

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS**

Éverton Fernandes da Cunha

Redes hipercomplexas

São Carlos

2021

Éverton Fernandes da Cunha

Redes hipercomplexas

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada

Opção: Física Computacional

Orientadora: Prof. Dr. Luciano da Fontoura Costa

Versão corrigida

(versão original disponível na Unidade que aloja o Programa)

São Carlos

2021

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Cunha, Éverton Fernandes da
Redes hipercomplexas / Éverton Fernandes da Cunha;
orientador Luciano da Fontoura Costa - versão corrigida --
São Carlos, 2021.
78 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Física Aplicada Computacional) -- Instituto de Física de
São Carlos, Universidade de São Paulo, 2021.

1. Complexidade. 2. Redes complexas. 3. Geração de
redes. 4. Topologia de redes. 5. Redes hipercomplexas. I.
Costa, Luciano da Fontoura, orient. II. Título.

Dedico este trabalho a todos os brasileiros que tanto estão sofrendo nesta pandemia de Covid-19, em especial os profissionais da saúde, pesquisadores e comunicadores científicos que continuam se sacrificando diariamente.

AGRADECIMENTOS

O avanço do conhecimento da humanidade apenas é possível coletivamente e o desenvolvimento deste trabalho científico não é uma exceção. A realização desta dissertação apenas foi possível graças ao suporte de inúmeras pessoas durante estes mais de dois anos de mestrado. Sendo assim, utilizo este espaço para expressar meus sinceros agradecimentos a todos que, de alguma forma durante esse tempo, participaram do meu desenvolvimento como pessoa e pesquisador que, por fim, resultou nesta dissertação. Infortunadamente, aqui não há espaço suficiente e minha memória certamente me impedirá de lembrar de todos, mas a quem eu não mencionar, tenha ciência que sou grato por ter participado do meu desenvolvimento.

Agradeço, primeiramente, aos meus pais, Valdeci e Haidê, pelo incondicional e incessante suporte em toda a minha vida. Apenas por tê-los comigo já é um imenso privilégio, porém, além disso, reconheço o fundamental mérito deles em todas as minhas conquistas. O acolhimento inexorável que me deram, durante o período da pandemia de Covid-19, com absoluta certeza foi o alicerce deste texto, não consigo imaginar como seria finalizar este trabalho neste momento tão devasto sem tê-los como refúgio.

Agradeço ao meu irmão Evandro, por me proporcionar um ambiente de pesquisa desde cedo, por sempre acreditar em mim como pesquisador e também como pessoa, e por sempre me proporcionar pontos de vistas diferentes durante nossas discussões. Também agradeço ao meu irmão Diego que, com suas limitações, me mostra como o ser humano é complexo e ao mesmo tempo sensível, sendo uma constante oportunidade de reflexão sobre como ser uma pessoa melhor. Nós três crescemos e mudamos muito com o passar dos anos e, apesar das dificuldades que enfrentamos, está sendo uma divertida e rica experiência dividir esta viagem com vocês.

Agradeço ao meu orientador Luciano Costa por todo apoio e conhecimento compartilhado, além de toda a experiência de pesquisa fundamental na minha formação como pesquisador. Também sou grato pela compreensão dos diversos problemas que passei durante este trabalho, isso contribuiu para que me sentisse com liberdade de pensar, pesquisar e ter o tempo que fosse necessário para poder fazer um trabalho de qualidade.

Agradeço ao meu amigo Henrique Arruda por inúmeros e inestimáveis conselhos durante a realização deste trabalho. Além disso, agradeço profundamente pela amizade, pelo apoio em momentos difíceis e por ser uma das pessoas que mais acreditam no meu trabalho e na minha capacidade como pesquisador, isso tudo foi fundamental para minha formação como cientista.

Agradeço aos meus colegas de grupo de pesquisa pois as nossas discussões foram

imprescindíveis para que eu aprendesse tudo o que aprendi durante esses anos além de me proporcionarem um ambiente em que eu me sentisse livre para levantar ideias e questionamentos. Agradeço ao Eric Tokuda por pacientemente lidar com diversas intromissões minhas em seu trabalho para falar sobre dúvidas de programação, problemas na pesquisas e diversos assuntos pessoais, eu não tive culpa de sentarmos próximos na mesma sala para ter que me aguentar, mas saiba que seu sempre presente suporte foi muito importante para mim. Agradeço ao Paulo Burke por inúmeras e prazerosas conversas sobre os mais variados assuntos, além de todo suporte computacional e de programação que foram fundamentais para que eu obtivesse os resultados deste trabalho. Agradeço ao Bruno Messias pelos inúmeros momentos de companhia, além das diversas conversas e apoio. Agradeço ao Filipi Nascimento que, mesmo em outro país, me ajudou muito em diversas questões relacionados a pesquisa e vida de pesquisador principalmente com uma visão internacional desses assuntos. Agradeço ao Thomas Peron e ao Paulo Ventura por muitas conversas e discussões, por sempre serem pacientes em discutir minhas visões e tirarem minhas dúvidas. Agradeço ao Roberto Furuta e ao Guilherme Domingues por várias questões que compartilhamos durante nossos mestrados. Também sou grato a diversos outros pesquisadores do meu grupo que, apesar de não termos tido tanto contato, eventualmente me proporcionaram ricas discussões importantes durante todo o mestrado.

Agradeço a Mara, minha professora de inglês, não apenas por tanto me auxiliar na língua, mas por ser sempre alguém disposta a ter boas conversas que me fez ter diversas importantes reflexões.

Em especial, eu gostaria de agradecer a um grupo de amigos que me acompanharam nesses últimos anos, impactando fortemente no que sou atualmente. Agradeço ao Guilherme Passos por compartilhar o meu inefável prazer de ter uma simples conversa permeada de curiosidade por coisas que nunca seremos capazes de compreender completamente. Agradeço ao Hudson Soares pelo companheirismo que se iniciou na minha graduação, mas que se estende até hoje, sempre me dando forças para que eu continue pesquisando. Agradeço ao Aislan de Borba por sempre compartilhar algum conhecimento ou experiência que eu não esperava me fazendo constantemente lembrar que existe muita coisa interessante para se conhecer. Agradeço ao Guilherme Rubel Porn por simplesmente ser quem ele é, é um prazer poder conviver com sua forma peculiar de viver, sempre sincero e amável. Agradeço a minha amiga Carolina Carvalho por me acompanhar nesses últimos meses, me fazendo ver o mundo de outras maneiras e por dividirmos as nossas felicidades e problemas cotidianos, com ela esses meses foram bem mais leves e prazerosos. Apesar das distâncias, espero que eu e todas essas pessoas possamos ter muitas histórias e experiências pela frente. Citando tio Sagan novamente ‘Diante da vastidão do tempo e da imensidão do universo, é um imenso prazer para mim dividir um planeta e uma época com você[s]’.

Agradeço ao CNPq (830717/1999-4 — 134181/2019-0) pelo financiamento deste

trabalho através da concessão de bolsa de mestrado.

Por último, mas não menos importante, agradeço às iniciativas que democratizam e tornam mais acessíveis o conhecimento da humanidade em especial a ave negra que com a chave em seu bico abre as portas das pesquisas e a grande biblioteca que disponibiliza o saber disponível em livros. O conhecimento é um direito humano fundamental e absolutamente nenhuma pessoa deveria ser privada de poder aprender, independente de qualquer que sejam suas condições. O prazer de aprender algo novo é uma felicidade que não deveria ser negada a ninguém. Matar a curiosidade e a esperança de conhecer coisas novas de uma pessoa é uma das maiores desumanidades que alguém pode cometer. Não sei se a vida é bela, mas sei que ela é uma viagem só de ida, cabendo a nós apenas experienciá-la da melhor forma possível e acredito que sendo amável, curioso e espalhando o fascínio pela diversidade é uma das melhores formas de viajar por ela.

*'The mystery of life isn't a problem to solve, but a reality to experience.
A process cannot be understood by stopping it. Understanding must
move with the flow of the process, must join it and flow with it'*
Frank Herbert, Dune

RESUMO

CUNHA, É. F. da **Redes hipercomplexas**. 2021. 78p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2021.

As redes complexas têm recebido muita atenção desde o início de seu desenvolvimento, porém, um dos conceitos centrais da área, a ‘complexidade’, ainda não está completamente definido. Tradicionalmente, compreende-se esse termo como a expressão da heterogeneidade dos graus dos nós de uma rede. Entretanto, dado que a distribuição de graus não é capaz de caracterizar suficientemente toda a topologia de uma rede complexa, faz-se necessário considerar medidas adicionais para obter uma descrição de complexidade mais abrangente. No presente trabalho, apresenta-se um índice de complexidade baseado na heterogeneidade de valores de um diverso grupo de medidas topológicas, incluindo grau dos nós, caminho mínimo, coeficiente de aglomeração local, *betweenness centrality*, *matching index*, autovalor Laplaciano e grau hierárquico. Utilizando esse índice, objetiva-se obter um método para aumentar a complexidade de redes complexas, gerando um novo modelo de redes complexas, nomeada de redes hipercomplexas - HC, que apresenta alta complexidade em comparação a um conjunto de modelos teóricos de referência (Erdős-Rényi, Barabási-Albert, Waxman, *Random Geometric Graph* e Watts-Strogatz). Esse método é proposto da seguinte forma: inicia-se com uma rede complexa, que então é submetida a um processo de otimização que altere sua configuração topológica enquanto busca maximizar o seu índice de complexidade. O processo de otimização à princípio aqui adotado consiste na reconexão de arestas de forma uniformemente aleatória enquanto é aplicado o algoritmo de descida de gradiente a cada iteração. Como resultado, observou-se um real incremento na complexidade das redes, em que as HCs, no final do processo, apresentaram índices de complexidade aproximadamente quatro vezes maiores do que os índices das redes do modelo Erdős-Rényi, que foram escolhidas como redes iniciais nesse experimento. Além disso, durante o processo de maximização, as redes hipercomplexas apresentaram uma tendência de gerar características diferentes das presentes nos modelos teóricos considerados. De fato, o incremento da complexidade das HCs resultou em um aumento da diversidade dos valores das medidas de três a quatro vezes em relação às redes iniciais. Por fim, numa abordagem qualitativa, observou-se que esse aumento na diversidade das medidas gerou características topológicas interessantes como o surgimento de longos ramos periféricos e vários *hubs* com conectividades semelhantes tanto ao iniciar o método com redes do modelo Erdős-Rényi quanto também do modelo Barabási-Albert.

Palavras-chave: Complexidade. Redes complexas. Geração de redes. Topologia de redes. Redes hipercomplexas.

ABSTRACT

CUNHA, É. F. da **Hypercomplex networks**. 2021. 78p. Dissertation (Master in Science) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2021.

Complex networks have been receiving a plenty of attention since the beginning of their development, however, one of the central concepts in the field, ‘complexity’, is not yet fully defined. Traditionally, this term is understood as the expression of the heterogeneity of the node degrees of a network. However, since the degree distribution is not enough to sufficiently characterize the overall topology of a complex network, it is necessary to consider additional measures to obtain a more comprehensive description of complexity. In this work, we present a complexity index based on the heterogeneity of values of a diverse group of topological measures, including degree of nodes, shortest path length, local cluster coefficient, betweenness centrality, matching index, Laplacian eigenvalue and hierarchical node degree. Employing this index, we aim at obtaining a method to increase the complexity of complex networks, generating a new model of complex networks, named hypercomplex networks - HC, which presents high complexity compared to a group of theoretical reference models (Erdős-Rényi, Barabási-Albert, Waxman, Random Geometric Graph e Watts-Strogatz). This method is proposed as follows: it starts with a complex network, which is then submitted to an optimization process that changes its topological configuration while seeking to maximize its complexity index. The optimization process at first adopted here consists of rewiring edges in a uniformly random way while applying the gradient descent algorithm to each iteration. As a result, there was a real increase in the complexity of the networks, in which the HCs, at the end of the process, presented complexity indices about four times greater than the network index of Erdős-Rényi model, that were chosen as initial networks in this experiment. Furthermore, during the maximization process, hypercomplex networks tended to generate characteristics different from those present in other theoretical models considered. Indeed, the complexity increment of HCs produced an increase in the diversity of the values of the measurements of three to four times compared with the initial networks. Finally, in a qualitative approach, it was observed that this enhancement in the diversity of measurements generated interesting topological features, such as the emergence of long peripheral branches and several hubs with similar connectivity both when starting the method with networks from the Erdős-Rényi model and also from the Barabási-Albert model.

Keywords: Complexity. Complex networks. Network generation. Network topology. Hypercomplex networks.

LISTA DE FIGURAS

- Figura 1 – Exemplo de uma rede com 20 nós com grau 5. Apesar de ser uma rede perfeitamente regular em sua distribuição de graus, apresenta uma complicada estrutura, não podendo ser intuitivamente considerada uma rede simples. 28
- Figura 2 – Representações de rede complexas. **(a)** Exemplo de uma rede não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas. **(b)** Representação em forma de matriz de adjacência caracterizando o exemplo de rede. Cada linha e cada coluna (realçadas em azul) representam um nó na rede. Quando $a_{ij} = 1$, existe uma aresta entre n_i e n_j e não existe quando $a_{ij} = 0$. Por ser uma rede não direcionada e sem peso a matriz é simétrica e composta apenas por 0s e 1s. **(c)** Representação em forma de lista de adjacência caracterizando o exemplo de rede em que a lista Ω_i guarda a informação de todos os nós em que n_i está conectado. . . . 32
- Figura 3 – Exemplo de uma rede complexa evidenciando a medida de grau. **(a)** Exemplo de uma rede não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas, com o valor do grau respectivo a cada nó. **(b)** Matriz de adjacência respectiva a rede complexa, porém com a expansão de uma linha e de uma coluna evidenciando, em vermelho, o grau de cada nó como o somatório das linhas e das colunas. **(c)** Distribuição dos graus da rede complexa, apresentando 3 nós com grau 2, 2 nós com grau 3 e 1 nó com grau 4. Assim, o grau médio da rede dá-se por $\langle k \rangle = 16/6 = 2,66$. 34
- Figura 4 – Exemplo de uma rede complexa evidenciando a medida de caminho mínimo. **(a)** Exemplo de uma rede não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas. Em vermelho, estão os dois caminhos mínimos possíveis entre os nós n_3 e n_5 , ambos passando inicialmente por n_0 , porém cada caminho passando por dois nós diferentes, n_4 e n_1 , contabilizando $l_{35} = 3$. Também pode ser observado que por ser uma rede não direcionada, tem-se $l_{35} = l_{53}$, alterando apenas a ordem do caminho. **(b)** Matriz de distâncias mínimas respectiva a rede complexa em que cada elemento l_{ij} assume o valor do caminho mínimo entre os respectivos nós. Também é possível perceber que a matriz é simétrica, informando novamente que $l_{35} = l_{53}$, e que não há autoligação já que $l_{ii} = 0$. Assim, o caminho mínimo médio dá-se por $\langle l \rangle = 48/30 = 1,6$. . . 35

Figura 5 – Exemplo de uma rede complexa evidenciando a medida coeficiente de aglomeração sendo não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas. Em vermelho, estão os valores de coeficiente de aglomeração para cada um dos nós da rede. Como exemplo, considere c_0 : o nó n_0 é vértice de 2 triângulos, sendo (n_0, n_3, n_2) e (n_0, n_4, n_1) , e o máximo de conexões possíveis que esse nó poderia fazer é $N_3 = 4 \times (4 - 1) \times 2 = 6$, portanto, $c_0 = 2/6 = 0,33$. Levando em consideração a rede toda, é possível calcular a sua média do coeficiente de aglomeração local por $\langle c \rangle = 4,67/6 = 0,78$ 36

Figura 6 – Exemplo de uma rede complexa não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas sendo evidenciado em vermelho a medida *betweenness centrality*. **(a)** Em vermelho, estão os valores normalizados de *betweenness centrality* para os nós. Como exemplo, considere B_1 : pelo nó n_1 se passa 3 caminhos mínimos, l_{05} , l_{25} e l_{35} , entretanto, para cada par de nós destes caminhos, existe mais um outro caminho mínimo passando por n_4 , portanto, tem-se $B_1 = 1/2 + 1/2 + 1/2 = 4,5$, com a normalização resulta-se $B_1 = 4,5/(5 \times 4/2) = 4,5/10 = 0,45$. **(b)** Em vermelho, estão os valores normalizados de *betweenness centrality* para as arestas. Para exemplificar, observe B_{15} : além do próprio caminho l_{15} , há mais 3 pares de nós em que seus caminhos mínimos passam pela aresta e_{15} porém também passam pela aresta e_{45} , portanto, tem-se $B_{15} = 1 + 1/2 + 1/2 + 1/2 = 2,5$, aplicando a normalização encontra-se $B_{15} = 2,5/(6 \times 5/2) = 0,17$ 37

Figura 7 – Exemplo de uma rede complexa não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas sendo evidenciado em vermelho a medida *matching index*. Como exemplo, considere a aresta e_{14} que conecta os nós n_1 e n_4 que por sua vez compartilham 2 vizinhos, os nós n_0 e n_5 , e seus graus são $k_1 = k_4 = 3$, portanto, tem-se $\mu_{14} = 2/(3 + 3 - 2) = 0,5$ 38

Figura 8 – Exemplo de uma rede complexa não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas evidenciando a medida grau hierárquico. Como exemplo, observá-se os níveis hierárquicos para o nó n_0 , definido pelos círculos concêntricos vermelho para o primeiro nível e verde para o segundo nível, e seus respectivos graus hierárquicos $k_0(1) = 4$ e $k_0(2) = 1$. 40

Figura 9 – Exemplo da distribuição de grau de uma rede do modelo Erdős-Rényi gerada com $N = 500$ nós, apresentando $\langle k \rangle = 7,88$ 41

- Figura 10 – Exemplo de uma rede do modelo Erdős-Rényi gerada com $N = 500$ nós e p calculado para que $\langle k \rangle = 8$. Valor real de $\langle k \rangle$ para essa rede é 7,88. A barra de cor representa o grau dos nós e com ela é possível constatar que os valores dos graus são aproximadamente iguais a média, demonstrando uma conectividade estatisticamente regular. 41
- Figura 11 – Exemplo de duas redes do modelo Watts-Strogatz gerada com $N = 500$ nós $\langle k \rangle = 8$. A rede superior foi gerada com o valor crítico de $p = 0,1$, ou seja, 10% de reconexão, sendo então um exemplo de rede com o efeito pequeno mundo. A rede inferior foi gerada com $p = 1$, ou seja, 100% de reconexão, sendo o caso extremo tornando-se uma rede completamente aleatória. A barra de cor representa o grau dos nós e com ela é possível constatar que em ambos os casos não há muita variação no valor dos graus. Porém, enquanto a rede superior apresenta praticamente todos os nós com o mesmo valor de grau abaixo de 10, a inferior contém diversos nós com grau um pouco acima de 10, tendo uma distribuição de grau levemente mais diversa que no primeiro caso. 43
- Figura 12 – Exemplo da distribuição de grau de uma rede do modelo Barabási-Albert gerada com $N = 500$ nós, apresentando $\langle k \rangle = 7,96$ 44
- Figura 13 – Exemplo de uma rede do modelo Barabási-Albert gerada com $N = 500$ nós e p calculado para que $\langle k \rangle = 8$. Valor real de $\langle k \rangle$ para essa rede é 7,96. A barra de cor representa o grau dos nós. Observa-se a existência de apenas 3 *hubs* com conectividade acima de 60, enquanto a imensa maioria dos nós apresenta grau abaixo da média e apenas alguns poucos nós estão entre estes dois extremos. 46
- Figura 14 – Exemplo de uma rede da segunda versão do modelo Waxman gerada com $N = 500$ nós e parâmetros calculados para que a rede tenha $\langle k \rangle$ próximo de 8, $\beta = 1$, $alpha = 0,04$. Valor real de $\langle k \rangle$ para essa rede é 8,08. A barra de cor representa o grau dos nós e com ela é possível observar que a conectividade varia entre 1 e aproximadamente 25, porém com a maioria dos nós apresentando grau próximo da média. 47
- Figura 15 – Exemplo de uma rede do modelo *Random Geometric Graph* gerada com $N = 500$ nós e $R = 0,0745$ calculado para que a rede tenha $\langle k \rangle$ próximo de 8. Valor real de $\langle k \rangle$ para essa rede é 7,92. A barra de cor representa o grau dos nós. É possível observar vários pequenos agrupamentos em volta de nós com grau um pouco acima da média, porém a maioria dos nós tem a conectividade aproximadamente ou menor que a média. 48

Figura 16 – Diagrama que ilustra os elementos adotados para a definição do índice de complexidade. Cada linha do esquemático corresponde ao desvio padrão de cada uma das medidas topológicas estandardizadas, σ_m , de uma dada rede complexa que terá sua complexidade mensurada. O desvio padrão máximo, σ_{max} , que aqui está sendo exemplificado por σ_2 , define a delimitação de um retângulo que representa o máximo desvio padrão efetivo, β , e também serve para calcular os valores da diferença δ_m para cada medida. A complexidade relativa será máxima, $\omega = 1$, quando as diferenças não existirem, resultando em $\Delta = 0$, ou seja, será o caso em que todos os σ_m estarão no limite desse retângulo. 51

Figura 17 – Gerador de redes hipercomplexas. Projetou-se o algoritmo para gerar redes hipercomplexas como um maximizador de complexidade através de pequenas alterações topológicas. Para que isso seja feito, é necessário que seis parâmetros sejam definidos: **(a)** escolher uma rede que terá sua complexidade maximizada; **(b)** selecionar um grupo de redes complexas, podendo ser modelos teóricos ou não, para que sirvam de referência; **(c)** escolher um grupo de medidas topológicas para que a complexidade possa ser calculada; **(d)** selecionar a medida de complexidade a ser maximizada, neste trabalho definimos o IC com esse propósito, mas outras podem ser utilizadas; **(e)** escolher a modificação topológica que será aplicada na rede para que aumente sua complexidade; **(f)** selecionar o algoritmo de otimização que maximizará a complexidade da rede. É importante notar que a decisão de um parâmetro pode influenciar na decisão de outros, por exemplo o IC utilizado aqui necessita que as medidas sejam estandardizadas. 54

Figura 18 – Média e desvio padrão do IC de 100 exemplares de redes HCs com $N = 500$ e $\langle k \rangle = 8$ durante o período de maximização de suas complexidades, com valores a cada 100 iterações. Apesar dos valores de ICs apresentarem uma dispersão maior da iteração 1500 em diante, fica bem claro que a complexidade aumentou consideravelmente com as sucessivas iterações. De fato, a média dos ICs das redes HC-4000 é mais do que 4 vezes maior do que a média da complexidade das redes ER (HC-0) e mesmo os menores ICs das redes HC-4000s ainda são mais do que 3,5 vezes maiores do que os maiores valores de complexidades do modelo ER. . . 58

Figura 19 – Análise visual da representação do espaço PCA das 600 redes de referência, 100 de cada modelo teórico, e 400 redes hipercomplexas divididas em 4 estágios de evolução do modelo, sendo HC-1000, HC-2000, HC-3000 e HC-4000. O mapa de calor apresenta os valores do IC referente a cada rede. Pode ser observado que as redes hipercomplexas se distanciam dos outros modelos conforme sua complexidade é aumentada. Além disso, elas descrevem uma tendência a ir para uma região de baixa densidade de redes, principalmente longe do modelo BA, indicando que o aumento da complexidade gera propriedades topológicas diferentes das presentes nos modelos considerados.	59
Figura 20 – Média e desvio padrão das 10 medidas topológicas calculadas a cada 100 passos de otimização para as redes hipercomplexas, evoluindo das redes iniciais, do modelo ER, até se tornarem redes HC-4000s.	61
Figura 21 – Exemplo de evolução de duas redes hipercomplexa de $N = 500$ nós e $\langle k \rangle = 8$ em que a sequência de redes da esquerda inicia-se com o modelo ER e a da direita com o modelo BA. Evolução até 4000 iterações, mas continuando na Figura 22. A escala do mapa de calor da esquerda é referente aos nós das redes hipercomplexas iniciadas no modelo ER, semelhantemente para BA na direita.	64
Figura 22 – Continuação da Figura 21 da evolução de duas redes hipercomplexa de $N = 500$ nós e $\langle k \rangle = 8$ em que a sequência de redes da esquerda inicia-se com o modelo ER e a da direita com o modelo BA. Evolução da iteração 4000 até 8000. A escala do mapa de calor da esquerda é referente aos nós das redes hipercomplexas iniciadas no modelo ER, semelhantemente para BA na direita.	65

LISTA DE ABREVIATURAS E SIGLAS

BA	modelo Barabási-Albert
ER	modelo Erdős-Rényi
HC	modelo de rede HiperComplexa
IC	Índice de Complexidade
PCA	<i>Principal Component Analysis</i>
RGG	modelo <i>Random Geometric Graph</i>
WAX	modelo Waxman
WS	modelo Watts-Strogatz

LISTA DE SÍMBOLOS

p	Probabilidade de conexão
N	Número de nós da rede
\mathcal{G}	Objeto matemático grafo
\mathcal{N}	Conjunto de nós de uma rede
\mathcal{E}	Conjunto de arestas de uma rede
E	Número de arestas de uma rede
n_i	Nó qualquer i
$e_{i,j}$	Aresta entre os nós i e j
A	Matriz de adjacência
D	Matriz diagonal em que os seus elementos são os graus dos respectivos nós
Ω_i	Lista de adjacência do nó i
a_{ij}	Elemento de A respectivo aos nós i e j
k_i	Grau do nó i
$\langle \cdot \rangle$	Média de um conjunto de valores
l_{ij}	Distância do caminho mínimo entre os nós i e j
$N_{\Delta}(i)$	Número de triângulos em que o nó i participa como um nó
$N_3(i)$	Número máximo possível de arestas que o nó i poderia ter
c_i	Coefficiente de aglomeração local do nó i
B_u	<i>Betweenness centrality</i> do elemento u
μ_{ij}	<i>Matching index</i> dos nós i e j
$\epsilon(i, j)$	Quantidade de vizinhos mútuos entre os nós i e j
L	Matriz Laplaciana
I	Matriz identidade

λ	Autovalores
v	Autovetores
h	Nível hierárquico
$k_i(h)$	Grau de hierarquia h do nó i
κ	Número de ligações iniciais do modelo Watts-Strogatz
τ	Quantidade de conexões que um novo nó faz na geração do modelo Barabási-Albert
$d_E(n_i, n_j)$	Distância Euclidiana entre os nós i e j
$\sigma(\cdot)$	Quantidade de caminhos mínimos
σ_m	Desvio padrão de medidas
μ_m	Média de medidas
β	Máximo desvio padrão efetivo
ω	Complexidade relativa
δ_m	Diferença individual da dispersão da medida m
Δ	Somatório das diferenças das dispersões

SUMÁRIO

1	INTRODUÇÃO	27
1.1	Objetivos	29
1.2	Contribuições	29
1.3	Organização da Dissertação	30
2	REDES COMPLEXAS	31
2.1	Representações de redes complexas	31
2.2	Medidas de redes	33
2.2.1	Grau	33
2.2.2	Caminho mínimo	33
2.2.3	Coeficiente de Aglomeração	34
2.2.4	<i>Betweenness Centrality</i>	35
2.2.5	<i>Matching Index</i>	37
2.2.6	Autovalores Laplacianos	38
2.2.7	Graus Hierárquicos	39
2.3	Modelos de redes	39
2.3.1	Erdős-Rényi	39
2.3.2	Watts-Strogatz	42
2.3.3	Barabási-Albert	43
2.3.4	Waxman	46
2.3.5	<i>Random Geometric Graph</i>	48
3	COMPLEXIDADE	49
3.1	Índice de Complexidade	49
3.2	Redes hipercomplexas	53
4	RESULTADOS E DISCUSSÃO	57
4.1	Análise quantitativa das redes hipercomplexas	57
4.2	Análise visual das redes hipercomplexas	62
5	CONCLUSÕES	67
5.1	Considerações finais	67
5.2	Trabalhos futuros	68
	REFERÊNCIAS	71

1 INTRODUÇÃO

A ciência das redes complexas tem se desenvolvido fervorosamente desde o início da sua formalização, em meados dos anos 2000, tanto teoricamente quanto em diversas aplicações. (1–3) Essa sua ampla utilização pode ser responsabilizada, em grande parte, por alguns motivos como: (i) a facilidade com que seus conceitos básicos podem ser entendidos e utilizados advindo da teoria de grafos (4–6) e mecânica estatística (7–10) e (ii) ser a representação de qualquer sistema complexo que possa ser discretizado. Portanto, apresenta-se como uma área essencialmente multidisciplinar, abrangendo, por exemplo, diversos tipos de processamento de texto (11–18), diversas áreas da biologia (19–25), dinâmicas de propagação de doenças, *e.g.*, pandemia (26–33) e inúmeras outras áreas. (34–44)

Apesar da grande quantidade de pesquisas relacionadas a redes complexas, o próprio conceito do que é uma rede complexa se mantém como um desafio a ser entendido. (45) A questão principal desse problema advém do termo ‘complexidade’, que, no contexto de redes complexas, se mantém formalmente indefinido. A interpretação comumente utilizada do que é uma rede complexa é dada através da caracterização relativa a modelos teóricos bem conhecidos, ou seja, as redes são comparadas com modelos considerados *simples* e *complexos*. (46)

Da parte dos modelos simples, tem-se os modelos de redes aleatórias (5), em especial o modelo Erdős-Rényi (ER) (47), que são assim considerados por serem caracterizados apenas pela sua conectividade *homogênea*. A distribuição do grau dos nós desses modelos pode ser razoavelmente prevista meramente pelo valor de sua média que, por sua vez, é controlada apenas por um parâmetro: a probabilidade de conexão, p . Portanto, entende-se uma rede como simples aquela que pode ser caracterizada pela regularidade em sua distribuição de grau, tendo a rede completamente regular o seu caso limite. Nesse sentido, modelos uniformemente aleatórios, como ER, podem ser entendidos como estatisticamente regulares. (46)

No outro extremo, os modelos considerados complexos são aqueles em que a distribuição de grau dos seus nós é *heterogênea* o suficiente para que suas propriedades não possam ser preditas por apenas a média dos graus. Como um exemplo comumente considerado na literatura, tem-se o modelo Barabási-Albert, BA (48), conhecido por ser uma rede livre de escala pois sua distribuição de grau obedece a uma lei de potência. De fato, as redes desse modelo apresentam características mais diversas como poucos nós com uma alta conectividade, os *hubs*, e também muitos nós com poucas conexões. Sendo assim, assume-se comumente que uma rede é complexa quando sua conectividade é pouco uniforme. (46)

Entretanto, essa concepção se mostra insuficiente para caracterizar a complexidade de uma rede, pois é possível existir redes que sejam classificadas como regulares, devido à uniformidade na distribuição de grau de seus nós, porém apresentarem características topológicas não regulares. (45, 49) A Figura 1 apresenta um exemplo desse aparente paradoxo em que todos os 20 nós desta rede estão conectados a exatos 5 outros nós. Apesar dessa ser uma rede totalmente regular na sua distribuição de grau, ela aparenta ter uma topologia bastante complicada, sendo difícil de saber o quão complexa é apenas observando o grau de seus nós. Dessa forma, fica evidente a necessidade de levar em consideração outras medidas topológicas, ou ao menos mais do que apenas a distribuição de grau, na caracterização da complexidade de uma rede.

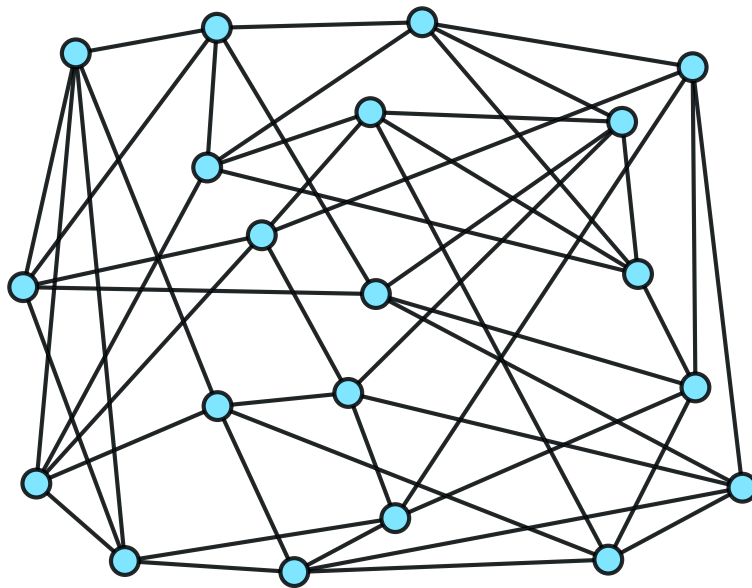


Figura 1 – Exemplo de uma rede com 20 nós com grau 5. Apesar de ser uma rede perfeitamente regular em sua distribuição de graus, apresenta uma complicada estrutura, não podendo ser intuitivamente considerada uma rede simples.

Fonte: Elaborada pelo autor.

Além dessa análise qualitativa, alguns esforços foram feitos no intuito de propriamente medir a complexidade de uma rede. As medidas propostas mais conhecidas e utilizadas se baseiam no conceito de *entropia* (50–65), principalmente a entropia de Shannon (64, 66, 67) (também chamada de *informação topológica* (50, 51)). Inclusive, alguns desses trabalhos descrevem a limitação de analisar apenas o grau e abordaram a complexidade por outras medidas como caminhos mínimos (63) e dimensão fractal. (61) Por exemplo, as medidas baseadas em entropia sofrem algumas limitações (68) e, dentre elas, a dificuldade de estimar propriamente a função de probabilidade da variável. (69–72) Além disso, essas formas de medir a complexidade não tiveram como objetivo combinar diversas propriedades topológicas das redes. Com isso, medir a complexidade de uma rede

complexa de uma maneira que considere diversas características topológicas se mantém como um problema em aberto.

Diante dessas questões, no presente trabalho foi proposto o desenvolvimento de um índice de complexidade de rede de uma forma mais abrangente do que as encontradas na literatura e, até onde se estende os conhecimentos dos autores deste trabalho, com uma abordagem completamente nova. Portanto, a medida de complexidade aqui definida se baseia na heterogeneidade de múltiplas medidas topológicas de redes complexas. Além disso, buscou-se investigar qual seria a rede mais complexa que pode ser obtida. Assim, desenvolveu-se um método para que seja possível aumentar a complexidade de redes complexas através da maximização dessa propriedade, tratando o índice aqui proposto como uma função a ser otimizada em um grande espaço de busca de reconfigurações topológicas possíveis.

1.1 Objetivos

O objetivo geral deste trabalho é desenvolver uma medida que mensure a complexidade de redes complexas e também gerar redes com alta complexidade. Dessa forma, os objetivos específicos são:

- **Definição de um Índice de Complexidade:** Propor um índice de complexidade que seja capaz de abranger diversos aspectos topológicos de redes complexas com base na heterogeneidade de um amplo conjunto de medidas;
- **Definição do método de geração de Redes Hipercomplexas:** Considerando a complexidade como uma propriedade de uma rede complexa e que pode ser maximizada, propomos um método que gera redes com alta complexidade através da aplicação de algoritmos de otimização e alterações topológicas das redes.

1.2 Contribuições

Nesta seção são apresentadas as contribuições desta dissertação que foram publicadas em periódico no artigo DA CUNHA, É. F.; COSTA, L. F. On Hypercomplex Networks, *Physica A*, v. 591, p. 126714, abr. 2021. DOI: 10.1016/j.physa.2021.126714:

- Caracterização de complexidade de redes complexas através da heterogeneidade de suas medidas;
- Definição de um índice de complexidade de redes complexas;
- Desenvolvimento de um método de geração de redes com alta complexidade;
- Geração de um novo modelo de rede complexa.

1.3 Organização da Dissertação

Esta dissertação está dividida da seguinte maneira. No Capítulo 2, é apresentado o referencial teórico do trabalho contendo as definições básicas sobre redes complexas assim como as medidas topológicas e modelos teóricos utilizados nos experimentos. No Capítulo 3, é apresentada a definição do índice de complexidade e a metodologia de geração das redes hipercomplexas. O Capítulo 4 apresenta os resultados dos experimentos utilizando o método e índice propostos. No Capítulo 5, são apresentadas as conclusões deste trabalho e, por fim, são descritas as perspectivas futuras desta pesquisa.

2 REDES COMPLEXAS

Esta dissertação se dedica ao estudo de complexidade em redes complexas abordando diversas medidas topológicas e modelos de rede. Assim sendo, este capítulo é responsável por apresentar os pontos fundamentais sobre esses tópicos para a compreensão do desenvolvimento deste trabalho. Dessa forma, a Seção 2.1 concentra os conceitos básicos sobre redes complexas em que as futuras definições serão baseadas, como as representações matemáticas de *grafo*, *matriz de adjacência* e *lista de adjacência*, bem como conceitos de *direcionalidade* e *peso* em redes. A Seção 2.2 apresenta as definições de algumas medidas topológicas e nomenclaturas importantes para o entendimento do restante deste texto. Por fim, na Seção 2.3, são introduzidos alguns modelos teóricos de redes que são clássicos no estudo de topologias de redes complexas, apresentando diversas características necessárias para a compreensão de complexidade baseada em heterogeneidade topológica.

2.1 Representações de redes complexas

Uma *rede complexa* é formalmente definida como um *grafo* e, conseqüentemente, essa representação demonstra o potencial de caracterizar matematicamente diversos sistemas que possam ser discretizados. Assim sendo, fica evidente que a ciência de redes desenvolveu-se sobre as definições e conceitos da teoria de grafos. (4–6, 73) Dessa forma, as redes complexas são compostas por dois elementos discretos, sendo eles os *nós*, também conhecidos como *vértices*, e as suas relações ou conexões, as *arestas*, também conhecidas como *ligações* ou *edges*. (2, 3, 34, 74–78) Uma rede complexa é definida como um objeto matemático $\mathcal{G}(\mathcal{N}, \mathcal{E})$, ou seja, um grupo \mathcal{G} formado pelo conjunto \mathcal{N} de N nós, sendo $\mathcal{N} = \{n_i \mid i \in \mathbb{Z} \text{ e } 0 \leq i \leq N\}$, e pelo conjunto \mathcal{E} com E arestas, sendo $\mathcal{E} = \{e_{ij} = (n_i, n_j) \mid n_i, n_j \in \mathcal{N}\}$.

Provavelmente, o conceito mais básico a ser construído sobre os elementos nós e arestas é a *adjacência*. Quando existe uma aresta conectando dois nós, é definido que esses dois nós são *adjacentes*, ou seja, são *vizinhos* e, com isso, tem-se as informações necessárias para reproduzir uma rede. As informações sobre as conexões entre os nós, que permitem a reprodução de uma rede, podem ser representadas em uma matriz, conhecida como *matriz de adjacência*, A , que na maioria dos casos assume dimensões $N \times N$. Outra representação de redes, que por muitas vezes é equivalente a matriz, é a *lista de adjacência*, Ω_i , que se constitui de um conjunto de listas, uma para cada nó n_i , guardando apenas os índices dos vizinhos de n_i . (1, 79)

Apesar da matriz de adjacência ser a representação mais conhecida e intuitiva, ela nem sempre é utilizada nos cálculos computacionais. Isso ocorre principalmente para redes grandes e esparsas, que são os casos em que existem poucas conexões entre os nós em comparação ao máximo possível, populando a matriz com muitos valores nulos referentes

aos nós não conectados. (3) Assim, a lista de adjacência resolve esse problema fornecendo apenas as informações entre os nós que apresentam uma conexão.

As arestas de uma rede apresentam duas características importantes, a *direcionalidade*, que informa se a relação do nó n_i para o nó n_j é a mesma do n_j para o n_i ou não; e o *peso*, que informa a intensidade das conexões. (79) A Figura 2 ilustra um exemplo de rede não direcionada, sem peso, e suas respectivas matriz e lista de adjacência. Os elementos da matriz de adjacência, a_{ij} , apresentam apenas os valores 0 e 1, demonstrando, respectivamente, se há ou não há conexão entre os nós n_i e n_j . Caso a rede tivesse pesos, a intensidade das conexões dos nós substituiria os valores 1 na matriz pelo respectivo peso. Além disso, por ser uma rede não direcionada, a matriz se torna simétrica, ou seja $a_{ij} = a_{ji}$. Portanto, caso ela fosse direcionada, haveria arestas que conectariam o nó n_i em n_j , mas não necessariamente o contrário, $a_{ij} \neq a_{ji}$. Também é interessante observar que, por essa rede não ter nós que se conectam a si mesmos, o que formariam as autoligações (*self-loops*), a diagonal principal é composta por 0s, ou seja, $a_{ii} = 0$. Já a lista de adjacência não apresenta informações sobre o peso da rede e caso ela fosse direcionada e/ou tivesse autoligações, apenas se acrescentaria o índice das conexões nas listas, sendo, portanto, uma representação mais concisa. (3)

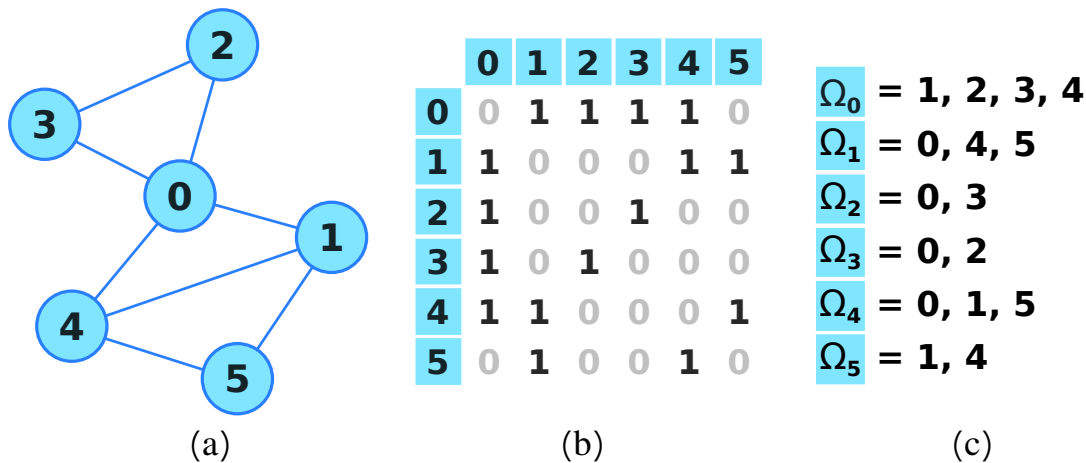


Figura 2 – Representações de rede complexas. (a) Exemplo de uma rede não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas. (b) Representação em forma de matriz de adjacência caracterizando o exemplo de rede. Cada linha e cada coluna (realçadas em azul) representam um nó na rede. Quando $a_{ij} = 1$, existe uma aresta entre n_i e n_j e não existe quando $a_{ij} = 0$. Por ser uma rede não direcionada e sem peso a matriz é simétrica e composta apenas por 0s e 1s. (c) Representação em forma de lista de adjacência caracterizando o exemplo de rede em que a lista Ω_i guarda a informação de todos os nós em que n_i está conectado.

2.2 Medidas de redes

Por se tratar de um trabalho que considera apenas redes não direcionadas e sem peso, todas as medidas topológicas necessárias para a compreensão deste trabalho serão definidas levando em consideração essas delimitações. Além disso, as redes também não apresentam autoligações e nem *múltiplas arestas* entre os nós, ou seja, entre qualquer par de nós conectados só existe uma conexão.

2.2.1 Grau

A partir da quantidade de vizinhos de um nó é possível obter a medida topológica de rede mais simples, o *grau* (também conhecida como *conectividade*) k_i de um dado nó n_i . (3) Utilizando a matriz de adjacência, é possível obter o grau somando os elementos da linha ou da coluna correspondente a esse nó, ou seja:

$$k_i = \sum_j^N a_{ij} = \sum_j^N a_{ji}. \quad (2.1)$$

Tendo a conectividade de cada nó, é possível obter uma medida mais geral da rede, o *grau médio da rede*, $\langle k \rangle$. Isso é feito somando o grau de todos os nós da rede e dividindo pelo seu número total de nós, N , definido então por:

$$\langle k \rangle = \frac{1}{N} \sum_i^N k_i. \quad (2.2)$$

Além disso, é possível obter a distribuição de grau da rede, que é um histograma das ocorrências dos graus nos nós da rede, sendo comumente a primeira análise exploratória que é feita em uma rede de interesse. (1) A Figura 3 ilustra a mesma rede da seção anterior, mas apresentando o grau de cada nó tanto na visualização da rede quanto em sua matriz de adjacência e também o histograma de sua distribuição de graus.

2.2.2 Caminho mínimo

Uma rede que não apresenta nenhum componente desconectado sempre apresentará no mínimo uma sequência de nós que criará um *caminho* entre quaisquer dois nós da rede. Entretanto, é possível haver mais de um caminho entre dois nós e o caminho que apresenta menos arestas é conhecido como *caminho mínimo* ou *caminho geodésico* em que l_{ij} é a *distância geodésica* entre os nós n_i e n_j . (3, 80) Dessa forma, cada nó tem uma distância geodésica para cada outro nó da rede e essa informação pode ser armazenada numa matriz $N \times N$, como na Figura 4. Esta matriz nada mais é do que um conjunto de vetores de caminhos mínimos de cada nó. Ou seja, para cada nó, tem-se uma linha de valores em que é possível extrair diversas outras medidas, *e.g.*, *média dos caminhos mínimos* e *desvio padrão dos caminhos mínimos*.

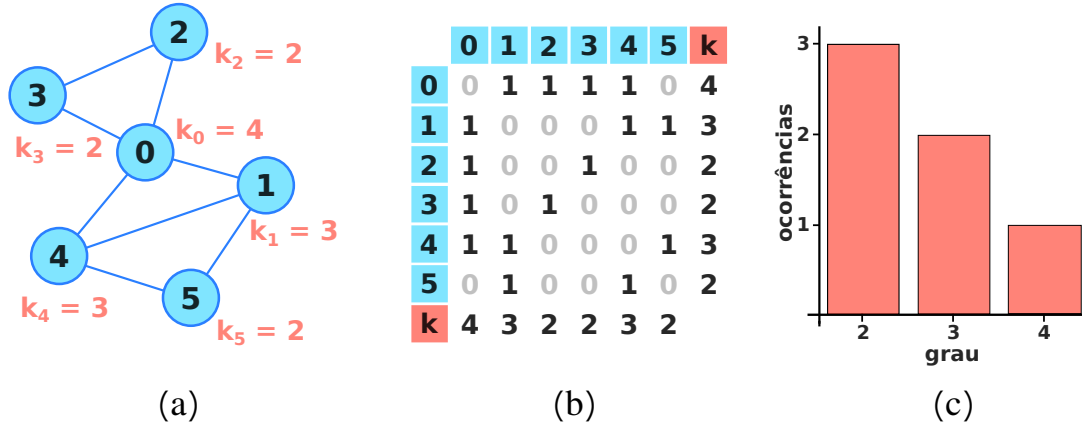


Figura 3 – Exemplo de uma rede complexa evidenciando a medida de grau. (a) Exemplo de uma rede não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas, com o valor do grau respectivo a cada nó. (b) Matriz de adjacência respectiva a rede complexa, porém com a expansão de uma linha e de uma coluna evidenciando, em vermelho, o grau de cada nó como o somatório das linhas e das colunas. (c) Distribuição dos graus da rede complexa, apresentando 3 nós com grau 2, 2 nós com grau 3 e 1 nó com grau 4. Assim, o grau médio da rede dá-se por $\langle k \rangle = 16/6 = 2,66$.

Fonte: Elaborada pelo autor.

Além disso, considerando todas as distâncias mínimas é possível extrair uma medida geral da rede, o *caminho mínimo médio*, $\langle l \rangle$. (3) Essa medida é calculada somando o comprimento de todos os caminhos mínimos entre todos os nós e dividindo pelo número máximo de arestas que a determinada rede poderia ter, sendo então definido por:

$$\langle l \rangle = \frac{1}{N(N-1)} \sum_{i \neq j}^N l_{ij}, \quad (2.3)$$

em que l_{ij} é a distância mínima entre qualquer nó n_i e n_j , e N é o número total de nós da rede.

2.2.3 Coeficiente de Aglomeração

A vizinhança de um nó também pode ser utilizado para informar sobre a conectividade local de uma rede complexa. Quando todos os vizinhos de um nó estão totalmente conectados entre si, tem-se um *clique*. (1, 3) Entretanto, esse é o caso extremo já que comumente uma região de uma rede não se apresenta completamente conectada. A partir disso, os pesquisadores Watts e Strogatz (81) definiram a medida *coeficiente de aglomeração local* (*local cluster coefficient*), c_i , de um nó n_i como sendo a taxa entre a real quantidade de conexões entre os vizinhos desse nó e o total de arestas possíveis se eles estivessem totalmente conectados. Sendo assim, tem-se:

$$c_i = \frac{N_{\Delta}(i)}{N_3(i)}, \quad (2.4)$$

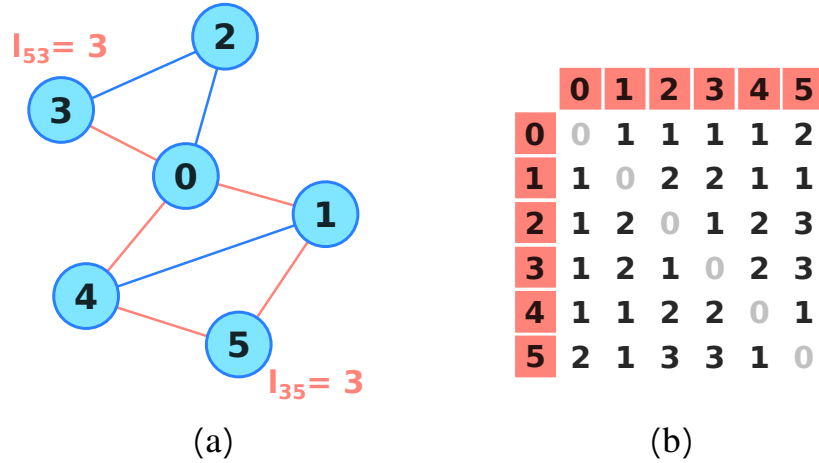


Figura 4 – Exemplo de uma rede complexa evidenciando a medida de caminho mínimo. **(a)** Exemplo de uma rede não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas. Em vermelho, estão os dois caminhos mínimos possíveis entre os nós n_3 e n_5 , ambos passando inicialmente por n_0 , porém cada caminho passando por dois nós diferentes, n_4 e n_1 , contabilizando $l_{35} = 3$. Também pode ser observado que por ser uma rede não direcionada, tem-se $l_{35} = l_{53}$, alterando apenas a ordem do caminho. **(b)** Matriz de distâncias mínimas respectiva a rede complexa em que cada elemento l_{ij} assume o valor do caminho mínimo entre os respectivos nós. Também é possível perceber que a matriz é simétrica, informando novamente que $l_{35} = l_{53}$, e que não há autoligação já que $l_{ii} = 0$. Assim, o caminho mínimo médio dá-se por $\langle l \rangle = 48/30 = 1,6$.

Fonte: Elaborada pelo autor.

em que $N_{\Delta}(i)$ é a quantidade de arestas entre os vizinhos de n_i , ou seja, o número de triângulos, (n_i, n_j, n_k) , em que o nó n_i participa como um vértice, e $N_3(i) = k_i(k_i - 1)/2$ é o número máximo de arestas possíveis de n_i , sendo k_i o grau de n_i . (1) Um exemplo dessa medida é ilustrada na Figura 5. A partir dos coeficiente de aglomeração de cada nó é possível definir uma medida geral para a rede complexa, a *média do coeficiente de aglomeração local* (81), em que todos os coeficientes são somados e divididos pelo número total de nós N , sendo então definida por:

$$\langle c \rangle = \frac{1}{N} \sum_{i=0}^N c_i$$

2.2.4 Betweenness Centrality

Pensando em termos de caminhos geodésicos de uma rede, é intuitivo pensar que se vários caminhos passarem por um determinado ponto, esse ponto poderá ser considerado de grande importância para o sistema. Analogamente, pode-se pensar em uma ilha conectada ao continente por apenas uma ponte, como o caso da cidade de Florianópolis, a capital do estado do Espírito Santo, Brasil. Todo caminho terrestre que liga qualquer ponto no

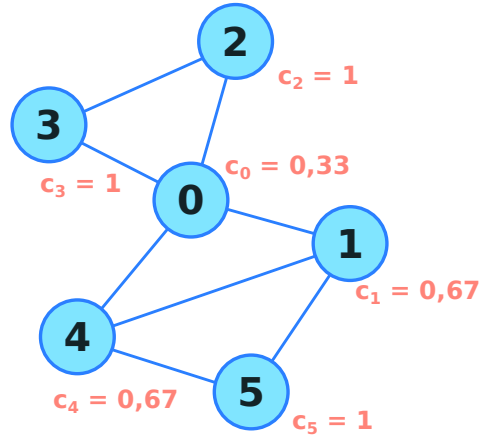


Figura 5 – Exemplo de uma rede complexa evidenciando a medida coeficiente de aglomeração sendo não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas. Em vermelho, estão os valores de coeficiente de aglomeração para cada um dos nós da rede. Como exemplo, considere c_0 : o nó n_0 é vértice de 2 triângulos, sendo (n_0, n_3, n_2) e (n_0, n_4, n_1) , e o máximo de conexões possíveis que esse nó poderia fazer é $N_3 = 4 \times (4 - 1) \times 2 = 6$, portanto, $c_0 = 2/6 = 0.33$. Levando em consideração a rede toda, é possível calcular a sua média do coeficiente de aglomeração local por $\langle c \rangle = 4,67/6 = 0,78$.

Fonte: Elaborada pelo autor.

continente a qualquer ponto da ilha, passará por essa ponte, logo, essa ponte é um ponto extremamente importante para um sistema de transporte terrestre continente-ilha, tanto que a ruptura dessa aresta deixará a ilha como um componente descontado. Sendo assim, é evidente a relevância de se mensurar essa importância. Uma das formas de se fazer isso é encontrando a razão entre a quantidade de caminhos mínimos entre dois nós n_i e n_j que passam por um ponto u e a quantidade de caminhos mínimos totais entre esses dois nós. A essa medida dá-se o nome de *betweenness centrality* (82) e é definida como:

$$B_u = \sum_{ij, i \neq j \neq u} \frac{\sigma(n_i, u, n_j)}{\sigma(n_i, n_j) \times 2}, \quad (2.5)$$

em que $\sigma(n_i, u, n_j)$ representa a quantidade de caminhos mínimos existentes entre os nós n_i e n_j que passam pelo nó ou aresta u e $\sigma(n_i, n_j)$ é a quantidade total de caminhos geodésicos entre estes dois nós considerando também outros caminhos além dos que passam por u . Vale ressaltar que existem algumas variações dessa definição na literatura, porém, Newman (3) argumenta que essas variações tipicamente não são importantes pois usualmente o interesse está na magnitude relativa da centralidade e não em seus valores absolutos. Portanto, a definição aqui apresentada é a que foi utilizada neste trabalho. Sendo assim, tem-se que a divisão por 2 é para compensar a direcionalidade na contagem de caminhos mínimos, não há autoligação ($n_i \neq n_j$) e quando u é um nó, ele não é considerado nem no início nem fim do caminho ($i \neq j \neq u$), porém quando u é uma aresta, e_{ij} , ele é contado como um caminho. (83–85)

Além disso, medida B_u é comumente utilizada de forma normalizada pelo número total de conexões possíveis na rede, entretanto, para o caso do *betweenness centrality* para nó, não é contado o nó n_i na normalização. Desta forma, a normalização da *betweenness centrality* para arestas é feita dividindo a medida por $N \times (N - 1)/2$, e na normalização da *betweenness centrality* para o nó divide-se a medida por $(N - 1) \times (N - 2)/2$. (83–85) A Figura 6 apresenta exemplos para os dois casos.

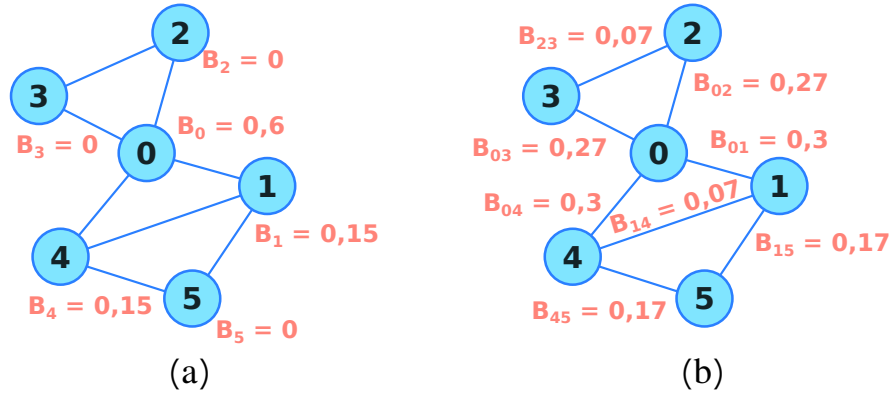


Figura 6 – Exemplo de uma rede complexa não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas sendo evidenciado em vermelho a medida *betweenness centrality*. (a) Em vermelho, estão os valores normalizados de *betweenness centrality* para os nós. Como exemplo, considere B_1 : pelo nó n_1 se passa 3 caminhos mínimos, l_{05} , l_{25} e l_{35} , entretanto, para cada par de nós destes caminhos, existe mais um outro caminho mínimo passando por n_4 , portanto, tem-se $B_1 = 1/2 + 1/2 + 1/2 = 1,5$, com a normalização resulta-se $B_1 = 1,5 / (5 \times 4/2) = 1,5/10 = 0,15$. (b) Em vermelho, estão os valores normalizados de *betweenness centrality* para as arestas. Para exemplificar, observe B_{15} : além do próprio caminho l_{15} , há mais 3 pares de nós em que seus caminhos mínimos passam pela aresta e_{15} porém também passam pela aresta e_{45} , portanto, tem-se $B_{15} = 1 + 1/2 + 1/2 + 1/2 = 2,5$, aplicando a normalização encontra-se $B_{15} = 2,5 / (6 \times 5/2) = 0,17$.

Fonte: Elaborada pelo autor.

2.2.5 Matching Index

Nas topologias de redes complexas, são encontradas algumas estruturas específicas, alguns tipos de padrões, que podem ser utilizados para obter informações da rede. Por exemplo, o coeficiente de aglomeração leva em consideração os triângulos que são formados entre os nós para mensurar agrupamentos. Além de triângulos, existem outros tipos de estruturas e, conseqüentemente, outros índices que os avaliam. (1, 3)

Surgindo inicialmente no contexto de redes cerebrais (86, 87), a medida *matching index* quantifica a semelhança estrutural entre duas regiões de uma rede. Basicamente, essa medida calcula para dois nós ligados por uma aresta a similaridade topológica entre os subgrafos formados por cada um dos nós. Isso é feito através da razão entre o número de

nós vizinhos compartilhados por ambos os nós considerados e o número total de conexões desses dois nós. Sendo assim, a partir da matriz de adjacência, define-se como:

$$\mu_{ij} = \frac{\sum_{k \neq i,j} a_{ik} a_{jk}}{\sum_{k \neq j} a_{ik} + \sum_{k \neq i} a_{jk}}, \quad (2.6)$$

sendo a_{ik} o elemento da matriz de adjacência em que $a_{ik} = 1$ significa que os nós n_i e n_k estão conectados, $a_{ik} = 0$ quando estão desconectados e de forma equivalente para a_{jk} , sendo n_j e n_k vizinhos de n_i . De forma mais intuitiva, é possível reescrever a Equação 2.6 como:

$$\mu_{ij} = \frac{\epsilon(i, j)}{k_i + k_j - 2}, \quad (2.7)$$

em que $\epsilon(i, j)$ representa a quantidade de vizinhos mútuos dos nós n_i e n_j , sendo k_i e k_j seus respectivos graus. Subtrai-se 2 para desconsiderar as arestas e_{ij} e e_{ji} na contagem. A Figura 7 ilustra esta medida.

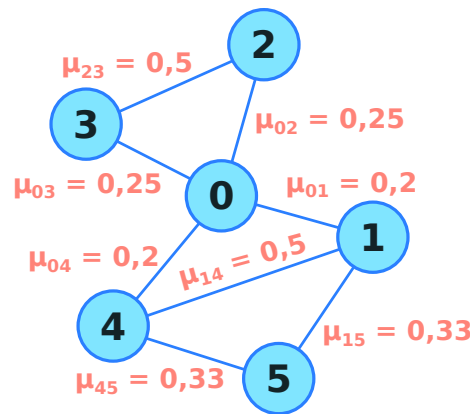


Figura 7 – Exemplo de uma rede complexa não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas sendo evidenciado em vermelho a medida *matching index*. Como exemplo, considere a aresta e_{14} que conecta os nós n_1 e n_4 que por sua vez compartilham 2 vizinhos, os nós n_0 e n_5 , e seus graus são $k_1 = k_4 = 3$, portanto, tem-se $\mu_{14} = 2/(3 + 3 - 2) = 0,5$.

Fonte: Elaborada pelo autor.

2.2.6 Autovalores Laplacianos

Como uma das bases da ciência de redes, mutualmente a teoria de grafos, temos a mecânica estatística. (7) Por diversas vezes uma área contribuiu com a outra na resolução de problemas. Este é o caso da dinâmica de *difusão* que descreve o processo do movimento de um gás indo de uma região de maior densidade para uma de menor através de vários caminhos. Analogamente, uma dinâmica semelhante também foi observada em grafos, porém, enquanto na difusão dos gases existe o operador Laplaciano, ∇^2 , em redes, se tem uma matriz, que justamente por ser análoga ao operador Laplaciano, recebe o nome de

matriz Laplaciana (3, 88–90) e é definida por:

$$L = D - A, \quad (2.8)$$

em que A é a matriz de adjacência e D é matriz diagonal em que seus elementos são os graus dos respectivos nós, como ilustrado na Figura 3 (b) da Seção 2.2.1. Dentre as diversas medidas que podem ser extraídas da matriz Laplaciana, tem-se os *autovetores* e *autovalores Laplacianos* encontrados resolvendo a equação:

$$(L - \lambda I)v = 0, \quad (2.9)$$

em que L é a matriz Laplaciana, I a matriz identidade de mesma dimensão de L , λ e v são, respectivamente, os autovalores e os autovetores de L .

2.2.7 Graus Hierárquicos

Uma rede complexa é composta de elementos unitários, os nós, e suas conexões, as arestas, porém nada impede que seja levado em consideração toda a vizinhança de um nó como um elemento unitário. Ou seja, é possível considerar o nó n_i como um elemento de hierarquia zero, sua vizinhança, como o seu primeiro nível hierárquico, e os primeiros vizinhos desta vizinhança como o seu segundo nível hierárquico e assim sucessivamente. Dessa forma, as medidas definidas para apenas um nó, podem ser estendidas para sua vizinhança, definindo medidas hierárquicas. (1, 91)

A medida mais direta a ser considerada é o *grau hierárquico*. O grau do nó n_i é a quantidade de seus nós vizinhos e, portanto, é considerado o primeiro grau hierárquico. Para calcular o segundo grau hierárquico, basta considerar toda a vizinhança do nó n_i , o primeiro nível hierárquico, como se fosse um nó e contar os nós vizinhos a essa vizinhança, ou seja, contar os vizinhos dos vizinhos desconsiderando os nós do mesmo nível e de níveis anteriores. O mesmo pode ser feito para os graus de níveis maiores, dependendo apenas do interesse da pesquisa e limitado pelo tamanho da rede. Sendo assim, é possível definir o grau do nó n_i na hierarquia h por $k_i(h)$. A Figura 8 ilustra os graus hierárquicos considerando círculos concêntricos em volta do nó delimitando dois níveis hierárquicos.

2.3 Modelos de redes

Esta seção apresenta os modelos teóricos de redes complexas utilizadas ao longo do trabalho. Vale ressaltar que as definições apresentadas dizem respeito a redes não direcionadas, sem peso, sem autoconexão e sem múltiplas ligações.

2.3.1 Erdős-Rényi

Um importante marco para a teoria de grafos foi a incorporação da aleatoriedade em seus modelos e análises. A esses modelos deu-se o nome de *grafos aleatórios* que, de uma

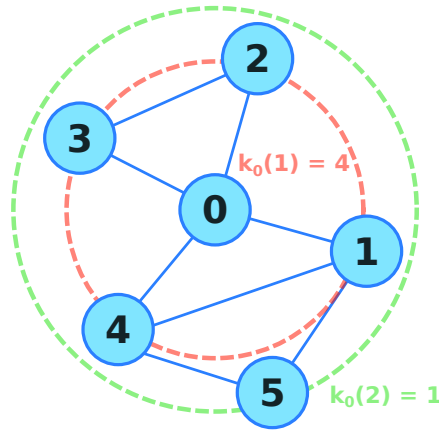


Figura 8 – Exemplo de uma rede complexa não direcionada e sem peso com $N = 6$ nós e $E = 8$ arestas evidenciando a medida grau hierárquico. Como exemplo, observá-se os níveis hierárquicos para o nó n_0 , definido pelos círculos concêntricos vermelho para o primeiro nível e verde para o segundo nível, e seus respectivos graus hierárquicos $k_0(1) = 4$ e $k_0(2) = 1$.

Fonte: Elaborada pelo autor.

forma geral, são grafos com alguns parâmetros de valores fixos, porém também apresentam características aleatórias. (3) Dentre eles, o modelo de grafo aleatório mais estudado é o Erdős-Rényi (ER). (47)

No modelo ER, não se determina valores fixos para os graus dos nós, como no caso de *grades regulares*, mas sim uma *probabilidade de conexão*, p . Desta forma, cada nó tem uma probabilidade, determinada por uma distribuição estatística, de estar conectado a outro nó. Sendo assim, para que uma rede ER seja construída, inicia-se o grafo com N nós desconectados e para cada par de nós, (n_i, n_j) , uma aresta e_{ij} é criada com a probabilidade p até que todos pares tenham sido considerados.

Apesar da aleatoriedade, esse modelo pode ser considerado como estatisticamente regular pois todos os nós da rede têm igual probabilidade de se conectar uns aos outros. Isto resulta em uma distribuição de grau que tende a uma curva de Poisson (1) conforme o número de nós aumenta, como ilustrado na Figura 9. A regularidade estatística na sua conectividade, é o primeiro fator que faz com que o modelo ER seja considerado um modelo simples. O segundo fator é por ser possível caracterizá-lo por apenas dois parâmetros que são o número de nós N e a probabilidade uniforme de conexão p e, com isso, é possível descrever o seu grau médio, $\langle k \rangle$, simplesmente pela equação:

$$\langle k \rangle = p(N - 1). \quad (2.10)$$

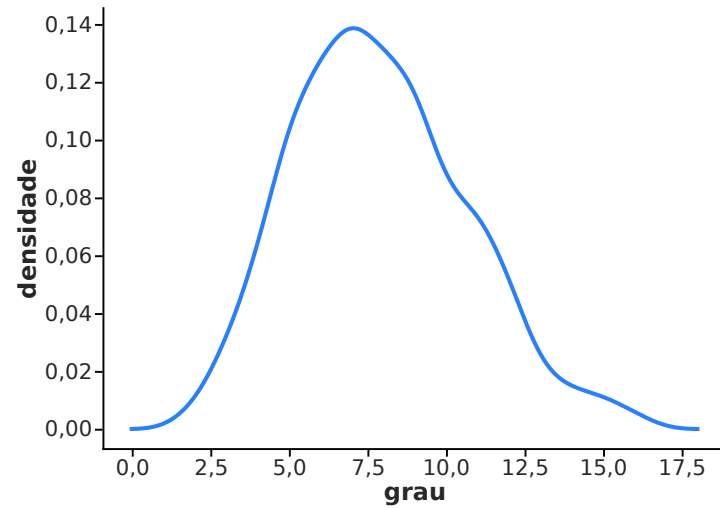


Figura 9 – Exemplo da distribuição de grau de uma rede do modelo Erdős-Rényi gerada com $N = 500$ nós, apresentando $\langle k \rangle = 7,88$.

Fonte: Elaborada pelo autor.

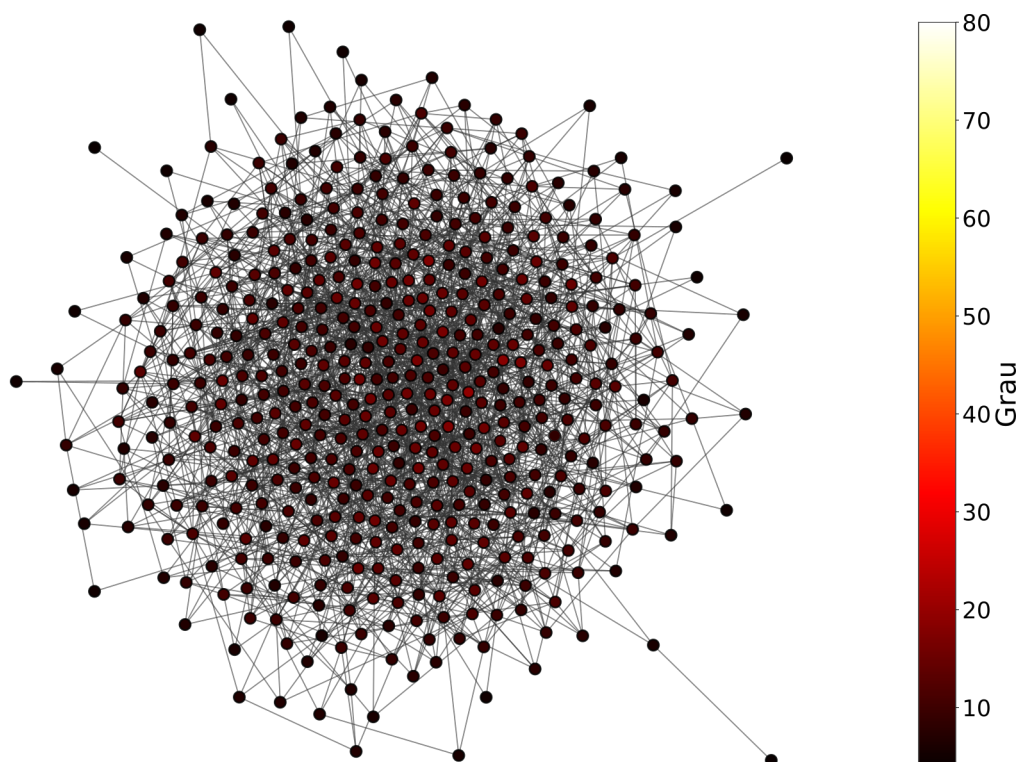


Figura 10 – Exemplo de uma rede do modelo Erdős-Rényi gerada com $N = 500$ nós e p calculado para que $\langle k \rangle = 8$. Valor real de $\langle k \rangle$ para essa rede é 7,88. A barra de cor representa o grau dos nós e com ela é possível constatar que os valores dos graus são aproximadamente iguais a média, demonstrando uma conectividade estatisticamente regular.

Fonte: Elaborada pelo autor.

2.3.2 Watts-Strogatz

A incorporação da aleatoriedade nas conexões de grafos foi um passo importante para que a teoria de grafos fornecesse ferramentas para o estudo das redes reais. (5) Entretanto, conforme os dados das redes reais foram sendo analisados, os pesquisadores perceberam que os grafos aleatórios não eram suficientes para explicar dois fenômenos encontrados nessas redes: o efeito *mundo pequeno* (*small world effect*) e um alto índice aglomeração local. (92)

O efeito mundo pequeno é caracterizado pela existência de distâncias geodésicas l curtas, acarretando em um baixo valor do caminho mínimo médio da rede, $\langle l \rangle$. Isso significa que, em geral, um nó qualquer de uma rede com esse efeito pode ser acessado por qualquer outro nó através de uma pequena quantidade de nós intermediários. Esse conceito popularizou-se no experimento de Stanley Milgram que descobriu que dois cidadãos americanos escolhidos aleatoriamente estão separados, em média, por menos de seis pessoas conhecidas intermediárias.

Além do efeito mundo pequeno, redes reais apresentam uma quantidade muito maior de caminhos fechados de ordem três do que em redes aleatórias com mesma quantidade de nós e arestas. Ou seja, essas redes apresentam muitos casos de nós conectados tendo vizinhos em comum, formando triângulos, o que resulta em um alto valor no índice de aglomeração local médio $\langle c \rangle$, o que não é observado nas redes aleatórias estudadas até então. (81)

Com o conhecimento destas questões, os pesquisadores Duncan J. Watts e Steven Strogatz desenvolveram um modelo, conhecido como *modelo Watts-Strogatz* (*modelo WS*), que incorpora essas características. (81) O modelo WS inicia-se como uma rede regular de N nós com κ ligações iniciais, comumente em formato de anel. Para incorporar a aleatoriedade no modelo, cada par de nós, (n_i, n_j) , tem uma probabilidade fixa p de ter sua aresta reconectada. Essa reconexão é feita trocando o nó n_j por um nó n_k qualquer que já não seja vizinho do nó n_i e que $n_j \neq n_k$, dessa forma, o grau médio da rede não sofre alterações.

Variando o valor de p de 0 a 1, o modelo WS se altera de uma rede completamente regular, com valores de $\langle c \rangle$ e $\langle l \rangle$ altos, para uma rede totalmente aleatória (porém diferente de uma ER) com valores de $\langle c \rangle$ e $\langle l \rangle$ pequenos. (92) Entretanto, o interessante desse modelo se encontra entre esses extremos. Com baixos valores de p , o $\langle l \rangle$ diminui drasticamente e em $p = 0,01$ acontece a emergência do regime pequeno mundo, ou seja, a rede WS passa a ter valores de $\langle l \rangle$ pequenos, semelhantes a grafos aleatórios, e valores de $\langle c \rangle$ grandes, se assemelhando as redes regulares. (81,92) A Figura 11 ilustra duas redes do modelo WS exemplificando dois casos com diferentes probabilidades de reconexão, quando surge o efeito de pequeno mundo e quando se torna completamente aleatória.

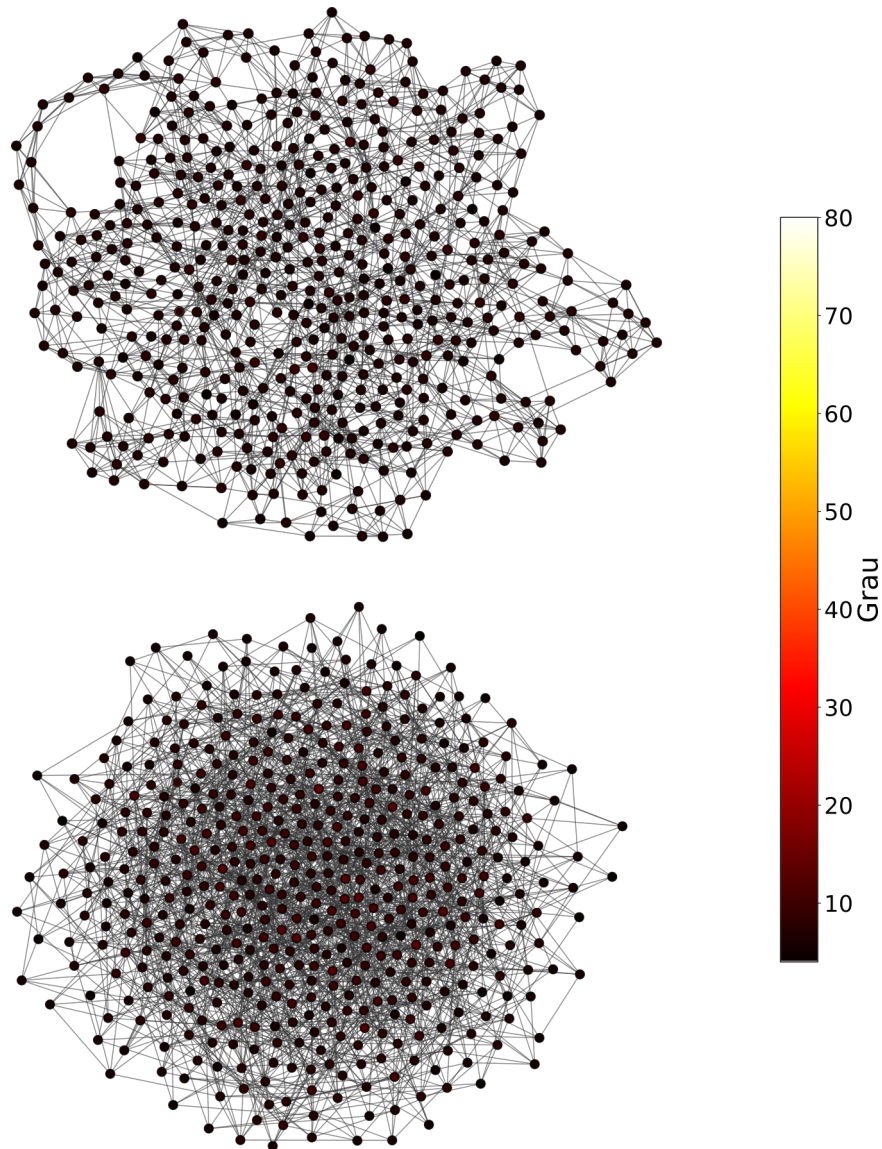


Figura 11 – Exemplo de duas redes do modelo Watts-Strogatz gerada com $N = 500$ nós $\langle k \rangle = 8$. A rede superior foi gerada com o valor crítico de $p = 0,1$, ou seja, 10% de reconexão, sendo então um exemplo de rede com o efeito pequeno mundo. A rede inferior foi gerada com $p = 1$, ou seja, 100% de reconexão, sendo o caso extremo tornando-se uma rede completamente aleatória. A barra de cor representa o grau dos nós e com ela é possível constatar que em ambos os casos não há muita variação no valor dos graus. Porém, enquanto a rede superior apresenta praticamente todos os nós com o mesmo valor de grau abaixo de 10, a inferior contém diversos nós com grau um pouco acima de 10, tendo uma distribuição de grau levemente mais diversa que no primeiro caso.

Fonte: Elaborada pelo autor.

2.3.3 Barabási-Albert

Ao observar as distribuições de graus de algumas redes reais, percebeu-se que elas eram semelhantes as encontradas em medidas de diversos outros eventos como frequência

de palavras (93), magnitude de terremotos (94), internet (95) e vários outros. (7, 48, 96) A distribuição estatística teórica comumente utilizada para caracterizar essas distribuições é a *lei de potência* (*power law*) (97), também conhecida por *Zipf's law* ou *Pareto distribution*, apesar de terem pequenas diferenças. (98) Dessa forma, observou-se que a distribuição de graus de diversas redes reais divergiam dos modelos teóricos estudados. Além disso, o interesse em estudar essa novidade cresceu principalmente por essas redes não apresentarem a regularidade estatística conhecida até então. Portanto, para representar essa diversidade na conectividade das redes reais, caracteriza-se a distribuição de graus pela lei de potência como uma função $P(k)$ dependente de k regida pelo decrescimento do expoente γ , ou seja:

$$P(k) = k^{-\gamma}. \quad (2.11)$$

A lei de potência tem a característica de apresentar um crescimento ou decrescimento exponencial nos seus valores que, no caso dos graus das redes complexas, isto se reflete criando uma curva de distribuição de graus com um pico à esquerda e com uma cauda longa no eixo dos graus. Isto é resultado de duas características marcantes de diversas redes reais: a presença de *hubs*, *i.g.*, nós com graus extremamente maiores do que a maioria dos nós da rede; e a presença de muitos nós com conectividade muito baixa, como pode ser visto na Figura 12. Por não ter um valor típico em que as medidas possam ser associadas, como o valor médio na Poisson no caso do modelo ER, as redes com distribuição lei de potência são chamadas de *livre de escala*. (98)

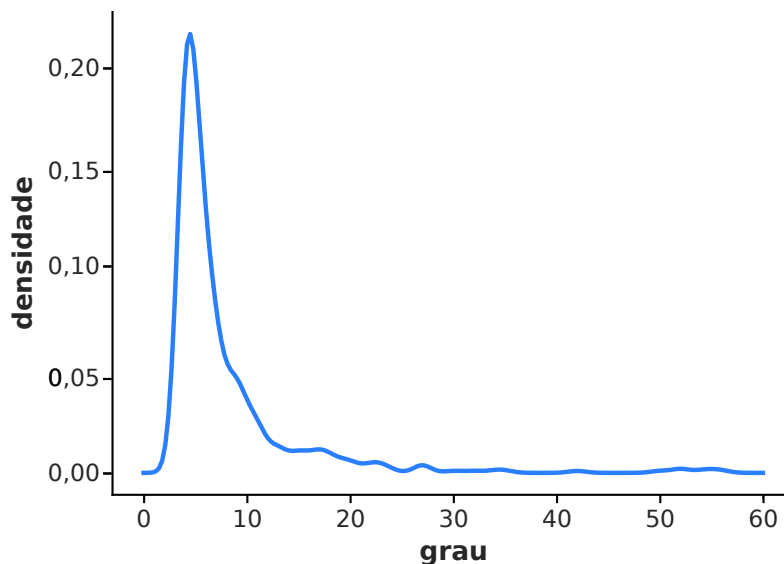


Figura 12 – Exemplo da distribuição de grau de uma rede do modelo Barabási-Albert gerada com $N = 500$ nós, apresentando $\langle k \rangle = 7,96$.

Fonte: Elaborada pelo autor.

Para explicar a distribuição livre de escalas de diversas redes reais, alguns modelos foram propostos considerando mais do que simplesmente o número de nós N e a probabilidade de conexão p . Como o primeiro modelo conhecido, o pesquisador Price, através do estudo de redes de citação de artigos, apresentou um modelo livre de escala para redes direcionadas, o *modelo Price*, propondo uma regra na formação das redes, a *vantagem cumulativa*. (99) Posteriormente, Barabási e Albert apresentaram um modelo similar para redes não direcionadas, o *modelo Barabási-Albert (BA)*, e renomeou a regra para *conexão preferencial*, sendo este o nome mais usado na literatura. (48) Essa regra altera a probabilidade que cada nó receberá uma nova conexão. Ao invés de todos os nós terem a mesma probabilidade, agora os mais conectados tem uma probabilidade maior de receber novos vizinhos do que nós com poucas conexões.

Além da conexão preferencial, Barabási e Albert acrescentaram mais uma característica para obter a distribuição de lei de potência, o *crescimento da rede*, ou seja, o acréscimo de novos nós com o passar do tempo obedecendo a conexão preferencial. (48) Sendo assim, o modelo BA inicia-se como uma rede tendo um pequeno número de nós, $\tau_{t=0} = \tau_0$, conectados e sendo acrescido de um novo nó a cada intervalo de tempo até atingir N nós desejados. Cada novo nó se conectará a τ nós que já estão na rede, não se conectando mais de uma vez ao mesmo nó. Devido a essa restrição, quando o número de nós na rede naquele momento, τ_t , for menor que τ , o novo nó se conectará apenas a todos os nós antigos, que é o que ocorre no início do processo. Quando $\tau_t > \tau$ a probabilidade de um nó antigo receber uma nova conexão, $p(k_i)$, é a razão entre ao seu grau e somatório do grau de todos os nós, ou seja:

$$p(k_i) = \frac{k_i}{\sum_j k_j}, \quad (2.12)$$

sendo k_i e k_j , respectivamente, o grau de um nó n_i qualquer e o grau de todos os outros nós n_j da rede. Isso faz com que nós muito conectados possuam uma probabilidade crescente de receber novos vizinhos enquanto que nós pouco conectados tenham cada vez menos chance de receber novas conexões, um efeito também conhecido como *rico fica mais rico*.

Em contrapartida a simplicidade atribuída ao modelo ER, o modelo BA, ilustrado na Figura 13, é tido como um modelos com alta complexidade. Essa atribuição é feita principalmente pela sua diversidade na distribuição de grau, apresentando uma topologia mais heterogenia do que redes ER.

Vale lembrar que algumas discussões pertinentes a essa seção como as distribuições que melhor explicam as redes reais; detalhes sobre o modelo BA e da lei de potência; e quais trabalhos abordaram primeiro esses temas, foram abordadas aqui de forma superficial. Elas são longas, surgindo por volta do início do século passado, e extensa, abrangendo diversas áreas como economia, biologia, ecologia, astronomia e diversas outras. (3, 8, 97, 98, 100, 101)

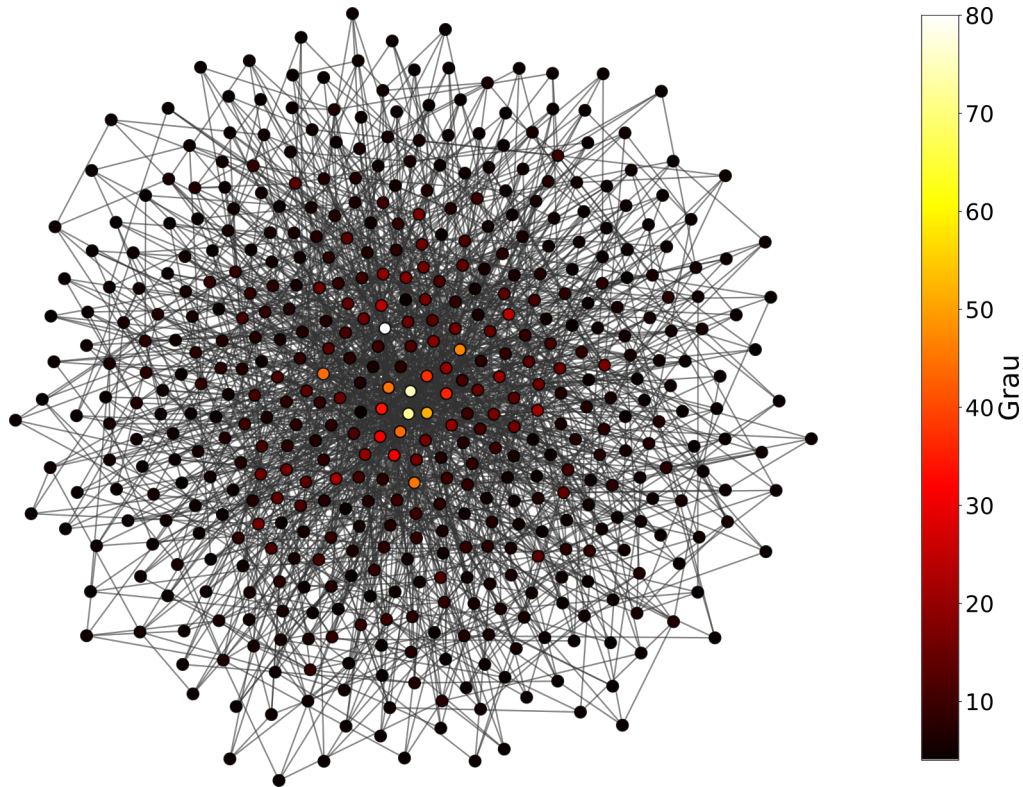


Figura 13 – Exemplo de uma rede do modelo Barabási-Albert gerada com $N = 500$ nós e p calculado para que $\langle k \rangle = 8$. Valor real de $\langle k \rangle$ para essa rede é 7,96. A barra de cor representa o grau dos nós. Observa-se a existência de apenas 3 *hubs* com conectividade acima de 60, enquanto a imensa maioria dos nós apresenta grau abaixo da média e apenas alguns poucos nós estão entre estes dois extremos.

Fonte: Elaborada pelo autor.

2.3.4 Waxman

Diferentemente dos modelos anteriores, o *modelo Waxman* (WAX) não foi apresentado com a intenção de estudar características de redes reais. (102) Na verdade, ele foi projetado pensando em sistemas de rotas de comunicação de multipontos para testar soluções do problema de árvore de Steiner (102), um problema de geometria estudado por Fermat (103) e posteriormente por Steiner (104, 105) e aplicado a grafos por Karp. (106) Uma característica importante desse modelo é que ele pertence a classe de redes *geográficas*, ou *espaciais*, *i.e.*, grafos em que seus nós tem posições definidas no espaço, porém com conexão aleatória. (107)

Existem duas versões do modelo WAX que se diferenciam apenas na forma em que se calcula a probabilidade de conexão dos nós. (102) Ambas versões iniciam-se com N nós distribuídos de forma aleatoriamente uniforme em uma área retangular e as arestas conectam dois nós (n_i, n_j) com a probabilidade p_{ij} proporcional a distância Euclidiana, $d_E(n_i, n_j)$, entre os dois nós. Vale ressaltar que essa distância é relativa as posições dos nós

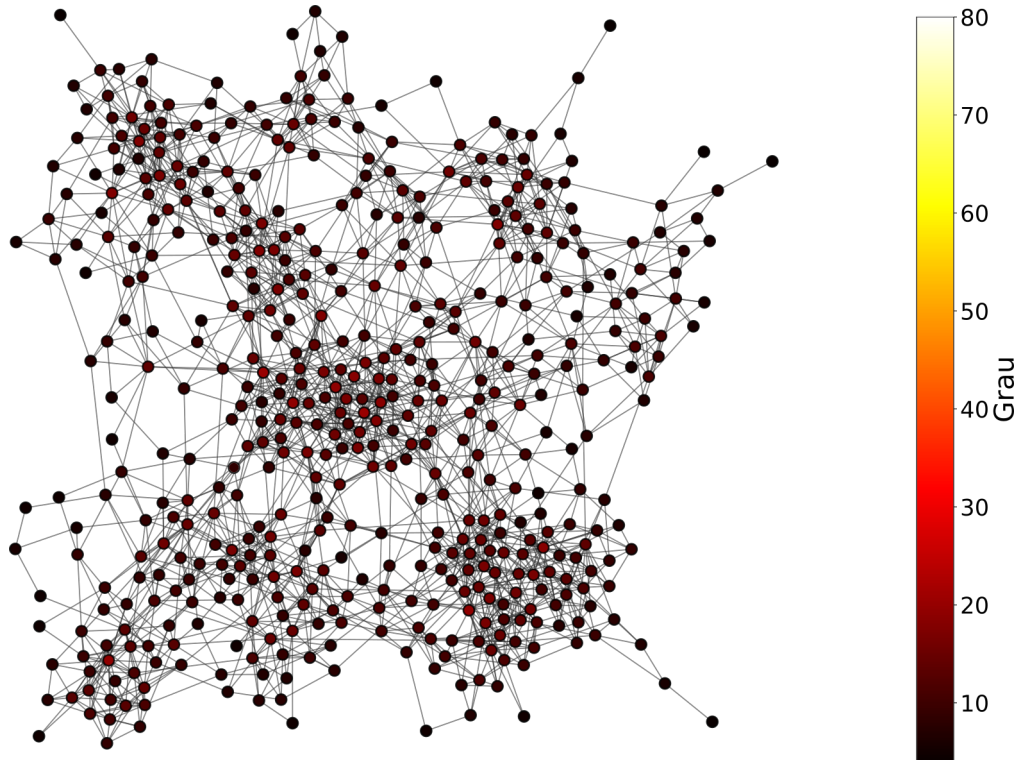


Figura 14 – Exemplo de uma rede da segunda versão do modelo Waxman gerada com $N = 500$ nós e parâmetros calculados para que a rede tenha $\langle k \rangle$ próximo de 8, $\beta = 1$, $\alpha = 0,04$. Valor real de $\langle k \rangle$ para essa rede é 8,08. A barra de cor representa o grau dos nós e com ela é possível observar que a conectividade varia entre 1 e aproximadamente 25, porém com a maioria dos nós apresentando grau próximo da média.

Fonte: Elaborada pelo autor.

no espaço e não deve ser confundida com distância de nós em grafos que diz respeito ao número de nós em um caminho entre dois nós. Dessa forma, a probabilidade de conexão é definida como:

$$p_{ij} = \beta \exp\left(\frac{-d_E(v_i, v_j)}{\alpha L}\right), \quad (2.13)$$

em que α e β são parâmetros de controle da probabilidade de conexão e assumem valores dentro de $(0, 1]$. Valores altos de β farão o grafo ter uma alta densidade de conexões e para valores pequenos de α , o grafo terá uma densidade maior de conexões espacialmente próximas do que longas. (102) L é uma distância entre nós da rede (número de nós intermediários) que assume duas possibilidades, uma para cada versão do modelo WAX. Na primeira, L é a maior distância entre dois nós na rede, também conhecido como *diâmetro* da rede. (3) Porém, na segunda, escolhe-se um valor, L_{max} , para que L seja sorteado em $(0, L_{max}]$ de forma uniformemente aleatória. Assim, na primeira versão, nós com a mesma distância terão probabilidade de conexão iguais, já na segunda, pares de nós que tenham a mesma distância, provavelmente terão p diferentes e serão limitados

pelo valor L sorteado. (102) Um exemplo de rede do modelo WAX, utilizando a segunda versão, é ilustrado na Figura 14.

2.3.5 *Random Geometric Graph*

Outro modelo de rede com topologia espacial (107) é a *rede geométrica aleatória*, do inglês *random geometric graph (RGG)*. (96, 108) Semelhantemente ao modelo WAX, o modelo RGG inicia-se com N nós distribuídos de forma uniformemente aleatória em um espaço pre-definido. Porém, suas conexões não são aleatórias: todos os nós que estiverem a uma distância Euclidiana menor que uma distância R de um determinado nó, n_i , serão conectados. Ou seja, para cada um dos nós da rede, tem-se uma área circular de raio R centrada neles e todos os demais nós dentro desta área se conectarão a ele. (109) Um exemplo de uma rede do modelo RGG é ilustrado na Figura 15.

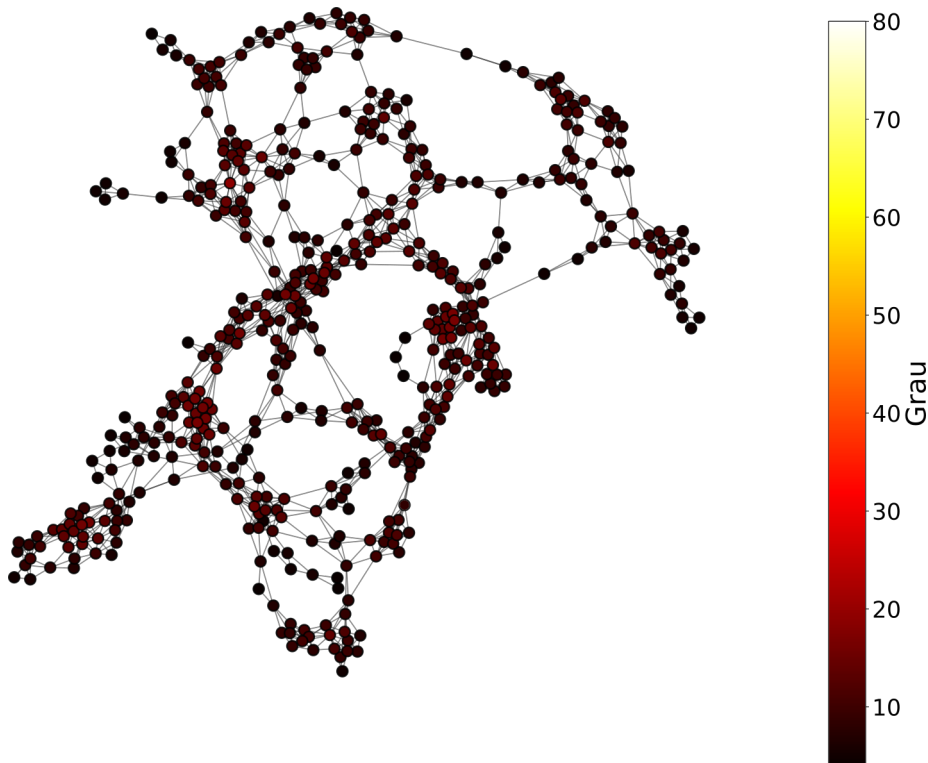


Figura 15 – Exemplo de uma rede do modelo *Random Geometric Graph* gerada com $N = 500$ nós e $R = 0,0745$ calculado para que a rede tenha $\langle k \rangle$ próximo de 8. Valor real de $\langle k \rangle$ para essa rede é 7,92. A barra de cor representa o grau dos nós. É possível observar vários pequenos agrupamentos em volta de nós com grau um pouco acima da média, porém a maioria dos nós tem a conectividade aproximadamente ou menor que a média.

Fonte: Elaborada pelo autor.

3 COMPLEXIDADE

Este capítulo discorre sobre o estudo de complexidade em redes complexas desenvolvido neste trabalho, apresentando as definições e métodos propostos. Primeiro, na Seção 3.1, apresenta-se a necessidade da formalização de uma medida de complexidade de redes complexas baseada na heterogeneidade de medidas topológicas. Com isso, desenvolve-se o Índice de Complexidade que será utilizado nos experimentos do próximo capítulo. Em seguida, na Seção 3.2, é apresentada a possibilidade de se desenvolver redes com alta complexidade topológica. A partir desta motivação, é apresentada a definição do método de geração de redes hipercomplexas bem como suas limitações e possíveis variações.

3.1 Índice de Complexidade

Como previamente descrito no Capítulo 1, apenas a heterogeneidade da distribuição de grau não é suficiente para expressar a complexidade de uma rede complexa. Dessa forma, para que a representação de complexidade seja mais abrangente, é necessário levar em consideração diversas medidas que descrevam múltiplas características topológicas da rede. Além disso, com a intenção de mensurar a complexidade de uma dada rede, para que seja possível realizar comparações e discussões quantitativas, é importante que seja desenvolvido um índice de complexidade. Portanto, partimos do princípio de que quanto maior for a diversidade nos valores de múltiplas medidas topológicas, ou seja, quanto mais heterogenia as medidas forem, mais complexa a rede será considerada.

Intencionando mensurar a complexidade de uma rede com base na heterogeneidade de um grupo de medidas topológicas, definimos o *Índice de Complexidade*, IC, com base no desvio padrão de cada uma das medidas, σ_m . Isso é feito devido ao desvio padrão ser uma medida estatística capaz de expressar a dispersão de um determinado conjunto de valores. Sendo assim, consideramos o desvio padrão como a representação da heterogeneidade de cada medida, uma quantidade que exprime a diversidade das propriedades topológicas.

É importante ressaltar que as medidas topológicas a serem consideradas devem apresentar um conjunto de valores que demonstrem a sua diversidade e não apenas um único valor para a rede toda. Portanto, deve ser considerado medidas que sejam relativas a cada nó e a cada aresta da rede. Para exemplificar, comparemos as medidas distribuição de graus e grau médio. A primeira apresenta um conjunto de valores, um para cada nó da rede, ou seja N valores, permitindo a caracterização da diversidade de valores de grau de uma determinada rede. Já a segunda medida é apenas um número que não informa sobre a diversidade de graus da rede, comunicando apenas um valor geral e, portanto, não pode ser considerada para o cálculo da complexidade.

Neste trabalho em específico, as dez medidas topológicas utilizadas para o cálculo da complexidade foram: (i) os graus dos nós; (ii) média e (iii) desvio padrão dos caminhos mínimos de cada nó; (iv) coeficiente de aglomeração para cada nó; *betweenness centrality* para (v) nós e (vi) arestas; (vii) *matching index* para cada par de nó que compartilham uma aresta; (viii) autovalores Laplacianos; e (ix) segundo e (x) terceiro grau hierárquico para cada nó.

No caso deste trabalho, ao mensurar a complexidade de uma dada rede, se faz importante ressaltar que esse conceito é relativo a outras redes. Ou seja, pode-se apenas dizer que uma rede é mais complexa que outra, porém continua sendo um problema expressar a complexidade de uma rede de forma isolada. Isso significa que, ao calcular o IC, deve-se considerar um conjunto de redes a serem comparadas. Além disso, a escolha das redes que servirão de referência não é fixa, necessitando que elas sejam determinadas para cada estudo em específico. Sendo assim, neste presente trabalho, para o cálculo do IC, consideramos como redes de referência os seguintes modelos: Erdős-Rényi, Barabási-Albert, segunda versão do Waxman, *Random Geometric Graph* e Watts-Strogatz com duas configurações de reconexão.

Vale lembrar que a dispersão de diferentes medidas topológicas podem apresentar escalas de valores bastante distintas, dificultando a adequação em apenas um índice. Por exemplo, em uma rede sem componentes isolados, os valores que o grau de um nó pode assumir são maiores do que 1, porém os valores do coeficiente de aglomeração de cada nó está compreendido entre 0 e 1. Dessa forma, o desvio padrão dessas duas medidas apresentarão escalas distintas, impedindo a comparação de suas heterogeneidades. Para resolver esse problema, é necessário pré-processar as medidas, normalizando-as para que os seus respectivos σ_m tenham escalas similares. Especificamente, aplicamos a standardização com a seguinte equação:

$$z_m = \frac{m - \mu_m}{\sigma_m}; \quad (3.1)$$

em que μ_m e σ_m são, respectivamente, a média e o desvio padrão de cada medida m . Após a standardização, todas as medidas terão média igual 0 e desvio padrão igual a 1, podendo ser incorporada em um único índice. Também é importante ressaltar que a standardização é aplicada considerando não apenas a rede em que temos interesse em calcular a complexidade, mas sim todo o conjunto de redes de referência. No caso desse trabalho, são utilizados 100 exemplares de cada um dos modelos citados e todas as redes apresentam o mesmo número de nós e mesmo grau médio.

Tendo o desvio padrão de cada medida, deve-se combiná-los para que seja possível obter um único índice. Idealmente, uma rede terá máxima complexidade quando todos os seus σ_m forem iguais e, dessa forma, todos assumirão valores iguais ao *desvio padrão máximo*, $\sigma_{max} = \max_{m=1, M}(\sigma_m)$. Consequentemente, isso significa que qualquer diferença entre os σ_m e σ_{max} reduzirá a complexidade da rede. Esses conceitos são ilustrados na Figura 16.

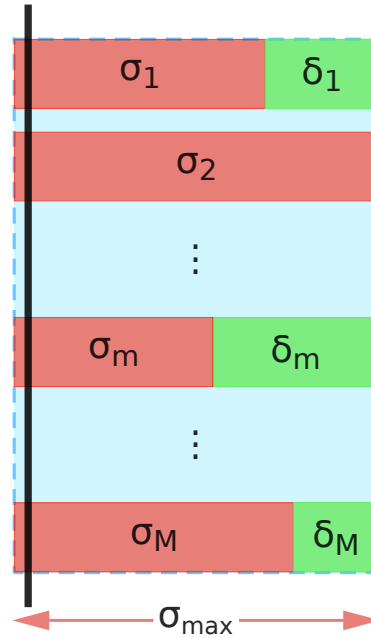


Figura 16 – Diagrama que ilustra os elementos adotados para a definição do índice de complexidade. Cada linha do esquemático corresponde ao desvio padrão de cada uma das medidas topológicas estandardizadas, σ_m , de uma dada rede complexa que terá sua complexidade mensurada. O desvio padrão máximo, σ_{max} , que aqui está sendo exemplificado por σ_2 , define a delimitação de um retângulo que representa o máximo desvio padrão efetivo, β , e também serve para calcular os valores da diferença δ_m para cada medida. A complexidade relativa será máxima, $\omega = 1$, quando as diferenças não existirem, resultando em $\Delta = 0$, ou seja, será o caso em que todos os σ_m estarão no limite desse retângulo.

Fonte: Elaborada pelo autor.

Com base no desvio padrão das medidas, é possível ilustrar o caso em que todas as M medidas têm a mesma dispersão, ou seja, o caso de máxima complexidade da rede, calculando *máximo desvio padrão efetivo*, β , por:

$$\beta = M\sigma_{max}. \quad (3.2)$$

Também é possível calcular o quanto a dispersão de cada medida contribui para a diminuição da complexidade em relação à complexidade máxima da rede. Para isso, calcula-se a diferença entre o desvio padrão máximo observado e o desvio padrão de cada medida como:

$$\delta_m = \sigma_{max} - \sigma_m. \quad (3.3)$$

Obtendo as diferenças individuais da dispersão de cada medida, δ_m , calcula-se o somatório das diferenças das dispersões, Δ , para expressar a diminuição de complexidade total em relação ao máximo de complexidade possível. Sendo assim, tem-se:

$$\Delta = \sum_{m=1}^M \delta_m. \quad (3.4)$$

Dessa forma, pode-se calcular o quanto uma rede é complexa em relação à máxima complexidade que ela poderia apresentar. Para isso, definimos um índice preliminar, ω , que expressa a *complexidade relativa* da rede de interesse, como:

$$\omega = \frac{\beta - \Delta}{\beta} = 1 - \frac{\Delta}{\beta}. \quad (3.5)$$

A complexidade relativa será máxima, $\omega = 1$, quando $\Delta = 0$, indicando que todas as medidas têm dispersões iguais, ou seja, $\delta_m = \delta_{max}, \forall m$. Entretanto, qualquer variação das dispersões em relação a δ_{max} contribuirá para o aumento do valor de Δ , o que acarretará na diminuição de ω . A complexidade relativa atingirá o seu valor mínimo quando todos os δ_m forem zero, com exceção de apenas um valor.

Por fim, definimos o *Índice de Complexidade*, IC , como sendo o valor de complexidade absoluta da rede, ω , sendo calculada pela multiplicação da complexidade relativa da rede de interesse e a média dos desvios padrões de suas medidas:

$$IC = \omega \langle \sigma \rangle. \quad (3.6)$$

Para evidenciar a diferença entre as complexidades, relativa e absoluta, considere um exemplo com duas redes complexas. A dispersão das medidas de cada rede são iguais entre si, porém são diferentes entre as duas redes, com a primeira rede assumindo $\sigma_m = 10, \forall m$ e a segunda $\sigma_m = 100, \forall m$. Dessa forma, as duas redes terão a máxima complexidade relativa, $\omega = 1$. Entretanto, a segunda rede é mais complexa do que a primeira, já que a dispersão de suas medidas são maiores do que as da primeira rede. Ou seja, a complexidade relativa expressa o quanto a rede é complexa relativamente a complexidade máxima que ela poderia assumir se todas suas dispersões fossem máximas, mas não permite a comparação de complexidade entre diferentes redes. Porém, a complexidade absoluta resolve esse problema permitindo essa comparação. Ao fazer o cálculo do IC das duas redes tendo cada uma a sua complexidade relativa máxima, a diferença será dada pela média das dispersões de suas medidas. Portanto, a primeira rede será menos complexa, com $IC = 1 \times 10 = 10$, do que a segunda rede, com $IC = 1 \times 100 = 100$.

Por último, deve ser observado que a definição aqui proposta de um índice que mensure a complexidade de uma rede pode ser utilizada para caracterizar qualquer rede complexa respectivamente a um conjunto de medidas topológicas e um conjunto de redes

de referência. Porém, devem ser levadas em consideração as limitações que aqui foram discutidas. Dessa forma, dadas redes complexas genéricas a serem estudadas, é necessário selecionar um conjunto de medidas que reflita o interesse específico da pesquisa, podendo ser as que foram adotadas aqui ou incorporar medidas alternativas (*e.g.*, (110–114)). De forma semelhante, pode ser feita a escolha das redes a serem consideradas como referência. Novamente, é possível que seja usado o mesmo grupo de modelos aqui adotado como também alguns modelos podem ser subtraídos ou adicionados, dependendo do contexto e das questões de cada pesquisa.

A priori não vemos evidências de que o IC não possa ser adaptado para calcular a complexidade de redes que estejam fora das limitações aqui previamente impostas. Isso significa que provavelmente deve ser possível realizar alterações nas medidas topológicas para que seja calculada a complexidade de redes direcionadas, com pesos, com componentes isolados entre outras. Entretanto, apesar dessas alterações serem interessantes de serem estudadas, elas fogem do escopo deste trabalho de propor uma nova forma de calcular complexidade de redes complexas.

3.2 Redes hipercomplexas

Partindo do princípio de que a complexidade de uma rede complexa pode ser considerada uma propriedade mensurável e também comparativa, é intuitivo especular sobre qual seria a rede que apresentaria a maior complexidade. Porém, sendo essa uma questão virtualmente impossível de se responder, devido à impossibilidade de medir a complexidade de todas as redes existentes, alternativamente pode-se pensar em aumentar a complexidade das redes conhecidas, principalmente as de modelos teóricos que suas propriedades já foram bastante estudadas. Dessa forma, é possível abordar a busca por alta complexidade como sendo uma otimização ocorrendo em um grande espaço de possíveis reconfigurações.

Neste trabalho, abordou-se a procura por redes com alta complexidade através de um processo de maximização, fazendo com que a sua complexidade seja uma função a ser calculada a cada passo de otimização. Para isso, propôs-se um método para aumentar a complexidade de qualquer rede de interesse e o nomeou-se de *gerador de redes hipercomplexas* em que as redes resultantes são as *redes hipercomplexas* (HC). O fluxograma da Figura 17 ilustra o método, especificando os seis parâmetros que podem ser alterados de diversas formas, possibilitando um grande número de possibilidades de geração de redes hipercomplexas.

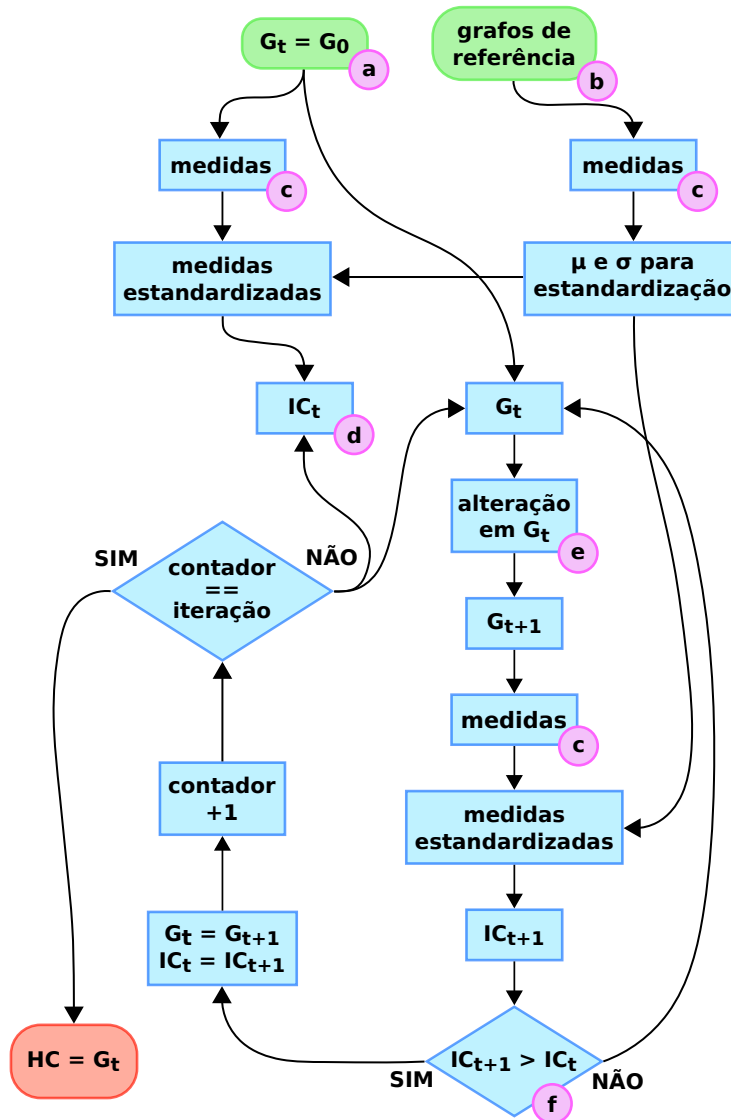


Figura 17 – Gerador de redes hipercomplexas. Projetou-se o algoritmo para gerar redes hipercomplexas como um maximizador de complexidade através de pequenas alterações topológicas. Para que isso seja feito, é necessário que seis parâmetros sejam definidos: **(a)** escolher uma rede que terá sua complexidade maximizada; **(b)** selecionar um grupo de redes complexas, podendo ser modelos teóricos ou não, para que sirvam de referência; **(c)** escolher um grupo de medidas topológicas para que a complexidade possa ser calculada; **(d)** selecionar a medida de complexidade a ser maximizada, neste trabalho definimos o IC com esse propósito, mas outras podem ser utilizadas; **(e)** escolher a modificação topológica que será aplicada na rede para que aumente sua complexidade; **(f)** selecionar o algoritmo de otimização que maximizará a complexidade da rede. É importante notar que a decisão de um parâmetro pode influenciar na decisão de outros, por exemplo o IC utilizado aqui necessita que as medidas sejam estandarizadas.

Fonte: Elaborada pelo autor.

O processo de geração de redes hipercomplexas é feito da seguinte maneira. Inicialmente, é necessário escolher a rede em que se deseja estudar o aumento de sua complexidade. *A priori* qualquer rede ou modelo pode ser escolhido. Neste trabalho, escolheu-se estudar principalmente a maximização da complexidade do modelo ER. Esse modelo foi escolhido justamente por ser considerado um modelo simples e estatisticamente regular, fornecendo um ponto de partida sem vieses que já o considerasse uma rede complexa, como é o caso da presença de *hubs* no modelo BA. Isso foi feito para que fosse possível observar quais caminhos o método evoluirá a rede para torná-la mais complexa sem interferências iniciais. Apesar disso, alguns experimentos foram feitos utilizando o modelo BA para comparar a topologia gerada pelo método tendo esse enviesamento inicial, mas o foco do trabalho foi essencialmente o modelo ER. Entretanto, qualquer outro modelo teórico ou rede real pode ser utilizada, apenas necessitando realizar possíveis adequações dependendo do tipo de rede, como escolher medidas que considerem pesos para redes que apresentam intensidades em suas conexões.

Após a escolha da rede a ser maximizada, é necessário determinar a medida de complexidade a ser utilizada, bem como as medidas de rede consideradas para que ela seja definida. O método foi desenhado para que, a princípio, qualquer medida de complexidade possa ser aplicada, guardando as respectivas adequações e limitações que cada definição de complexidade possa exigir. No caso do presente trabalho, utilizou-se o Índice de Complexidade definido na Seção 3.1. Para que o IC seja utilizado, escolheu-se o mesmo grupo de medidas de rede selecionadas para definir o índice e que foram apresentadas na seção anterior. A escolha do conjunto de redes de referência também foi a mesma apresentada na definição do IC na seção anterior, fornecendo, então, um conjunto razoavelmente abrangente de características topológicas para que as redes hipercomplexas possam ser comparadas.

Para que a maximização da complexidade de uma rede seja feita, é necessário que a rede sofra alterações em sua topologia. Essas alterações podem ser das mais diversas, podendo ser aumentar ou diminuir o número de nós ou de arestas, exclusão ou adição de *hubs*, ou alterações seguindo alguma regra que evolua a rede de alguma forma. Como a intenção é comparar a complexidade de redes com o mesmo tamanho, escolhe-se não fazer mudanças no número de elementos da rede. Dessa forma, optou-se pela alteração mais simples possível: selecionou-se de forma uniformemente aleatória uma aresta (e_{ij}), constituída dos nós n_i e n_j , deletou-se essa conexão e reconectou-se o nó n_i com outro nó escolhido de forma uniformemente aleatória, não permitindo autoligações, conexões com nós que já são seus vizinhos e nem que o nó n_j se torne um componente isolado. Com isso, manteve-se tanto o mesmo número de elementos, quanto o grau médio das redes durante todo o processo de otimização.

O processo de otimização também pode ser de diversos tipos, podendo ser métodos de gradientes, métodos estocásticos, algoritmos genéticos e assim por diante. Fez-se o uso do método de descida de gradiente (*gradient descent*) da sua forma mais simples. Inicialmente, calculou-se o IC da rede ER, que considerou-se como rede atual, $G_t = G_{t=0}$, e a cada nova alteração na rede, gerando G_{t+1} , calculou-se o da rede alterada e comparou-se ambas complexidades. Se a complexidade da nova rede for maior do que a atual, a rede nova se torna a atual, $G_t = G_{t+1}$, e passará pelo próximo passo de otimização. Porém, se o IC da rede nova não for maior que a da atual, descarta-se a rede alterada e o processo é repetido na rede atual até que a complexidade da nova rede seja maior do que a atual.

Dois critérios de parada foram adotados para o método de geração de redes hipercomplexas: (i) o número de passos aceitos atinge o valor delimitado previamente; (ii) o número de alterações nas redes atingem um valor limite para impedir que o processo ocorra infinitamente. O número limite comumente adotado foi de dez vezes o valor do número de passos aceitos, porém ele não foi atingido em nenhum dos casos avaliados. Outro critério que também pode ser adotado é o tempo de execução. Nos experimentos da Seção 4.2, em que a intenção era de visualizar a evolução estrutural de poucos exemplares de redes complexas, optou-se por evoluir a rede por volta de 3 dias antes de parar o processo. A rede que resulta deste processo é nomeada pela combinação de HC e o número de iterações aceitas, ou seja, uma rede que foi otimizada mil vezes será nomeada como HC-1000.

Vale ressaltar que o processo de geração de redes hipercomplexas, da forma como foi descrito aqui, é razoavelmente custoso. Isso é devido, principalmente, a grande quantidade de operações computacionais necessárias para que as medidas de redes complexas sejam realizadas a cada cálculo de complexidade, e conseqüentemente, a cada passo de otimização. Como esta pesquisa foi desenvolvida utilizando a linguagem *Python*, que é uma linguagem conhecida por não ter uma boa eficiência computacional, diversos artifícios foram considerados para que os experimentos pudessem ser realizados em um tempo que possibilitasse a pesquisa. A principal escolha foi basear o projeto na biblioteca de redes complexas em Python, *graph-tool*, que, além de ter sido desenvolvida para ser uma biblioteca eficiente, tem a opção de fazer as operações em paralelo de forma automática. Todos os algoritmos das medidas de redes utilizadas ou foram as já disponíveis na *graph-tool* ou foram escritas, ou adaptadas, utilizando a funções dessa biblioteca. Além disso, os demais algoritmos foram escritos utilizando vetorização das bibliotecas numéricas *NumPy* e *SciPy*, e alguns usando a biblioteca de otimização *Numba*. Todos os experimentos aqui relatados levaram dias mesmo com todos esses artifícios, alguns levando mais de uma semana. Entretanto, os códigos dessa pesquisa podem ser otimizados e, além disso, devido ao tempo do experimento ser fortemente relacionado as medidas utilizadas, a utilização de outras medidas poderá acelerar os processos.

4 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os principais resultados obtidos durante este trabalho. A Seção 4.1 apresenta tanto o aumento da complexidade das redes hipercomplexas como os resultados que derivam desse aumento. Também é feita a comparação com modelos teóricos evidenciando as alterações topológicas que o método aqui proposto causou. Por fim, a Seção 4.2 apresenta as visualizações das evoluções das redes hipercomplexas através do exemplo de duas redes, uma do modelo Erdős-Rényi e outra do modelo Barabási-Albert.

4.1 Análise quantitativa das redes hipercomplexas

Objetivando comparar as redes hipercomplexas com redes de referência, utilizou-se o mesmo grupo de modelos apresentados nos capítulos anteriores. Dessa forma, levou-se em consideração os modelos Erdős-Rényi (ER), Barabási-Albert (BA), Waxman (WAX), *Random Geometric Graph* (RGG) e duas variações de Watts-Strogatz geradas com 10% e 100% de probabilidades de reconexão (WS-10 e WS-100). Todas as redes foram geradas com $N = 500$ nós, não havendo elementos isolados e com $\langle k \rangle = 8$ com $\pm 5\%$ de tolerância. Para cada um desses modelos, foram utilizados 100 exemplares, totalizando 600 redes, com a intenção de aumentar a significância estatística da representação de cada modelo.

O método de geração de redes hipercomplexas foi aplicado em 100 redes, ou seja, gerando 100 exemplares de HCs, para que fosse possível estudar as características do processo e das novas redes com alguma confiabilidade estatística. Dessa forma, gerou-se 100 exemplares do modelo ER para que fossem utilizadas como ponto de partida para a geração das HCs. Esse modelo foi escolhido por ser estatisticamente regular, não apresentando vieses topológicos iniciais. Evoluiu-se essas redes até 4000 iterações que consecutivamente aumentaram o valor de seus respectivos ICs, como pode ser visto na Figura 18.

Como primeira abordagem para observar os resultados do método de geração de redes hipercomplexas, fez-se o uso da análise dos componentes principais (*principal component analysis* - PCA). (115, 116) Sendo cada rede como um elemento representado em um espaço com 10 dimensões (10 medidas topológicas utilizadas), aplicou-se o PCA para reduzir a dimensionalidade para duas dimensões, permitindo a análise visual. O PCA foi aplicado tanto nas redes dos seis modelos de referência quanto para 4 grupos de HCs (100 exemplares cada) sendo HC-1000, HC-2000, HC-3000 e HC-4000, como podem ser conferidos na Figura 19. Observou-se uma explicação de 96,16% da variância dos dados, corroborando a relevância da projeção do PCA nesse conjunto de redes. A média dos CIs para cada grupo de redes é apresentado com a coloração do mapa de calor.

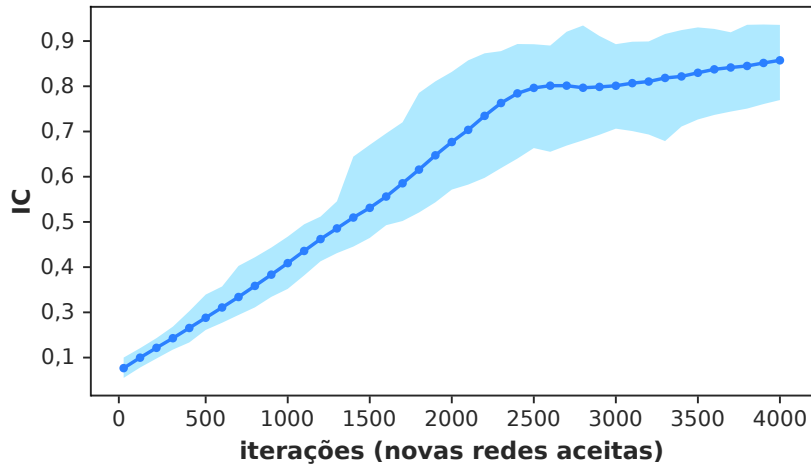


Figura 18 – Média e desvio padrão do IC de 100 exemplares de redes HCs com $N = 500$ e $\langle k \rangle = 8$ durante o período de maximização de suas complexidades, com valores a cada 100 iterações. Apesar dos valores de ICs apresentarem uma dispersão maior da iteração 1500 em diante, fica bem claro que a complexidade aumentou consideravelmente com as sucessivas iterações. De fato, a média dos ICs das redes HC-4000 é mais do que 4 vezes maior do que a média da complexidade das redes ER (HC-0) e mesmo os menores ICs das redes HC-4000s ainda são mais do que 3,5 vezes maiores do que os maiores valores de complexidades do modelo ER.

Fonte: Elaborada pelo autor.

Diversos resultados interessantes podem ser inferidos a partir da análise da Figura 19. É possível observar que os modelos de redes complexas formaram agrupamentos bem definidos, com as redes bem próximas dos seus respectivos modelos e bem separados dos outros. Há, entretanto, algumas exceções.

No caso dos modelos ER e WS-100, que estão sobrepostos, isso ocorre provavelmente por ambos serem redes totalmente aleatórias. Porém, é interessante observar que apesar dos dois modelos formarem um só agrupamento, as ICs médias de cada grupo são razoavelmente diferentes. Especula-se que, apesar de serem redes totalmente aleatórias, suas propriedades topológicas divergem razoavelmente devido à diferença no processo de geração de cada modelo.

Outros grupos de redes que apresentam sobreposição são as redes hipercomplexas. Nesse caso, as sucessivas mudanças alteraram a topologia das redes, porém, devido às alterações topológicas serem pequenas, as mudanças não foram muito grandes, gerando sobreposições entre os modelos. Isso aparentemente demonstra que, apesar de que na média exista uma tendência de se tornar um modelo cada vez mais diferente, diversos exemplares individuais de estágios diferentes da evolução compartilham características topológicas semelhantes. Isso significa que há a possibilidade de existir algumas redes HC-3000s que são mais complexas que algumas redes HC-4000s. Essa observação pode

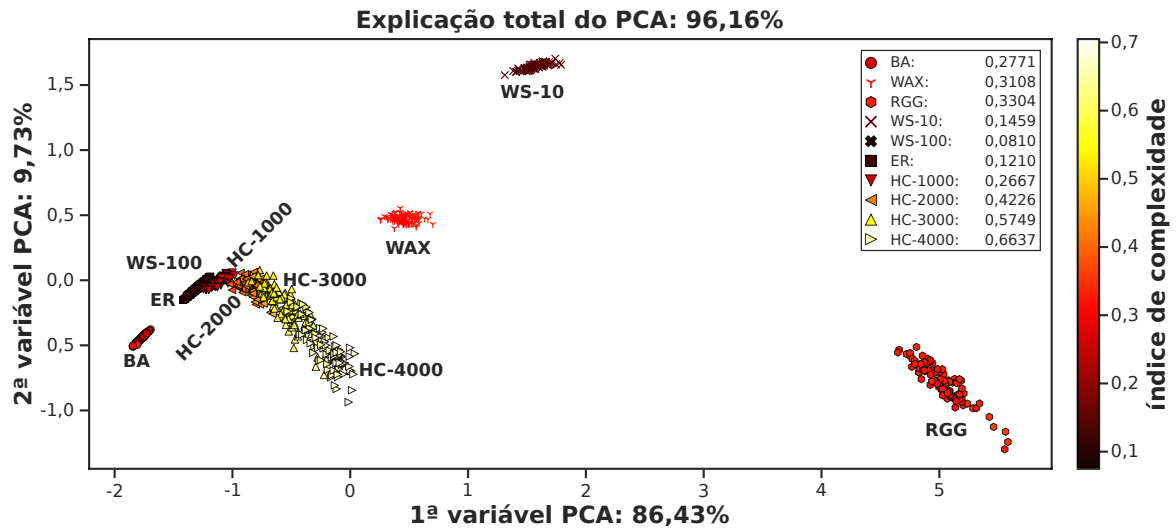


Figura 19 – Análise visual da representação do espaço PCA das 600 redes de referência, 100 de cada modelo teórico, e 400 redes hipercomplexas divididas em 4 estágios de evolução do modelo, sendo HC-1000, HC-2000, HC-3000 e HC-4000. O mapa de calor apresenta os valores do IC referente a cada rede. Pode ser observado que as redes hipercomplexas se distanciam dos outros modelos conforme sua complexidade é aumentada. Além disso, elas descrevem uma tendência a ir para uma região de baixa densidade de redes, principalmente longe do modelo BA, indicando que o aumento da complexidade gera propriedades topológicas diferentes das presentes nos modelos considerados.

Fonte: Elaborada pelo autor.

ser corroborada pela Figura 18 já que diversos valores de ICs de redes HC-2000s, ou com mais iterações, são maiores do que o menor valor de IC das redes HC-4000. Ou seja, existe uma espécie de gradiente em que, na média, existe um aumento na complexidade e uma diferenciação topológica das redes hipercomplexas, porém isso não acontece de forma tão brusca e delimitada.

Além da presença dos agrupamentos e, mais importante, como consequência dos subsequentes passos de otimização e com o aumento da complexidade, as redes HCs definiram uma progressão, iniciando no agrupamento do modelo ER, da qual foram derivadas, e traçando uma ‘trajetória’ em uma região vazia da projeção PCA. Esse distanciamento dos outros modelos de referência, que esta trajetória formou, definem uma tendência de que quanto maior a maximização, mais diferente as topologias das redes HCs se tornam dos modelos teóricos considerados. Em especial, o distanciamento do modelo BA com o aumento da complexidade, sugere que as características deste modelo não são suficientes para caracterizar redes com alta complexidade. Entretanto, vale ressaltar que essa sugestão é baseada na definição de complexidade adotada neste trabalho, podendo ser diferente se outra definição for considerada.

Os resultados obtidos corroboram que o modelo HC não só é mais complexo que as redes consideradas (considerando o índice de complexidade aqui definido), mas que também as estruturas topológicas das redes HCs se deslocam em direção a uma configuração substancialmente diferente dos outros modelos. Sendo então as HCs diferentes dos modelos utilizados, temos evidências de que o método de geração de redes hipercomplexas geram redes diferentes tanto das que são comparadas quanto das que são derivadas. Para que melhor sejam entendidas as subsequentes mudanças que as redes HCs sofreram e as propriedades associadas, apresentamos na Figura 20 a média do desvio padrão, bem como a dispersão máxima e mínima, de cada medida topológica a cada 100 passos de maximização aplicados para que as redes HCs sejam obtidas.

Verifica-se, pela Figura 20, que o processo dinâmico de otimização gerou uma tendência a aumentar o desvio padrão da maioria das medidas adotadas, como era esperado pelo aumento do IC já apresentado. Na verdade, todas as medidas tiveram um aumento em seu desvio padrão quando comparado com o modelo ER, que foram as redes iniciais. Esse resultado corrobora a efetividade desse método em derivar redes mais complexas respectivamente a todas as medidas topológicas e não apenas ao grau de seus nós.

Além disso, analisando o crescimento do desvio padrão das medidas relativo a suas próprias escalas, vê-se que as medidas relacionadas ao grau, apresentaram um aumento menor do que algumas outras quando comparadas com os seus valores iniciais. Enquanto as medidas como o grau dos nós e o segundo e terceiro grau hierárquico, apresentaram aumento por volta de 2 vezes o valor inicial, as medidas coeficiente de aglomeração, *betweenness centrality* para os nós e para arestas e o *matching index* apresentaram um aumento de aproximadamente 6 vezes. Ou seja, isso indica que com o processo de aumento da complexidade, as medidas relacionadas ao grau dos nós tiveram uma contribuição menor do que outras medidas para o valor de IC.

Entretanto, as medidas grau dos nós, autovalores Laplacianos e segundo grau hierárquico, apresentaram um leve decréscimo por volta da iteração 2000, mas voltaram a crescer por volta da iteração 3000, o que pode ser o início de uma dinâmica de oscilação. Além disso, a diversidade das demais medidas apresentaram dinâmicas de crescimento bastante individuais. Tanto a média quanto o desvio padrão do caminho mínimo demonstraram um crescimento constante por boa parte do tempo, porém tiveram um aumento repentino em um determinado momento (iteração 3000 para a média e 2000 para o desvio padrão), e uma aparente leve diminuição do crescimento próximo à iteração 4000. Já o coeficiente de aglomeração e o *matching index* percorreram um aumento quase totalmente linear com apenas uma pequena diminuição na taxa de crescimento por volta da iteração 2500, porém não indicando uma diminuição no crescimento. Por outro lado, as medidas *betweenness centrality* para os nós e arestas e o terceiro grau hierárquico parecem estar entrando em um limiar de saturação em que não haverá grandes aumentos.

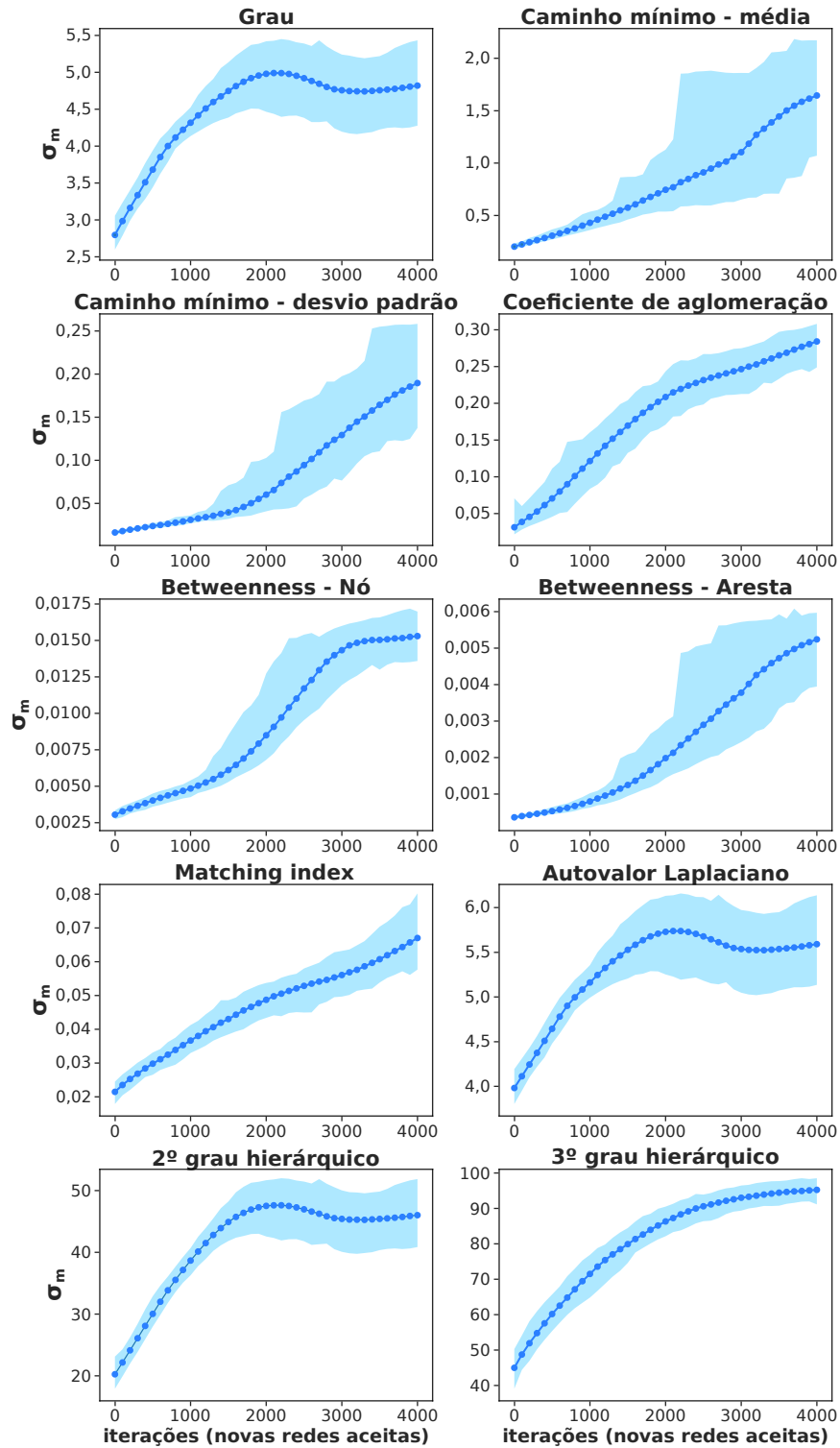


Figura 20 – Média e desvio padrão das 10 medidas topológicas calculadas a cada 100 passos de otimização para as redes hipercomplexas, evoluindo das redes iniciais, do modelo ER, até se tornarem redes HC-4000s.

Fonte: Elaborada pelo autor.

Duas observações devem ser feitas. A primeira é que essa análise está limitada a apenas 4000 passos de maximização e é difícil prever se a evolução de cada uma

das medidas se manterá a mesma se for aplicado um número maior de iterações. Em segundo lugar, é importante levar em consideração que essas análises foram feitas com base na média dos desvios padrões, porém, é possível observar que, conforme o número de iterações aumenta, maior é a dispersão dos valores. Isso sugere a presença de diversas redes com maior complexidade, mas que apresentam um certo grau de diferenciação entre si. Portanto, para uma melhor extrapolação dos resultados, são necessários estudos que levem em consideração essas duas observações.

4.2 Análise visual das redes hipercomplexas

Para que seja feita a análise visual das características das redes hipercomplexas, levou-se em consideração apenas um exemplo de rede, ao invés de 100, de um mesmo modelo teórico. O processo é exatamente o mesmo dos últimos experimentos, ou seja, mesmo conjunto de medidas para cálculo do IC, mesmo conjunto de redes e método de otimização. Entretanto, optou-se por evoluir duas redes iniciais, uma do modelo ER e uma do modelo BA, ambas com $N = 500$ e $\langle k \rangle = 8$. Isso foi feito para que seja possível realizar uma primeira comparação visual entre as características que surgirão de ambos casos e assim observar o impacto que a mudança nas redes iniciais causam nas redes hipercomplexas finais. Além disso, evoluiu-se as redes HCs até 8 mil passos. As Figuras 21 e 22 apresentam as visualizações das evoluções das redes ER e BA até redes HC-8000s com diversos estágios intermediários.

A partir da Figura 21, da evolução dos modelos ER e BA até 4000 iterações, já é possível identificar algumas características interessantes. Nas iterações 1000 e 2000, tanto a evolução de ER quanto BA aparenta uma tendência de aumentar aglomeração de alguns nós no centro da rede enquanto na periferia surgem sequências de nós com poucas conexões. Essa tendência se acentua nas iterações 3000 e 4000, surgindo aglomerações cada vez mais densas e ramos periféricos cada vez mais longos. É interessante notar que essas tendências foram mais expressivas na HCs do modelo BA do que do modelo ER. Aparentemente, a presença de *hubs* e vários nós com poucos vizinhos do modelo BA, enviesou a sua evolução.

As tendências de alta aglomeração no centro da rede e longos ramos periféricos continuaram existindo na Figura 22 que ilustra a continuação da evolução das redes até a iteração 8000. De fato, os ramos ficaram cada vez mais longos e com baixa densidade de conexões. Entretanto, a rede hipercomplexa do modelo ER não aparenta grandes evoluções, apresentando apenas um longo ramo periférico e que aumentou pouco o seu tamanho até a iteração 8000. Já no caso do modelo BA, também surge um grande ramo, porém existem alguns outros pequenos, mas maiores do que os que surgem no modelo ER. Além disso, o longo ramo de BA aparenta estar criando novos sub-ramos, como pode ser observado da iteração 7000 para 8000. Por último, pode ser notado que a característica do ER ser mais homogêneo e do BA ser mais heterogêneo aparenta persistir. Enquanto a HC do primeiro

tem um largo núcleo com nós de graus semelhantes, com exceção de alguns poucos *hubs*, a HC do segundo apresenta um ‘núcleo de *hubs*’ cercado de um grande número de nós com baixa conectividade.

Com esses resultados, pode-se observar que as redes HCs de ambos modelos criaram um contorno com baixo coeficiente de aglomeração, definindo um gradiente de grau dos nós que se estende do centro das redes com direção as periferias. Também é interessante notar que algumas estruturas de árvore surgiram nas bordas das redes HCs e foram se estabelecendo, e crescendo, conforme as redes foram evoluindo. O surgimento dessas características estruturais aparentam ser os responsáveis pelo aumento da complexidade das redes, ainda mais que nenhuma delas existiam previamente nem nos modelos ER e BA nem nos outros modelos de referência.

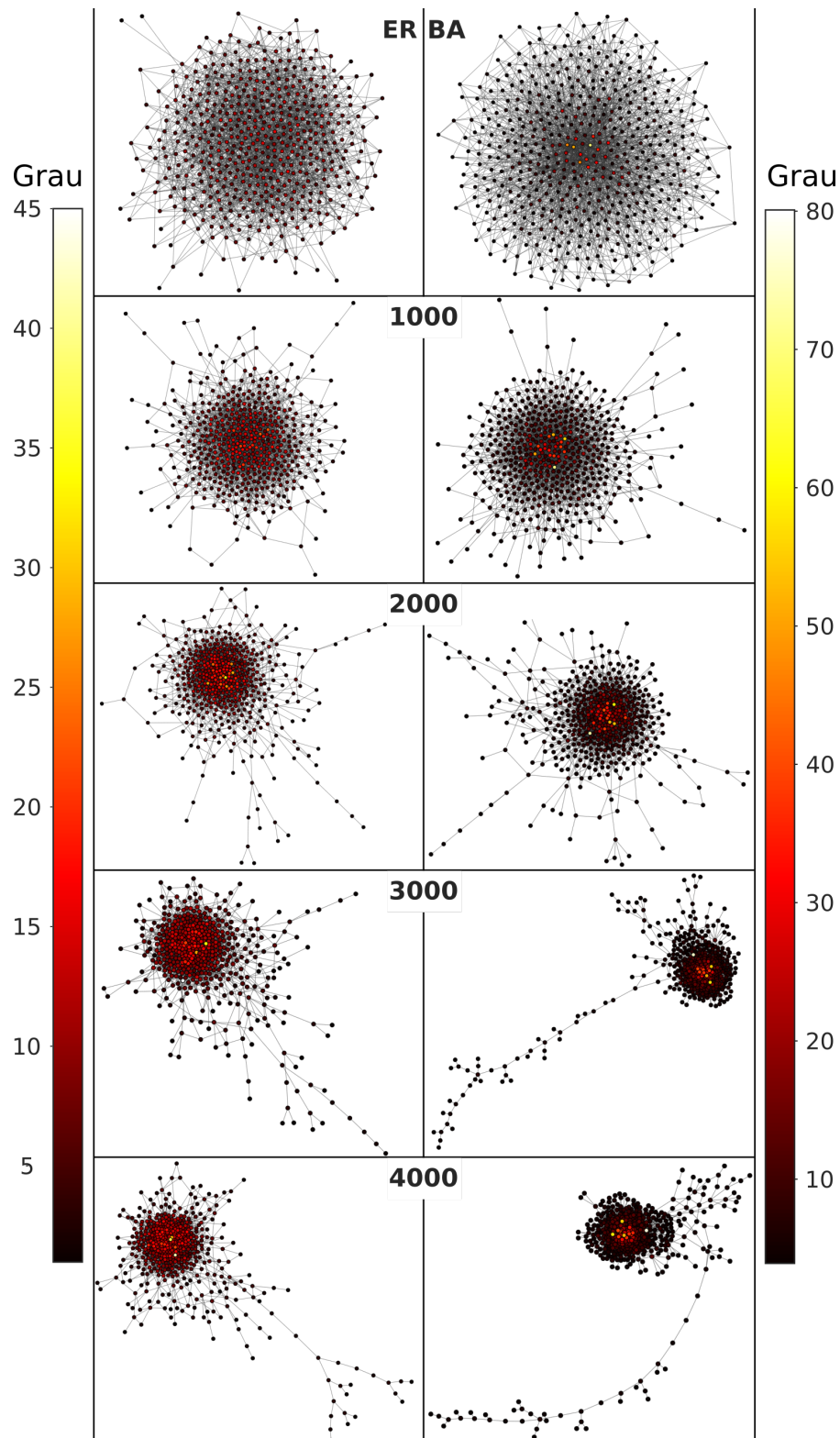


Figura 21 – Exemplo de evolução de duas redes hipercomplexa de $N = 500$ nós e $\langle k \rangle = 8$ em que a sequência de redes da esquerda inicia-se com o modelo ER e a da direita com o modelo BA. Evolução até 4000 iterações, mas continuando na Figura 22. A escala do mapa de calor da esquerda é referente aos nós das redes hipercomplexas iniciadas no modelo ER, semelhantemente para BA na direita.

Fonte: Elaborada pelo autor.

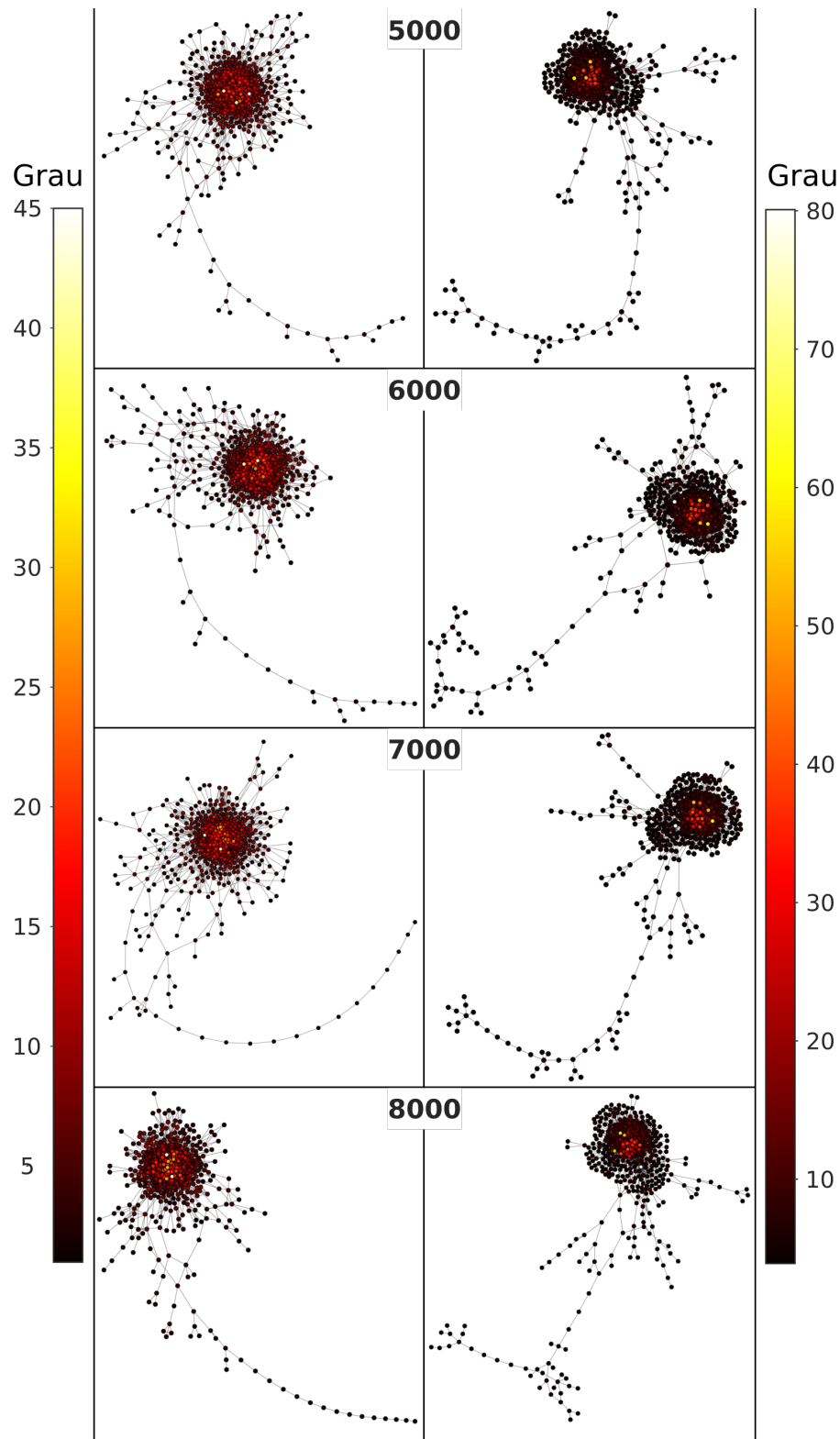


Figura 22 – Continuação da Figura 21 da evolução de duas redes hipercomplexa de $N = 500$ nós e $\langle k \rangle = 8$ em que a sequência de redes da esquerda inicia-se com o modelo ER e a da direita com o modelo BA. Evolução da iteração 4000 até 8000. A escala do mapa de calor da esquerda é referente aos nós das redes hipercomplexas iniciadas no modelo ER, semelhantemente para BA na direita.

Fonte: Elaborada pelo autor.

5 CONCLUSÕES

Este capítulo apresenta as considerações finais desta pesquisa e as perspectivas de trabalhos futuros. Na Seção 5.1, estão presentes as conclusões obtidas com a metodologia e resultados dessa dissertação. Por fim, na Seção 5.2, são descritas as ideias e especulações que podem ser investigadas futuramente utilizando este trabalho como base.

5.1 Considerações finais

De certa forma, é possível dizer que a área de ciência de redes desenvolveu-se a partir da motivação do estudo da complexidade. Ou seja, enquanto os grafos aleatórios já vinham sendo estudados por várias décadas, foi apenas com a observação de que alguns sistemas reais apresentam características que não eram tratadas anteriormente que a área de redes complexas começou a florescer. Na verdade, o próprio conceito de redes complexas advém da comparação com grafos que são considerados ‘simples’, como o modelo ER, pois apresentam nós que tendem a ter valores similares ao seu grau médio, apresentando homogeneidade e regularidade estatística. Sendo mais específico, é possível considerar a caracterização da distribuição de nós seguindo uma *lei de potência* no modelo BA como um marco que influenciou o início do estudo de complexidade em redes complexas por volta dos anos 2000. Dessa forma, com a expansão das análises da teoria de grafos para o estudo de sistemas reais com características bastante distintas das já estudadas que deu origem ao campo de pesquisa de redes complexas.

Neste trabalho, objetivou-se discutir alguns dos conceitos relativos ao termo *complexidade* em redes complexas atualmente adotados na literatura da área. Mais especificamente, intencionou-se aumentar a complexidade das redes complexas. Para isso, levou-se em consideração o conceito comumente adotado sobre complexidade que é a caracterização da diversidade da rede relativa a sua distribuição de grau ou algumas outras medidas topológicas. Entretanto, expandiu-se este conceito considerando que à complexidade é uma característica relativa a heterogeneidade de diversas medidas topológicas, não apenas uma e nem apenas a distribuição de grau, portanto, sendo caracterizada como uma medida de diversidade topológica da rede.

Para que fosse possível aumentar a complexidade de uma rede, primeiro foi necessário medi-la e para isso definiu-se o Índice de Complexidade, IC. Este índice foi projetado para que refletisse a heterogeneidade de diversas medidas topológicas de uma dada rede. Mais especificamente, o IC expressa a dispersão relativa de cada uma das medidas consideradas em respeito a maior valor de dispersão, atingindo o valor máximo quando todas as medidas tiverem o mesmo valor máximo.

Após a definição de um índice que mede a complexidade de uma rede, fez-se necessário o desenvolvimento de um método que aumentasse a sua complexidade. Para isso, definiu-se o gerador de redes hipercomplexas como um método que considera a complexidade como uma função a ser maximizada. Dessa forma, aplicou-se um procedimento de otimização nas redes de interesse, utilizando consecutivas alterações topológicas em conjunto com a descida de gradiente, para que fosse possível maximizar os valores de IC. As redes resultantes desse processo foram nomeadas como *redes hipercomplexas* - HCs, sendo substancialmente mais complexas que as redes que foram inicialmente derivadas. Vale ressaltar que este trabalho verteu-se principalmente ao estudo do modelo ER como exemplares de redes iniciais a serem evoluídas, portanto a maioria dos resultados e discussões são restritos a esse modelo.

Diversos resultados interessantes foram obtidos. Mais especificamente, tem-se que, conforme o número de passos de otimização aumenta, as redes HC apresentam uma tendência a performar uma trajetória bem definida no espaço bidimensional do PCA obtido a partir de diversas medidas topológicas. Foi observado que essa trajetória seguiu direcionada a uma área de baixa densidade no espaço PCA, apresentando um afastamento de todas as outras redes teóricas consideradas. Essa observação é uma evidência de que as redes HCs realmente apresentam características que não estão presentes nos outros modelos considerados, sugerindo, portanto, que se trata de um novo modelo de redes complexas. Além disso, como era esperado, o IC dessas redes tenderam a aumentar progressivamente acompanhando a movimentação desta trajetória.

A partir da visualização das redes HCs foi possível constatar que de fato esse método tem a característica de gerar redes que apresentam algumas propriedades topológicas adicionais que não estão presentes em nenhuma das redes comparadas, *e.g.*, ramos periféricos com baixa aglomeração. Portanto, essa diferenciação na topologia das redes HCs em relação aos outros modelos é a provável razão para que os valores de IC sejam maiores do que das outras redes consideradas. Ou seja, a alta complexidade das HCs em relação as outras redes é de fato devido ao aumento da diversidade de várias medidas topológicas e não apenas a distribuição de grau.

5.2 Trabalhos futuros

Este trabalho abre caminho para uma série de desenvolvimentos futuros tanto em investigações que considerem alterações no que foi aqui proposto, quanto no estudo que considere especificamente alternativas para as limitações desta presente pesquisa. Primeiramente, seria interessante estudar mais profundamente o impacto de cada medida topológica na caracterização da complexidade pelo Índice de Complexidade. Isso deve ser feito tanto para que seja possível compreender quais características topológicas têm maior contribuição para sua complexidade nas medidas já consideradas, quanto em novas

medidas. Além disso, também é interessante estudar a complexidade com um conjunto de medidas que caracterizam especificamente a mesma propriedade, *e.g.*, centralidade pelas medidas de grau, *betweenness centrality*, centralidade de autovetor, *pagerank*, centralidade de Katz, centralidade de proximidade e etc. Com isso, talvez seja possível apresentar uma definição de complexidade relativa a uma propriedade específica da rede.

Um segundo caminho de pesquisa a ser seguido é focar nas alterações que podem ser feitas em cada uma dos passos do método de geração de redes hipercomplexas. No primeiro passo, na escolha da rede a ser maximizada, deve ser interessante iniciar o método com redes teóricas que não sejam do modelo ER. É esperado que esses outros modelos, sendo menos homogêneos que o ER, possam implicar em um enviesamento inicial que pode ou não fazer com que a maximização da complexidade se torne tendenciosa. Além disso, esse viés pode inclusive retardar a busca por alta complexidade caso o método necessite desfazer a configuração topológica tendenciosa para depois aumentar a complexidade ou acelerar a maximização, como parece ocorrer com o experimento com o modelo BA.

A escolha da medida de complexidade e as propriedades topológicas a serem consideradas em sua definição podem apresentar resultados inesperados na geração de redes hipercomplexas. Além disso, a evolução das redes durante a maximização de sua complexidade, pode seguir caminhos muito diversos justamente por existir um grande número de dinâmicas de mudanças que possam ser aplicadas nas redes e também uma infinidade de métodos de otimização. Para exemplificar o quão diverso podem ser os caminhos possíveis para evolução de uma rede pode-se considerar duas alterações simples (aumentar e diminuir o número de arestas) e dois algoritmos de otimização (descida de gradiente estocástico e *simulated annealing*). Apenas com esses exemplos, tem-se quatro possibilidades de encontrar HCs que tenham propriedades completamente diferentes umas das outras.

Um terceiro caminho de pesquisa pode ser os estudos das propriedades dessas novas redes (tanto a deste trabalho quanto dos trabalhos sugeridos), como as sub-redes com ramos periféricos e outras que não foram observados. Uma quarta sugestão de pesquisa é considerar redes reais ao invés de modelos teóricos para serem usadas como redes iniciais. Por último, deve se obter resultados bastante surpreendentes ao adaptar essa pesquisa, e essas sugestões de trabalhos, para redes direcionadas, com peso, com componentes desconectados, dentre outras, além de também considerar com outros tamanhos.

REFERÊNCIAS

- 1 COSTA, L. F. *et al.* Characterization of complex networks: a survey of measurements. **Advances in Physics**, v. 56, n. 1, p. 167–242, Jan. 2007. DOI: 10.1080/00018730601170527.
- 2 COSTA, L. F. *et al.* Analyzing and modeling real-world phenomena with complex networks: a survey of applications. **Advances in Physics**, v. 60, n. 3, p. 329–412, June 2011. DOI: 10.1080/00018732.2011.572452.
- 3 NEWMAN, M. **Networks**. Oxford: Oxford University Press, 2018. ISBN 978-0-19-880509-0.
- 4 BOLLOBÁS, B. **Modern graph theory**. New York: Springer, 1998. (Graduate texts in mathematics, v. 184). ISBN 978-0-387-98488-9.
- 5 BOLLOBÁS, B. **Random graphs**. 2 nd. ed. Cambridge: Cambridge University Press, 2001. ISBN 978-0-511-81406-8.
- 6 MAJEED, A.; RAUF, I. Graph theory: a comprehensive survey about graph theory applications in computer science and social networks. **Inventions**, v. 5, n. 1, p. 10, Mar. 2020. DOI: 10.3390/inventions5010010.
- 7 ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, v. 74, n. 1, p. 47–97, Jan. 2002. DOI: 10.1103/RevModPhys.74.47.
- 8 DOROGOVTSEV, S. N.; MENDES, J. F. F. Evolution of networks. **Advances in Physics**, v. 51, n. 4, p. 1079–1187, June 2002. DOI: 10.1080/00018730110112519.
- 9 CASTELLANO, C.; FORTUNATO, S.; LORETO, V. Statistical physics of social dynamics. **Reviews of Modern Physics**, v. 81, n. 2, p. 591–646, May 2009. DOI: 10.1103/RevModPhys.81.591.
- 10 FORTUNATO, S. Community detection in graphs. **Physics Reports**, v. 486, n. 3, p. 75–174, Feb. 2010. DOI: 10.1016/j.physrep.2009.11.002.
- 11 FERRER I CANCHO, R.; SOLÉ, R. V.; KÖHLER, R. Patterns in syntactic dependency networks. **Physical Review E**, v. 69, n. 5, p. 051915, May 2004. DOI: 10.1103/PhysRevE.69.051915.
- 12 ANTIQUEIRA, L. *et al.* A complex network approach to text summarization. **Information Sciences**, v. 179, n. 5, p. 584–599, Feb. 2009. DOI: 10.1016/j.ins.2008.10.032.
- 13 AMANCIO, D. R.; OLIVEIRA JUNIOR., O. N.; COSTA, L. F. Structure–semantics interplay in complex networks and its effects on the predictability of similarity in texts. **Physica A**, v. 391, n. 18, p. 4406–4419, Sept. 2012. DOI: 10.1016/j.physa.2012.04.011.
- 14 SEGARRA, S.; EISEN, M.; RIBEIRO, A. Authorship attribution through function word adjacency networks. **IEEE Transactions on Signal Processing**, v. 63, n. 20, p. 5464–5478, Oct. 2015. DOI: 10.1109/TSP.2015.2451111.

- 15 AMANCIO, D. R. A complex network approach to stylometry. **PLoS ONE**, v. 10, n. 8, p. e0136076, Aug. 2015. DOI: 10.1371/journal.pone.0136076.
- 16 ARRUDA, H. F.; COSTA, L. F.; AMANCIO, D. R. Using complex networks for text classification: discriminating informative and imaginative documents. **Europhysics Letters**, v. 113, n. 2, p. 28007, Jan. 2016. DOI: 10.1209/0295-5075/113/28007.
- 17 ARRUDA, H. F. *et al.* Representation of texts as complex networks: a mesoscopic approach. **Journal of Complex Networks**, v. 6, n. 1, p. 125–144, Feb. 2018. DOI: 10.1093/comnet/cnx023.
- 18 OLIVA, S. Z. *et al.* Text structuring methods based on complex network: a systematic review. **Scientometrics**, v. 126, n. 2, p. 1471–1493, Feb. 2021. DOI: 10.1007/s11192-020-03785-y.
- 19 KITANO, H. Systems biology: a brief overview. **Science**, v. 295, n. 5560, p. 1662–1664, Mar. 2002. DOI: 10.1126/science.1069492.
- 20 SHEN-ORR, S. S. *et al.* Network motifs in the transcriptional regulation network of *Escherichia coli*. **Nature Genetics**, v. 31, n. 1, p. 64–68, May 2002. DOI: 10.1038/ng881.
- 21 KITANO, H. Biological robustness. **Nature Reviews Genetics**, v. 5, n. 11, p. 826–837, Nov. 2004. DOI: 10.1038/nrg1471.
- 22 DIAMBRA, L.; COSTA, L. F. Complex networks approach to gene expression driven phenotype imaging. **Bioinformatics**, v. 21, n. 20, p. 3846–3851, Oct. 2005. DOI: 10.1093/bioinformatics/bti625.
- 23 COSTA, L. F.; RODRIGUES, F. A.; TRAVIESO, G. Protein domain connectivity and essentiality. **Applied Physics Letters**, v. 89, n. 17, p. 174101, Oct. 2006. DOI: 10.1063/1.2363142.
- 24 WALTEMATH, D. *et al.* Toward community standards and software for whole-cell modeling. **IEEE Transactions on Biomedical Engineering**, v. 63, n. 10, p. 2007–2014, Oct. 2016. DOI: 10.1109/TBME.2016.2560762.
- 25 BURKE, P. E. P. **Simulation of biochemical systems using constraint-based methods and complex networks**. 2021. 99 p. Tese (Doutorado em Interunidades Bioinformática) — Instituto de Física de São Carlos, Instituto de Matemática e Estatística, Universidade de São Paulo, São Carlos, 2021.
- 26 MORENO, Y.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic outbreaks in complex heterogeneous networks. **European Physical Journal B**, v. 26, n. 4, p. 521–529, Apr. 2002. DOI: 10.1140/epjb/e20020122.
- 27 NEWMAN, M. E. J. Threshold effects for two pathogens spreading on a network. **Physical Review Letters**, v. 95, n. 10, p. 108701, Sept. 2005. DOI: 10.1103/PhysRevLett.95.108701.
- 28 MARCEAU, V. *et al.* Modeling the dynamical interaction between epidemics on overlay networks. **Physical Review E**, v. 84, n. 2, p. 026105, Aug. 2011. DOI: 10.1103/PhysRevE.84.026105.

-
- 29 KARRER, B.; NEWMAN, M. E. J. Competing epidemics on complex networks. **Physical Review E**, v. 84, n. 3, p. 036106, Sept. 2011. DOI: 10.1103/PhysRevE.84.036106.
- 30 SANZ, J. *et al.* Dynamics of interacting diseases. **Physical Review X**, v. 4, n. 4, p. 041005, Oct. 2014. DOI: 10.1103/PhysRevX.4.041005.
- 31 DEMERTZIS, K.; TSIOTAS, D.; MAGAFAS, L. Modeling and forecasting the COVID-19 temporal spread in Greece: an exploratory approach based on complex network defined splines. **International Journal of Environmental Research and Public Health**, v. 17, n. 13, p. 4693, Jan. 2020. DOI: 10.3390/ijerph17134693.
- 32 TSIOTAS, D.; MAGAFAS, L. The effect of anti-COVID-19 policies on the evolution of the disease: a complex network analysis of the successful case of Greece. **Physics**, v. 2, n. 2, p. 325–339, June 2020. DOI: 10.3390/physics2020017.
- 33 REHMAN, A. U.; SINGH, R.; AGARWAL, P. Modeling, analysis and prediction of new variants of covid-19 and dengue co-infection on complex network. **Chaos, Solitons & Fractals**, v. 150, p. 111008, Sept. 2021. DOI: 10.1016/j.chaos.2021.111008.
- 34 NEWMAN, M. E. J. The structure and function of complex networks. **SIAM Review**, v. 45, n. 2, p. 167–256, Jan. 2003. DOI: 10.1137/S003614450342480.
- 35 BASSETT, D. S.; BULLMORE, E. Small-world brain networks. **Neuroscientist**, v. 12, n. 6, p. 512–523, Dec. 2006. DOI: 10.1177/1073858406293182.
- 36 ARENAS, A. *et al.* Synchronization in complex networks. **Physics Reports**, v. 469, n. 3, p. 93–153, Dec. 2008. DOI: 10.1016/j.physrep.2008.09.002.
- 37 GONZÁLEZ, M. C.; HIDALGO, C. A.; BARABÁSI, A.-L. Understanding individual human mobility patterns. **Nature**, v. 453, n. 7196, p. 779–782, June 2008. DOI: 10.1038/nature06958.
- 38 ARIANOS, S. *et al.* Power grid vulnerability: a complex network approach. **Chaos**, v. 19, n. 1, p. 013119, Mar. 2009. DOI: 10.1063/1.3077229.
- 39 LYNALL, M.-E. *et al.* Functional connectivity and brain networks in schizophrenia. **Journal of Neuroscience**, v. 30, n. 28, p. 9477–9487, July 2010. DOI: 10.1523/JNEUROSCI.0333-10.2010.
- 40 BASSETT, D. S. *et al.* Dynamic reconfiguration of human brain networks during learning. **Proceedings of the National Academy of Sciences**, v. 108, n. 18, p. 7641–7646, May 2011. DOI: 10.1073/pnas.1018985108.
- 41 NEWMAN, M. E. J. Complex systems: a survey. **American Journal of Physics**, v. 79, p. 800–810, Aug. 2011. DOI: 10.1119/1.3590372.
- 42 CHU, C.-C.; IU, H. H.-C. Complex networks theory for modern smart grid applications: a survey. **IEEE Journal on Emerging and Selected Topics in Circuits and Systems**, v. 7, n. 2, p. 177–191, June 2017. DOI: 10.1109/JETCAS.2017.2692243.
- 43 BARABÁSI, A.-L.; GULBAHCE, N.; LOSCALZO, J. Network medicine: a network-based approach to human disease. **Nature Reviews Genetics**, v. 12, n. 1, p. 56–68, Jan. 2011. DOI: 10.1038/nrg2918.

- 44 SUN, G.; CHEN, C.-C.; BIN, S. Study of cascading failure in multisubnet composite complex networks. **Symmetry**, v. 13, n. 3, p. 523, Mar. 2021. DOI: 10.3390/sym13030523.
- 45 COSTA, L. F. **What is a complex network? (CDT-2)**. 2018. Disponível em: <http://rgdoi.net/10.13140/RG.2.2.10450.04804/2>. Acesso em: 23 jan. 2021.
- 46 COSTA, L. F.; RODRIGUES, F. A. Seeking for simplicity in complex networks. **Europhysics Letters**, v. 85, n. 4, p. 48001, Feb. 2009. DOI: 10.1209/0295-5075/85/48001.
- 47 ERDŐS, P.; RÉNYI, A. On the strength of connectedness of a random graph. **Acta Mathematica Academiae Scientiarum Hungaricae**, v. 12, n. 1-2, p. 261–267, Mar. 1964. DOI: 10.1007/BF02066689.
- 48 BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n. 5439, p. 509–512, Oct. 1999. DOI: 10.1126/science.286.5439.509.
- 49 KIM, J.; WILHELM, T. What is a complex graph? **Physica A**, v. 387, n. 11, p. 2637–2652, Apr. 2008. DOI: 10.1016/j.physa.2008.01.015.
- 50 RASHEVSKY, N. Life, information theory, and topology. **Bulletin of Mathematical Biophysics**, v. 17, n. 3, p. 229–235, Sept. 1955. DOI: 10.1007/BF02477860.
- 51 TRUCCO, E. A note on the information content of graphs. **Bulletin of Mathematical Biophysics**, v. 18, n. 2, p. 129–135, June 1956. DOI: 10.1007/BF02477836.
- 52 MOWSHOWITZ, A. Entropy and the complexity of graphs: IV. entropy measures and graphical structure. **Bulletin of Mathematical Biophysics**, v. 30, n. 4, p. 533–546, Dec. 1968. DOI: 10.1007/BF02476673.
- 53 MOWSHOWITZ, A. Entropy and the complexity of graphs: III. graphs with prescribed information content. **Bulletin of Mathematical Biophysics**, v. 30, n. 3, p. 387–414, Sept. 1968. DOI: 10.1007/BF02476603.
- 54 MOWSHOWITZ, A. Entropy and the complexity of graphs: II. the information content of digraphs and infinite graphs. **Bulletin of Mathematical Biophysics**, v. 30, n. 2, p. 225–240, June 1968. DOI: 10.1007/BF02476692.
- 55 MOWSHOWITZ, A. Entropy and the complexity of graphs: I. an index of the relative complexity of a graph. **Bulletin of Mathematical Biophysics**, v. 30, n. 1, p. 175–204, Mar. 1968. DOI: 10.1007/BF02476948.
- 56 TSALLIS, C. Possible generalization of Boltzmann-Gibbs statistics. **Journal of Statistical Physics**, v. 52, n. 1-2, p. 479–487, July 1988. DOI: 10.1007/BF01016429.
- 57 TSALLIS, C. Entropic nonextensivity: a possible measure of complexity. **Chaos, Solitons & Fractals**, v. 13, n. 3, p. 371–391, Mar. 2002. DOI: 10.1016/S0960-0779(01)00019-4.
- 58 DEHMER, M.; MOWSHOWITZ, A. A history of graph entropy measures. **Information Sciences**, v. 181, n. 1, p. 57–78, Jan. 2011. DOI: 10.1016/j.ins.2010.08.041.
- 59 MOWSHOWITZ, A.; DEHMER, M. Entropy and the complexity of graphs revisited. **Entropy**, v. 14, n. 3, p. 559–570, Mar. 2012. DOI: 10.3390/e14030559.

-
- 60 CAO, S.; DEHMER, M.; SHI, Y. Extremality of degree-based graph entropies. **Information Sciences**, v. 278, p. 22–33, Sept. 2014. DOI: 10.1016/j.ins.2014.03.133.
- 61 ZHANG, Q. *et al.* Tsallis information dimension of complex networks. **Physica A**, v. 419, p. 707–717, Feb. 2015. DOI: 10.1016/j.physa.2014.10.071.
- 62 NIE, T. *et al.* Using mapping entropy to identify node centrality in complex networks. **Physica A**, v. 453, p. 290–297, July 2016. DOI: 10.1016/j.physa.2016.02.009.
- 63 XU, Z.; PU, C.; YANG, J. Link prediction based on path entropy. **Physica A**, v. 456, p. 294–301, Aug. 2016. DOI: 10.1016/j.physa.2016.03.091.
- 64 LEI, M.; LIU, L.; WEI, D. An improved method for measuring the complexity in complex networks based on structure entropy. **IEEE Access**, v. 7, p. 159190–159198, 2019. DOI: 10.1109/ACCESS.2019.2950691.
- 65 WEN, T.; JIANG, W. Measuring the complexity of complex network by Tsallis entropy. **Physica A**, v. 526, p. 121054, July 2019. DOI: 10.1016/j.physa.2019.121054.
- 66 DEHMER, M. Information processing in complex networks: graph entropy and information functionals. **Applied Mathematics and Computation**, v. 201, n. 1-2, p. 82–94, July 2008. DOI: 10.1016/j.amc.2007.12.010.
- 67 HAN, L. *et al.* Graph characterizations from von Neumann entropy. **Pattern Recognition Letters**, v. 33, n. 15, p. 1958–1967, Nov. 2012. DOI: 10.1016/j.patrec.2012.03.016.
- 68 MORZY, M.; KAJDANOWICZ, T.; KAZIENKO, P. On measuring the complexity of networks: Kolmogorov complexity versus entropy. **Complexity**, v. 2017, p. e3250301, Nov. 2017. DOI: 10.1155/2017/3250301.
- 69 BASHARIN, G. P. On a statistical estimate for the entropy of a sequence of independent random variables. **Theory of Probability & Its Applications**, v. 4, p. 333–336, Jan. 1959. DOI: 10.1137/1104033.
- 70 HARRIS, B. **The statistical estimation of entropy in the non-parametric case**. Madison, Wisconsin: National Technical Information Service, 1975. 47 p. MRC Technical Summary Report #1605.
- 71 WOLPERT, D. H.; WOLF, D. R. Estimating functions of probability distributions from a finite set of samples. **Physical Review E**, v. 52, n. 6, p. 6841–6854, Dec. 1995. DOI: 10.1103/PhysRevE.52.6841.
- 72 HAN, Y.; JIAO, J.; WEISSMAN, T. Adaptive estimation of Shannon entropy. *In*: INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY (ISIT), 2015, Hong Kong. **Proceedings [...]**. Hong Kong: IEEE, 2015. p. 1372–1376. DOI: 10.1109/ISIT.2015.7282680.
- 73 WEST, D. **Introduction to graph theory**. 2 nd. ed. New York: Prentice Hall, 2001. ISBN 978-0-13-014400-3.
- 74 STROGATZ, S. H. Exploring complex networks. **Nature**, v. 410, n. 6825, p. 268–276, Mar. 2001. DOI: 10.1038/35065725.

- 75 ZWEIG, K. A. **Network analysis literacy**: a practical approach to the analysis of networks. Vienna: Springer, 2016. (Lecture notes in social networks). ISBN 978-3-7091-0740-9 978-3-7091-0741-6.
- 76 LATORA, V.; NICOSIA, V.; RUSSO, G. **Complex networks**: principles, methods and applications. Cambridge: Cambridge University Press, 2017. ISBN 978-1-107-10318-4 978-1-316-21600-2.
- 77 MENCZER, F.; FORTUNATO, S.; DAVIS, C. A. **A first course in network science**. Cambridge: Cambridge University Press, 2020. ISBN 978-110-865394-7.
- 78 COSCIA, M. **The Atlas for the aspiring network scientist**. 2021. Disponível em: <http://arxiv.org/abs/2101.00863>. Acesso em: 30 abr. 2021.
- 79 BOCCALETTI, S. *et al.* Complex networks: structure and dynamics. **Physics Reports**, v. 424, n. 4-5, p. 175–308, Feb. 2006. DOI: 10.1016/j.physrep.2005.10.009.
- 80 DIJKSTRA, E. W. A note on two problems in connexion with graphs. **Numerische Mathematik**, v. 1, n. 1, p. 269–271, Dec. 1959. DOI: 10.1007/BF01386390.
- 81 WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **Nature**, v. 393, n. 6684, p. 440–442, June 1998. DOI: 10.1038/30918.
- 82 FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, v. 40, n. 1, p. 35, Mar. 1977. DOI: 10.2307/3033543.
- 83 BRANDES, U. A faster algorithm for betweenness centrality. **Journal of Mathematical Sociology**, v. 25, n. 2, p. 163–177, June 2001. DOI: 10.1080/0022250X.2001.9990249.
- 84 ADAMIC, L. A.; GLANCE, N. **The political blogosphere and the 2004 U.S. election**: divided they blog. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/1134271.1134277>. Acesso em: 01 nov. 2021.
- 85 BRANDES, U.; PICH, C. Centrality estimation in large networks. **International Journal of Bifurcation and Chaos**, v. 17, n. 07, p. 2303–2318, July 2007. DOI: 10.1142/S0218127407018403.
- 86 HILGETAG, C. C. *et al.* Computational methods for the analysis of brain connectivity. *In*: ASCOLI, G. A. (ed.). **Computational neuroanatomy**: principles and methods. Totowa, NJ: Humana Press, 2002. p. 295–335. DOI: 10.1007/978-1-59259-275-3_14.
- 87 KAISER, M.; HILGETAG, C. C. Edge vulnerability in neural and metabolic networks. **Biological Cybernetics**, v. 90, n. 5, p. 311–317, May 2004. DOI: 10.1007/s00422-004-0479-1.
- 88 POTHEEN, A.; SIMON, H. D.; LIOU, K.-P. Partitioning sparse matrices with eigenvectors of graphs. **SIAM Journal on Matrix Analysis and Applications**, v. 11, n. 3, p. 430–452, July 1990. DOI: 10.1137/0611030.
- 89 FIEDLER, M. Algebraic connectivity of graphs. **Czechoslovak Mathematical Journal**, v. 23, n. 2, p. 298–305, 1973. DOI: 10.21136/CMJ.1973.101168.
- 90 NEWMAN, M. E. J. Finding community structure in networks using the eigenvectors of matrices. **Physical Review E**, v. 74, n. 3, p. 036104, Sept. 2006. DOI: 10.1103/PhysRevE.74.036104.

-
- 91 COSTA, L. F.; SILVA, F. N. Hierarchical characterization of complex networks. **Journal of Statistical Physics**, v. 125, n. 4, p. 841–872, Nov. 2006. DOI: 10.1007/s10955-006-9130-y.
- 92 WATTS, D. J. **Small worlds: the dynamics of networks between order and randomness**. Princeton: Princeton University Press, 1999. ISBN 978-0691117041.
- 93 ALTMANN, E. G.; PIERREHUMBERT, J. B.; MOTTER, A. E. Beyond word frequency: bursts, lulls, and scaling in the temporal distributions of words. **PLoS ONE**, v. 4, n. 11, p. e7678, Nov. 2009. DOI: 10.1371/journal.pone.0007678.
- 94 ABE, S.; SUZUKI, N. Scale-free network of earthquakes. **Europhysics Letters**, v. 65, n. 4, p. 581, Feb. 2004. DOI: 10.1209/epl/i2003-10108-1.
- 95 PERCACCI, R.; VESPIGNANI, A. Scale-free behavior of the Internet global performance. **European Physical Journal B**, v. 32, n. 4, p. 411–414, Apr. 2003. DOI: 10.1140/epjb/e2003-00123-6.
- 96 HAYASHI, Y. A review of recent studies of geographical scale-free networks. **Information and Media Technologies**, v. 1, n. 2, p. 1136–1145, 2006. DOI: 10.11185/imt.1.1136.
- 97 MITZENMACHER, M. A brief history of generative models for power law and lognormal distributions. **Internet Mathematics**, v. 1, n. 2, p. 226–251, Jan. 2004. DOI: 10.1080/15427951.2004.10129088.
- 98 NEWMAN, M. E. J. Power laws, Pareto distributions and Zipf's law. **Contemporary Physics**, v. 46, n. 5, p. 323–351, Sept. 2005. DOI: 10.1080/00107510500052444.
- 99 PRICE, D. D. S. A general theory of bibliometric and other cumulative advantage processes. **Journal of the American Society for Information Science**, v. 27, n. 5, p. 292–306, 1976. DOI: <https://doi.org/10.1002/asi.4630270505>.
- 100 MANDELBROT, B. B. **Fractals and scaling in finance: discontinuity, concentration, risk**. selecta volume E. New York: Springer-Verlag, 1997. ISBN 978-0-387-98363-9.
- 101 GABAIX, X. Zipf's law for cities: an explanation. **The Quarterly Journal of Economics**, v. 114, n. 3, p. 739–767, Aug. 1999. DOI: 10.1162/003355399556133.
- 102 WAXMAN, B. M. Routing of multipoint connections. **IEEE Journal on Selected Areas in Communications**, v. 6, n. 9, p. 1617–1622, Dec. 1988. DOI: 10.1109/49.12889.
- 103 COXETER, H. S. M. **Introduction to geometry**. 2 nd. ed. New York: Wiley, 1969. ISBN 978-0471504580.
- 104 GILBERT, E. N.; POLLAK, H. O. Steiner minimal trees. **SIAM Journal on Applied Mathematics**, v. 16, n. 1, p. 1–29, Jan. 1968. DOI: 10.1137/0116001.
- 105 COURANT, R.; ROBBINS, H. **What is mathematics?: an elementary approach to ideas and methods**. 2 nd. ed. Oxford: Oxford University Press, 1996. ISBN 978-0195105193.

- 106 KARP R., M. Reducibility among combinatorial problems. *In*: MILLER, R. E.; THATCHER, J. W.; BOHLINGER, J. D. (ed.). **Complexity of computer computations**: the IBM research symposia series. Boston, MA: Springer, 1972. p. 85–103. DOI: 10.1007/978-1-4684-2001-2_9.
- 107 BARTHÉLEMY, M. Spatial networks. **Physics Reports**, v. 499, n. 1, p. 1–101, Feb. 2011. DOI: 10.1016/j.physrep.2010.11.002.
- 108 DALL, J.; CHRISTENSEN, M. Random geometric graphs. **Physical Review E**, v. 66, n. 1, p. 016121, July 2002. DOI: 10.1103/PhysRevE.66.016121.
- 109 PENROSE, M. **Random geometric graphs**. Oxford: Oxford University Press, 2003. ISBN 978-0-19-850626-3.
- 110 KIM, H.-J.; KIM, J. M. Cyclic topology in complex networks. **Physical Review E**, v. 72, n. 3, p. 036109, Sept. 2005. DOI: 10.1103/PhysRevE.72.036109.
- 111 ESTRADA, E.; RODRÍGUEZ-VELÁZQUEZ, J. A. Subgraph centrality in complex networks. **Physical Review E**, v. 71, n. 5, p. 056103, May 2005. DOI: 10.1103/PhysRevE.71.056103.
- 112 DANGALCHEV, C. Residual closeness in networks. **Physica A**, v. 365, n. 2, p. 556–564, June 2006. DOI: 10.1016/j.physa.2005.12.020.
- 113 LI, C. *et al.* Identification of influential spreaders based on classified neighbors in real-world complex networks. **Applied Mathematics and Computation**, v. 320, p. 512–523, Mar. 2018. DOI: 10.1016/j.amc.2017.10.001.
- 114 WANG, J.; LI, C.; XIA, C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. **Applied Mathematics and Computation**, v. 334, p. 388–400, Oct. 2018. DOI: 10.1016/j.amc.2018.04.028.
- 115 WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 2, n. 1-3, p. 37–52, 1987. DOI: 10.1016/0169-7439(87)80084-9.
- 116 GEWERS, F. L. *et al.* Principal component analysis: a natural approach to data exploration. **ACM Computing Surveys**, v. 54, n. 4, p. 1–34, July 2021. DOI: 10.1145/3447755.