

Uma introdução à Ciência de Redes e Teoria de Grafos

An introduction to Network Science and Graph Theory

Paulo Freitas Gomes^{*1}

¹Universidade Federal de Jataí, Instituto de Ciências Exatas e Tecnológicas, Jataí, GO, Brasil.

Recebido em 29 de maio de 2024. Revisado em 19 de agosto de 2024. Aceito em 19 de agosto de 2024.

A Ciência de Redes é um formalismo matemático essencial na descrição de sistemas complexos. Este artigo é uma breve introdução dos conceitos principais da Teoria de Grafos, a base da Ciência de Redes. Começando com uma introdução histórica apresentamos os elementos básicos para definição de grafos como grau médio, matriz adjacência e componentes. Em seguida apresentamos alguns modelos básicos e famosos de redes mostrando suas definições e propriedades. Finalizamos com uma apresentação do fenômeno de percolação em rede que mostra a sinergia entre Ciência de Redes e conceitos mais tradicionais da física como transição de fase e criticalidade.

Palavras-chave: Redes, grafos, grau, componentes, rede aleatória, rede livre de escala.

Network Science emerged as an essential mathematical framework for the description of complex systems. This paper is a concise introduction to the main concepts of Graph Theory, the foundation of the Network Science. Starting with an historical introduction, we present the basic definition of graph and its properties such as average degree, adjacency matrix and components. Then we present some famous graph models showing its definitions and properties. We finish this paper with a description of percolation on networks showing the synergy between Network Science and Physics, while dealing with concepts like phase transition and criticality.

Keywords: Network, graph, degree, components, random network, scale-free network.

1. Introdução

Recentemente, a denominação Ciência de Redes tem sido aplicada às mais diversas aplicações da Teoria de Grafos nos mais diversos ramos da Ciência. A definição mais simples de uma rede (ou grafo) é bastante intuitiva: uma coleção de pontos e linhas, sendo que cada linha conecta dois pontos [1, 2]. Pode haver pontos desconectados, linhas com sentido definido, pontos com várias linhas e até linhas que saem e voltam ao mesmo ponto. Os pontos são chamados de vértices (ou nós) e as linhas de conexões (ou arestas). Repare que a função das linhas é apenas conectar dois pontos de forma que sua trajetória (reta ou curva) em uma representação gráfica não tem importância. Teoria de Grafos é o ramo da matemática que descreve as propriedades de grafos. Já a denominação “Redes Complexas” é comumente usado para designar modelos de redes aplicados no estudo de sistemas complexos [3–5]. Os termos rede e grafo são usados para o mesmo objeto, mas dependente do contexto. Diversos sistemas e fenômenos reais podem ser descritos utilizando o ferramental da Ciência de Redes [6], por exemplo:

- Redes biológicas: vários sistemas biológicos podem ser modelados utilizando redes [7]. A rede neuronal é a rede de conexões entre os neurônios no

cérebro [8]. Já as redes bioquímicas representam a interação entre as moléculas nos diversos processos metabólicos em uma célula. Há também as redes ecológicas que representam as interações entre as diferentes espécies em um dado ecossistema.

- Redes sociais: são redes nas quais pessoas ou grupos de pessoas são os vértices e suas interações são as conexões. As relações entre as pessoas formam as comunidades e sociedades, e elas tem sido a base de estudos de fenômenos como disseminação de cultura e opinião, processos cooperativos, aprendizado de linguagem, colaboração científica e caracterização de grupos sociais [9]. Toda essa área é também chamada de Dinâmica Social [10]. As redes sociais apresentam muitas características interessantes, como a propriedade chamada *mun*do *pequeno* na qual o diâmetro (máxima distância entre dois vértices) cresce logaritmicamente com o tamanho da rede, ao invés de linearmente [11, 12]. Este é o conceito por trás do famoso resultado conhecido como “6 graus de separação”: não importa o quão grande uma rede social seja, é necessário um número pequeno (~ 6) conexões para se conectar quaisquer duas pessoas dessa rede [13, 14].
- Redes tecnológicas e de transporte: diversos avanços tecnológicos presentes na sociedade formam redes [2]. A internet conecta dispositivos no mundo inteiro, enquanto a rede elétrica distribui energia elétrica nas cidades (veja Figura 1). Já a malha

*Endereço de correspondência: paulofreitasgomes@ufj.edu.br

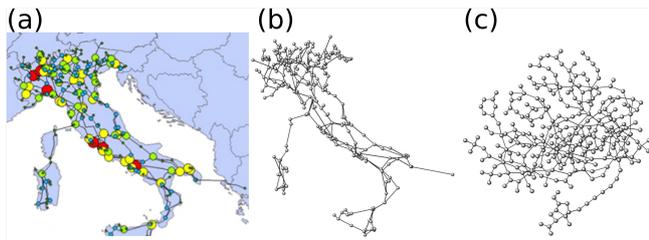


Figura 1: (a) Rede elétrica italiana justaposta com o mapa geográfico da Itália. (b) Rede mantendo a organização espacial dos vértices. (c) Rede sem a organização espacial mas mantendo as propriedades da topologia. Adaptado de [1, 19].

rodoviária, ferroviária, rotas aéreas e mapas de cidades formam as chamadas redes de transporte. No caso da última, os cruzamentos são os vértices e ruas são as conexões [15].

- Texto: um texto pode ser representado como uma rede na qual os vértices são as palavras e as conexões determinam a sequência. Diversas propriedades podem ser obtidas com essa representação [16, 17].

Redes sociais remetem quase que imediatamente as plataformas online de interação social como Facebook, Instagram, etc... Porém, o estudo das interações entre pessoas em um dado grupo é bem mais antigo do que isso. Um primeiro trabalho sobre interações sociais pode ser atribuído ao psiquiatra Jacob Moreno, que publicou seu livro em 1934 [18], no qual relata seus estudos sobre interação em diversos grupos de pessoas. Uma rede utilizada no livro está representada na Figura 2, mostrando a rede de interações entre um grupo de bebês.

1.1. Euler

Curiosamente, podemos estabelecer com precisão a primeira utilização de grafos como ferramenta matemática.

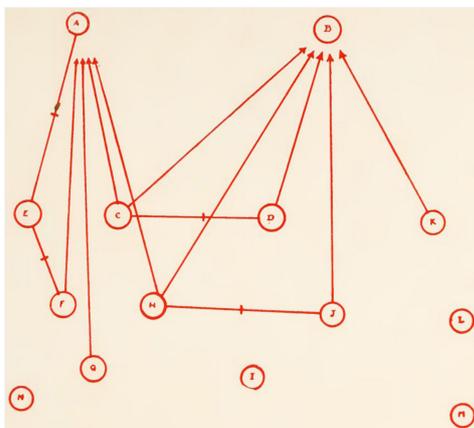


Figura 2: Representação da rede usada no estudo das interação entre bebês. Cada círculo representa um bebê e as linhas entre eles são as suas interações. Adaptado de [18], pág. 33.

Em 1735, o matemático suíço Leonard Euler resolveu um famoso problema, batizado de “As sete pontes de Königsberg”. Poucas áreas de pesquisa podem ter seu nascimento definido de maneira tão precisa no tempo e espaço: em 1735 na cidade de Königsberg (hoje Kaliningrado - Rússia). A cidade é cortada pelo antigo rio Pregel (agora chamado de rio Pregolya) que permitiu construir 7 pontes, como mostra a Figura 3(a), sendo 5 delas conectadas com a ilha Kneiphof. Essa configuração gerou um quebra-cabeça: é possível atravessar todas as pontes sem nunca atravessar uma mesma ponte duas vezes? O problema só foi resolvido em 1735 quando Euler ofereceu uma prova rigorosa mostrando que não havia solução [20]. Euler considerou os pedaços de terra como os vértices A, B, C e D e as pontes como conexões. Isso forma o grafo mostrado na Figura 3(b). Euler percebeu que para haver um caminho contendo todos os vértices, e passando por cada conexão apenas uma vez, apenas o primeiro e o último vértice poderiam ter um número ímpar de conexões. Mas nesse caso, todos os vértices tem um número ímpar de conexões, impossibilitando tal caminho [21]. A solução de Euler foi a primeira vez que alguém resolveu um problema matemático utilizando o

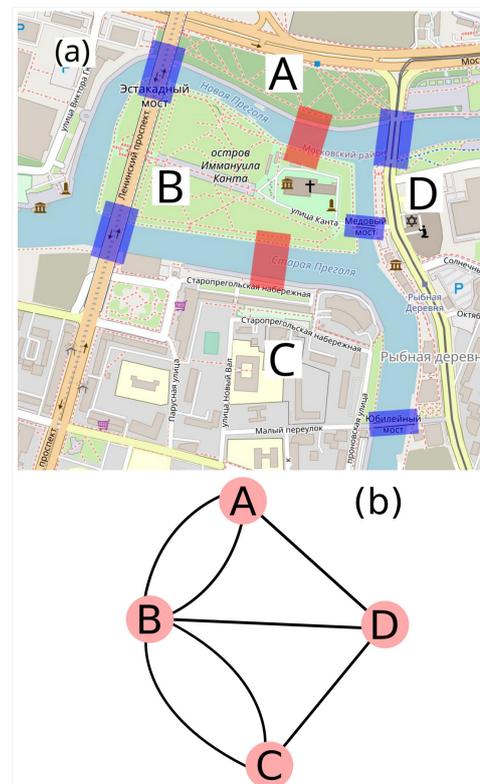


Figura 3: (a) Mapa das pontes de Königsberg. Os retângulos em azul são as pontes originais que ainda existem. Os retângulos em vermelhos indicam as posições de duas pontes originais que não existem mais hoje. Imagem adaptada a partir do Open Street Map (<https://www.openstreetmap.org/#map=17/54.70638/20.51040>). (b) Grafo equivalente onde os vértices são as regiões e as conexões são as pontes.

conceito de grafo. Com isso, ele mostrou que alguns problemas podem ser tornar mais simples quando utilizamos Teoria de Grafos. E além disso, a solução desse problema não era uma questão tecnológica ou de imaginação, e sim uma propriedade topológica do grafo em questão. Atualmente, a configuração das pontes na região está diferente e esse problema tem solução já que os nós A e C têm um número par de vértices.

1.2. Consolidação

A Ciência de Redes se consolidou como uma nova área de conhecimento no século XXI, e isso se deve a dois fatores principais [1]:

1. Capacidade de obter, organizar e processar uma grande quantidade de informação. Essa capacidade cresceu junto com o desenvolvimento dos computadores e da internet. Computadores oferecem meios de armazenamento e processamento das informações, permitindo sua organização enquanto a internet possibilitou a obtenção dos mais diferentes tipos de informação. Redes neuronais e de pessoas já existiam muito antes do século XXI, mas a falta de dados organizados para se criar um mapa das rede das interações impossibilitava o estudo sistemático de suas propriedades.
2. Universalidade das redes: existem vários modelos de redes aplicados em diversos fenômenos e sistemas reais. Porém, há uma universalidade nas características das redes, permitindo a definição de características básicas e sua observação em sistemas completamente diferentes.

A Ciência de Redes é uma ciência interdisciplinar tendo como base o formalismo matemático chamado de Teoria de Grafos [2]. Tem também um aspecto computacional quando aplicada na análise de enormes quantidades de dados. Essa análise refere-se a organização das redes a partir dos dados brutos, como também a análise em si das redes. Há uma grande oferta de ferramentas computacionais bem estabelecidas, para análise de dados (Ciência de Dados) e de redes diretamente. Há opções gratuitas, comerciais, plataformas independentes, pacotes para serem usados em uma determinada linguagem de programação, dentre outras. Vale destacar aqui as linguagens de programação Python (<https://www.python.org/>) e R (<https://www.r-project.org/>), as quais tem inúmeros pacotes de funções bem estabelecidos tanto para análise de dados quanto análise de redes. Ambas as opções são gratuitas e de código aberto. Há também inúmeros repositórios abertos com dados de redes reais, disponíveis para interessados utilizá-los em estudos e análises (uma busca no Google retorna várias opções).

1.3. Conteúdo

Este trabalho está organizado da seguinte forma. Na seção 2 fazemos uma rápida descrição do formalismo

de Teoria de Grafos, elencando alguns conceitos principais. Em seguida descrevemos com alguns detalhes 3 modelos de redes que tem sido muito utilizados nas últimas décadas: na seção 3 descrevemos a chamada rede aleatória, na seção 4 descrevemos uma rede livre de escala, finalizando com a rede geométrica na seção 5. Já na seção 6 descrevemos um exemplo de processo muito estudado em rede e com diversas aplicações, percolação, e finalizaremos com as conclusões.

2. Formalismo

Uma rede G contendo N vértices é definida como um conjunto $G(N, C)$. Aqui, C corresponde a um conjunto de M pares ordenados (i, j) , que representa as conexões entre vértices [5]. O conjunto C é chamado de lista de adjacências. Os vértices são indexados por $i = 1, 2, 3, \dots, N$, enquanto cada conexão é definida pelos dois vértice que ela liga: a conexão (i, j) liga os vértices i e j . Quando as conexões (i, j) e (j, i) são as mesmas, dizemos que a conexão é não direcionada. Por exemplo no Facebook se uma pessoa é amiga de outra, a outra é amiga da primeira. Se a ordem importa a conexão é chamada de direcionada: exemplos são as redes no Instagram e X (antigo Twitter): é possível que uma pessoa A siga uma pessoa B sem que B reciprocamente siga A. Por outro lado, se a conexão puder ser mais intensa ou fraca, dizemos que ela é ponderada ou tem peso [22]. Relações de amizade podem ser representada por conexões com peso: um indivíduo é muito amigo de um segundo indivíduo mas não tão amigo de um terceiro.

A lista de adjacências C é uma forma eficiente de se representar uma rede. A rede da Figura 4(a) é definida por $N = 6$ e pela lista $C = \{(1 : 2), (1 : 3), (1 : 4), (1 : 5), (3 : 4)\}$. Para uma rede não direcionada podemos usar o padrão $(i : j)$ com $i < j$ para ordenamento. Como já dito anteriormente, quando uma rede é representada graficamente a trajetória das linhas que representam as conexões não tem importância. O mesmo é válido para as posições em que os vértices são desenhados. Ao se criar uma representação gráfica a posição dos vértices é definida de forma conveniente para se realçar a propriedade estudada na rede. As linhas das conexões em geral são retas mas podem ser curvas caso desejado para uma visualização mais intuitiva. Por exemplo, as 3 redes desenhadas na Figura 4 são idênticas, como também são as redes da Figura 1.

2.1. Matriz adjacência

No estudo analítico das propriedades de rede, é conveniente usar a matriz adjacência \mathbf{A} onde cada elemento é definido da forma [2]:

$$\mathbf{A} = \begin{cases} a_{ij}, & \text{peso da conexão entre } i \text{ e } j \\ 0, & \text{se não houver conexão.} \end{cases}$$

Uma rede sem peso tem $a_{ij} = 1$, enquanto em uma rede com peso a_{ij} pode ser um número inteiro, ou

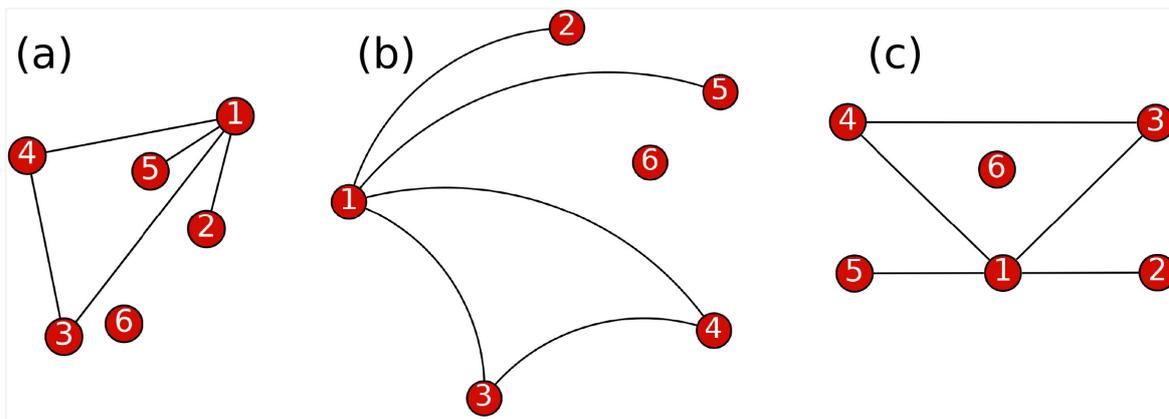


Figura 4: Ilustração de uma rede simples (não ponderada) e não direcionada representada de 3 formas diferentes. Apesar dos vértices estarem em posições diferentes as 3 redes são idênticas pois não há alteração das conexões nem do número de vértices. Parâmetros: $N = 6$ e $M = 5$. A matriz adjacência dessa rede está definida pela Eq. 1 enquanto que a lista de adjacências é $C = \{(1 : 2), (1 : 3), (1 : 4), (1 : 5), (3 : 4)\}$. Número de componentes $\mathcal{N}_c = 2$ e tamanho da maior componente $\mathcal{S}_c = 5$.

real, diferente de 1.0. Se a rede é não direcionada tem-se que $A_{ij} = A_{ji}$, já no caso direcionada pode-se ter $A_{ij} \neq A_{ji}$ [23]. No geral tem-se também $A_{ii} = 0$, mas pode haver redes com uma auto-conexão: uma aresta que sai de um vértice e volta nele mesmo, indicada por $A_{ii} \neq 0$. A matriz adjacência das 3 redes da Figura 4 é:

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

2.2. Propriedades básicas

O grau k_i de um vértice i é o número de conexões que ele tem com outros vértices, ou seja, sua quantidade de vizinhos. Em uma rede não direcionada com N vértices, temos que [24]:

$$k_i = \sum_{j=1}^N A_{ij}. \quad (2)$$

Seja M o número de conexões. Em uma rede não direcionada, o total de conexões é metade da soma de todos os graus:

$$M = \frac{1}{2} \sum_{i=1}^N k_i. \quad (3)$$

Cada conexão é contada duas vezes: uma vez para cada um dos dois vértices que ela conecta. Por isso há o fator $1/2$. Podemos expressar M em função da matriz adjacência \mathbf{A} usando a Eq. 2:

$$M = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N A_{ij}.$$

Já o grau médio de uma rede é a média dos graus de todos os vértices da rede:

$$\mu = \frac{1}{N} \sum_{i=1}^N k_i.$$

Combinando esse resultado com a Eq. 3 temos $\mu = 2M/N$.

O número máximo de conexões em uma rede é o número de pares distintos possíveis: $M_d = N(N-1)/2$. Em geral, para redes reais, temos que $M \ll M_d$, ou seja, há bem menos conexões do que o máximo possível. Nesse caso, a matriz adjacência será do tipo esparsa, ou seja, terá grande quantidade de zeros. Da mesma forma a rede também é chamada de esparsa. Há várias aproximações que deixam os cálculos analíticos mais simples quando a rede é esparsa. Uma rede que não é esparsa é a rede completa, totalmente conectada, na qual todos os vértices tem grau $k_i = N-1$ (todos conectados com todos).

A distribuição de grau $P(k)$ indica a probabilidade de haver um vértice com grau k . Impondo sua normalização teremos que $\sum_{k=1}^{\infty} P(k) = 1$. Essa distribuição é calculada a partir de $P(k) = N_k/N$, onde N_k é o número de vértices com grau k . O grau médio, dentre outras grandezas, pode ser calculado em função desta distribuição: $\mu = \sum_{k=0}^{\infty} kP(k)$.

2.3. Coeficiente de clustering

Matematicamente, uma dada relação “o” é dita transitiva se “ $a \circ b$ ” e “ $b \circ c$ ” implicam em “ $a \circ c$ ”. O caso mais simples é a igualdade: $a = b$ e $b = c$ implicam em $a = c$. Em Ciência de Redes, essa transitividade pode ser estudada em várias propriedades entre vértices. Como a propriedade mais básica é a conexão, vale a pergunta: se (u, v) e (v, w) são conexões entre os vértices u, v e w , qual a chance de haver a conexão (u, w) ? Há formação de um triângulo fechado se as 3 conexões existirem: é

o caso já conhecido “o amigo do meu amigo também é meu amigo”. A resposta a esta questão é fornecida pelo coeficiente de aglomeração (ou coeficiente de *clustering*) e indicado por C . Esse coeficiente pode ser definido globalmente para a rede ou localmente para um dado vértice i :

$$C_i = \frac{\text{qtd. de triângulos fechados envolvendo o vértice } i}{\text{qtd. total de triângulos possíveis}}.$$

O total de triângulos possíveis inclui os triângulos abertos: quando dois vizinhos j e k de i não são conectados. Seja m_i o numerador dessa expressão enquanto o denominador é o total de pares entre k_i vértices $k_i(k_i - 1)/2$. Logo [25]:

$$C_i = \frac{2m_i}{k_i(k_i - 1)}.$$

O caso extremo $C_i = 0$ ocorre quando só houver triângulos abertos $m_i = 0$. O outro extremo é quando só há triângulos fechados formando um grafo completo (totalmente conectado), logo $C_i = 1.0$. A média de C_i para todos os vértices será o coeficiente médio de *clustering* da rede. Esse coeficiente é uma propriedade importante, com várias aplicações em modelos de rede [26, 27].

2.4. Componentes

Uma outra característica importante de redes é a sua distribuição de componentes. Uma componente é um aglomerado conectado de vértices e o seu tamanho é o número de vértices contidos nela [2]. Se uma componente tem dois ou mais vértices, isso implica que todo vértice i dentro dessa componente tem $k_i > 0$. Um vértice isolado é uma componente de tamanho 1.

Uma rede pode conter mais de uma componente: cada componente é um subconjunto independente de vértices sem qualquer conexão com outra componente. Uma rede com uma componente apenas é chamada de rede conectada. Nesse caso, é possível ir de qualquer vértice da rede para qualquer outro vértice da rede percorrendo suas conexões. A rede representada na Figura 4 tem duas componentes: uma com 5 vértices e outra com apenas 1 vértice. Logo, trata-se de uma rede desconectada. Já a rede representada na Figura 5 tem 3 componentes, enquanto a rede quadrada representada na Figura 6(a) é conectada. Quando há uma conexão que, se destruída, desconecta componentes de uma rede, essa conexão é chamada de ponte.

A matriz adjacência pode ser usada para se detectar se uma rede é conectada ou não. Os números atribuídos aos vértices podem ser distribuídos tal que em uma rede desconectada a matriz fica formada por blocos quadrados diagonais contendo elementos 1's e 0's, enquanto o resto é formado por zeros. Em redes pequenas, uma simples inspeção visual é suficiente para determinar suas componentes. Porém, em redes grandes com milhões de

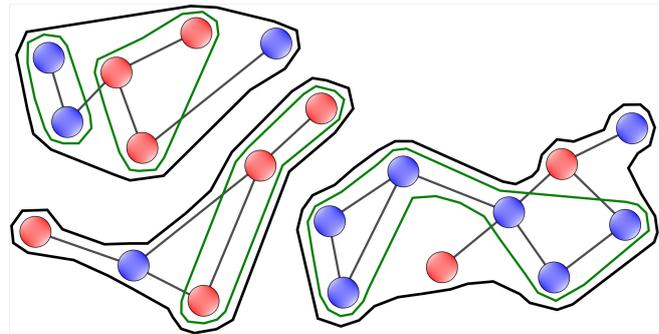


Figura 5: Ilustração de componentes e domínios de uma rede. O estado dos vértices é indicado pela cor: vermelho e azul. O contorno verde delimita um domínio enquanto o contorno preto delimita a componente. Vértice de grau 1 ligado a um vértice de cor diferente é um domínio de tamanho 1. Nesta rede temos: $N = 20$ vértices, $\mathcal{N}_c = 3$ componentes, a maior componente tem $\mathcal{S}_c = 9$ vértices, há $\mathcal{N}_d = 10$ domínios e o maior domínio (no caso com estado azul) tem $\mathcal{S}_d = 6$ vértices.

vértices é necessária a implementação computacional de algoritmos otimizados, e mesmo assim, pode demorar um tempo considerável. Um algoritmo muito utilizado é o chamado BF (do inglês *breadth-first*) [1]. A ideia é intuitiva:

1. Selecione um vértice aleatório e aplique o rótulo 1.
2. Identifique seus vizinhos e aplique o rótulo 1 a eles.
3. Identifique os vizinhos dos vizinhos e aplique novamente o rótulo 1.
4. Repita o passo 3 até atingir todos os vértices possíveis esgotando a lista de vizinhos dos vizinhos.
5. Se a quantidade de vértices rotulados for N (tamanho da rede) significa que a rede tem apenas uma componente e o processo é encerrado.
6. Caso contrário, selecione outro vértice arbitrário não rotulado e aplique o rótulo 2. Repita os passos 2 e 3.
7. Se a quantidade de vértices já rotulado considerando as duas componentes já criadas for N , significa que a rede tem apenas duas componentes.
8. Se ainda houver vértices não rotulados, selecione um deles e aplique o rótulo 3 e repita os passos 2 e 3.
9. Repita o passo 8 enquanto houver vértices não rotulados, sempre numerando a próxima componente como sendo o próximo inteiro.

Chamamos a quantidade de vértices contidos em uma componente de seu tamanho. Vamos considerar \mathcal{N}_c como o número de componentes de uma rede, e \mathcal{S}_c o tamanho da sua maior componente. Uma rede conectada tem apenas uma componente, logo: $\mathcal{N}_c = 1$ e $\mathcal{S}_c = N$. Quando analisamos as componentes de uma rede, é comum utilizar as densidades desses parâmetros de forma que redes de tamanhos diferentes possam ser

comparadas:

$$n_c = \frac{\mathcal{N}_c}{N}, \quad s_c = \frac{\mathcal{S}_c}{N}.$$

2.5. Domínios

Além das componentes, pode-se definir também os domínios: subconjunto de vértices dentro de uma componente tendo o mesmo estado ou atributo. Os estados possíveis são definidos pelo modelo físico em questão quando a formulação de rede é aplicada em algum tipo de problema. Se esse estado é representado por um número inteiro ou alguma característica discreta, é possível identificar diferentes vértices com o mesmo estado. Assim, um grupo de vértices dentro da mesma componente com o mesmo estado formam um domínio. Uma componente pode conter vários domínios. A Figura 5 mostra uma rede com 3 componentes e vários domínios, enquanto os vértices tem dois estados possíveis definidos pela cor. Por exemplo, no modelo de Axelrod de disseminação cultural [28], o estado de cada vértice define sua carga cultural representada por um vetor de F componentes, onde cada componente assume valores inteiros de 1 até Q [29]. Neste modelo, uma componente com apenas um domínio representa uma população monocultural. Vários domínios representam uma população multicultural. Já em modelos epidêmicos, os estados possíveis também são representados por categorias distintas [30]: saudável, infectado, vulnerável, imune, etc. . . , de forma que também é possível definir e calcular a distribuição de domínios.

Da mesma forma que feita para as componentes, define-se a quantidade de domínios \mathcal{N}_d e o tamanho do maior domínio \mathcal{N}_d da rede completa como os parâmetros importantes. As densidades são:

$$n_d = \frac{\mathcal{N}_d}{N}, \quad s_d = \frac{\mathcal{S}_d}{N}.$$

Vale destacar que estes dois parâmetros são calculados considerando a rede completa: analisando todas as componentes e todos os estados.

As conexões em uma rede também podem ter estados, apesar de ser menos comum [31], como está ilustrado na Figura 6(b). O estado da conexão determina a natureza dessa conexão: duas pessoas podem ter uma interação de amizade ou de inimizade.

3. Rede Aleatória

Tradicionalmente, a rede quadrada (veja Figura 6(a)) tem sido amplamente utilizada em simulações numéricas de Monte Carlo de diferentes modelos devido a sua fácil implementação [32]. Essa rede é fixa e não permite o estudo da influência da topologia nas propriedades do modelo estudado. A pergunta natural então é: como estudar a influência da topologia da rede nas propriedades de um dado modelo físico? Uma maneira de testar uma dada topologia seria criar duas redes artificiais, uma com

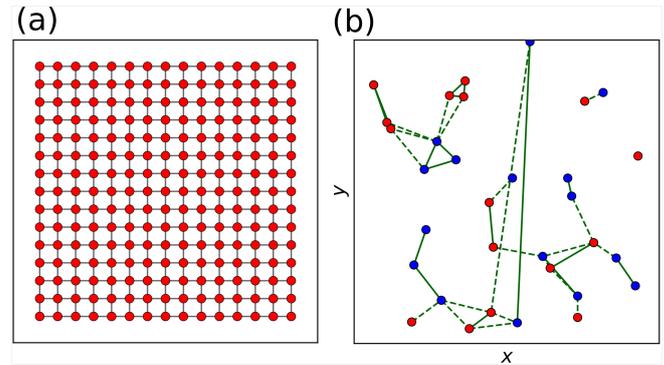


Figura 6: (a) Rede quadrada com 225 vértices. Cada vértice dos cantos tem apenas 2 vizinhos enquanto os vértices das laterais tem 3 vizinhos. Já os vértices do interior tem 4 vizinhos. (b) Rede do tipo RGG (com condições de contorno periódicas) com diferentes estados para os vértices (indicados pelas cores azul e vermelha) e para as conexões (indicados por linha sólida ou tracejada). Parâmetros: $N = 32$ vértices, raio $r = 1.0$ e $\mu = 2.38$.

e outra sem a topologia em questão, e analisar suas diferenças. Melhor ainda seria criar um grande número de redes diferentes de cada classe de topologia para avaliar estatisticamente as diferenças entre as duas classes. Isso é exatamente a motivação por trás do modelo de rede aleatória [2], que será descrito nesta seção. Neste modelo, as redes tem uma propriedade específica de interesse, definida em termos de uma probabilidade. Dessa forma, as redes geradas serão aleatórias mas seguirão uma determinada distribuição estatística. Há vários modelos diferentes de redes seguindo diferentes distribuições de probabilidades. Nesta seção descrevemos o primeiro modelo proposto deste tipo de rede, chamada rede Erdős-Rényi (ER), as vezes também chamada simplesmente de rede aleatória, como faremos aqui.

Duas instâncias de uma rede aleatória serão diferentes mas suas propriedades seguem a mesma distribuição aleatória. Por ser o modelo básico de rede aleatória, este modelo tem sido amplamente utilizado em vários sistemas diferentes [33]. Esta rede tem duas definições equivalentes, sendo a definição mais operacional a seguinte: uma rede com N vértices, onde a chance de um dado par (i, j) arbitrário de vértices ser conectado é β . N e β são os parâmetros de controle da rede $G(N, \beta)$. Tradicionalmente a criação dessa rede é creditada aos matemáticos Pál Erdős and Alfréd Rényi, os quais contribuíram enormemente no entendimento de suas propriedades, [34] apesar deles não terem sido os primeiros estudar essa rede [35]. O roteiro para criação dessa rede é simples [25]:

- Crie N vértices e escolha uma probabilidade β tal que $0.0 \leq \beta \leq 1.0$.
- Faça uma varredura em todas as duplas (i, j) distintas de vértices: $N(N - 1)/2$ no total.

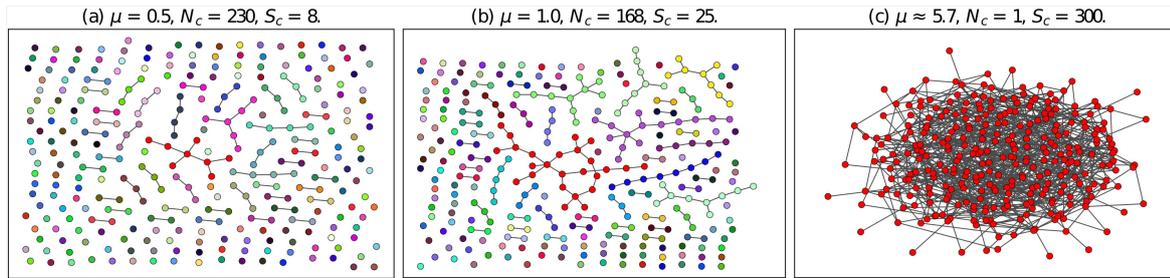


Figura 7: Rede Aleatória com $N = 300$ vértices e diferentes valores de grau médio μ . Os valores de N_c e S_c estão indicados no topo de cada gráfico. Cada cor representa uma componente distinta, enquanto a maior componente está sempre indicada pela cor vermelha. (a) $\mu = 0.5$. (b) $\mu = 1.0$. (c) $\mu = \ln N \approx 5.7$.

- Para cada dupla (i, j) sorteie um número aleatório Γ obtido de uma distribuição uniforme no intervalo $0.0 \leq \Gamma < 1.0$. Se $\Gamma < \beta$, conecte essa dupla.

Ao final, cerca de $\beta N(N - 1)/2$ pares estarão conectados.

Não é possível prever o número exato da quantidade de conexões L ou do grau da rede para cada amostra. Porém, é possível calcular analiticamente o valor médio a partir da definição $G(N, \beta)$. Há no total $N(N - 1)/2$ conexões, sendo que cada conexão liga dois vértices diferentes. Logo, o número médio de conexões será $\langle M \rangle = \beta N(N - 1)/2$. Já o grau médio será:

$$\mu = \langle k \rangle = \frac{2 \langle M \rangle}{N} = \beta(N - 1).$$

Variando o parâmetro β de 0.0 a 1.0, o número médio de conexões $\langle M \rangle$ vai de 0.0 até $N(N - 1)/2$. Já o grau médio μ vai de 0.0 até $N - 1$. O caso $\beta = 1.0$ corresponde à rede totalmente conectada. Usando a expressão anterior podemos escolher um grau médio μ e determinar o β necessário para retornar esse grau médio, de forma que pode-se usar μ como parâmetro de controle da rede ao invés de β . A vantagem é que μ tem um valor mais intuitivo e conveniente do que β . Na Figura 7 está uma representação gráfica de 3 exemplos de uma rede aleatória realçando a influência do grau médio μ .

3.1. Distribuição binomial

Para esta rede aleatória, a probabilidade de um dado vértice ter k conexões segue a distribuição binomial [1]:

$$P(k) = \binom{N-1}{k} \beta^k (1 - \beta)^{N-1-k}, \quad (4)$$

onde o coeficiente binomial é definido como:

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}.$$

Os parâmetros N e β são os parâmetros de controle da rede aleatória $G(N, \beta)$. Para uma distribuição arbitrária $P(x)$, a média (primeiro momento), o segundo momento

e o desvio padrão são definidos como:

$$\langle x \rangle = \sum_{x=0}^N xP(x), \quad \langle x^2 \rangle = \sum_{x=0}^N x^2P(x),$$

$$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}.$$

A distribuição binomial é aquela que permite calcular a chance de obtermos x caras em sequência ao longo de N lançamentos de uma moeda [36]. A média esperada e a largura do pico para a distribuição $P(k)$ da rede aleatória (Eq. 4) são:

$$\mu = \sum_{k=0}^N kP(k) = \beta(N - 1),$$

$$\sigma_k = \beta(1 - \beta)(N - 1).$$

Por esse motivo, esta rede aleatória também é chamada de rede binomial.

3.2. Distribuição de Poisson

A maioria das redes reais é do tipo esparsa, ou seja, com poucas conexões em relação ao tamanho da rede. Mais precisamente, seguem a condição $\mu \ll N$. Nesse limite a distribuição binomial da Eq. 4 pode ser aproximada por uma distribuição de Poisson:

$$P(k) = \frac{\mu^k}{k!} e^{-\mu}.$$

Repare que, agora, o grau médio μ é um parâmetro de entrada da distribuição. Logo, a distribuição de graus dos nós de uma rede aleatória sempre é descrita por uma distribuição binomial que, nos casos apontados acima, pode ser aproximada por uma distribuição de Poisson. Por esse motivo também a rede em questão as vezes é chamada de rede de Poisson. Ambas as distribuições são centradas em $\mu = \langle k \rangle$, mas a dispersão (largura da distribuição) depende de β no caso da binomial e de μ no caso da distribuição de Poisson.

Apesar da distribuição de Poisson ser apenas uma aproximação, é uma aproximação válida na maioria dos casos de interesse [1]. Além disso, o grau médio μ é um

parâmetro de entrada, e o desvio padrão é $\sqrt{\mu}$. Essas expressões são mais simples do que as análogas no caso da distribuição binomial.

3.3. Redes reais não são aleatórias

Esta rede aleatória tem sido amplamente utilizada na modelagem de sistemas em Mecânica Estatística. Porém, diversas redes reais simplesmente não seguem esse modelo de rede aleatória. Um estudo da colaboração acadêmica na área de matemática considerando publicações nos principais jornais da área entre 1991 e 1998 montou a rede resultante como tendo $N = 70975$ vértices e grau médio $\mu = 3.9$ [37]. Nessa rede, cada vértice é um pesquisador e a conexão significa que ambos os pesquisadores tem uma publicação juntos. O coeficiente médio de *clustering* da rede foi calculado em 0.59. Gerando uma rede aleatória ER com o mesmo N e μ , podemos calcular o *clustering* médio como sendo 5.4×10^{-5} , 4 ordens de grandeza menor! Além disso, a distribuição de grau da rede de colaboração também não segue a distribuição binomial nem de Poisson. Outro exemplo é a rede social Instagram, na qual as 5 contas com maior número de seguidores (pelo menos até a época do acesso) estão indicadas na Tabela 1. Repare que esses 5 primeiros tem mais de 395 milhões de seguidores mas, considerando todas as contas, 50 % tem menos de 10 mil seguidores e 26 % tem menos de mil [38]. Essas 5 contas (e várias outras nessa ponta do ranking) estão muito acima do valor médio. Isso não é compatível com o modelo de rede aleatória ER.

4. Rede Livre de Escala

Uma característica observada em várias redes reais é que sua distribuição de grau $P(k)$ segue uma lei de potência [39]:

$$P(k) = Ck^{-\gamma}, \quad (5)$$

onde C e γ são constantes. As redes que seguem esse comportamento são chamadas de redes livre de escala (do inglês *scale-free*). Essa propriedade já foi observado em várias redes reais, como redes metabólicas [40] rede de colaboração mencionada no parágrafo anterior [37],

Tabela 1: As 5 contas no Instagram com o maior número de seguidores. Fonte: https://en.wikipedia.org/wiki/List_of_most-followed_Instagram_accounts, acessado em 21 de Maio de 2024.

	Nome	# seguidores
# 1	Cristiano Ronaldo	629 milhões
# 2	Lionel Messi	502 milhões
# 3	Selena Gomez	428 milhões
# 4	Kylie Jenner	399 milhões
# 5	Dwayne “The Rock” Johnson	397 milhões

dentre outras. Obviamente, redes reais possuem tamanho finito e, portanto, não podem ser “verdadeiramente” livres de escala, devido aos efeitos de tamanho finito. Porém, essa descrição é adequada quando estamos observando fenômenos em várias ordens de grandeza.

O mecanismo mais famoso de geração de rede livre de escala é chamado de crescimento por anexação preferencial (do inglês *preferential attachment growth*). Porém, diversos outros mecanismos também geram esse tipo de rede [41], por exemplo:

- A rede é iniciada com m_0 vértices, formando uma rede totalmente conectada: $k_i = m_0 - 1$ para todo vértice i .
- Um vértice novo é adicionado a cada instante t de tempo e conectado com $m < m_0$ vértices da rede, seguindo uma certa distribuição de probabilidades $\Pi(i)$.

A chance $\Pi(i)$ do vértice novo se conectar com um dado vértice i da rede é proporcional ao grau k_i desse vértice [3, 42]:

$$\Pi(i) = \frac{k_i}{\sum_{j=1}^{D(t)} k_j},$$

onde $D(t)$ é o número total de vértices da rede no instante t . Veja que o denominador dessa expressão assegura que $0.0 < \Pi(i) < 1.0$. A adição de vértices continua até que seja obtido o tamanho N desejado para a rede. A partir dessa probabilidade, pode-se mostrar que a distribuição de grau final da rede será uma lei de potência, da forma [39]:

$$P_f(k) = \frac{2m^2}{k^3}. \quad (6)$$

Na Figura 8(a) está uma representação gráfica de uma rede livre de escala. Repare que apesar do grau médio ser 3.95 há pelo menos 5 vértices com grau maior que 20, bem maior que a média. Já a Figura 8(b) mostra a distribuição de grau obtida numericamente, simulando uma rede, e a função $P_f(k)$, e vemos uma boa concordância entre elas.

A Eq. 6 é uma lei de potência com expoente $\gamma = 3$. Redes reais modeladas como uma rede livre de escala podem ter outros valores de γ mas, no geral, é observado $\gamma > 2$ em redes reais [1].

4.1. Momentos da distribuição

O termo livre de escala é muito usado no estudo de transições de fase em Mecânica Estatística, na qual leis de potência aparecem com frequência. A origem desse termo pode ser melhor entendida com os momentos da distribuição de grau. Rigorosamente, o grau k dos vértices de uma rede é uma variável inteira entre 0 e $N - 1$, sendo N o tamanho da rede. Nesse caso sua distribuição é chamada de p_k . Porém, é comum

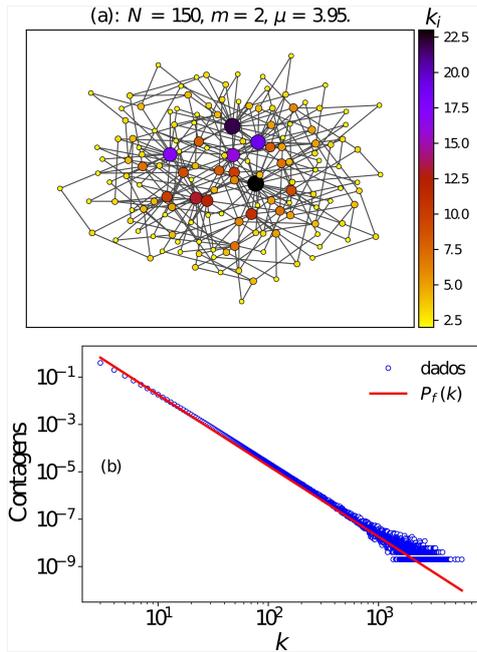


Figura 8: Rede livre de escala (scale-free). (a) Representação gráfica com $N = 150$ vértices, $m = 2$ e $\mu = 3.95$. A cor (escala lateral) e o tamanho de cada vértices são determinados pelo seu grau k_i . (b) Gráfico do histograma normalizado. A linha vermelha é a função $P_f(k)$ (Eq. 6) e os pontos em azul (indicado por dados) é a distribuição de grau obtida numericamente a partir de 500 amostras de uma rede gerada com os parâmetros $N = 10^6$ e $m = 3$.

considerar k como uma variável contínua, o que fornece resultados com boa aproximação principalmente para redes grandes. Nesse caso a distribuição é chamada de $p(k)$.

Vale relembrar o significado dos primeiros momentos de uma distribuição. O momento de ordem n de uma distribuição arbitrária $p(k)$ é definido como

$$\langle k^n \rangle = \sum_{k_m}^{\infty} k^n p_k \approx \int_{k_m}^{\infty} k^n p(k) dk,$$

onde k_m é o valor mínimo de k na distribuição. Os primeiros momentos tem as seguintes interpretações:

- $n = 1$: consiste no grau médio $\mu = \langle k \rangle$.
- $n = 2$: fornece a variância $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$ ou o desvio padrão $\sigma = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$.
- $n = 3$: é o coeficiente de assimetria, ou seja, indica o quão simétrica a distribuição é em torno da média $\langle k \rangle$.

Já considerando uma rede livre de escala, a distribuição contínua de grau é uma lei de potência dada pela Eq. 5. Assim, o momento de ordem n é dado por [1]

$$\langle k^n \rangle = \int_{k_m}^{k_M} k^n P(k) dk = C \frac{k_M^{n-\gamma+1} - k_m^{n-\gamma+1}}{n - \gamma + 1}, \quad (7)$$

onde k_M é o grau máximo observado na rede. O grau mínimo k_m em geral é fixo, mas o máximo k_M vai para infinito quando o tamanho da rede tende a infinito (diverge com N no limite termodinâmico).

Vamos analisar o comportamento dos três primeiros momentos no limite de grandes redes, fazendo $k_M \rightarrow \infty$. O valor limite de $\langle k^n \rangle$ vai depender do termo $k_M^{n-\gamma+1}$: o sinal do expoente $n - \gamma + 1$ pode fazer esse termo ir para infinito (sinal positivo) ou para zero (sinal negativo ou igual a zero).

- Se $n - \gamma + 1 \leq 0$, o termo $k_M^{n-\gamma+1}$ vai pra zero quando k_M aumenta. Logo todos os momentos que satisfazem $n \leq \gamma - 1$ são finitos.
- Se $n - \gamma + 1 > 0$ o termo $k_M^{n-\gamma+1}$ vai pra infinito quando $k_M \rightarrow \infty$. Logo todos os momentos que satisfazem $n > \gamma - 1$ divergem.

Fazendo $n = 1$ na Eq. 7, temos para o primeiro momento:

$$\langle k \rangle = C \frac{k_M^{1-\gamma+1} - k_m^{1-\gamma+1}}{1 - \gamma + 1}. \quad (8)$$

Para a maioria das redes reais é observado que $2 \leq \gamma \leq 3$ [1], logo o expoente $1 - \gamma + 1$ será negativo ou zero. Assim, de acordo com a Eq. 8 no limite $k_M \rightarrow \infty$ (redes muito grandes), o primeiro momento será finito.

Porém a situação é diferente para o segundo e terceiro momentos. Fazendo $n = 2$ e $n = 3$ na Eq. 7, temos:

$$\langle k^2 \rangle = C \frac{k_M^{2-\gamma+1} - k_m^{2-\gamma+1}}{2 - \gamma + 1},$$

$$\langle k^3 \rangle = C \frac{k_M^{3-\gamma+1} - k_m^{3-\gamma+1}}{3 - \gamma + 1}.$$

Para os expoentes, considerando ainda $2 \leq \gamma \leq 3$, temos

$$2 - \gamma + 1 \geq 0,$$

$$3 - \gamma + 1 > 0.$$

Logo, $\langle k^2 \rangle$ e $\langle k^3 \rangle$ divergem, sendo impossível determinar quão espalhado em torno da média um dado conjunto de amostras está. Essa divergência é a origem do termo “livre de escala”.

4.2. Livre de escala

Para uma distribuição gaussiana de grau k temos que $\sim 68\%$ das amostras estarão no intervalo $k = \langle k \rangle \pm \sigma$, onde σ é o desvio padrão. Já as amostras fora desse intervalo são menos prováveis porque a chance decai exponencialmente. Assim, o grau de um vértice arbitrário está dentro de uma escala. Da mesma forma em uma rede aleatória com distribuição de Poisson temos $\sigma_k = \sqrt{\langle k \rangle}$, o que também é menor que a média. Já na rede livre de escala, σ_k diverge pois $\langle k^2 \rangle$ diverge. Logo, ao observarmos o grau de um vértice arbitrário não há uma escala para a qual esperar esse valor do grau: pode ser próximo de 1 ou muito grande, da ordem de N .

Um exemplo é a rede direcionada WWW, formada pelas páginas da internet: o grau de uma página é o número de outras páginas que tem links direcionando para ela. A média dessa rede é $\langle k \rangle = 4.60$ e o expoente $\gamma = 2.1$ [1]. Assim, uma página arbitrária da internet pode ter poucos links (poucas páginas com link apontando para ela) enquanto outras tem milhões de links apontando para ela como as páginas do Google e Facebook.

5. Rede Aleatória Geométrica

Na maioria das redes, o critério espacial (posição dos vértices) não entra em sua definição. Uma rede que utiliza esse critério espacial é a chamada rede aleatória geométrica ou RGG (do inglês *random geometric graph*). Seu algoritmo de construção é [43]

1. Sejam N vértices distribuídos de maneira aleatória e uniforme em um quadrado de lado L . O raio de interação é r .
2. Dois vértices arbitrários (i, j) serão conectados se a distância euclidiana d_{ij} entre eles for menor que o raio: $d_{ij} < r$.

A Figura 9 ilustra este tipo de rede. Esta rede tem um caráter local que a rede aleatória não tem. Cada vértice só sente a interação dos vértices próximos a ele. Quanto menor o raio r menor será o grau médio μ e maior será o caráter local. O grau médio μ segue uma lei de potência em função do raio r . As componentes nessa rede mostram a distribuição espacial dos aglomerados de vértices. Redes do tipo RGG tem sido usadas em diversas aplicações como em processos cooperativos [44], modelo ZGB [45], redes com diferentes estados para as conexões [31].

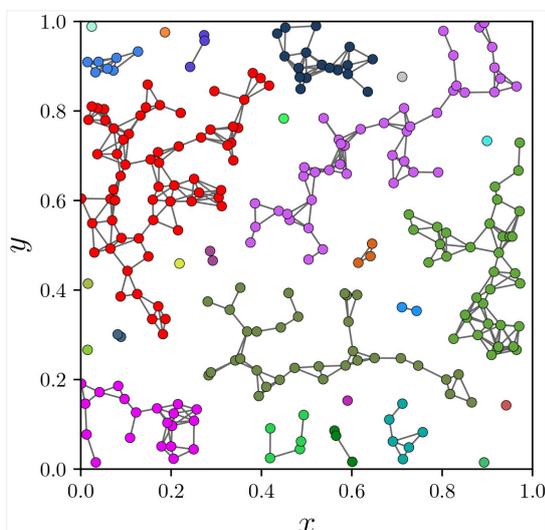


Figura 9: Rede RGG. Parâmetros: $N = 256$, $r = 0.07$, $L = 1.0$ e $\mu \approx 3.73$. Cada cor representa uma componente distinta.

6. Percolação

Diversos fenômenos e sistemas já foram estudados utilizando o formalismo de redes, como transição de fase, resiliência e robustez, sincronização, caminhada aleatória, disseminação de doenças, etc. . . [24]. Nesta seção iremos descrever de maneira resumida um desses processos: a chamada percolação [46]. Diz-se que há percolação quando alguma propriedade do sistema atravessa toda a sua dimensão. Em termos de Ciência de Redes, ocorre percolação quando a propriedade em questão atinge a maior parte dos vértices. Há várias formas e mecanismos de se descrever uma percolação e iremos adotar aqui a percolação como sendo a formação de uma componente gigante na rede.

6.1. Percolação na rede aleatória

Vamos analisar a percolação da rede aleatória $G(N, \beta)$ ou de forma equivalente $G(N, \mu)$. A distribuição das componentes varia com o parâmetro β :

- $\beta = 0.0$: a rede é totalmente desconectada tal que $\mu = 0.0$. Há N componentes ($\mathcal{N}_c = N$ e $n_c = N/N = 1$) e todas as componentes contém apenas um vértice ($\mathcal{S}_c = 1$ e $s_c = 1/N$). No limite de uma rede muito grande $N \rightarrow \infty$, teremos $s_c \rightarrow 0$ (chamado também de componente microscópica).
- $\beta = 1.0$: a rede é totalmente conectada tal que $\mu = N - 1$. Haverá uma componente apenas ($\mathcal{N}_c = 1$ e $n_c = 1/N$) englobando todos os vértices ($\mathcal{S}_c = N$ e $s_c = N/N = 1$). No limite de uma rede muito grande $N \rightarrow \infty$, teremos $n_c \rightarrow 0$ (componente macroscópica).

Vamos nos concentrar na densidade s_c . No limite de uma rede infinita $N \rightarrow \infty$, variando β de 0 a 1 (ou μ de 0 até $N - 1$), teremos que a componente sai de $s_c = 0$ até $s_c = 1$. Alguém poderia pensar que s_c aumentaria gradualmente com β . Porém, simulando um sistema grande o suficiente e obtendo-se s_c numericamente (veja Figura 10), observa-se que s_c se mantém

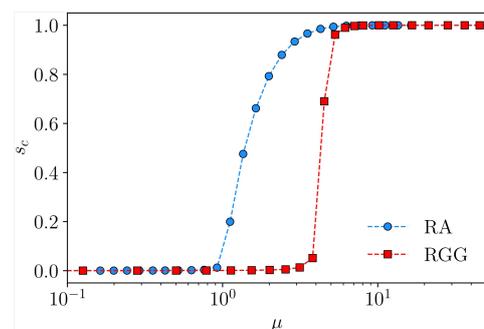


Figura 10: Variação da densidade da maior componente s_c em função do grau médio μ para a rede aleatória (círculo azul) e RGG (quadrado vermelho). Ambas as redes contém $N = 128^2$ vértices enquanto $L = \sqrt{N}$ para a RGG.

pequeno (aproximadamente zero quanto maior for a rede) até um valor crítico de β no qual $\mu \sim 1.0$ [34]. A partir desse valor, ocorre um rápido aumento de s_c , significando o aparecimento de uma componente gigante na rede, cuja quantidade de vértices é da ordem de N . Isso se diferencia dos casos em que, tipicamente, as componentes tem $\sim \log N$ vértices.

Repare que para a formação da componente gigante, é necessária pelo menos uma conexão por vértice na média. Mas e se o grau médio for exatamente 1.0, haverá a componente gigante?

6.2. Fases

Ao longo dessa transição de fase, a rede apresenta as duas fases, antes e depois de percolada, além do comportamento crítico na transição [1]:

- Fase não percolada $\mu < 1.0$: é o regime sub-crítico, mostrado na Figura 7(a). Há pequenas componentes, de forma que ainda temos $s_c \approx 0$ para redes grandes. Mais especificamente, temos que o tamanho da maior componente é $\mathcal{N}_c \sim \ln N$, que cresce bem mais lentamente que N . Assim, para redes grandes, \mathcal{N}_c/N tende a zero (componente microscópica mencionada anteriormente).
- Regime crítico $\mu \sim 1.0$: separa as duas fases (veja Figura 10). Nesta região temos $\beta = \mu/(N-1) = 1/(N-1)$ enquanto o tamanho da maior componente é $\mathcal{N}_c \sim N^{2/3}$, o que também cresce mais lentamente que N . Nessa vizinhança tem-se $\mathcal{N}_c \sim (\beta - \beta_c)N$, comportamento também observado perto da transição de fase de outros modelos. A maior componente ainda tende para zero para redes grandes mas a diferença de comportamento em relação a fase não percolada é evidente (veja Figura 7(b)). Pode-se ver isso considerando uma rede aleatória com $N = 10^{10}$ vértices. Na fase não percolada, a maior componente é $\mathcal{N}_c \sim \ln 10^{10} \approx 23$, enquanto que no regime crítico $\mathcal{N}_c \sim (10^{10})^{2/3} \approx 4.6 \times 10^6$, uma diferença de 5 ordens de magnitude! Além disso, \mathcal{N}_c segue uma lei de potência.
- Fase percolada $\mu > 1.0$: regime super-crítico, mostrado na Figura 7(c). É a fase mais relevante em termos de descrição de redes reais, já que há uma clara componente gigante. A partir de $\mu \sim \ln N$, temos $\mathcal{S}_c \sim N$, de forma que a rede se torna conectada (contém uma componente apenas). Veja que, mesmo assim, a rede ainda pode ser esparsa, já que pode haver muitos vértices com uma conexão apenas. A rede se torna completa (totalmente conectada) apenas quando $\mu = N - 1$.

No caso de redes geométricas (por exemplo RGG e rede quadrada), o aparecimento da componente gigante está relacionado com a formação de um caminho conectando um vértice em um extremo da rede até o outro extremo, o que caracteriza o aparecimento do fenômeno

de percolação (veja Figura 10). Mas diferente da rede aleatória, o valor crítico de μ para a RGG irá depender da densidade de vértices no quadrado N/L^2 .

7. Conclusões

Ciência de Redes é um campo muito extenso, com inúmeros conceitos, processos e aplicações fascinantes. Neste artigo, apenas arranhamos a superfície de uma introdução à área, elencando os conceitos iniciais. Apresentamos também alguns exemplos de modelos importantes de redes, além de um processo básico (percolação). Mas há vários outros modelos e vários outros processos importantes e bastante estudados nas últimas décadas.

Neste trabalho, a análise numérica e os gráficos foram feitos em ambiente Python. Os pacotes utilizados foram: Numpy para cálculo numérico [47], Matplotlib para geração dos gráficos [48], Pandas para análise de dados [49] e NetworkX para análise de redes [50]. Com exceção da Figura 4, que foi gerada com a plataforma Gephi (<https://gephi.org/>). Todos os dados e códigos estão disponíveis mediante solicitação.

Agradecimentos

P.F. Gomes agradece o auxílio financeiro da Fapeg (www.fapeg.go.gov.br/) e do CNPq (www.cnpq.br, processo 405508/2021-2). Esta pesquisa foi apoiada também pelo LaMCAD/UFG.

Referências

- [1] A.L. Barabási, *Network Science* (Cambridge University Press, Glasgow, 2016).
- [2] M.E.J. Newman, *Networks: an introduction* (Oxford University Press, Oxford, 2010).
- [3] R. Albert e A.L. Barabási, *Review of Modern Physics* **74**, 47 (2002).
- [4] L.F. Costa, *What is a Complex Network? (CDT-2)*, disponível em: <http://dx.doi.org/10.13140/RG.2.2.1.0450.04804/2>.
- [5] A.S. Mata, *Brazilian Journal of Physics* **50**, 658 (2020).
- [6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez e D.U. Hwang, *Physics Reports* **424**, 175 (2006).
- [7] C. Liu, Y. Ma, J. Zhao, R. Nussinov, Y.C. Zhang, F. Cheng e Z. Ke Zhang, *Physics Reports* **846**, 1 (2020).
- [8] R. Pazzini, O. Kinouchi e A.A. Costa, *Physical Review E* **104**, 014137 (2021).
- [9] M.A. Ribeiro, R.A. Vosgerau, M.L.P. Andruchiw e S.E.S. Pinto, *Revista Brasileira de Ensino de Física* **38**, 1304 (2016).
- [10] C. Castellano, S. Fortunato e V. Loreto, *Reviews of Modern Physics* **81**, 591 (2009).
- [11] D.J. Watts e S.H. Strogatz, *Nature* **393**, 440 (1998).
- [12] Y.L. Silva e A.A. Costa, *European Physical Journal B* **97**, 105 (2024).
- [13] S. Milgram, *Psychology Today* **2**, 60 (1967).

- [14] I. Samoylenko, D. Aleja, E. Primo, K. Alfaro-Bittner, E. Vasilyeva, K. Kovalenko, D. Musatov, A.M. Raigorodskii, R. Criado, M. Romance et al, *Physical Review X* **13**, 021032 (2023).
- [15] G.S. Domingues, F.N. Silva, C.H. Comin e L.F. Costa, *Journal of Statistical Mechanics* **8**, 083212 (2018).
- [16] G. Corso, C.R. Fossa e G.B. Oliveira, *Revista Brasileira de Ensino de Física* **27**, 389 (2005).
- [17] B.C. Souza, F.N. Silva, H.F. Arruda, G.D. Silva, L.F. Costa e D.R. Amancio, *Information Sciences* **641**, 119124 (2023).
- [18] J.L. Moreno, *Who shall survive?* (Nervous and Mental Disease Publishing, Chicago, 1934).
- [19] R.V. Solé, M. Rosas-Casals, B. Corominas-Murtra e S. Valverde, *Physical Review E* **77**, 026102 (2008).
- [20] L. Euler, *Commentarii Academiae Scientiarum Imperialis Petropolitanae* **8**, 128 (1741).
- [21] G. Alexanderson, *Bulletin of the American Mathematical Society* **43**, 567 (2006).
- [22] L.F. Costa, F.A. Rodrigues, G. Travieso e P.R. Villas Boas, *Advances in Physics* **56**, 167 (2007).
- [23] S.M. Reia, P.F. Gomes e J.F. Fontanari, *The European Physical Journal B* **92**, 205 (2019).
- [24] A. Barrat, M. Barthélemy e A. Vespignani *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
- [25] R. Oliveira, L. Sanz e R. Chaves, *Revista Brasileira de Ensino de Física* **44**, e20220189 (2022).
- [26] M.E.J. Newman e J. Park, *Physical Review E* **68**, 036122 (2003).
- [27] S.N. Soffer e A. Vázquez, *Physical Review E* **71**, 057101 (2005).
- [28] S.M. Reia, *International Journal of Modern Physics* **31**, 2050116 (2020).
- [29] S.M. Reia, P.F. Gomes e J.F. Fontanari, *Journal of Statistical Mechanics* **2020**, 033402 (2020).
- [30] H.W. Hethcote, *SIAM Review* **42**, 599 (2000).
- [31] P.F. Gomes, H.A. Fernandes e A.A. Costa, *Physica A* **597**, 127269 (2022).
- [32] J. Marro e R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, Cambridge, 1999).
- [33] L.E.C. Rocha, *Revista Brasileira de Ensino de Física* **31**, 3303 (2009).
- [34] P. Erdős e A. Rényi, *Publicationes Mathematicae* **6**, 290 (1959).
- [35] R. Solomonoff e A. Rapoport, *Bulletin of Mathematical Biophysics* **13**, 107 (1951).
- [36] M. Novaes, *Revista Brasileira de Ensino de Física* **44**, e20210424 (2022).
- [37] A.L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert e T. Vicsek, *Physica A* **311**, 590 (2002).
- [38] Instagram Engagement Report 2023, disponível em: <https://info.mention.com/en/instagram-engagement-report-2023/thank-you>.
- [39] A.L. Barabási e R. Albert, *Science* **286**, 509 (1999).
- [40] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai e A.L. Barabási, *Nature* **407**, 651 (2000).
- [41] S.N. Dorogovtsev e J.F.F. Mendes, *Advances in Physics* **51**, 1079 (2002).
- [42] R.A. Zara e E. Koehler, *Revista Brasileira de Ensino de Física* **26**, 53 (2004).
- [43] J. Dall e M. Christensen, *Physical Review E* **66**, 016121 (2002).
- [44] P.F. Gomes, S.M. Reia, F.A. Rodrigues e J.F. Fontanari, *Physical Review E* **99**, 032301 (2019).
- [45] E.B. Vilela, H.A. Fernandes, F.L.P. Costa e P.F. Gomes, *Journal of Computational Chemistry* **41**, 1964 (2020).
- [46] S.N. Dorogovtsev, A.V. Goltsev e J.F.F. Mendes, *Review on Modern Physics* **80**, 1275 (2008).
- [47] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith et al, *Nature* **585**, 357 (2020).
- [48] J.D. Hunter, *Computing in Science & Engineering* **9**, 90 (2007).
- [49] W. McKinney, em: *Proceedings of the 9th Python in Science Conference* (Texas, 2010).
- [50] A.A. Hagberg, D.A. Schult e P.J. Swart, em: *Proceedings of the 7th Python in Science Conference (SciPy2008)* (Pasadena, 2008).