -ésimo sítio estão totalmente desconectados entre si;
- $C_i = 1$ os sítios vizinhos do i -ésimo sítio estão totalmente conectados entre si;

Dessa forma, C_i representa a probabilidade de dois sítios com um vizinho em comum (i -ésimo sítio), possuam também ligações entre si. Tendo o coeficiente de agrupamento global dado pela média (BARABÁSI, 2013):

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i. \quad (2.8)$$

2.1.2 Comprimento médio do caminho mais curto e diâmetro

Na física, o conceito de distância possui uma grande relevância, contudo, quando trabalhamos com rede complexas, esse conceito precisa ser generalizado (BARABÁSI, 2013). Podemos observar essa característica ao usarmos como exemplo a rede de internet, onde dois computadores podem estar conectados através de um IP(número atribuído a cada dispositivo conectado a uma rede (POSTEL, 1980)), mesmo estando a milhares de quilômetros de distância. Para esse caso, faz-se necessário surgir o conceito de caminho e comprimento do caminho, onde por caminho tomamos a trajetória dentro da rede, entre, por exemplo, dois sítios i e j . Essa trajetória pode ser representada por uma sucessão de ligações, com número total representando o comprimento do caminho. Exemplificamos a ideia de caminho na Figura 2. Considere a distância entre os sítios s (fonte, do inglês *source*) e t (alvo, do inglês *target*) na rede representada na Figura 2. Existem dois caminhos possíveis, o primeiro representado de vermelho, com caminho $E = \{(s, 6), (6, 8), (8, 5), (5, 4), (4, 1), (1, t)\}$ e o segundo destacado de azul, com caminho $E = \{(s, 6), (6, 4), (4, 1), (1, t)\}$. Tendo o comprimento do caminho definido pelo símbolo ℓ_{st} com o sufixo representando o sítio de saída e chega, respectivamente. Teremos $\ell_{st} = 6$ para o caminho vermelho e $\ell_{st} = 4$ para o caminho azul. Observamos que mesmo ambos os caminhos saindo do mesmo sítio, tendo mesmo sítio de fonte, um deles possui um valor menor. Sendo esse caminho de menor valor conhecido com o comprimento do menor caminho.

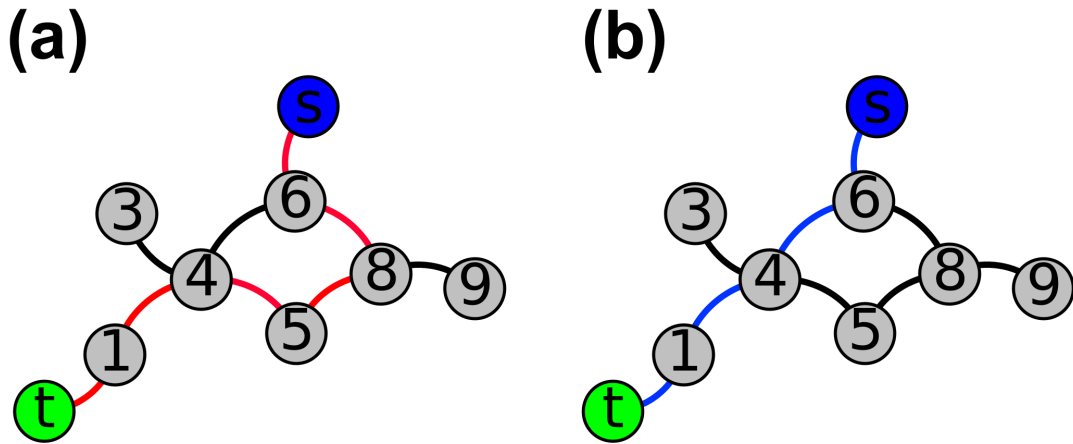
Dado um sítio inicial (fonte) e um sítio final (alvo), nós temos duas possibilidades de caminho para seguir. Contudo, um deles tem $\ell_{03} = 4$, enquanto no outro $\ell_{03} = 2$. O caminho mais curto tem bastante relevância no estudo de algumas propriedades da rede, sendo também chamado de caminho geodésico, representado por ℓ_{ij} de menor valor, dentre todos os caminhos possíveis. De forma geral, nós podemos calcular o comprimento médio do caminho mais curto por (BARABÁSI, 2013)

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{i,j \in G; i \neq j} \ell_{ij}, \quad (2.9)$$

com ℓ_{ij} sendo o comprimento do caminho mais curto entre o sítio i e o sítio j . Para o cálculo do conjunto de menores distâncias, ℓ_{ij} , diversos algoritmos foram desenvolvidos ao longo da história, dentre os quais, há o algoritmo de BFS (do inglês *Breadth-first search*).

Nessa rede expressa na Figura 2, nós temos $\ell_{03} = 2$, representando o menor caminho entre os sítios $i = 0$ e $j = 3$. O estudo do menor caminho leva em consideração apenas um par de sítios da rede, caso queiramos saber uma característica global com relação aos caminhos, o

Figura 2 – Os círculos enumerados de 0 até 4 representam sítios, com as conexões entre eles sendo as ligações. Tendo os conjuntos de caminhos dos sítios 0 e 3, destacados de verde e azul. Nós temos a sequência de ligações $E = \{(0, 1), (1, 2), (2, 4), (4, 3)\}$, com comprimento do caminho $\ell_{03} = 4$, para o caminho verde, enquanto o azul possui a sequência de ligações $E = \{(0, 1), (1, 3)\}$, com comprimento do caminho $\ell_{03} = 2$.



Fonte: elaborado pelo autor.

conceito de diâmetro e comprimento médio do caminho mais curto na rede são muito importantes. O diâmetro é o maior valor de todo o conjunto de menores comprimentos de caminhos mais curto da rede, ou seja, dado o conjunto de menores caminhos de toda rede, aquele de maior valor, será o diâmetro da rede, sendo representado pelo símbolo d_{max} . Para a rede representada na Figura 2, nós temos todos os conjuntos de distâncias representadas na tabela abaixo, com o diâmetro destacado em vermelho. Podemos observar a simetria nessa tabela, decorrente da rede ser não-direcionada.

	V_0	V_1	V_2	V_3	V_4
V_0	0	1	2	2	3
V_1	1	0	1	1	2
V_2	2	1	0	2	1
V_3	2	1	2	0	1
V_4	3	2	1	1	0

Tabela 1 – Conjunto de comprimentos dos menores caminhos entre os sítios da rede representado na Figura 2, com valor do diâmetro, $d_{max} = 3$, destacado em vermelho.

Fonte: elaborado pelo autor

A matriz que representa o conjunto de menores distância é simétrica, tendo diagonal principal nula. Para calcularmos o comprimento médio dos menores caminhos, basta somarmos os termos do triângulo superior ou inferior da matriz e dividir pelo número de termos, que é equivalente a expressão dada pela Eq. (2.9), nos dando o comprimento médio do caminho mais

curto da rede, para rede expressa na Figura 2 nós temos $\langle \ell \rangle = 1.6$. O diâmetro dessa rede é dada pelo caminho representado pela sequência de ligações $E = \{(0, 1), (1, 2), (2, 4)\}$.

O diâmetro da rede é definido como

$$d_{max} = \max(\ell_{ij}), \quad (2.10)$$

com ℓ_{ij} sendo o conjunto dos comprimentos dos caminhos mais curtos de toda rede. Para calcularmos esse conjunto de comprimentos, podemos utilizar o algoritmo BFS, sendo um dos mais simples e mais importantes algoritmos de rede (CORMEN CHARLES E. LEISERSON, 2001). Esse algoritmo foi inventado em 1959 pelo matemático e cientista computacional norte-americano Edward F. Moore, que usou para encontrar o menor caminho para fora de um labirinto (MOORE, 1959). O algoritmo funciona baseando-se numa navegação com conhecimento global da rede. Dada uma rede representada por $G = (V, E)$, é selecionado um sítio s dessa rede, o algoritmo BFS explora as ligações de G , para descobrir cada um dos sítios acessíveis a partir do sítio s , ou seja, todos os sítios que possuem algum grau de conexão com sítio s . É então calculado o comprimento da menor distância – caminho entre o sítio s e os demais sítios que possuem algum grau de conexão com sítio s , no qual haja o menor número de ligações (CORMEN CHARLES E. LEISERSON, 2001). O BFS é aplicado para cada um dos N sítios da nossa rede, cada uma das implementações nos retorna uma lista de distâncias. Para calcularmos o diâmetro da rede, basta armazenamos o maior valor de cada uma das implementações, selecionando o maior valor dentre todos.

2.1.3 Algoritmo BFS

Para implementação do algoritmo BFS nós utilizamos três listas: uma lista que nos dará as cores dos sítios, $color[u]$, outra lista que nos retornará os parentes do sítio (vizinhos), $\Pi[u]$, e, uma lista que guardará as distâncias do sítio fonte, s , representado por $d[u]$. Faremos também uso de um conjunto dinâmico de valores, chamados de fila (do inglês *queue*), aqui representado pela letra Q , que irá gerir o conjunto de sítios cinzas (CORMEN CHARLES E. LEISERSON, 2001). Abaixo nós temos um exemplo de implementação do algoritmo BFS.

Algoritmo BFS

- (i) Os sítios começam todos coloridos de branco;
- (ii) O sítio inicial, s , é pintado de cinza;

- (iii) Após colorir o sítio inicial de cinza, adicionamos ele à Q ;
- (iv) É feita uma checagem de todos os sítios na primeira vizinhança do sítio s , pintamos todos de cinza, com o sítio fonte sendo pintado de preto;
- (v) Atualizamos as listas $\text{color}[u]$, $\Pi[u]$, $d[u]$ e Q . Ao colorir o sítio fonte de s , ele é removido da lista Q ;
- (vi) O algoritmo repete os passos (iv) e (v), onde o sítio analisado passa a ser os novos sítios cinzas, checando seus vizinhos, até que todos os N sítios sejam analisados e pintados de pretos.

Para o caso em que um sítio não tenha vizinhos, podemos expressar os vizinhos desse sítio por $\Pi[u] = \text{NIL}$, um exemplo é caso onde $u = s$, isto é, a vizinhança do sítio fonte com ele mesmo.

2.1.3.1 Fenômeno de mundo pequeno

O conceito por trás do fenômeno de mundo pequeno está relacionado com a concepção de que numa rede complexa, há “pequenos caminhos” conectando dois sítios quaisquer (EASLEY; KLEINBERG, 2010). Esse fenômeno também é comumente conhecido como *seis graus de separação*, servindo de título para o livro de John Guare, tendo a frase dita nesse livro: “Eu li em algum lugar que todo mundo no planeta está separado apenas por seis outras pessoas. Seus graus de separação entre nós e qualquer outra pessoa no planeta”. Essa frase reflete a essência desse fenômeno. Tendo a rede social de todo mundo $N \sim 10^9$ indivíduos, nós teremos $\langle \ell \rangle$ muito pequeno, se comparado com N .

O primeiro experimento que tentou explicar esse fenômeno foi realizado pelo psicólogo social norte-americano, Stanley Milgran (MILGRAM, 1967). Nesse experimento, Stanley deu a um indivíduo um pacote. O objetivo era que esse pacote saísse desse indivíduo fonte e chegasse até outra pessoa alvo, sendo que ambos não se conheciam. O processo ocorria, com o indivíduo inicial enviando esse pacote para um conhecido seu, que ele julgasse estar mais perto do alvo, e esse, repetia o procedimento com outro. Esse ciclo era repetido até que o pacote fosse entregue ao alvo.

Baseando-se nesse fenômeno de mundo pequeno, Strogatz propôs um algoritmo para construção de redes com essa propriedade (STROGATZ, 1998). O algoritmo fundamenta-se na realização de uma interpolação de uma rede no formato de anel regular atingindo uma rede totalmente aleatória. O processo ocorreu por meio da escolha de um sítio da rede, selecionando

a ligação com o sítio mais próximo dele no sentido horário. Com uma probabilidade p , essa ligação sofre um processo de realocação com outro sítio da rede, selecionado de forma aleatória. Esse processo é repetido até percorrer todos os sítios da rede. Esse método de construção de rede ficou conhecido como *Modelo de Watts e Strogatz*, tendo a rede comumente representação na literatura pelo símbolo $G(N, k, p)$, onde N expressa o número de sítios da rede, $k = \langle k \rangle$ o grau médio da ligação entre eles e p a probabilidade de realocar a ligação.

Figura 3 – Exemplificação do processo de interpolação para o caso de $N = 20$ sítios e $\langle k \rangle = 4$. A rede inicia-se no formato regular regular, com os sítios conectados aos seus $\langle k \rangle = 4$ vizinhos mais próximos. É escolhido um sítio na rede, selecionando a ligação com seu sítios mais próximo no sentido horário. Com uma probabilidade p , essa ligação é realocada para outro sítio aleatório da rede. Esse processo é repetido até percorrer todos os sítios da rede. Podemos observar que para $p = 0$ a rede permanece a mesma, enquanto para o caso $p = 1$ todos os sítios são realocados, tornando a rede aleatória. Tendo as redes com fenômeno de mundo pequeno ocorrem num ponto de transição em que $p \in (0, 1)$.



Fonte: elaborado pelo autor.

Foi observado que $\langle \ell \rangle$ e $\langle C \rangle$ estão relacionados a N e $\langle k \rangle$ da seguinte forma:

$$p \rightarrow 0 \Rightarrow \begin{cases} \langle \ell \rangle \sim \frac{N}{2\langle k \rangle}, \\ \langle C \rangle \sim \frac{3}{4}. \end{cases} \quad (2.11)$$

$$p \rightarrow 1 \Rightarrow \begin{cases} \langle \ell \rangle \approx \langle \ell_{random} \rangle \sim \frac{\ln(N)}{\ln(\langle k \rangle)}, \\ \langle C \rangle \approx \langle C_{random} \rangle \sim \frac{\langle k \rangle}{N}. \end{cases} \quad (2.12)$$

A dependência de $\langle \ell \rangle$ com $\ln N$ para o caso das redes aleatórias reflete a característica fundamental por detrás do fenômeno de mundo pequeno. Para rede com muitos sítios, $N \gg 1$, nós temos que $N \gg \ln N$, isso implica que $\langle \ell \rangle$ pode ser ordens de magnitude menor do que o tamanho da rede. A relação $\langle \ell \rangle$ expressa na Eq. (2.12) pode ser entendida para o diâmetro da rede (BARABÁSI, 2013):

$$d_{max} \sim \frac{\ln(N)}{\ln(\langle k \rangle)}, \quad (2.13)$$

onde os valores obtidos a partir da Eq. (2.13) possuem uma pequena discrepância quando comparado com o caso de $\langle \ell \rangle$ devido ao valor de d_{max} representar um pequeno conjunto de valores de menores caminhos médio da rede.

2.1.4 Assortatividade

A assortatividade, assim como a distribuição de graus, o menor comprimento do caminho médio e o diâmetro, é uma das propriedades de interesse no estudo das redes. Essa propriedade está relacionada com a correlação entre os sítios da rede (NEWMAN, 2003), sendo uma métrica que representa a tendência de sítios um “tipo” se conectarem entre sítios de mesmo “tipo” (FOSTER *et al.*, 2011).

O conceito de “tipo” de sítio varia de acordo com a rede estudada. Numa rede social, podemos ter o caso de uma rede de relacionamentos amorosos entre indivíduos de diferentes etnias, com os diferentes tipos de sítios representando as diferentes etnias. No nosso caso, os tipos de sítios serão dados por sítios de diferentes graus.

A forma de mensurar a assortatividade da rede é através do coeficiente de assortatividade, comumente representado pela letra R na literatura. O cálculo desse coeficiente pode ser feito de duas forma, através da caracterização discreta e da caracterização por propriedades escalares da rede. Na caracterização discreta é levado em consideração propriedades que distingue ou separa um tipo de sítio, sendo essa propriedade inerente aos sítios, como por exemplo, numa rede de relacionamentos, podemos levar em consideração a linguagem ou a etnia. Enquanto na caracterização por propriedades escalar levamos em consideração uma ou mais propriedades escalar dos sítios da rede, como por exemplo, o grau de um sítio.

2.1.4.1 Caracterização discreta

Na caracterização discreta, a assortatividade pode ser descrita pela quantidade $e_{\alpha\beta}$, onde essa medida representa a fração de ligações da rede que conecta os sítios de tipo α com o tipo β . Numa rede não-direcionada teremos $e_{\alpha\beta} = e_{\beta\alpha}$. Além dessa medida, ainda temos as medidas a_α e b_β , representando a soma das frações de ligações que terminam no tipo de sítio α e β , respectivamente (NEWMAN, 2003).

Vale observar que, por $e_{\alpha\beta}$ representar frações das ligações da rede, a soma sobre todos os valores de α e β obrigatoriamente deve ser 1, e, a partir da definição descrita acima de a_α e b_β , nós teremos também:

$$\sum_{\alpha,\beta} e_{\alpha\beta} = 1, \quad \sum_{\beta} e_{\alpha\beta} = a_\alpha, \quad \sum_{\alpha} e_{\alpha\beta} = b_\beta. \quad (2.14)$$

Partindo dessas propriedades é possível calcular o coeficiente de assortatividade por meio da equação:

$$R = \frac{\sum_{\alpha} e_{\alpha\alpha} - \sum_{\alpha} a_{\alpha} b_{\alpha}}{1 - \sum_{\alpha} a_{\alpha} b_{\alpha}}, \quad (2.15)$$

o valor do coeficiente de assortatividade, R , determina se a rede possui ou não assortatividade, e, caso possua, qual o tipo de assortatividade ela tem. Caso os sítios de um tipo conectem-se mais a sítios de mesmo tipo, dizemos que a rede é assortativa. Caso os sítios de um tipo conectem-se mais a sítios de diferentes tipos, dizemos que a rede é dessortativa. E, por fim, caso os sítios de um tipo não possuam nenhuma tendência preferencial quanto ao tipo de sítio que ele vá se conectar, dizemos que a rede é não-assortativa. As redes podem ser classificadas de três formas com relação a assortatividade: rede assortativa, dessortativa e não-assortativa, tendo cada caso classificado abaixo:

1. Todos os sítios se conectam somente com mesmo tipo (rede completamente assortativa).

Nesse caso nós temos ($e_{\alpha\beta} = 0, \forall \alpha \neq \beta$), nos dando a seguinte matriz

$$e = \begin{pmatrix} e_{11} & 0 & 0 \\ 0 & e_{22} & 0 \\ 0 & 0 & e_{33} \end{pmatrix}, \quad (2.16)$$

observamos que $\sum_{\alpha} e_{\alpha\alpha} = 1$, reduzindo a Eq. (2.15) a forma

$$R = R_{max} = \frac{1 - \sum_{\alpha} a_{\alpha} b_{\alpha}}{1 - \sum_{\alpha} a_{\alpha} b_{\alpha}} = 1, \quad (2.17)$$

nos dando a situação onde $R = R_{max} = 1$, a rede é completamente assortativa.

2. Os sítios conectam-se apenas a sítios de tipos diferentes (rede totalmente dessortativa).

Nesse caso nós temos ($e_{\alpha\beta} = 0, \forall \alpha = \beta$), nos dando a seguinte matriz

$$e = \begin{pmatrix} 0 & e_{12} & e_{13} \\ e_{21} & 0 & e_{23} \\ e_{31} & e_{32} & 0 \end{pmatrix}, \quad (2.18)$$

observamos que $\sum_{\alpha} e_{\alpha\alpha} = 0$, reduzindo a Eq. (2.15) a forma

$$R = R_{min} = -\frac{\sum_{\alpha} a_{\alpha} b_{\alpha}}{1 - \sum_{\alpha} a_{\alpha} b_{\alpha}}. \quad (2.19)$$

Tendo que $a_{\alpha} b_{\alpha}$ são frações com soma de valor menor do que 1, o valor de nosso coeficiente de assortatividade, R , será negativo. Tendo uma matriz não direcionada, o menor valor possível que nós podemos ter é o caso de apenas dois sítios conectados, sendo eles de tipos distintos, dando por exemplo, a matriz expressa abaixo

$$e = \begin{pmatrix} 0 & e_{12} & 0 \\ e_{21} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (2.20)$$

numa rede não-direcionada, nós temos que $e_{\alpha\beta} = e_{\beta\alpha}$, nos dando assim

$$\begin{aligned} \sum_{\alpha} a_{\alpha} b_{\alpha} &= a_2 b_2 + a_3 b_3, \\ \sum_{\alpha} a_{\alpha} b_{\alpha} &= \frac{1}{4} + \frac{1}{4}, \end{aligned}$$

$$R_{min} = -\frac{\frac{1}{2}}{1 - \frac{1}{2}} = -1, \quad (2.21)$$

de forma geral, tendo a fração de sítios conectados a sítios de diferente tipos, valores iguais, isso nos dará, para um conjunto de M diferentes tipos de sítios

$$\sum_{\alpha} a_{\alpha} b_{\alpha} = (e_{12} + \dots + e_{1M})(e_{21} + \dots + e_{M1}) + \dots + (e_{M1} + \dots + e_{MM-1})(e_{1M} + \dots + e_{M-1M}), \quad (2.22)$$

sendo uma rede não-direcionada $e_{\alpha\beta} = e_{\beta\alpha}$, nos dando

$$\sum_{\alpha} a_{\alpha} b_{\alpha} = (e_{12} + \dots + e_{1M})^2 + \dots + (e_{M1} + \dots + e_{MM-1})^2, \quad (2.23)$$

as expressões dentro dos parênteses possuem $M - 1$ termos (tiramos a diagonal principal para o caso das redes completamente dessortativas), tendo M expressões. Como todas as frações $e_{\alpha\beta}$ são igualmente prováveis, teremos $e_{\alpha\beta} = 1/(M^2 - M)$, pois ela é quadrada de ordem M e removemos os M termos da diagonal principal, com isso, teremos

$$\sum_{\alpha} a_{\alpha} b_{\alpha} = M \left((M - 1) \frac{1}{M^2 - M} \right)^2, \quad (2.24)$$

usando a Eq. (2.24) na Eq. (2.19), nós obtemos

$$R = \begin{cases} -\frac{1}{M} & , \forall M > 2, \\ -1 & , M = 2. \end{cases} \quad (2.25)$$

A Eq. (2.25) nos diz que, se R estiver no intervalo $\in [-1, 0)$, a rede é completamente dessortativa. Vemos que uma rede completamente dessortativa, quando descrita pela caracterização discreta, ocorre num intervalo de valores de R , diferente do caso da rede completamente assortativa, que se dá no valor exato de $R = 1$.

Na matriz e , que representa o conjunto de elementos $e_{\alpha\beta}$, pode possuir cada vez mais valores, conforme nos aumentamos o número de diferentes tipos de sítios, ou seja, a ordem da matriz e , representada pela letra M . Para o caso da rede completamente assortativa, esse aumento não impacta no valor de R , pois haverá apenas elementos na diagonal principal, e, a soma dessa diagonal principal sempre será igual a 1. Contudo, como podemos observar no cálculo de R para uma rede completamente dessortativa expressa na Eq. (2.25), o aumento do número de tipos de ligações impacta diretamente no cálculo de R , e, mesmo com valores distintos, todos os casos representam uma rede completamente dessortativa. Para o caso da matriz representada na Eq. (2.18) nós temos $R = -0.5$, enquanto para rede representada na matriz expressa na Eq. (2.16), nós temos $R = -1$, sendo a matriz mais simples para rede completamente dessortativa.

3. Os sítios não possuem viés de conexão quanto ao tipo de sítio (rede não-assortativa).

Para esse caso, nós temos que ($e_{\alpha\beta} = a_{\alpha} b_{\beta}$). A partir dessa condição, vemos que nosso R , dado pela Eq. (2.15) será igual a 0.

Com todos esses casos, nós chegamos a seguinte conclusão

$$R = \begin{cases} x \in [-1, 0) & \implies \text{Rede completamente dessortativa,} \\ x = 0 & \implies \text{Rede não-assortativa,} \\ x \in (0, 1) & \implies \text{Rede assortativa,} \\ x = 1 & \implies \text{Rede completamente assortativa.} \end{cases} \quad (2.26)$$

Outra medida importante que nós podemos obter acerca da rede é o desvio padrão médio de R . Uma das formas de calcular essa medida é através do método *jackknife*. Esse método consiste em analisar uma amostragem do nosso conjunto de dados, retirando um elemento e analisando os elementos restantes. Nos repetimos esse processo n vezes, sendo n o tamanho da nossa mostragem (EFRON, 1979), para o nosso caso, nós teremos

$$\sigma_R^2 = \sum_{i=1}^M (R_i - R)^2, \quad (2.27)$$

onde R_i é o valor de R , com a i -ésima ligação removida, sendo M sendo o número de ligações da rede. Existe uma similaridade matemática entre o coeficiente R e o coeficiente de correlação interclasses usado na física estatística, que serve para comparar medidas entre diferentes grupos de classes, nós podemos adaptar o cálculo desse desvio através da relação expressa abaixo (NEWMAN, 2003) (FLEISS *et al.*, 1969) (FLEISS *et al.*, 2013):

$$\sigma_R^2 = \frac{1}{M} \frac{\sum_{\alpha} a_{\alpha} b_{\alpha} + \left[\sum_{\alpha} a_{\alpha} b_{\alpha} \right]^2 - \sum_{\alpha} a_{\alpha}^2 b_{\alpha} - \sum_{\alpha} a_{\alpha} b_{\alpha}^2}{1 - \sum_{\alpha} a_{\alpha} b_{\alpha}}. \quad (2.28)$$

2.1.4.2 Assortatividade por propriedades escalares

No caso da assortatividade por propriedades escalares, o cálculo da assortatividade depende de uma ou mais propriedades escalares dos sítios da rede (NEWMAN, 2003). Nesse caso, o coeficiente de assortatividade é calculado através do coeficiente de correlação de Pearson (BOSLAUGH, 2012), dado por

$$R_{xy} = \frac{\langle x \rangle \langle y \rangle - \langle xy \rangle}{\sigma_x \sigma_y}, \quad (2.29)$$

com $\langle x \rangle$, $\langle y \rangle$ e $\langle xy \rangle$ sendo as médias do conjunto de medidas x , y e do produto xy , e, σ_x e σ_y sendo o desvio padrão das medidas x e y . Podemos expressar a Eq. (2.29) em termos das frações $e_{\alpha\beta}$ e da somas das frações a_{α} , b_{β} como definidos anteriormente:

$$\langle x \rangle = \sum_{\alpha} x_{\alpha} e_{\alpha\beta}, \quad \langle y \rangle = \sum_{\beta} y_{\beta} e_{\alpha\beta}, \quad (2.30)$$

tendo que as medidas de x e y são variáveis independentes, podemos junta-las numa única soma

$$\langle x \rangle \langle y \rangle = \sum_{\alpha, \beta} x_i y_j e_{\alpha\beta}, \quad (2.31)$$

além disso, nós temos que a_α e b_β representam a soma das frações dos sítios terminados em i e j , respectivamente, com isso

$$\langle xy \rangle = \sum_{\alpha, \beta} a_\alpha b_\beta x_\alpha y_\beta, \quad (2.32)$$

com isso, podemos reescrever a expressão dada pela Eq. (2.29) na forma

$$R_{xy} = \frac{\sum_{\alpha, \beta} x_\alpha y_\beta (e_{\alpha\beta} - a_\alpha b_\beta)}{\sigma_x \sigma_y}. \quad (2.33)$$

Para que possamos utilizar a Eq. (2.33) para calcular a assortatividade por graus dos sítios, tomaremos $e_{\alpha\beta}$ como sendo a fração de ligações que conectam os sítios de grau i aos sítios de grau j . As grandezas a_i e b_j serão relacionadas com a distribuição de graus da rede, para isso, lembramos que a distribuição de graus da rede representa a probabilidade de selecionarmos aleatoriamente um sítio na rede de grau k . Para que possamos relacionar essas medidas, devemos nos atentar aos *graus remanescentes* dos sítios, que representa o número de ligações que saem dos sítios diferentes do sítio analisado no momento. Ou seja, caso estejamos observando o sítio 1 na rede representada na Figura 2, teremos do sítio 2 as ligações $L_1 = \{\{2,1\}, \{2,4\}\}$, do sítio 3 as ligações $L_3 = \{\{3,1\}, \{3,4\}\}$ e do sítio 0 a ligação $L_0 = \{0,1\}$. Essa nova distribuição é igual ao valor um a menos do que a distribuição total de graus p_k , logo, ela deve ser proporcional a $(k+1)p_{k+1}$ (NEWMAN, 2002). Normalizando essa distribuição, nós obtemos

$$q_k = \frac{(k+1)p_{k+1}}{\sum_j j p_j}, \quad (2.34)$$

com $\sum_j j p_j$ sendo o termo de normalização.

Nós podemos expressar a fração $e_{\alpha\beta}$ em termos do grau de um sítio analisado, k , obtendo assim a fração de sítios remanescentes e_{ik} , obedecendo as regras

$$\sum_{jk} e_{jk} = 1 \quad \text{e} \quad \sum_j e_{jk} = q_k, \quad (2.35)$$

tendo as condições acima, podemos observar que

$$a_i = q_i, \quad b_j = q_j, \quad (2.36)$$

fixando $i = k$ para que possamos expressar r em termo dos graus remanescentes, teremos

$$R = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k), \quad (2.37)$$

tendo o desvio padrão como definido na Eq. (2.28), após as devidas manipulações algébricas, nós podemos reescrever nosso R na forma

$$R = \frac{\sum_i^M j_i k_i - M^{-1} [\sum_i^M \frac{1}{2} (j_i + k_i)]^2}{\sum_i^M \frac{1}{2} (j_i^2 + k_i^2) - M^{-1} [\sum_i^M (j_i + k_i)]^2}, \quad (2.38)$$

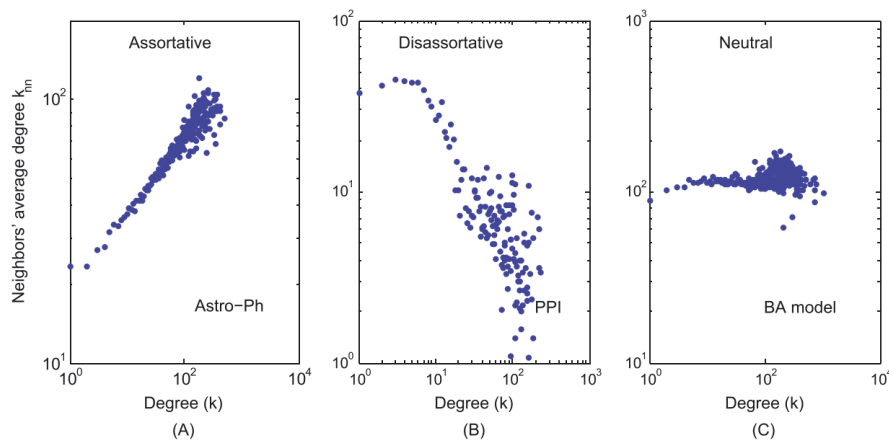
onde j_i, k_i são os graus dos sítios que terminam na i -ésima ligação, com $i = 1, \dots, M$.

Partindo da Eq. (2.37), nós podemos também calcular o grau médio dos sítios da vizinhança, k_{nn} , definido da forma (JIANG *et al.*, 2017)

$$k_{nn} = \sum_j j p_{j|k}, \quad (2.39)$$

onde $p_{j|k}$ é a probabilidade condicional de uma ligação pertencente a um sítio de grau k esteja conectado com um sítio de grau j . Analisando as Eqs. (2.38), (2.39), podemos observar que se existir uma correlação positiva de k_{nn} com k , o valor de R será positivo, nos dando uma rede assortativa ($0 < R \leq 1$), se a correlação for negativa nós teremos uma rede dessortativa ($-1 \leq R < 0$), se não houver uma correlação, nós teremos uma rede não-assortativa ($R \approx 0$).

Figura 4 – k_{nn} representa o grau médio dos sítios da vizinhança e k o grau dos sítios. Na imagem (A) nós temos o comportamento da curva para uma rede assortativa ($0 < R \leq 1$), representando a rede de afiliação na categoria Astro-ph do arXiv (LESKOVEC *et al.*, 2007). Na imagem (B) para uma rede dessortativa ($-1 \leq R < 0$), representando a rede de interação proteína-proteína (PPI, do inglês *protein-protein interactions*) (JEONG *et al.*, 2001). E na imagem (C) de uma rede não-assortativa ($R = 0$), representando o modelo de Barabási (BARABASI; ALBERT, 1999).



Fonte: Jiang, Jiaojiao, et al (JIANG *et al.*, 2017).

3 CONCEITOS DE FÍSICA ESTATÍSTICA

Para iniciarmos as fundamentações acerca da física estatística, devemos fundarmos as bases para a análise de quaisquer sistemas mecânico estatísticos em equilíbrio, podendo ser divididos em três etapas (SALINAS, 2008):

1. Caracterizar os estados microscópios do sistema, tendo o conjunto de estados definindo um *ensemble* estatístico;
2. Definir um postulado estatístico básico acerca *ensemble* e utilizar a teoria das probabilidades;
3. Estabelecer uma conexão com a termodinâmica, ou seja, definir de quais formas aquelas variáveis que descreve estados microscópicos (microestado) afetam os estados macroscópicos (macroestado), sendo esses estados categorizados por variáveis termodinâmicas.

Nós podemos estudar um sistema físico de partículas, com as leis da mecânica fornecendo-nos meios para que possamos determinar os conjuntos de estados microscópicos (*ensemble*). Contudo, para alguns casos é mais conveniente construirmos modelos específicos para lidar com o problema analisado em questão, como é o caso do *ensemble* microcanônico, onde expressa o sistema possui energia total fixa, e, com isso é proposto probabilidades iguais a priori. Ou seja, todos os microestados do sistemas são igualmente prováveis (SALINAS, 2008).

A diferença entre **microestado** e **macroestados** pode ser melhor compreendida com exemplo abaixo:

Suponhamos que um indivíduo coloque dentro de uma caixa um total de 50 dados idênticos, cada um deles com seis lados. Ele fecha a caixa com uma tampa e chacoalha ela por um período. Após isso, ele abre a caixa e observa a disposição dos dados, no caso, quais faces estão viradas para cima de cada um dos dados. Como a probabilidade dos dados estarem virados para cima são independentes, nós temos que o número possível de configurações possíveis para esse sistema é 6^{50} , pois cada dado possui 6 faces e há um total de 50 dados. Assumindo que todas as configurações são igualmente prováveis, nós temos que a probabilidade de obter uma dada configuração é simplesmente 6^{-50} . Cada um dessas configurações é conhecida como **microestado** desse sistema. A Figura 5 representa dois dos 6^{50} microestados do sistema. Nós podemos rotular cada um desses dados e checar em cada um dos microestados, qual valor cada um desses dados rotulados foi tirado no dado, ou, de forma mais eficiente, podemos contar o número de dados com os diferentes valores foi resultado. Essa forma de categorizar os dados é chamado de **macroestado**. Esse valor pode ser mensurado através de conceitos fundamentais da

probabilidade. Tendo:

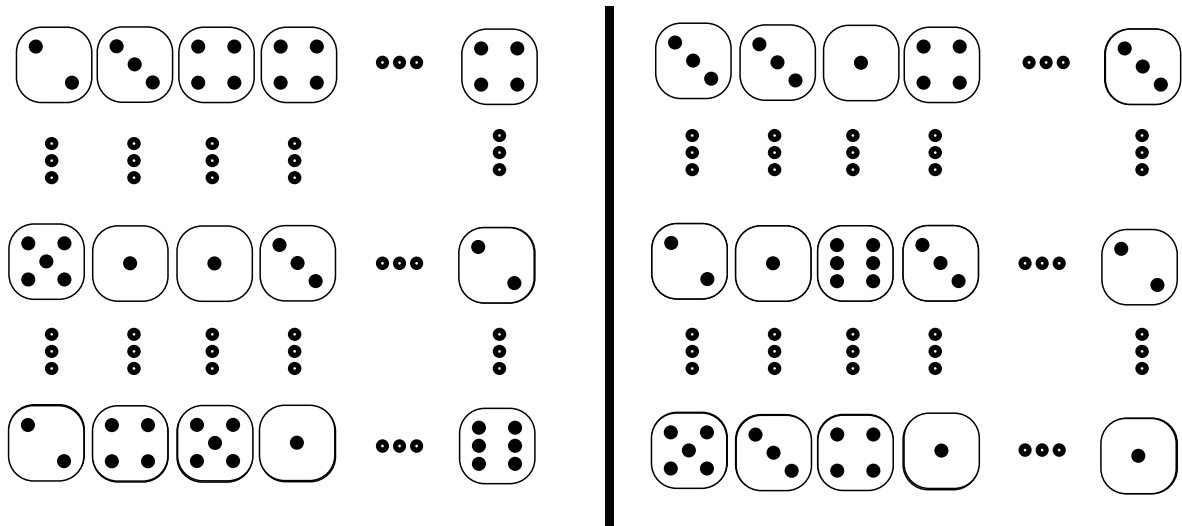
$$N = \sum_i^6 N_i = 50, \quad (3.1)$$

com N representando o número de dados e N_i representando o número de dados com valor $i = 1, 2, 3, 4, 5, 6$ voltado para cima. Assim, nós teremos que a configuração que representa o número N de dados virados para cima é dada por

$$\text{Número de dados com valor } N_i \text{ voltados para cima} = \frac{N!}{N_1!N_2!N_3!N_4!N_5!N_6!}, \quad (3.2)$$

com $i = (1, 2, 3, 4, 5, 6)$. Esse valor representando o **macroestado** do sistema.

Figura 5 – Do lado esquerdo nós temos um possível microestado do sistema, enquanto do lado direito temos outro microestado do mesmo sistema.



Fonte: elaborado pelo autor.

3.1 Temperatura, entropia e ensambles

Para que possamos estabelecer uma conexão da física estatística com a termodinâmica, primeiramente iremos precisar definir alguns conceitos fundamentais da termodinâmica. Além de fixarmos nossa atenção a sistemas em equilíbrio termodinâmico, ou seja, as propriedades macroscópicas não sofrem alterações com o tempo.

Para estudarmos sistemas físicos da perspectiva mecânica, devemos focar nossa atenção acerca da dinâmica que regem os objetos de interesse. Na termodinâmica, essa forma de trabalhar torna-se irracional, pois sistemas termodinâmicos apresentam um número de partículas da ordem da Avogrado ($N \sim 10^{24}$), caso trabalhemos da mesma forma como é trabalho na

mecânica, teríamos que encontrar a equação de movimento de todas as partículas, além de dar um significado a essas soluções (NUSSENZVEIG, 2006).

Nós podemos estudar a termodinâmica sobre duas perspectivas diferente: fenomenológica e postulada (estatística) (KITTEL; KROEMER, 1980). Iremos focar nossa atenção na perspectiva postulada, sendo baseada nos quatro postulados (CALLEN, 1985):

1. Existe um estado particular (chamado de estado de equilíbrio) do sistema simples que, macroscopicamente, são caracterizados completamente pela energia interna U , o volume V , e o número de mols N_1, N_2, \dots, N_r , dos componentes químicos;
2. Existe uma função (chamada entropia S) de parâmetros extensivos de qualquer composição do sistema, definida para todos os estados de equilíbrio e tem a seguinte propriedade: os valores assumidos pelos parâmetros extensivos na ausência de um vínculo interno são aqueles que maximizam a entropia sobre a variedade de estados de equilíbrio;
3. A entropia de um sistema composto é aditiva sobre os subsistemas constituintes. A entropia é contínua e diferencial e é uma função monotonicamente crescente com a energia interna;
4. A entropia de qualquer sistema é nula no estado para os quais

$$(\partial U / \partial S)_{V, N_1, \dots, N_r} = 0, \quad (3.3)$$

ou seja, na temperatura nula.

Partindo do primeiro postulado nós podemos encontrar o conceito de temperatura, que representa a média da energia interna das partículas do sistema, sendo representada por T , com o primeiro postulado podendo ser simplificado por: sistemas em equilíbrio termodinâmico, possuem temperaturas iguais.

A função entropia descrita no segundo postulado é essencial para o estudo de quaisquer sistemas termodinâmicos, estando associada a medida de desordem ou aleatoriedade em um sistema. É uma grandeza termodinâmica que descreve a quantidade de energia não disponível para realizar trabalho em um sistema ou, de maneira equivalente, a quantidade de caos ou desordem molecular em um sistema.

A conjectura fundamental da mecânica estatística estabelece que todos os estados microscópicos acessíveis a um sistema fechado em equilíbrio são igualmente prováveis (SALINAS, 2008). Baseando-se nisso, nós podemos definir uma relação direta entre a probabilidade $P(\{X_i\})$ de encontrar um sistema sujeito a um conjunto de vínculos $\{X_i\}$ e um sistema cujo número de microestados depende de E , V e N na forma:

$$P(\{X_i\}) \propto \Omega(E, V, E; \{X_i\}), \quad (3.4)$$

ou seja, a probabilidade é proporcional ao número de microestados acessíveis do sistema. O desenvolvimento estatístico é realizado partindo da relação dada pela Eq. (3.4).

Para simplificar nosso estudo, iremos trabalhar com sistemas simples, ou seja, sistemas contendo um único tipo de átomo, fazendo com que o número de mols expresso no postulado 2 seja resumido a N . No equilíbrio termodinâmico, todos os microestados são igualmente prováveis, caso nós tenhamos um sistema caracterizado pelos parâmetros E , V e N , sendo E a energia, V o volume e N o número de partículas, com E fixo, o conjunto de microestados acessíveis caracteriza o ensemble microcanônico. Partindo do segundo postulado, nós podemos realizar a conexão entre esses microestados e a entropia por meio da relação (SALINAS, 2008)

$$S(E, V, N) = k_B \ln \Omega(E, V, N), \quad (3.5)$$

sendo k_B a constante de Boltzmann e $\Omega(E, V, N)$ o número de microestados acessíveis do sistema. Nós podemos expressar a Eq. (3.5) numa forma mais generalizada na forma

$$S = k_B \ln \Omega, \quad (3.6)$$

onde S e Ω são funções das mesmas variáveis, que dependem dos vínculos expressos no sistema, com esses vínculos caracterizando o tipo de ensemble de nosso interesse. A Eq. (3.6) é conhecida como equação de Boltzmann.

3.2 Estatística de Boltzmann-Gibbs

Como pode ser observado pela Eq. (3.4), há uma relação direta entre os microestados do sistema e a probabilidade de encontrarmos o sistema naquele microestado. Essa relação é válida para quaisquer ensembles, ou seja, dado quaisquer sistemas termo-estatísticos, nós teremos sempre uma relação direta entre os microestados e as probabilidades associadas a ele. Baseando-se nisso, nós podemos escrever a entropia do sistema em termos da probabilidade.

Para o caso do ensemble microcanônico apresentado na seção anterior, nós teremos o vínculo da energia total ser fixa, ou seja, dado dois subsistemas isolados termicamente, caso a parede que separa os dois sistemas torne-se diatérmica, permitindo que a transição de energia entre ambos os sistemas, mas mantendo o volume e o número de partículas constante, a energia total do sistema será dada por

$$E_0 = E_1 + E_2 \implies E_1 = E_0 - E_2, \quad (3.7)$$

onde E_1 e E_2 representam a energia do sistema 1 e 2, respectivamente e E_0 a energia total do sistema, sendo constante.

Sendo o microestado de ambos os sistemas independentes, o microestado do sistema composto é dado pelo produto de cada um

$$\Omega = \Omega_1(E_1, V_1, N_1)\Omega_2(E_2, V_2, N_2), \quad (3.8)$$

como o volume e o número de partículas não varia, tendo a relação entre as energias, nós podemos simplificar a Eq. (3.8)

$$\Omega(E_1; E_0) = \Omega_1(E_1)\Omega_2(E_0 - E_1). \quad (3.9)$$

Utilizando a Eq. (3.4), nós teremos

$$P(E_1) = c\Omega(E_1, E_0) = c\Omega_1(E_1)\Omega_2(E_0 - E_1), \quad (3.10)$$

tendo assim a constante c que normaliza P como sendo

$$c = \frac{1}{\Omega_c} = \frac{1}{\Omega}, \quad (3.11)$$

sendo $\Omega_c = \Omega$ o microestado acessível do sistema composto, sendo uma função unicamente da energia.

Com isso, nossa entropia no caso do ensemble microcanônico fica na forma

$$S = -k_B \sum_j P_j \ln P_j, \quad (3.12)$$

para o caso onde há mais de um subsistema, com a soma sobre os todos microestados do sistema. A Eq. (3.12) é conhecida como equação de Boltzmann-Gibbs, e, partindo dela nós podemos derivar todas as outras grandezas termodinâmicas.

A equação de Boltzmann-Gibbs nos mostra a relação entre a entropia e a incerteza ou à falta de conhecimento sobre o estado microscópico exato do sistema. Quando ela é uniforme (estados igualmente prováveis), a entropia é máxima. Quando a distribuição está concentrada num único microestado (indicando incerteza sobre o estado do sistema), a entropia é mínima.

Como a estatística de Boltzmann-Gibbs descreve sistemas cuja energia total é fixa, ela é bastante útil para descrever sistemas homeostáticos (ecoequilíbrio de matéria) e isovolumétricos (invariância no volume).

3.3 Estatística de Tsallis

A entropia de Tsallis foi proposta por Constantino Tsallis como uma possível generalização a entropia de Boltzmann-Gibbs. Tsallis postulou sua entropia por utilizando o parâmetro q que é uma quantidade normalmente escalada em multifractais (TSALLIS, 1988). A entropia postulada por ele tem a forma:

$$S_q \equiv k \frac{1 - \sum_{i=1}^W p_i^q}{q-1}, \quad (3.13)$$

onde $q \in \mathbb{R}$, p_i são as probabilidades associadas a algum evento, $W \in \mathbb{N}$ representa o número total de possíveis configurações e k é uma constante real positiva conveniente ao problema. Nós iremos inicialmente definir as seguintes funções:

$$\ln_q(x) \equiv \frac{x^{1-q} - 1}{1-q} \quad (x > 0; \ln_1(x) = \ln(x)), \quad (3.14)$$

$$e_q^x \equiv [1 + (1-q)x]_+^{1/(1-q)} \quad (e_q^x = e^x), \quad (3.15)$$

sendo essas funções generalizações do logaritmo natural e da exponencial, respectivamente, sendo chamadas de q-logaritmo e q-exponencial.

Nós podemos reescrever a Eq. (3.13) na forma

$$S_q = k \frac{\sum_i p_i - \sum_i p_i^q}{(q-1)} = -k \sum_i \frac{p_i(p_i^{1-q} - 1)}{1-q} = -k \sum_i p_i^q \ln_q(p_i). \quad (3.16)$$

Partindo da Eq. (4.21) podemos facilmente perceber que no limite onde q tende a 1, nós obtemos a entropia de Boltzmann-Gibbs expressa na Eq. (3.12). Tendo a função que otimiza, sujeita as vínculos Eq. (V.1) e Eq. (V.2)

$$\sum_{i=1}^W p_i = 1, \quad (V.1)$$

$$\sum_{i=1}^W p_i \varepsilon_i = U_q, \quad (V.2)$$

a forma

$$p_i = \frac{[1 - \beta(q-1)\varepsilon_i]^{1/(q-1)}}{Z_q} = \frac{e_q^{\beta\varepsilon_i}}{Z_q}, \quad (3.17)$$

onde ε_i representa a energia associada ao i -ésimo evento, β está associado ao parâmetro dos multiplicadores de Lagrange e Z_q é a função de partição, que normaliza p_i , sendo dada por

$$Z_q \equiv \sum_{i=1}^W e_q^{\beta \varepsilon_i}. \quad (3.18)$$

Essa generalização proposta por Tsallis têm obtido sucesso em várias aplicações na matéria condensada em átomos frios (LIU; GOREE, 2008), vidro de spin (PICKUP *et al.*, 2009), sistemas biológicos (UPADHYAYA *et al.*, 2001) e supercondutores do tipo II (ANDRADE *et al.*, 2010).

4 O MODELO DE BARABÁSI-ALBERT

O modelo de Barabási-Albert foi desenvolvido com o intuito de estudar redes dinâmicas, cujo crescimento se dá forma preferencial. Para implementar essa característica na rede, Barabási propôs um mecanismo chamado de conexão preferencial, onde os novos sítios adicionados a rede possuíam uma maior probabilidade de conectar-se aos sítios com muitas conexões. Baseando-se nessa característica, Samuráí *et. al* escreveu uma modificação dessa conexão preferencial, possuindo pesos das distâncias entre os sítios (BRITO *et al.*, 2016).

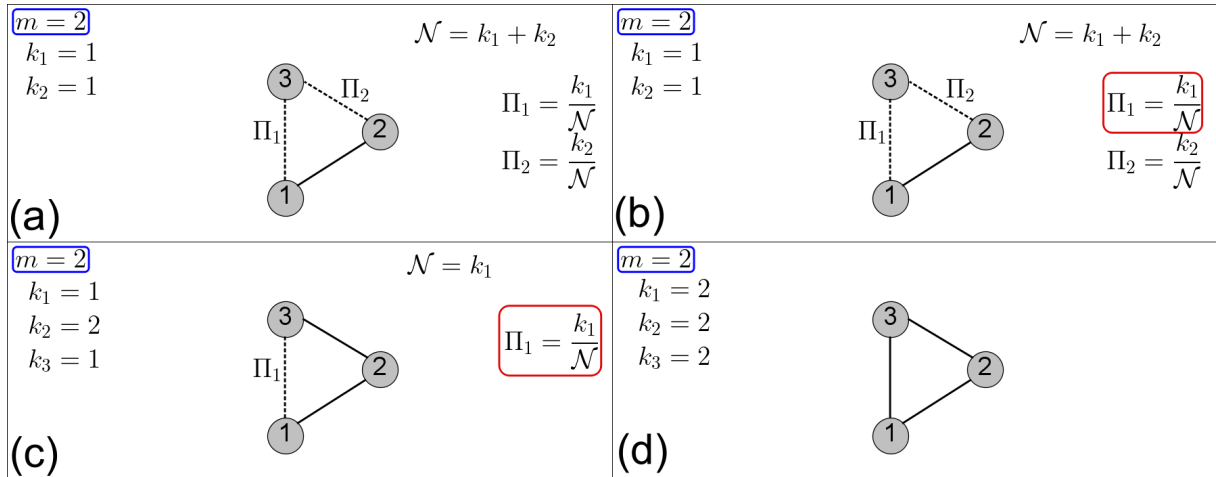
Para implementação do modelo, a rede inicial começa com m_0 sítios, o número de ligações iniciais pode ser escolhido de forma arbitrária, seguindo a restrição $m \leq m_0$. O desenvolvimento da rede segue duas etapas (BARABÁSI, 2013): para cada interação adiciona-se um novo sítio a rede, e m ligações entre ele e os demais sítios da rede. A restrição para os valores de m é que $m \leq m_0$, essa restrição impede que dois sítios, A e B por exemplo, tenham duas ligações entre si, isto é, restringimos a rede apenas a redes simples¹ e sem *loops*. A probabilidade de ligação desse novo sítio com relação aos sítios já existentes é dada por:

$$\Pi_i = \frac{k_i}{\sum_j k_j}, \quad (4.1)$$

onde k_i representa o grau do i -ésimo sítio, Π_i representa a probabilidade do novo sítio conectar-se ao sítio i . Essa probabilidade é o que caracteriza a conexão preferencial, um dos mecanismo probabilístico existentes do qual faz surgir os *hubs* da rede. Os parâmetros do modelo de Barabási-Albert são portanto m_0 , m e N . A figura 6 que exemplifica o processo por trás do modelo de Barabási-Albert, enquanto na Figura 7 nós temos dois exemplos de rede de Barabási, com $N = 500$ e diferentes valores de m e m_0 .

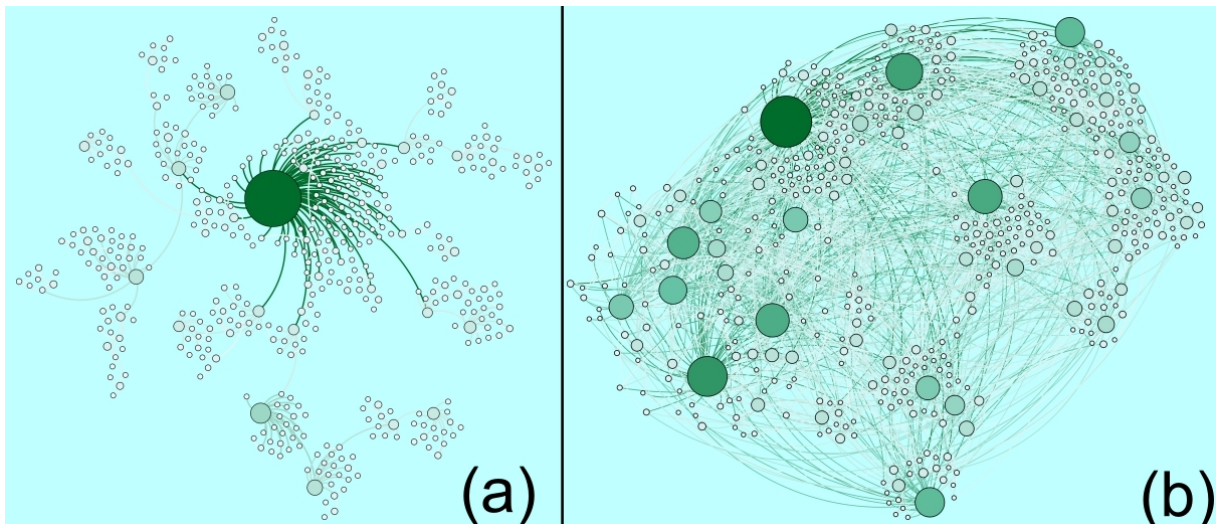
¹ Redes onde há apenas uma ligação entre cada um dos pares de sítios da rede.

Figura 6 – Os círculos enumerados cinza representam os sítios, com m representando o número de ligações que o novo sítio adicionado fará, $m_0 = 2$, \mathcal{N} a constante de normalização e k_i o grau do i -ésimo sítio. Na figura 6.(a) nós temos a rede inicial, com sítio 1 e 2 conectados. Na figura 6.(b) é realizado um sorteio do sítio que fará ligação com o novo sítio 3, tendo a probabilidade dada pela Eq (4.1), tendo sítio sorteado 1, realizamos uma ligação entre esses dois sítios. Na figura 6.(c) o sorteio é repetido, pois é necessário realizar $m = 2$ ligações. Na figura 6.(d) nós temos o novo sítio fazendo todas as ligações. O processo é repetido mais $N - 3$ vezes para obtermos uma rede com N sítios.



Fonte: elaborado pelo autor.

Figura 7 – Na figura abaixo nós temos dois exemplos de rede de Barabási. Os sítios são representados pelos círculos, a intensidade da cor e o tamanho do sítio representam seu grau. Na Figura 7.(a) a rede possui $m_0 = 2$ com $m = 1$, enquanto na Figura 7.(b) nós temos $m_0 = 4$ e $m = 3$. Como para representada em (b) possui valor de m maior, a rede possuirá mais ligações. Ambas as redes possuem $N = 500$ sítios.



Fonte: elaborado pelo autor.

4.1 Grau médio e distribuição de graus

De forma geral, a distribuição de graus de uma rede é dada pela Eq. (2.2), tendo como diferencial de cada rede a forma de N_k . Para o modelo de Barabási-Albert, nós podemos obter a distribuição de graus exata de forma analítica partindo da conexão preferencial dada pela Eq.(4.1) (BARABÁSI, 2013). A rede oriunda do modelo de Barabási é dinâmica, ou seja, ela cresce com o tempo, ciente disso e partindo da definição da distribuição de graus dada pela Eq. (2.2), nós teremos para o modelo de Barabási que o número de sítios com grau k é dado por $N(k, t)$ e o número de sítios total é dado por $N(t)$, onde a variável t representa o instante onde estamos calculando a distribuição de graus. A distribuição de graus será dada por:

$$P_k(t) = \frac{N(k, t)}{N(t)}. \quad (4.2)$$

A conexão preferencial pode ser escrita na forma

$$\Pi(k) = \frac{k}{\sum_j k_j}, \quad (4.3)$$

tendo o número de sítios totais após um tempo t dado por $N = t$, caso os novos sítios adicionados façam m ligações dentro da rede, nós teremos o número de ligações após t instantes de tempo dada por $2mt$, com o fator 2 representando o fato da rede ser não-direcionada, teremos assim

$$\Pi(k) = \frac{k}{2mt}, \quad (4.4)$$

nosso objetivo é calcular de que forma $N(k, t)$ varia, conforme nós adicionamos novos sítios na rede. Nós podemos analisar duas possibilidades ao adicionarmos um novo sítio da rede, com relação ao número de sítios com grau k , $N(k, t)$:

- (i) O novo sítio conecta-se a um sítio de grau k , tornando o grau daquele sítio em $k + 1$, diminuindo $N(k, t)$.
- (ii) O novo sítio conecta-se a um sítio de grau $k - 1$, tornando o grau daquele sítio em k , aumentando $N(k, t)$.

O número de ligações que é esperado para conectar sítios de grau k após adicionarmos um novo sítio é dado por:

$$\underbrace{\frac{k}{2mt}}_{\text{termo 1}} \times \underbrace{NP_k(t)}_{\text{termo 2}} \times \underbrace{m}_{\text{termo 3}} = \frac{k}{2} P_k(t), \quad (4.5)$$

onde o termo 1 na Eq. (4.5) representa a probabilidade de um novo sítio conectar-se a um sítio de grau k (conexão preferencial), o termo 2 o número total de sítios com grau k e o termo 3

representa o número de ligações que o novo sítio fará, quanto maior m , maior a chance do novo sítio de realizar uma ligação com os sítios de grau k . Tendo a Eq. (4.5), nós iremos aplica-la nos dois casos (i) e (ii):

(i') O número de sítios de grau k que adquirem uma nova ligação, aumentando o grau para $k + 1$:

$$\frac{k}{2}P_k(t). \quad (4.6)$$

(ii') O número de sítios de grau $(k - 1)$ que adquirem uma nova ligação, aumentando o grau para k é :

$$\frac{k-1}{2}P_{k-1}(t). \quad (4.7)$$

Combinando as Eqs. (4.6), (4.7), nós obtemos o número esperado de sítios com grau k após a adição de um novo sítio:

$$(N+1)P_k(t+1) = \underbrace{NP_k(t)}_{\text{termo 1}} + \underbrace{\frac{k-1}{2}P_{k-1}(t)}_{\text{termo 2}} - \underbrace{\frac{k}{2}P_k(t)}_{\text{termo 3}}, \quad (4.8)$$

o termo 1 na Eq. (4.8) representa o número total de sítios com grau k antes da adição do novo sítio, o termo 2 o número de sítios com grau k após a adição do novo sítio e o termo 3 os sítios removidos da contagem, devido ao fato de terem ganhado mais um grau e tornado seu grau em $k + 1$. A Eq. (4.8) é aplicado para todos os sítios com grau $k > m$. Nós precisamos expressar a Eq. (4.8) para os sítios de grau $k < m$. Para isso, iremos utilizar de argumentos similares ao usado para obtermos a Eq. (4.8), obtendo

$$(N+1)P_m(t+1) = NP_m(t) + \frac{m}{2}P_m(t). \quad (4.9)$$

Como foi demonstrado por Barabási, a rede oriunda do seu modelo possui distribuição de graus estacionária (BARABÁSI, 2013). Isso significa que no limite $N = t \rightarrow \infty$, $P_k(\infty) = P_k$. Usando esse fato nas Eqs. (4.8), (4.9) nós teremos

$$(N+1)P_k(t+1) - NP_k(t) \rightarrow NP_k(\infty) - P_k(\infty) - NP_k(\infty) = P_k(\infty) = P_k, \quad (4.10)$$

$$(N+1)P_m(t+1) - NP_m(t) \rightarrow P_m, \quad (4.11)$$

usando as Eqs. (4.10), (4.11) nas Eqs. (4.8), (4.9), respectivamente, nós obtemos

$$P_k = \frac{k-1}{k+1}p_{k-1} \quad k > m, \quad (4.12)$$

$$P_m = \frac{2}{m+1}. \quad (4.13)$$

Podemos ainda reescrever a Eq. (4.13) na forma

$$P_{k+1} = \frac{k}{k+3} p_k, \quad (4.14)$$

realizando a mudança de variável $k \rightarrow k+1$. Iremos encontrar a recursividade de P_k partindo do menor grau possível para os sítios da rede ($k = m$), logo em seguida, iremos utilizar a relação dada pela Eq. (4.14). Teremos portanto:

$$\begin{aligned} P_{m+1} &= \frac{m}{m+3} P_m = \frac{2m}{(m+2)(m+3)}, \\ P_{m+2} &= \frac{m+1}{m+4} P_{m+1} = \frac{2m(m+1)}{(m+2)(m+3)(m+4)}, \\ P_{m+3} &= \frac{m+2}{m+5} P_{m+2} = \frac{2m(m+1)}{(m+3)(m+4)(m+5)}. \end{aligned} \quad (4.15)$$

Substituindo $(m+3)$ com k no denominador nós obtemos a probabilidade de observar um sítio com grau k :

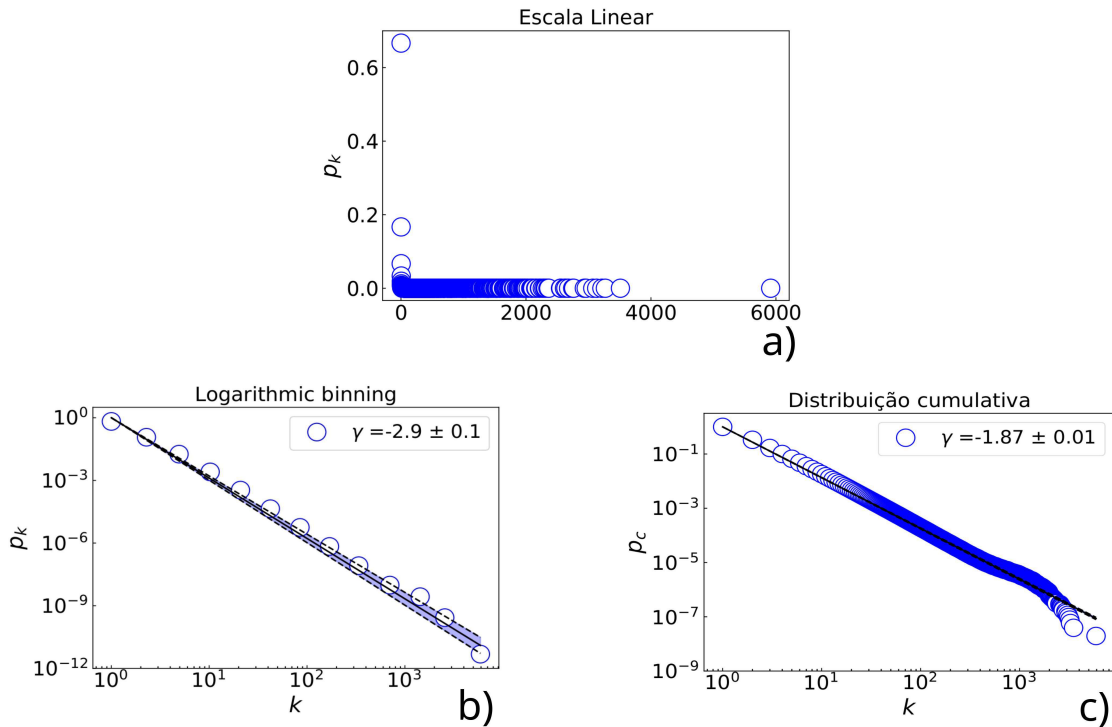
$$P_k = \frac{2m(m+1)}{k(k+1)(k+2)}, \quad (4.16)$$

sendo a solução exata para distribuição do modelo de Barabási. Podemos observar que no caso de $k \gg 1$ obtemos uma lei de potência

$$p_k \sim k^{-\gamma}, \quad (4.17)$$

com $\gamma = 3$.

Figura 8 – Os círculos coloridos de azul são os valores calculados a partir da distribuição de graus oriundo do modelo de Barabási. Foram utilizadas $M = 50$ amostras para o cálculo da distribuição, com valor de $m_0 = 1$ e $m = 1$. Enquanto a seta preta representa a regressão linear dos dados, dos quais podemos inferir o valor de γ a partir do coeficiente linear. As redes geradas possuem $N = 10^6$ sítios.



Fonte: elaborado pelo autor

Na Figura 8 mostramos quatro representações em escalas diferentes de $P(k)$ para uma rede gerada pelo modelo de Barabási-Albert com parâmetros $m = 1$, $m_0 = 1$, $N = 10^6$. Na Figura 8a é usado binagem linear, ou seja, N_k , o número de ligações entre k e $k + \Delta k$, é calculado com Δk constante. Na Figura 8b a binagem é logarítmica, ou seja, $\Delta k \sim \log k$. Na Figura 8c mostramos a distribuição cumulativa P_c com binagem linear.

Podemos observar que o valor de γ varia dependendo do método utilizado para calculá-lo, no caso do método *logarithmic binning*, o valor encontrado está dentro da margem do valor teórico, enquanto na cumulativa o valor encontra-se ligeiramente fora. Tendo que o coeficiente encontrado a partir da função cumulativa como sendo $\gamma_c = -1.87 \pm 0,01$, para o caso da distribuição cumulativa, devemos Eq. (2.6)

$$\gamma_c = -\gamma + 1, \quad (4.18)$$

onde γ_c é o coeficiente da distribuição cumulativa de graus e γ da distribuição de graus, com isso,

nós obtemos

$$\gamma = -2.87 \pm 0,01. \quad (4.19)$$

4.2 Caminho médio, menor caminho e diâmetro

Para o modelo de Barabási-Albert, nós temos que o diâmetro da rede depende de N , sendo N o número de sítios da rede. Para o $N \gg 1$ e $m > 1$, foi obtido de forma analítica que a dependência é da forma (ALBERT *et al.*, 1999)

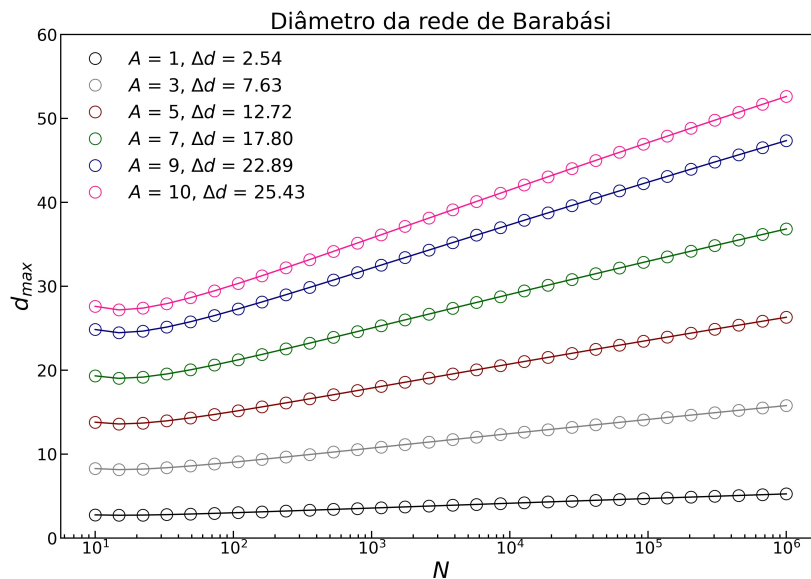
$$d_{max} \sim \frac{\ln N}{\ln \ln N}, \quad (4.20)$$

podemos observar pela Eq. (4.20) que o diâmetro da rede no modelo de Barabási-Albert cresce muito lentamente com N . Fazendo o prefator da Eq. (4.20) representado pela letra A , ou seja

$$d_{max} = A \frac{\ln N}{\ln \ln N}, \quad (4.21)$$

pela Figura 9 podemos observar o forma como o diâmetro cresce com relação ao tamanho da rede.

Figura 9 – Os círculos pretos, cinza, carmesim, verdes, azuis e rosas representam os valores dos diâmetros para $A = 1, 3, 5, 7, 9, 10$, respectivamente. Enquanto Δd representam a variação no diâmetro da rede. Podemos observar que para o prefator $A = 1$, uma variação de N na ordem de 10^6 representa uma variação de $\Delta d = 2.54$, mostrando que o diâmetro dessa rede cresce muito lentamente com o número de sítios.



Fonte: elaborado pelo autor.

A solução expressa na Eq. (4.20) foi obtida através de uma análise da rede em camadas, onde as camadas representam as divisões da rede em conjuntos de sítios a uma distância l , partindo da origem. Foi realizado uma contagem do número de sítios em cada uma dessas camadas, analisando os efeitos desses valores nos conjuntos de graus de toda rede e em sua topologia.

5 O MODELO DE CONEXÃO PREFERENCIAL ENVOLVENDO DISTÂNCIAS EUCLIDIANAS

Em trabalhos passados, o mecanismo de conexão preferencial proposto por Barabási foi modificado para representar sistemas físico-estatísticos, como foi feito para o caso do condensado de Bose-Einstein (BARABASI, 2001b). No trabalho de Soares *et. al.*, foi introduzido um fator geométrico na conexão preferencial, com sítios alocados num espaço euclidiano bidimensional (SOARES *et al.*, 2005). Baseando-se nesse modelo, Samurái *et. al* propôs uma generalização d -dimensional.

No modelo de Barabási-Albert que foi visto no capítulo anterior, nós temos uma dependência exclusiva do grau dos sítios já existentes para a evolução temporal da rede, isto é, não há uma dependência espacial na construção da rede, como pode ser visto na Eq. (4.1). Já no modelo proposto por Samurái, foi apresentada uma dependência da distância euclidiana na conexão preferencial (BRITO *et al.*, 2016). Essa dependência é expressa na equação abaixo:

$$\Pi_{ij} = \frac{k_i r_{ij}^{-\alpha_A}}{\sum_i k_i r_{ij}^{-\alpha_A}} \in [0, 1] \quad (\alpha_A \geq 0), \quad (5.1)$$

com r_{ij} sendo a distância euclidiana do sítio novo, j , aos sítios já existentes na rede, i .

O grau dos sítios já existentes da rede continuam influenciando na probabilidade de ligação, como ser visto o termo k_i . Tendo $\alpha_a \geq 1$, sendo um parâmetro que está relacionado com a forma da dependência da distância entre os sítios. Podemos observar que tomando o limite onde α_a tende a zero, a dependência da distância gradualmente perde sua relevância, e, no limite em que $\alpha_a = 0$ nós retornamos a expressão para conexão preferencial do modelo de Barabási-Albert, expressa na equação Eq. (4.1). Além da conexão preferencial, para que possamos dispor os sítios no espaço, devemos adicionar os novos sítios na rede numa posição, cuja distância para o centro de massa da rede já existente siga uma distribuição isotrópica d -dimensional (BRITO *et al.*, 2016)

$$p(r) \sim \frac{1}{r^{d+\alpha_g}}, \quad (5.2)$$

com $r \geq 1$, onde r representa a distância do sítio adicionado ao centro de massa da rede no instante anterior, d a dimensão em que a rede está sendo construída e α_g um parâmetro que está relacionado a agregação espacial dos sítios, isto é, quanto maior o valor de α_g , fixado d , maior será a probabilidade dos sítios adicionados estarem próximo ao centro de massa da rede já existente. A dimensão expressa na Eq. (5.2) determinará a forma como será realizado o cálculo

das distâncias na conexão preferencial expressa na Eq. (5.1), Para dimensão $d = 1$, nós teremos a distância dada por $r_{ij} = |x_i - x_j|$, para $d = 2$, nós teremos $r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, assim por diante para as dimensões $d = 3$ e $d = 4$. A distribuição isotrópica expressa na Eq. (5.2) deve ser normalizada, com isso, devemos impor valor acerca de α_g de forma que $p(r)$ seja normalizável. Ciente que $r \geq 1$, e, que os valores gerador por r serão truncados por um r_{max} , teremos:

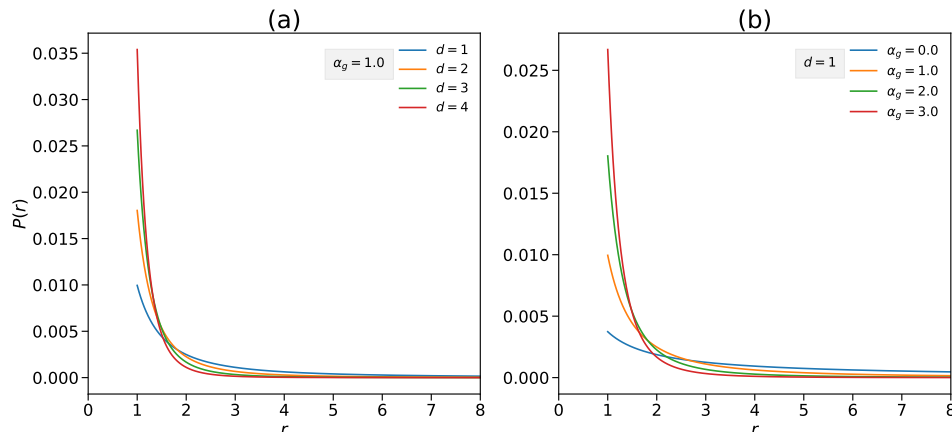
$$\int_1^{r_{max}} A r^{d-1} r^{-(d+\alpha_g)} dr = 1, \quad (5.3)$$

com A sendo a constante de normalização. Resolvendo essa integral, a constante de normalização é dada por

$$A = \begin{cases} 1/\ln(r_{max}) & \text{se } \alpha_g = 0, \\ \alpha_g r_{max}^{\alpha_g} / (r_{max}^{\alpha_g} - 1) & \text{se } \alpha_g > 0. \end{cases} \quad (5.4)$$

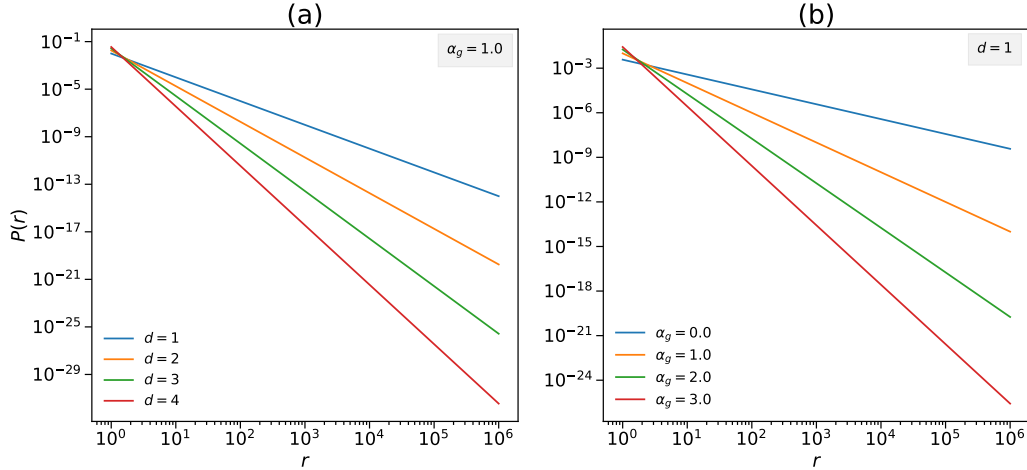
Podemos observa nas Figuras. 10 e 11 os efeitos da dimensão e do parâmetro α_g na distribuição $p(r)$. Fixado $\alpha_g = 1.0$ aumentando d é possível observar que a probabilidade de encontrar pequenos valores de r cresce consideravelmente. Enquanto fixado $d = 1$ e aumentando α_g é notável o mesmo efeito. Esse efeito fica ainda mais perspectivo ao analisarmos o plot $\log p(r)$ versus $\log r$ expresso na Figura 11. Podemos observar que conforme aumentamos a dimensão, fixado $\alpha_g = 1.0$, menor a probabilidade de encontrarmos grandes valores de r . Esse efeito também ocorre para o crescimento de α_g , fixado a dimensão $d = 1$.

Figura 10 – Nós temos r como sendo a distância e $p(r)$ a distribuição dada pela Eq. (5.3). Na figura (a) nós podemos observar os efeitos do parâmetro α_g sobre a distribuição dada pela Eq. (5.3), sendo $d = 1$ a dimensão. Enquanto na Figura (b) podemos observar os efeitos da dimensão, fixado $\alpha_g = 2.0$



Fonte: elaborado pelo autor.

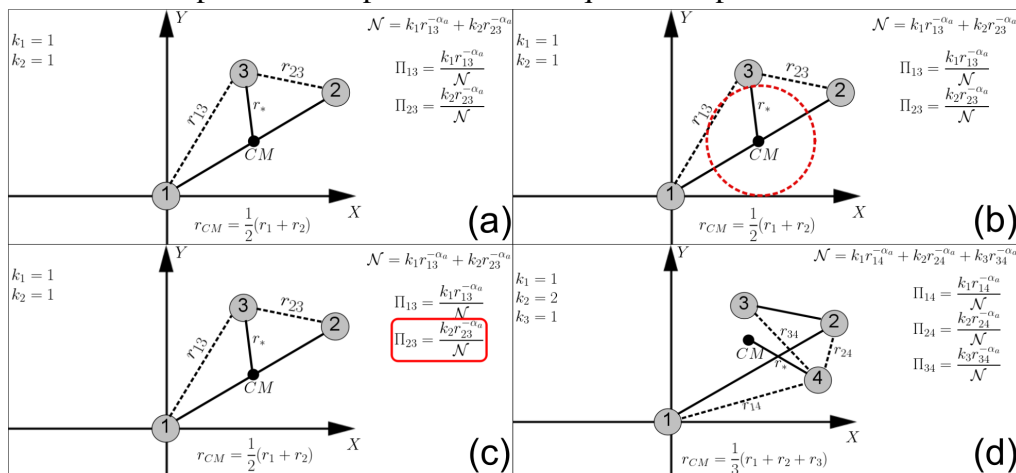
Figura 11 – Nós temos $\log r$ como sendo a distância e $\log p(r)$ a distribuição dada pela Eq. (5.3). Na figura (a) nós podemos observar os efeitos do parâmetro α_g sobre a distribuição dada pela Eq. (5.3), sendo $d = 1$ a dimensão. Enquanto na Figura (b) podemos observar os efeitos da dimensão, fixado $\alpha_g = 2.0$



Fonte: elaborado pelo autor.

A Figura 12 exemplifica o processo por trás do modelo de crescimento da conexão preferencial.

Figura 12 – Os círculos representam os sítios, enquanto k_i representa o grau do i -ésimo sítio, sendo \mathcal{N} a constante de normalização e r_{CM} o centro de massa. Na Figura 12.(a) nós temos a rede com os sítios 1 e 2 conectados entre si, a distância r_* foi gerada seguindo a distribuição dada pela Eq. (5.2), na Figura 12.(b) nós temos a região onde o novo sítio pode ser alocado, um círculo de raio r_* . Na Figura 12.(c) foi realizado um sorteio, sendo o sítio 2 selecionado para conectar ao sítio 3. Na Figura 12.(d) o sítio 2 e 3 são conectados, atualizando o centro de massa e o grau dos sítios. O processo é repetido $N - 3$ até que a rede possua N sítios.



Fonte: elaborado pelo autor.

5.1 Distribuição de graus da rede

A conexão entre as redes oriundas do modelo de crescimento da conexão preferencial e a estatística de Tsallis é advindo da relação entre a entropia de Tsallis e a função que otimiza ela (SOARES *et al.*, 2005). A entropia de Tsallis é expressa na Eq. (3.13), impondo os vínculos $\langle k \rangle = \sum_i k_i P(k_i) = C$ e $\sum_i P(k_i) = 1$, sendo C uma constante, nós obtemos diretamente a distribuição de graus (BRITO *et al.*, 2016)

$$P(k) = P(0)e_q^{-k/\lambda} = P(0) \left[1 + (q-1) \frac{k}{\lambda} \right]^{1/(1-q)}, \quad (5.5)$$

onde $e_q^{-k/\lambda}$ representa q-exponencial e $P(0)$ é a constante de normalização. Tomando $\varepsilon_i = k_i$ na Eq. (V.2), nós teremos $U_q \rightarrow \langle k \rangle = C$, sendo C uma constante.

6 RESULTADOS

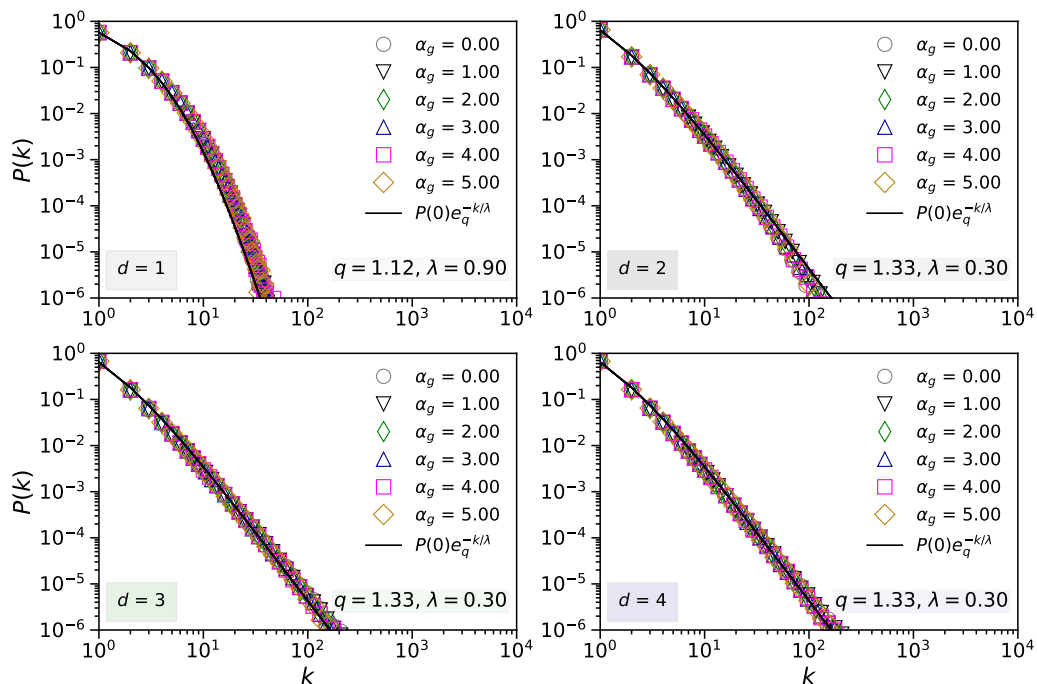
Os resultados a seguir foram obtidos através dos estudos do modelo de crescimento da conexão preferencial. A análise do efeito da dimensionalidade na distribuição de graus das redes oriundas já fora realizado (BRITO *et al.*, 2016), sendo reproduzido nesse trabalho. Enquanto os efeitos da dimensionalidade das propriedades: assortatividade, comprimento médio do menor caminho e diâmetro da rede fora realizado nesta dissertação.

6.1 Distribuição de graus da rede

Podemos observar que o modelo de crescimento da conexão preferencial possui três parâmetros: α_g e d , que são parâmetros relacionados as distâncias dos novos sítios para o centro de massa do sistema, sendo d a dimensão, e, α_a que está relacionado com o peso das distâncias na conexão preferencial. A influência desses parâmetros na distribuição de graus da rede podem ser observadas nas Figuras 14 e 13.

Figura 13 – No gráfico expresso, nós temos que d representa a dimensão, enquanto α_g é o parâmetro relacionado com a distância dos novos sítios para o centro de massa e α_a o parâmetro relacionado as distâncias na conexão preferencial. Para todos os conjuntos fixamos $\alpha_a = 2.0$. Foram utilizadas $M = 82$ amostras com $N = 10^5$ sítios.

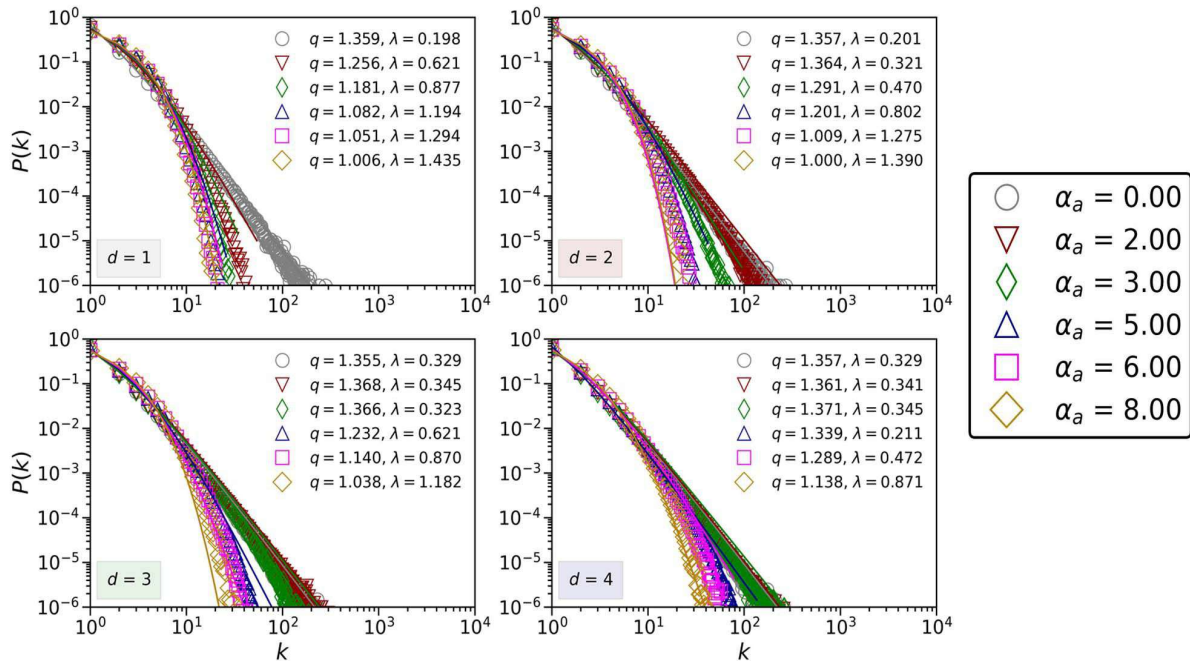
Distribuição de graus para múltiplos α_g



Fonte: elaborado pelo autor

Figura 14 – No gráfico expresso, nós temos que d representa a dimensão, enquanto α_g é o parâmetro relacionado com a distância dos novos sítios para o centro de massa e α_a o parâmetro relacionado as distâncias na conexão preferencial. Para todos os valores fixamos $\alpha_g = 2.0$. Foram utilizadas $M = 82$ amostras com $N=10^5$ sítios.

Distribuição de graus para múltiplos α_a



Fonte: elaborado pelo autor

Podemos observar pela Figura 13 que para diversos valores de α_g , fixado $\alpha_a = 2.0$, a distribuição de graus é exatamente a mesma. Indicando assim que a distribuição de graus é independente de α_g . Fixando $\alpha_g = 2.0$, variamos α_a . É possível verificar através da Figura 14 que a forma da distribuição, $P(k)$, varia para cada α_a e d diferentes, mostrando assim a dependência de $P(k)$ por esses dois parâmetros. A dependência da distribuição com α_a e d é explicitada nos parâmetros q e λ , sendo dada por (BRITO *et al.*, 2016):

$$\lambda(\alpha_a/d) = \begin{cases} 0.3 & \text{se } 0 \leq \alpha_a/d \leq 1, \\ -1.15e^{1-\alpha_a/d} + 1.45 & \text{se } \alpha_a/d > 1, \end{cases} \quad (6.1)$$

com q dado por

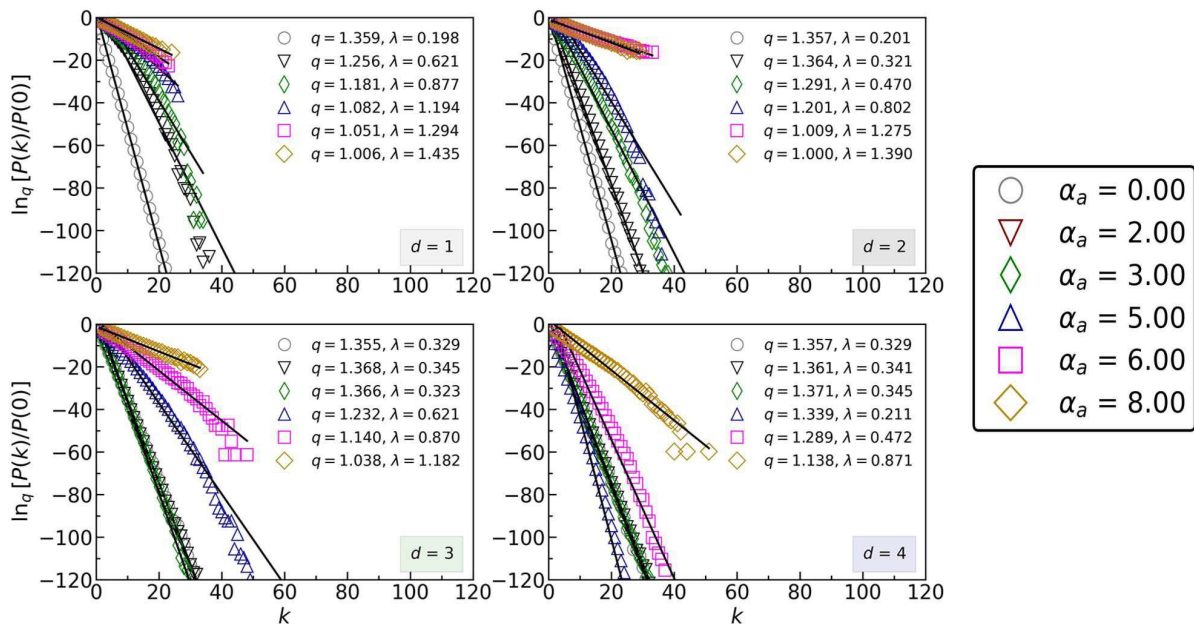
$$q(\alpha_a/d) = \begin{cases} 4/3 & \text{se } 0 \leq \alpha_a/d \leq 1, \\ \frac{1}{3}e^{1-\alpha_a/d} + 1 & \text{se } \alpha_a/d > 1. \end{cases} \quad (6.2)$$

Através das Eqs. (6.1), (6.2), vemos que no limite em que α_a tende ao infinito, q tende a 1 e λ tende a 1.45. Ou seja, existe um valor máximo e um valor mínimo para q e λ .

$q \in [4/3, 1]$ e $\lambda \in [0.3, 1.45]$. Com os valores obtidos a partir das Eqs. (6.1), (6.2) nós podemos encontrar os melhores q, λ para cada valor de α_a, d , nos dando o resultado expresso na Figura 15. Podemos observar que a regressão para alguns conjuntos de α_a e d possuem um platô. Esse platô é devido ao fato de não estarmos analisando os dados no limite termodinâmico, $N \gg 1$.

Figura 15 – No gráfico expresso, nós temos que d representa a dimensão, fixamos o valor de $\alpha_g = 2.0$ para todas as curvas, enquanto variamos α_a . Foram utilizadas $M = 82$ amostras para redes com $N = 10^5$ sítios. As retas pretas representam a curva que melhor se ajusta ao conjunto de dados, com os valores de q e λ para cada uma das curvas expressa na Tabela 2. O ajuste foi realizado numa região dos dados, devido ao platô nas curvas oriundo da grande quantidade de sítios com mesmo grau e poucos sítios com grande valor de grau. Esse platô desaparece no limite termodinâmico, onde $k \gg 1$.

Distribuição de graus da rede



Fonte: elaborado pelo autor

Tabela 2 – Os parâmetros q e λ foram calculados utilizando o algoritmo LM (do inglês *Levenberg-Marquardt algorithm*) (GAVIN, 2019), enquanto R^2 é o coeficiente de determinação, que representa precisão dos dados mensurados com relação a função hipótese (S.A. SLINKER B.K., 2016), no nosso caso, a função hipótese é a distribuição expressa na Eq. (5.5). $R^2 \in [0, 1]$, quanto mais próximo de 1, maior a tendência da função hipótese descrever aqueles dados.

Dimensão	α_a	q	λ	R^2
$d = 1$	0.0	1.3588 ± 0.0003	0.1979 ± 0.0005	0.999998
	2.0	1.256 ± 0.004	0.621 ± 0.007	0.9999704
	3.0	1.181 ± 0.006	0.88 ± 0.01	0.9999451
	5.0	1.082 ± 0.006	1.19 ± 0.01	0.9999551
	6.0	1.051 ± 0.006	1.29 ± 0.01	0.9999632
	8.0	1.006 ± 0.003	1.433 ± 0.009	0.9999845
$d = 2$	0.0	1.3573 ± 0.0003	0.2008 ± 0.0004	0.9999993
	2.0	1.364 ± 0.001	0.3205 ± 0.0006	0.9962342
	3.0	1.2915 ± 0.0009	0.4598 ± 0.0005	0.9999755
	5.0	1.201 ± 0.005	0.80 ± 0.01	0.9999617
	6.0	1.01 ± 0.01	1.28 ± 0.02	0.9985588
	8.0	1.00 ± 0.01	1.39 ± 0.02	0.9989897
$d = 3$	0.0	1.3552 ± 0.0002	0.3293 ± 0.0004	0.9955482
	2.0	1.3680 ± 0.0001	0.3451 ± 0.0007	0.9895939
	3.0	1.3658 ± 0.0006	0.3226 ± 0.0007	0.9956309
	5.0	1.232 ± 0.002	0.621 ± 0.002	0.9998642
	6.0	1.140 ± 0.007	0.869 ± 0.006	0.9994284
	8.0	1.04 ± 0.01	1.18 ± 0.01	0.9987625
$d = 4$	0.0	1.3571 ± 0.0003	0.3294 ± 0.0005	0.9953050
	2.0	1.3611 ± 0.0001	0.3407 ± 0.0005	0.9920349
	3.0	1.3714 ± 0.0001	0.3451 ± 0.0008	0.9889740
	5.0	1.339 ± 0.001	0.2110 ± 0.0005	0.9969615
	6.0	1.289 ± 0.0009	0.4717 ± 0.0005	0.9999766
	8.0	1.141 ± 0.007	0.871 ± 0.006	0.9994152

6.1.1 Comprimento médio do menor caminho e diâmetro da rede

Através dos resultados da seção anterior, pudemos observar que os parâmetros α_a e d afetam diretamente a distribuição de graus, tendo uma independência direta do parâmetro α_g . Como a distribuição de graus está relacionado com a forma como os graus estão distribuídos em todos os sítios da rede, essa independência do α_g indica que o diâmetro, a assortatividade e o comprimento médio do menor caminho também podem ser independente de α_g . Além do mais, como foi mostrado no modelo de Barabási, o diâmetro da rede tinha uma relação direta com o tamanho da mesma, tendo também uma relação entre redes com fenômeno de mundo pequeno, uma relação direta com o tamanho das redes. Isso nos dá que o estudo do efeito do tamanho da rede nessas propriedades é bastante relevante.

Figura 16 – No gráfico expresso, nós temos que $\langle \ell \rangle$ representa o comprimento médio dos menores caminhos da rede e α_g é o parâmetro relacionado as distâncias dos novos sítios para o centro de massa, foi utilizado $\alpha_a = 1.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de $\langle \ell \rangle$ para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios.

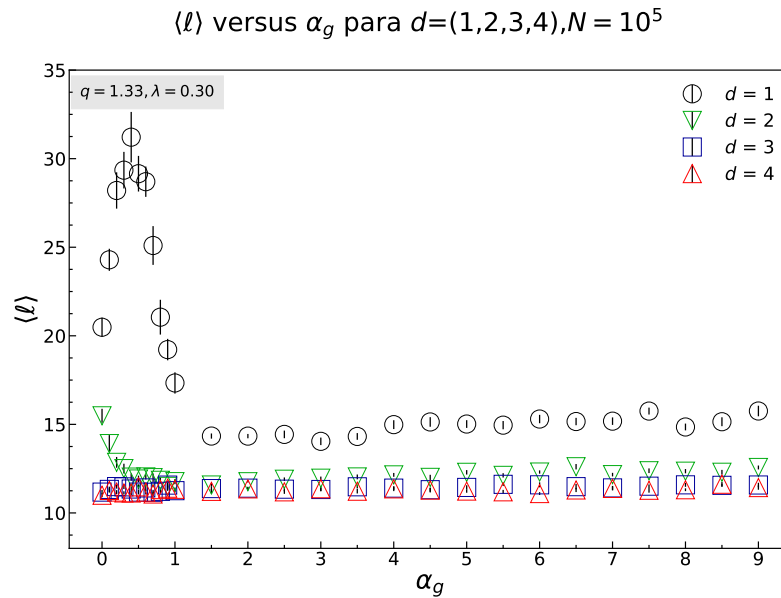


Figura 17 – No gráfico expresso, nós temos que $\langle \ell \rangle$ representa o comprimento médio dos menores caminhos da rede e α_a é o parâmetro relacionado a conexão preferencial modificada, foi utilizado $\alpha_g = 1.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de $\langle \ell \rangle$ para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios.

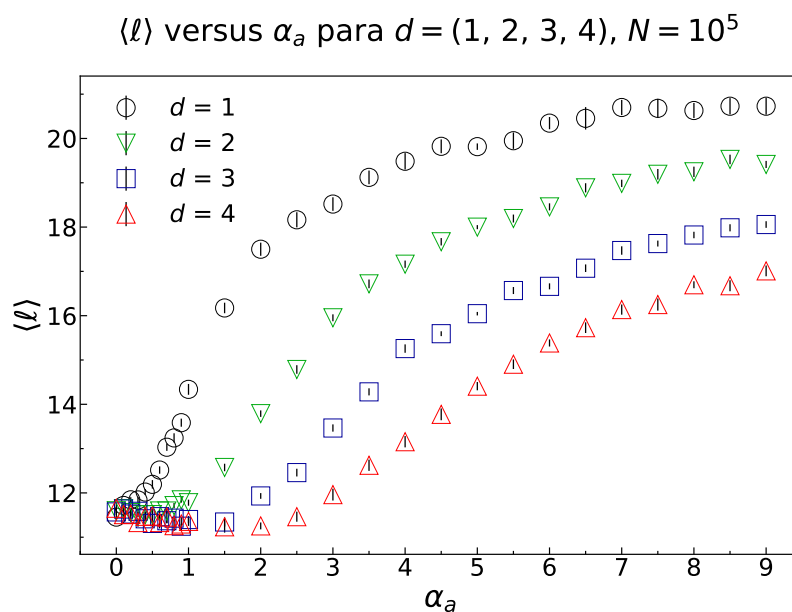


Figura 18 – No gráfico expresso, nós temos que d_{max} representa o diâmetro da rede e α_g é o parâmetro relacionado as distâncias dos novos sítios para o centro de massa, foi utilizado $\alpha_a = 1.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de d_{max} para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios.

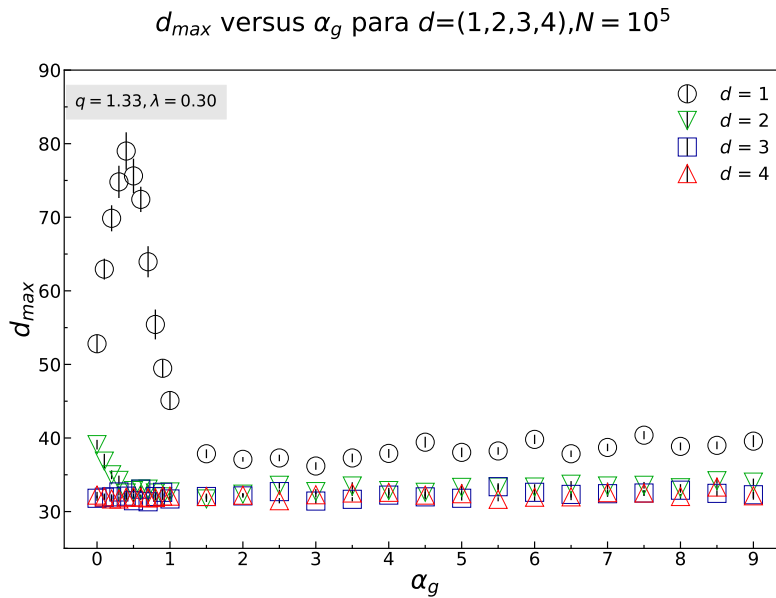


Figura 19 – No gráfico expresso, nós temos que d_{max} é o diâmetro da rede e α_a é o parâmetro relacionado as distâncias dos novos sítios para o centro de massa, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras. Os círculos preto é o valor de d_{max} para dimensão $d = 1$, triângulos verde $d = 2$, quadrados azul $d = 3$ e triângulos vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios.

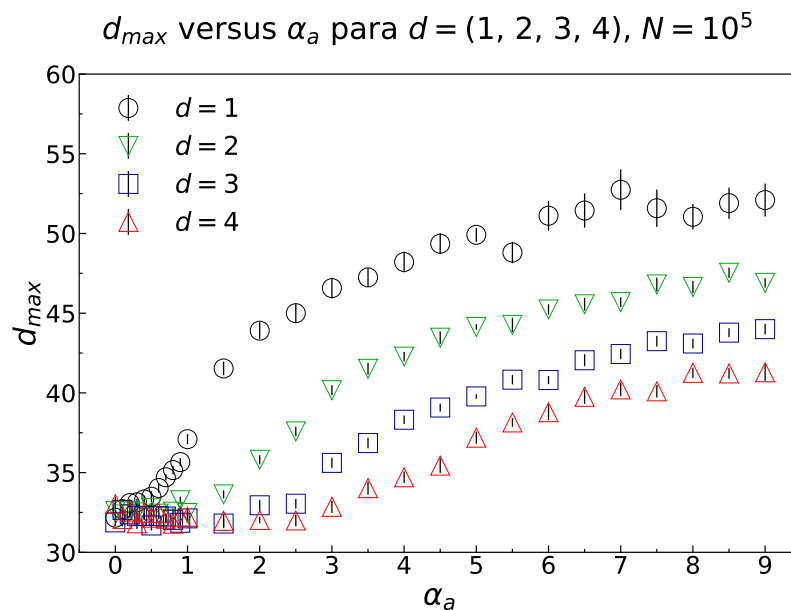
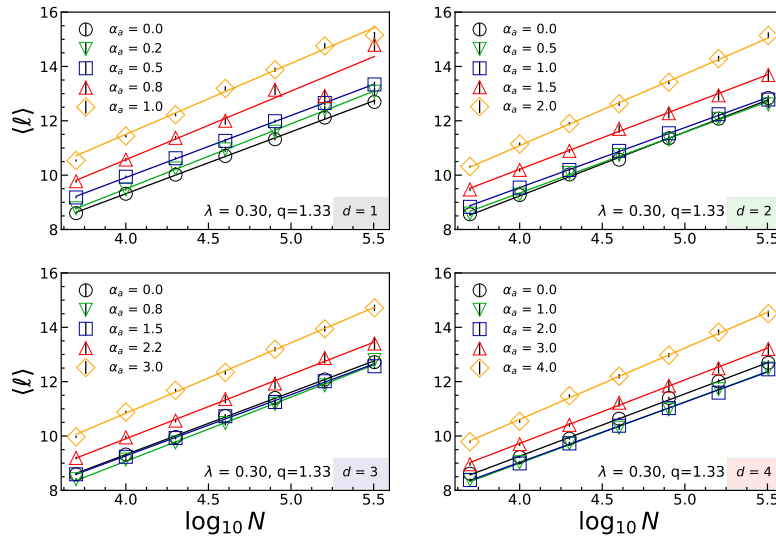


Figura 20 – No gráfico expresso, nós temos que $\langle \ell \rangle$ representa o comprimento médio do menor caminho da rede, N é o tamanho da rede, d a dimensão e α_a o parâmetro relacionado as distâncias na conexão preferencial, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 402, 252, 202, 92, 92, 118, 75, 12$ amostras, para as redes com $N = 5.000, 10.000, 20.000, 40.000, 80.000, 160.000, 320.000$ sítios, respectivamente.

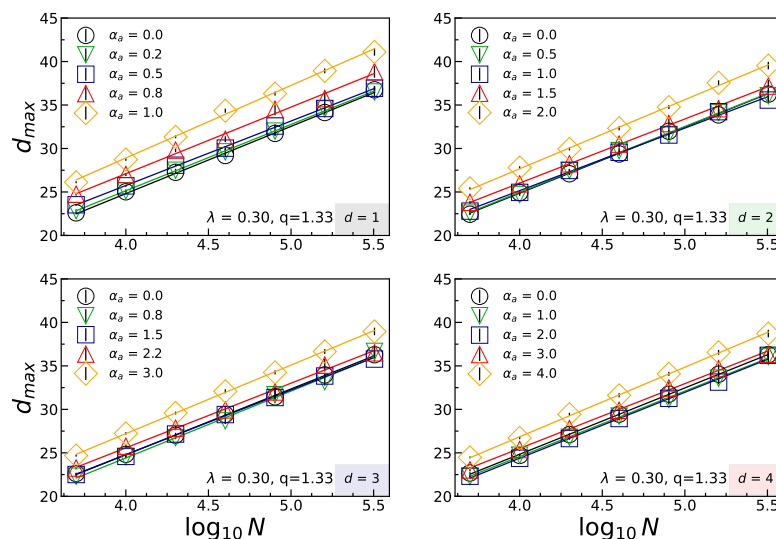
$\langle \ell \rangle$ versus N para $d=(1,2,3,4)$, α_a variado



Fonte: elaborado pelo autor

Figura 21 – No gráfico expresso, nós temos que d_{max} representa o diâmetro da rede, N é o tamanho da rede, d a dimensão e α_a o parâmetro relacionado as distâncias na conexão preferencial, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 402, 252, 202, 92, 92, 118, 75, 12$ amostras, para as redes com $N = 5.000, 10.000, 20.000, 40.000, 80.000, 160.000, 320.000$ sítios, respectivamente.

d_{max} versus N para $d=(1,2,3,4)$, α_a variado



Fonte: elaborado pelo autor

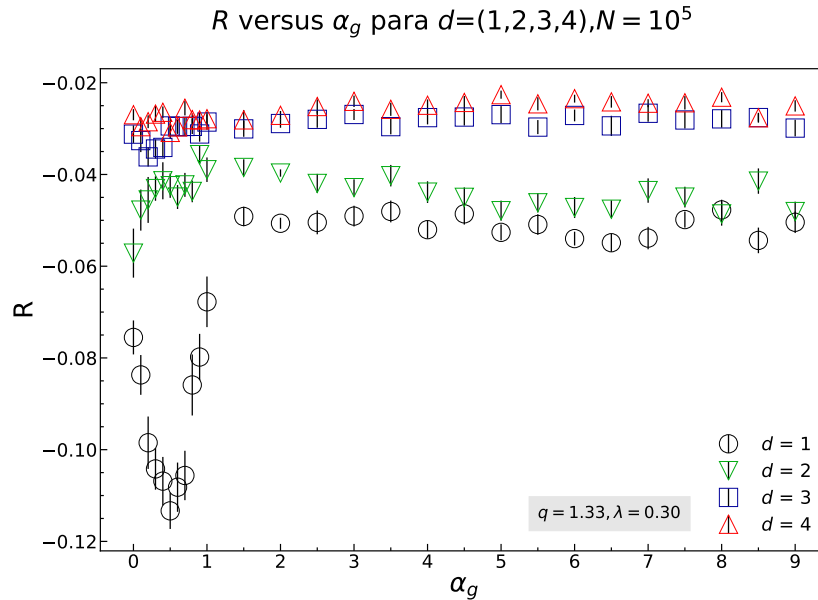
Para as dimensão $d=(1, 2)$, os valores $\langle \ell \rangle$, na Figura 16, sofrem uma variação em seus valores, tendo uma máximo no caso de $d = 1$ e um decaimento em $d = 2$ na região $\alpha_g \in [0.1]$. Com valor mantendo-se constante na região $\alpha_g > 1$. Como a distribuição de graus é independente de α_g era esperado um valor constante de $\langle \ell \rangle$ para qualquer α_g , como podemos notar para os casos de $d = (3, 4)$. Pelas Figura 18 podemos observar que esse mesmo comportamento também ocorre para d_{max} .

Na Figura 17, podemos perceber a dependência direta de $\langle \ell \rangle$ com α_a e d , tendo que os valores de $\langle \ell \rangle$ são maiores, conforme nós aumentamos a dimensão na qual a rede está imbuída, comportamento esse que também é observado no diâmetro na Figura 19. Na região $\alpha_a \in [0.1]$ a variação $\langle \ell \rangle$ é bastante sutil, tendo uma elevação nos valores na região $\alpha_a > 1$. Através das Figuras 20 e 21 podemos observar uma relação linear entre o comprimento médio dos menores caminhos da rede e do diâmetro com o logaritmo do tamanho da rede. Essa relação linear é esperada para redes com fenômeno de mundo pequeno.

6.1.2 Assortatividade

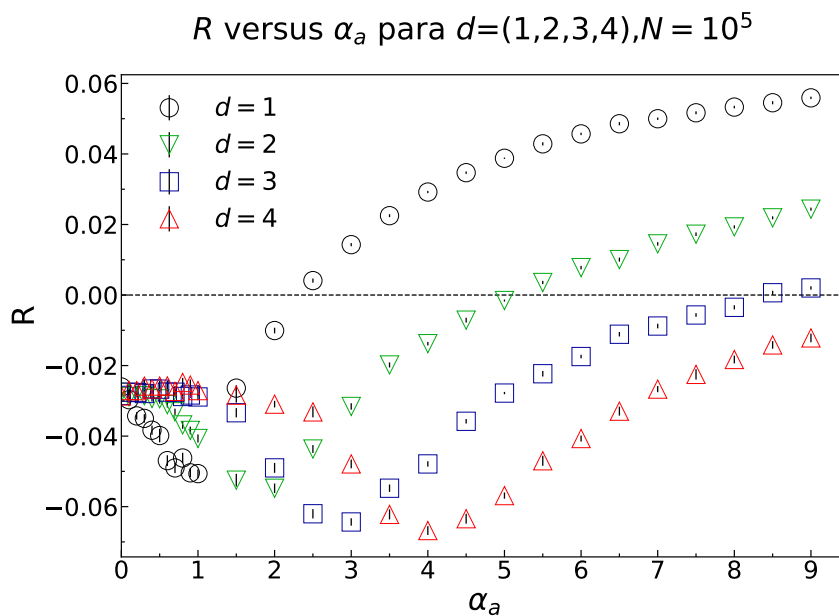
A assortatividade da rede é uma propriedade que caracteriza a natureza das ligações na rede. Ou seja, através da assortatividade, nós podemos mensurar a tendência da ligação entre os sítios, no nosso caso, se sítios de mesmo grau tendem a conecta-se entre si ou a sítios de diferentes graus. Há duas formas equivalentes de analisar a assortatividade da rede. A primeira forma é através do comportamento da curva no plot k_{nn} versus k , sendo k_{nn} o grau médio da vizinhança do sítio e k o grau do sítio. Esse plot mostra a correlação entre os graus dos sítios e os graus dos sítios da sua vizinhança, caso a curva possua um comportamento crescente, ela é assortativa, quando possui comportamento decrescente é dessortativa, se não apresentar crescimento e nem decrescimento, ela é tida como não-assortativa. A outra forma de analisar a assortatividade da rede é através do coeficiente de assortatividade, obtido através do coeficiente de Pearson, expresso na Eq. (2.38).

Figura 22 – No gráfico expresso, nós temos que R representa o coeficiente de assortatividade e α_g o parâmetro as distâncias para o centro de massa do sistema, foi utilizado $\alpha_a = 1.0$ para todos os conjuntos de amostras. Os círculos pretos é o valor de R para dimensão $d = 1$, estrelas verdes $d = 2$, triângulos azuis $d = 3$ e quadrado vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios.



Fonte: elaborado pelo autor

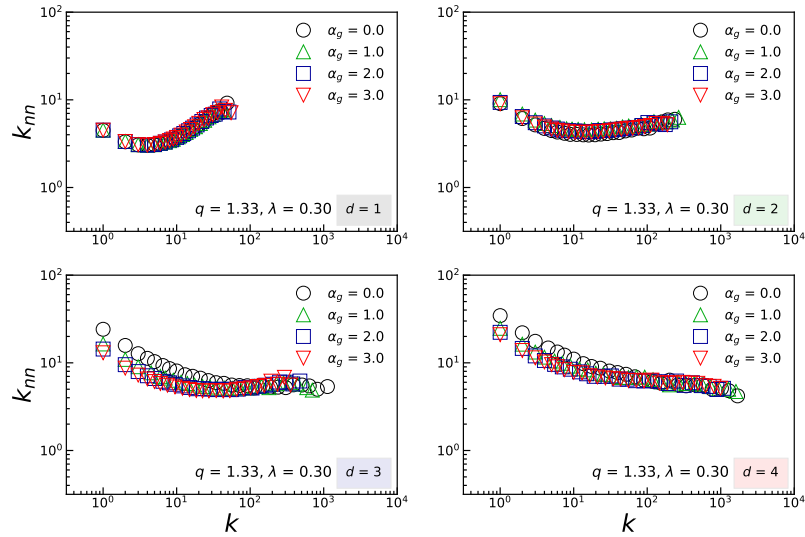
Figura 23 – No gráfico expresso, nós temos que R representa o coeficiente de assortatividade e α_a o parâmetro relacionado as distâncias do centro de massa da rede, foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras. Os círculos pretos é o valor de R para dimensão $d = 1$, estrelas verdes $d = 2$, triângulos azuis $d = 3$ e quadrado vermelho $d = 4$. Foram utilizados $M = 82$ amostras para redes com $N = 10^5$ sítios.



Fonte: elaborado pelo autor

Figura 24 – No gráfico expresso, nós temos que k_{nn} representa o grau da vizinhança, k o grau do sítio, α_g o parâmetro relacionado as distâncias geradas com relação ao centro de massa do sistema e d a dimensão. Para todos os casos foram utilizados $\alpha_a = 2.0$. Foram utilizados $M = 41$ amostras para redes com $N = 10^5$ sítios. Foi realizado *Logarithmic binning* nos dados.

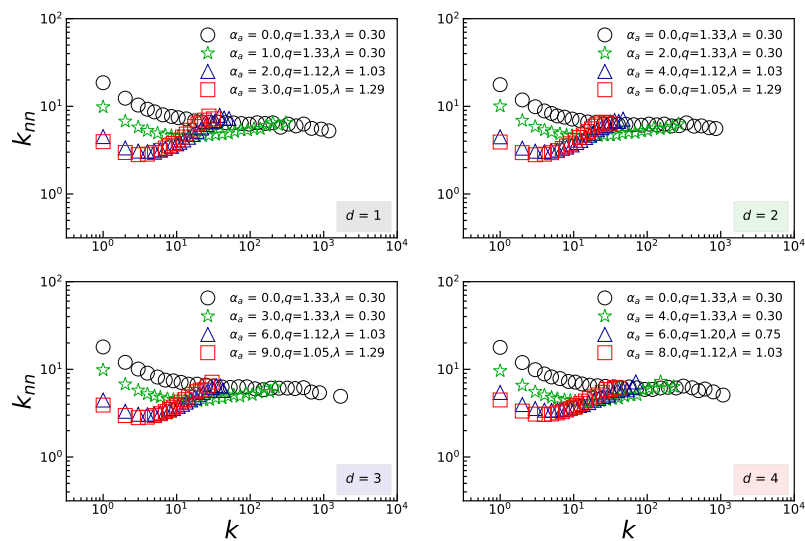
k_{nn} versus k para $d=(1,2,3,4)$, α_g variado



Fonte: elaborado pelo autor

Figura 25 – No gráfico expresso, nós temos que k_{nn} representa o grau da vizinhança, k o grau do sítio, α_a o parâmetro relacionado as distâncias na conexão preferencial e d a dimensão. Para todos os casos foram utilizados $\alpha_g = 2.0$. Foram utilizados $M = 41$ amostras para redes com $N = 10^5$ sítios. Foi realizado *Logarithmic binning* nos dados.

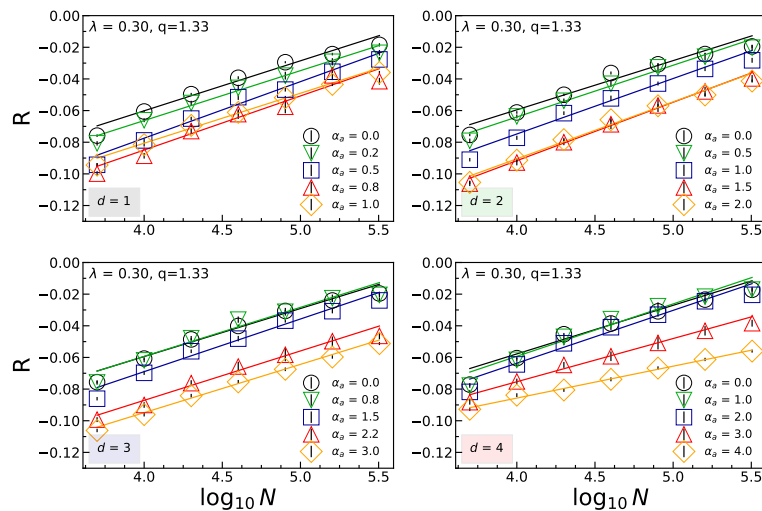
k_{nn} versus k para $d = (1,2,3,4)$, α_a variado



Fonte: elaborado pelo autor

Figura 26 – No gráfico expresso, nós temos que R é o coeficiente de assortatividade, N o número de sítios na rede, α_a o coeficiente relacionado as distâncias na conexão preferencial, d a dimensão, λ e q os parâmetros da distribuição de graus da rede. Os diferentes marcadores representam diferentes valores de α_a , foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 402, 252, 202, 92, 92, 118, 75, 12$ amostras, para as redes com $N = 5.000, 10.000, 20.000, 40.000, 80.000, 160.000, 320.000$ sítios, respectivamente.

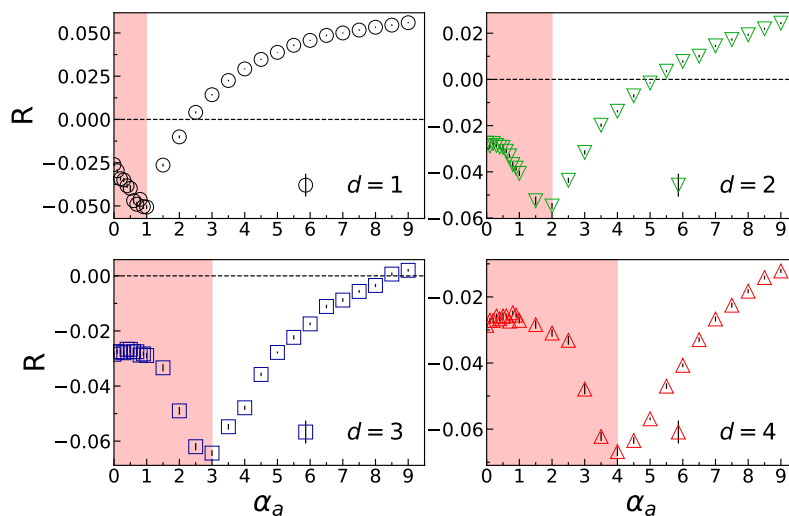
R versus N para $d=(1,2,3,4)$, α_a variado



Fonte: elaborado pelo autor

Figura 27 – No gráfico expresso, nós temos que R é o coeficiente de assortatividade, α_a o coeficiente relacionado as distâncias na conexão preferencial, d a dimensão. Os diferentes marcadores representam diferentes valores de α_a , foi utilizado $\alpha_g = 2.0$ para todos os conjuntos de amostras, tendo $M = 81$ amostras para redes com $N = 10^5$ sítios. As regiões avermelhadas representam a área onde $q = 1.33$, $\lambda = 0.30$, fora dessa região os valores de q decresce e λ aumenta, segundo as Eqs. (6.2), (6.1).

R versus α_a para $d = (1, 2, 3, 4)$, $N = 10^5$



Fonte: elaborado pelo autor

Podemos observar através da Figura 22, que para regiões $\alpha_g \in (0, 1)$, existe uma variação no coeficiente de assortatividade em $d = (1, 2)$, tendo um pico com valor mínimo próximo de $\alpha_g = 0.5$, comportamento similar ao que ocorreu no comprimento médio dos menores caminhos da rede e com o diâmetro da rede. Para dimensão $d = 2$, há um aumento no valor de R , até atingir um máximo próximo ao valor de $\alpha_g = 0.5$, com os valores decaindo após o valor de $\alpha_g = 0.5$. Para todos os valores de α_g , o coeficiente de assortatividade sempre é negativo, ou seja, a rede sempre adota uma característica dessortativa.

Por meio da Figura 23 é possível observar que há uma mudança no comportamento da assortatividade em $\alpha_a = d$, onde d representa a dimensão no qual a rede é construída. A linha tracejada no gráfico delimita a região de transição na assortatividade da rede, podemos notar que para redes com dimensão $d = (3, 4)$, para múltiplos valores de α_a , as redes tendem a manter-se dessortativas. Contudo, para redes em $d = (1, 2)$, nós temos um valor de $\alpha_{a,T}$ em que ocorre a transição na assortatividade da rede. Para $d = 1$ esse valor é $\alpha_{a,T} \approx 2.5$ e para $d = 2$ é $\alpha_{a,T} \approx 5.1$. É possível observar essa mudança na assortatividade para $d = (1, 2)$, através da mudança na inclinação das curvas na Figura 25. Outro fenômeno por trás do coeficiente de assortatividade é a mudança no seu comportamento, quando aumentamos os valores de q e λ . Através da Figura 27 podemos notar que conforme os valores de q decresce e λ aumenta. Esse ponto de transição ocorre em $\alpha_a \approx d$, onde d é a dimensão onde a rede é construída.

Outro comportamento interessante que podemos observar acerca da assortatividade é seu crescimento linear com $\log N$ através da Figura 26. Isso mostra que tanto o comprimento médio do menor caminho, quanto o diâmetro e assortatividade crescem lentamente com o tamanho da rede.

7 CONCLUSÃO

Podemos observar através dos resultados que a assortatividade das redes possuem um valor mínimo para os valores de $\alpha_a \approx d$, sendo d a dimensão sobre a qual a rede é construída. E, que dependendo da dimensão da rede, ocorre uma transição de uma rede dessortativa para uma rede assortativa. Além disso, foi constatado que o menor caminho médio e o diâmetro da rede possuem sempre valores maiores, quanto maior a dimensão onde a rede esta imbuída, para múltiplos valores de α_a . Os efeitos do parâmetro α_g nessas propriedades possuem um comportamento inesperado para as baixas dimensões, sendo um possível assunto para estudos futuros.

REFERÊNCIAS

- ALBERT, R.; JEONG, H.; BARABASI, A.-L. **The diameter of the world wide web**. [S. l.]: arXiv, 1999.
- ANDRADE, J. S.; SILVA, G. F. T. da; MOREIRA, A. A.; NOBRE, F. D.; CURADO, E. M. F. **Thermostatistics of Overdamped Motion of Interacting Particles**. [S. l.]: American Physical Society (APS), 2010.
- BAIRON, S. **O que é hipermídia**. [S. l.: s. n.], 2012.
- BARABASI, A.-L. **The physics of the Web**. 2001. 33 p.
- BARABASI, A.-L. **Statistical mechanics is offering new insights into the structure and dynamics of the Internet, the World Wide Web and other complex interacting systems The physics of the Web**. 2001.
- BARABÁSI, A.-L. Network science. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, The Royal Society Publishing, v. 371, n. 1987, p. 20120375, 2013.
- BARABASI, A.-L.; ALBERT, R. **Emergence of Scaling in Random Networks**. 1999. 509-512 p. Disponível em: <https://www.science.org/doi/abs/10.1126/science.286.5439.509>.
- BOAVENTURA, S. J. N. P. O. **Grafos: Introdução e prática**. 2. ed. [S. l.]: Blucher, 2017. ISBN 8521211333; 9788521211334.
- BOSLAUGH, S. **Statistics in a Nutshell**. 2. ed. Sebastopol, CA: O'Reilly Media, 2012.
- BRITO, S.; SILVA, L. D.; TSALLIS, C. **Role of dimensionality in complex networks**. [S. l.]: Nature Publishing Group, 2016. 1–8 p.
- CALLEN, H. B. **Thermodynamics and an introduction to thermostatistics; 2nd ed**. New York, NY: Wiley, 1985.
- CORMEN CHARLES E. LEISEN, R. L. R. C. S. T. H. **Introduction to algorithms**. 2. ed. [S. l.]: The MIT Press, 2001. ISBN 0262032937; 9780262032933; 0262531968; 9780262531962; 0070131511; 9780070131514.
- DEISENROTH, M. P.; FAISAL, A. A.; ONG, C. S. **Mathematics for Machine Learning**. [S. l.]: Cambridge University Press, 2020.
- EASLEY, D. A.; KLEINBERG, J. M. **Networks, Crowds, and Markets: Reasoning about a highly connected world**. [S. l.]: Cambridge University Press, 2010. I-XV, 1-727 p. ISBN 978-0-521-19533-1.
- EFRON, B. **Computers and the Theory of Statistics: Thinking the unthinkable**. [S. l.]: Society for Industrial Applied Mathematics (SIAM), 1979. 460–480 p.
- FLEISS, J. L.; COHEN, J.; EVERITT, B. S. **Large sample standard errors of kappa and weighted kappa**. [S. l.]: American Psychological Association (APA), 1969. 323–327 p.
- FLEISS, J. L.; LEVIN, B.; PAIK, M. C. **Statistical methods for rates and proportions**. [S. l.]: John Wiley & Sons, 2013.

FOSTER, D. V.; FOSTER, J. G.; GRASSBERGER, P.; PACZUSKI, M. Clustering drives assortativity and community structure in ensembles of networks. **Physical Review E**, American Physical Society (APS), v. 84, n. 6, dec 2011.

GAVIN, H. P. **The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems**. 2019.

GOLDSTEIN, H. **Classical Mechanics**. [S. l.]: Addison-Wesley, 1980.

JEONG, H.; MASON, S. P.; BARABÁSI, A.-L.; OLTVAI, Z. N. **Lethality and centrality in protein networks**. [S. l.]: Springer Science and Business Media LLC, 2001. 41–42 p.

JIANG, J.; WEN, S.; YU, S.; XIANG, Y.; ZHOU, W.; HASSAN, H. **The structure of communities in scale-free networks**. [S. l.]: John Wiley and Sons Ltd, 2017.

JOHNSTON, J.; ANDERSEN, T. **Random Processes with High Variance Produce Scale Free Networks**. [S. l.]: arXiv, 2022.

KITTEL, C.; KROEMER, H. **Thermal Physics**. [S. l.]: W. H. Freeman, 1980. ISBN 9780716710882.

LESKOVEC, J.; KLEINBERG, J.; FALOUTSOS, C. **Graph evolution: Densification and shrinking diameters**. 2007.

LI, G.; REIS, S. D.; MOREIRA, A. A.; HAVLIN, S.; STANLEY, H. E.; ANDRADE, J. S. **Optimal transport exponent in spatially embedded networks**. 2013.

LIU, B.; GOREE, J. **Superdiffusion and Non-Gaussian Statistics in a Driven-Dissipative 2D Dusty Plasma**. [S. l.]: American Physical Society (APS), 2008.

MILGRAM, S. **The Small-World Problem**. 1967. 61–67 p.

MOORE, E. **The Shortest Path Through a Maze**. [S. l.]: Bell Telephone System., 1959. (Bell Telephone System. Technical publications. monograph).

NEWMAN, M. E. **Assortative mixing in networks**. [S. l.]: APS, 2002. 208701 p.

NEWMAN, M. E. J. **Mixing patterns in networks**. [S. l.]: American Physical Society (APS), 2003.

NUSSENZVEIG, H. M. **Curso de física básica, 2: fluídos, oscilações e ondas, calor**. 4.ed.. ed. São Paulo: Edgard Blücher,, 2006. Inclui bibliografia e índice.

ORABY, T.; BALOGH, A. **Modeling the Effect of Observational Social Learning on Vaccination and Diseases using Household Networks**. [S. l.]: arXiv, 2022.

PICKUP, R. M.; CYWINSKI, R.; PAPPAS, C.; FARAGO, B.; FOUQUET, P. **Generalized Spin Glass Relaxation**. 2009.

POSTEL, J. **DoD standard Internet Protocol**. [S. l.]: RFC Editor, 1980. RFC 760. (Request for Comments, 760).

S.A. SLINKER B.K., N. T. G. **Primer of applied regression and analysis of variance**. 3. ed. [S. l.]: McGraw-Hill, 2016. ISBN 9780071824118; 0071824111; 9780071822442; 0071822445.

SALINAS, S. **Introdução à Física Estatística**. 2nd. ed. [S. l.]: Edusp, 2008. ISBN 9788531403866; 8531403863.

SOARES, D. J.; TSALLIS, C.; MARIZ, A. M.; SILVA, L. R. da. **Preferential attachment growth model and nonextensive statistical mechanics**. [S. l.]: IOP Publishing, 2005. 70 p.

STROGATZ, D. J. W. . S. H. **Collective dynamics of ‘small-world’ networks**. 1998.

TSALLIS, C. **Possible generalization of Boltzmann-Gibbs statistics**. 1988.

UPADHYAYA, A.; RIEU, J.-P.; GLAZIER, J. A.; SAWADA, Y. **Anomalous diffusion and non-Gaussian velocity distribution of Hydra cells in cellular aggregates**. 2001. 549-558 p.

APÊNDICE A – ALGORITMO: MODELO BARABÁSI-ALBERT

Para construção da rede foram utilizadas duas listas, intituladas K_{list} e K_{filter} , e uma lista de listas, denominada de $Nodes$. A lista K_{list} guardará o grau do i -ésimo sítio, se repetindo o número de vezes igual ao grau daquele sítio. Enquanto K_{filter} será uma lista filtrada do K_{list} e $Nodes$ guardará as ligações da rede, em tuplas. O algoritmo segue da seguinte forma.

Algoritmo para gerar uma rede de Barabási

- (i) A rede começa inicialmente com m_0 sítios totalmente conectados, tendo as listas K_{list} , K_{filter} e a lista de listas $Nodes$ dadas por

$$K_{list} = [0, \dots, 0, 1, \dots, 1, \dots, l, \dots, l, \dots, m_0, \dots, m_0] \quad (\text{A.1})$$

$$K_{filter} = [0, \dots, 0, 1, \dots, 1, \dots, l, \dots, l, \dots, m_0, \dots, m_0], \quad (\text{A.2})$$

$$Nodes = [(0, 1), \dots, (0, m_0), \dots, (l, 1), \dots, (m_0, 1), \dots, (m_0, m_0 - 1)]. \quad (\text{A.3})$$

- (ii) É adicionado um novo sítio $m_0 + 1$ a rede. É realizada uma seleção com distribuição de probabilidade uniforme na lista expressa na Eq. (A.1). Essa seleção segue a probabilidade de ligação expressa na Eq. (4.1). Ao ser sorteado o sítio l , é realizada a ligação do novo sítio $m_0 + 1$ ao sítio l . Atualizamos as listas K_{list} , K_{filter} e $Nodes$. Da lista K_{filter} é removido todas as aparições do sítio sorteado, l , para evitar que esse sítio possa efetuar ligações duplas com o novo sítio.

$$K_{list} = [0, \dots, 0, 1, \dots, 1, \dots, l, \dots, l, \dots, m_0, \dots, m_0, l, m_0 + 1] \quad (\text{A.4})$$

$$K_{filter} = [0, \dots, 0, 1, \dots, 1, \dots, m_0, \dots, m_0], \quad (\text{A.5})$$

$$Nodes = [(0, 1), \dots, (0, m_0), \dots, (l, 1), \dots, (m_0, 1), \dots, (m_0, m_0 - 1), (m_0 + 1, l)]. \quad (\text{A.6})$$

- (iii) O processo é repetido m vezes, até que o sítio novo adicionado $m_0 + 1$ realize as m ligações com a rede já existente. Ao fim das m ligações atualizamos K_{filter} , fazendo $K_{filter} = K_{list}$.
- (iv) Repetimos as etapas (ii) e (iii) até que cheguemos aos N sítios totais da rede, onde $N = m_0 + t$, sendo t o número de sítios adicionados e m_0 o número de sítios iniciais.

APÊNDICE B – ALGORITMO: MODELO DO CRESCIMENTO DA CONEXÃO PREFERENCIAL

A rede é iniciada com um sítio no centro. Para que possamos dispor os sítios em posição arbitrárias, com r satisfazendo a Eq. (5.2). Abaixo podemos observar como funcionará o algoritmo para o modelo de crescimento da conexão preferencial.

Algoritmo para o modelo de crescimento da conexão preferencial

- (i) É colocado o primeiro sítio na origem do sistema de referência, e, atualizado a posição do centro de massa inicial da rede, tendo assim

$$\begin{cases} \vec{r}_0 = (x_0, y_0, z_0, q_0) = (0, 0, 0, 0), \\ \vec{r}_{CM_0} = \vec{r}_0, \end{cases} \quad (\text{B.1})$$

atualizando também a lista K

$$K = [0]. \quad (\text{B.2})$$

- (ii) É gerado um valor aleatório de r seguindo a probabilidade dada pela Eq. (5.2), logo em seguida geramos uma uma posição numa na superfície de uma hiperesfera unitária (se $d = 4$), ou na superfície de uma esfera unitária (se $d = 3$), ou numa circunferência unitária (se $d = 2$), ou numa reta unitária (se $d = 1$). Expressamos a posição desse sítio relativo a origem do sistema de referencia, utilizando a Eq. (??), e, atualizamos o centro de massa

$$\begin{cases} \hat{r}_1 = (x_1, y_1, z_1, q_1), \\ \vec{r}'_1 = r\hat{r}_1 + \vec{r}_{CM_0}, \\ \vec{r}_{CM_1} = \frac{1}{2}(\vec{r}'_1 + \vec{r}_0). \end{cases} \quad (\text{B.3})$$

- (iii) O primeiro sítio adicionado se conectará ao sítio já existente, com isso, atualizaremos as listas de ligações, $Nodes$, e a lista dos graus K

$$\begin{cases} Nodes = [(1, 2)], \\ K = [1, 1], \end{cases} \quad (\text{B.4})$$

onde a tupla (1, 2) em *Nodes* nos diz que o sítio de índice 1 se conecta ao sítio de índice 2, enquanto a posição na lista *K* representa o índice do sítio, e o valor o seu grau.

- (iv) Geramos um outro sítio no espaço adequado, seguindo as restrições impostas em (ii), atualizando \vec{r}' e \vec{r}_{CM}

$$\begin{cases} \hat{r}_2 = (x_2, y_2, z_2, q_2), \\ \vec{r}'_2 = r\hat{r}_2 + \vec{r}_{CM_1}, \\ \vec{r}_{CM_2} = \frac{1}{3}(\vec{r}'_2 + \vec{r}_{CM_1}). \end{cases} \quad (\text{B.5})$$

- (v) Para que possamos utilizar a probabilidade de ligação da rede, dada pela Eq. (5.1), definimos o termo de normalização, N_{norm} , como sendo

$$N_{norm} = \sum_i K_i \cdot dist(r_i, r_j)^{-\alpha_A}, \quad (\text{B.6})$$

onde K_i é o grau do i -ésimo sítio expresso na lista *K* e $dist(r_i, r_j)$ é a distância euclidiana, dada por

$$dist(r_i, r_j) = \sqrt{((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 + (q_i - q_j)^2)}, \quad (\text{B.7})$$

sendo r_j a posição do novo sítio e r_i a posição dos sítios já existentes na rede.

- (vi) Geramos um número aleatório com distribuição de probabilidade uniforme no intervalo $R \in [0, 1)$. Calculando a probabilidade de ligação dos sítios já existentes na rede, i , com o novo sítio adicionado, j , através da Eq. (5.2). Começando com primeiro índice $i = 1$,

$$\Pi_{1j} = \frac{1}{N_{norm}} K_1 \cdot dist(r_1, r_j)^{-\alpha_A}, \quad (\text{B.8})$$

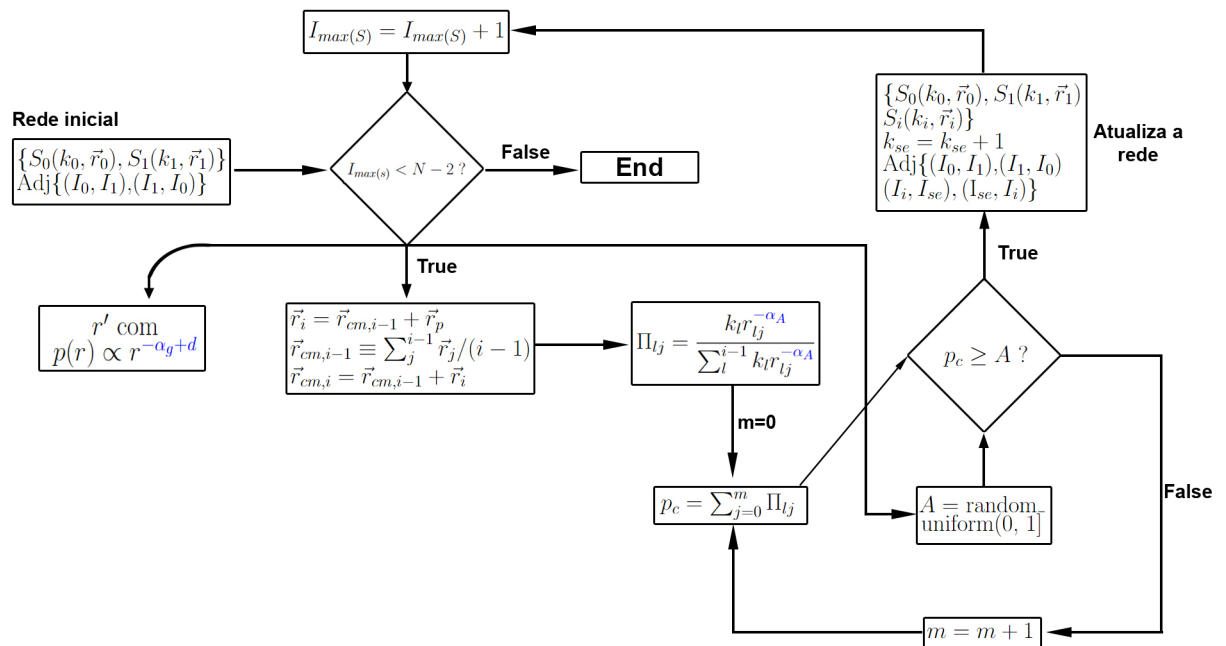
adicionando Π_{1j} ao valor de S , $S = S + \Pi_{1j}$. Caso $S < R$, calculamos Π_{2j} e adicionamos ao valor de S . Esse processo é repetido até que a condição $S \geq R$ seja satisfeita. Quando essa condição for satisfeita, selecionamos o último sítio adicionado que ocasionou nessa condição, sítio de índice 2, por exemplo, e adicionamos uma ligação dele com o novo sítio,

3. Atualizando as listas *K* e *Nodes*

$$\begin{cases} K = [1, 2, 1], \\ Nodes = [(1, 2), (2, 3)]. \end{cases} \quad (\text{B.9})$$

- (vii) As etapas (iv)-(vi) é repetido $N - 1$ vezes, para que possamos obter uma rede de tamanho N .

Figura 28 – Fluxograma de um algoritmo para implementação do modelo de crescimento da conexão preferencial. Os símbolos destacados em azul serão os parâmetros que iremos variar ao longo do trabalho.



Fonte: elaborado pelo autor

APÊNDICE C – OTIMIZAÇÃO DA ENTROPIA DE TSALLIS

Sistemas que possuem a entropia de Tsallis são não-extensivos, devido a essa entropia ser não-aditiva. Cientes dessa característica, nós podemos otimizar a entropia dada pela Eq. (3.13) sujeitas aos vínculos

$$\sum_{i=1}^W p_i = 1, \quad (\text{V.1})$$

$$\sum_{i=1}^W p_i \varepsilon_i = U_q, \quad (\text{V.2})$$

onde o vínculo dado pela Eq. (V.1) representa a normalização da distribuição das possíveis configurações, enquanto o vínculo dado pela Eq. (V.2) nos diz respeito a energia interna do sistema, sendo ε_i o espectro de energia, chamada de espectro generalizado e U_q a energia interna generalizada, sendo chamada de energia interna generalizada.

Utilizando a Eq. (3.14), podemos reescrever a entropia dada pela Eq. (3.13) na forma

$$S_q = k \frac{\sum_i^W p_i - \sum_i^W p_i^q}{(q-1)} = -k \sum_i^W \frac{p_i(p_i^{1-q} - 1)}{1-q} = -k \sum_i^W p_i^q \ln_q(p_i). \quad (\text{C.1})$$

Nosso objetivo é encontrar uma função que optimize a entropia descrita na Eq. (V.1), para isso, usaremos a técnica dos multiplicadores de Lagrange (GOLDSTEIN, 1980). Nós queremos encontrar o conjunto de parâmetros λ_1, λ_2 que maximizem nossa função L , dada por:

$$L = S_q + \lambda_1 \underbrace{\left(\sum_i^W p_i - 1 \right)}_{\text{vínculo 1}} + \lambda_2 \underbrace{\left(\sum_i^W \varepsilon_i p_i - U_q \right)}_{\text{vínculo 2}}, \quad (\text{C.2})$$

devemos encontrar os valores de λ_1, λ_2 que satisfaça $\partial L / \partial p_j = 0$. Aplicando a derivada parcial na Eq. (C.2), teremos

$$\left. \frac{\partial L}{\partial p_j} \right|_{\forall j} = \sum_i^W \frac{\partial}{\partial p_j} \left[S_q + \lambda_1 (p_i - 1) + \lambda_2 (\varepsilon_i p_i - U_q) \right] = 0. \quad (\text{C.3})$$

Para prosseguirmos com as contas, devemos encontrar a forma explícita da derivada parcial da entropia. Nós teremos:

$$\frac{\partial}{\partial p_j} S_q = -k \frac{\partial}{\partial p_j} \left(\sum_i^W p_i \ln_q(p_i) \right) \quad (\text{C.4})$$

$$\frac{\partial}{\partial p_j} S_q = k \left[\underbrace{\sum_i^W \ln_q(p_i) \frac{\partial}{\partial p_j} p_i^q}_{I_1} + \underbrace{\sum_i^W p_i^q \frac{\partial}{\partial p_j} \ln_q(p_i)}_{I_2} \right], \quad (\text{C.5})$$

separando a conta em dois termos separados, I_1 e I_2 , para a expressão I_1 , teremos:

$$I_1 = \sum_i^W \left(q p_i^{q-1} \ln_q(p_i) \frac{\partial}{\partial p_j} p_i \right), \quad (\text{C.6})$$

usando a definição de \ln_q dada pela Eq. (3.14)

$$I_1 = \frac{1}{1-q} \sum_i^W \left(q p_i^{q-1} (p_i^{1-q} - 1) \frac{\partial}{\partial p_j} p_i \right), \quad (\text{C.7})$$

$$I_1 = \frac{q}{1-q} \sum_i^W \left((1 - p_i^{q-1}) \frac{\partial}{\partial p_j} p_i \right), \quad (\text{C.8})$$

$$I_1 = \frac{q}{1-q} \left[\sum_i^W \frac{\partial}{\partial p_j} p_i - \sum_i^W p_i^{q-1} \frac{\partial}{\partial p_j} p_i \right], \quad (\text{C.9})$$

devido ao vínculo dado pela Eq. (V.1), nós teremos

$$\sum_i^W p_i = 1 \implies \frac{\partial}{\partial p_j} \sum_i^W p_i = 0, \quad (\text{C.10})$$

$$\boxed{I_1 = \frac{q}{q-1} \sum_i^W p_i^{q-1} \frac{\partial}{\partial p_j} p_i}, \quad (\text{C.11})$$

Para I_2 nós teremos

$$I_2 = \sum_i^W p_i^q \frac{\partial}{\partial p_j} \ln_q(p_i), \quad (\text{C.12})$$

$$I_2 = \frac{1}{1-q} \sum_i^W p_i^q \frac{\partial}{\partial p_j} (p_i^{1-q} - 1), \quad (\text{C.13})$$

$$I_2 = \frac{1}{1-q} \sum_i^W p_i^q (1-q) p_i^{-q} \frac{\partial}{\partial p_j} p_i, \quad (\text{C.14})$$

$$\boxed{I_2 = \frac{1}{1-q} \sum_i^W \frac{\partial}{\partial p_j} p_i = 0.} \quad (\text{C.15})$$

Utilizando as Eqs. (C.11), (C.15) na Eq. (C.5), teremos

$$\boxed{\frac{\partial}{\partial p_j} S_q = - \sum_i^W \frac{qk}{q-1} p_i^{q-1} \frac{\partial}{\partial p_j} p_i.} \quad (\text{C.16})$$

Voltamos para o cálculo do diferencial de L dado pela Eq. (C.3). Nós teremos

$$\begin{cases} \frac{\partial}{\partial p_j} U_q = 0, \\ \frac{\partial}{\partial p_j} 1 = 0, \end{cases} \quad (\text{C.17})$$

fazendo com que a Eq. (C.3) fique na forma

$$\left. \frac{\partial L}{\partial p_j} \right|_{\forall j} = \sum_i^W \frac{\partial}{\partial p_j} [S_q + \lambda_1 p_i + \lambda_2 \varepsilon_i p_i] = 0, \quad (\text{C.18})$$

$$\implies \left[\frac{\partial}{\partial p_j} (S_q) + \lambda_1 \frac{\partial}{\partial p_j} (p_i) + \lambda_2 \frac{\partial}{\partial p_j} (\varepsilon_i p_i) \right]_{\forall i} = 0, \quad (\text{C.19})$$

usando a Eq. (C.16) na Eq. (C.19), teremos

$$\implies \left[\left(\frac{-kq}{q-1} p_i^{q-1} + \lambda_1 + \lambda_2 \varepsilon_i \right) \frac{\partial p_i}{\partial p_j} \right]_{\forall i} = 0, \quad (\text{C.20})$$

como existem valores nos quais $\partial p_i / \partial p_j \neq 0$, teremos

$$\implies \frac{-kq}{q-1} p_i^{q-1} + \lambda_2 \varepsilon_i + \lambda_1 = 0, \quad (\text{C.21})$$

isolando p_i , obtemos

$$p_i = \left[(\lambda_1 + \lambda_2 \varepsilon_i) \frac{q-1}{kq} \right]^{1/(q-1)}, \quad (\text{C.22})$$

fazendo $\lambda_2 = -\beta kq \lambda_1 (q-1)$, teremos

$$p_i = (\lambda_1 + \lambda_2 \varepsilon_i) \frac{q-1}{kq}, \quad (\text{C.23})$$

$$p_i = \left(\frac{(q-1)\lambda_1}{kq} \right)^{1/(q-1)} \left[1 - \beta \varepsilon_i (q-1) \right]^{1/(q-1)}, \quad (\text{C.24})$$

$$p_i = \left(\frac{(q-1)\lambda_1}{kq} \right)^{1/(q-1)} e^{\beta \varepsilon_i}, \quad (\text{C.25})$$

por meio do vínculo expresso na Eq. (V.1), tendo p_i dado pela Eq. (C.25), teremos

$$\sum_i \left(\frac{(q-1)\lambda_1}{kq} \right)^{1/(q-1)} e^{\beta \varepsilon_i} = 1, \quad (\text{C.26})$$

$$\left(\frac{(q-1)\lambda_1}{kq} \right)^{1/(q-1)} \sum_l^W e^{\beta \varepsilon_l} = 1 \implies \left(\frac{(q-1)\lambda_1}{kq} \right)^{1/(q-1)} = \frac{1}{\sum_l^W e^{\beta \varepsilon_l}}, \quad (\text{C.27})$$

usando a Eq. (C.27) na Eq. (C.25), obtemos

$$\boxed{p_i = \frac{e^{\beta \varepsilon_i}}{\sum_l^W e^{\beta \varepsilon_l}}}. \quad (\text{C.28})$$