

2 Redes complexas

Conforme relatado na introdução, o interesse pela teoria de redes complexas aumentou consideravelmente ao longo dos últimos anos, graças à sua capacidade de explorar e oferecer novos horizontes no estudo de sistemas reais. Embora suas origens remontem à teoria de grafos (Séc. XVIII), foi somente a partir do final da década de 1990 que a área de redes complexas se consolidou, graças em grande parte aos trabalhos de Watts e Strogatz (37), e Barabási e Albert (38), onde no primeiro é proposto um modelo teórico de rede que ajuda a compreender o paradigma de “pequeno-mundo”, observado por exemplo, em redes sociais (estabelecidas através de relações entre pessoas e seus conhecidos), fenômeno já observado na década de 1960 pelo estudo sociológico de Milgram (39), de onde surgiu o conceito dos “seis graus de separação”; e no segundo, também através da proposição de um modelo, os autores apresentam uma explicação plausível para o surgimento da alta heterogeneidade observada em muitos sistemas, como por exemplo a Internet (13), onde a interconectividade entre os elementos da rede (roteadores ou domínios) obedece uma lei de potências, onde enquanto a maioria dos elementos conectam-se com poucos outros, há também certos elementos que estabelecem conexões com uma grande quantidade dos demais. Tais trabalhos representaram o pontapé inicial para uma série de estudos relacionados tanto à caracterização da topologia, ou seja, à estrutura de sistemas complexos, quanto ao seu comportamento dinâmico, com base em processos se desenvolvendo espaço-temporalmente nos mesmos. Em redes complexas, a dinâmica é fortemente influenciada pela topologia das mesmas, uma vez que a evolução da primeira depende da interconectividade dos elementos que formam o sistema. A caracterização temporal de processos dinâmicos, por sua vez, pode oferecer um retrato da estrutura dos sistemas. Nesta seção, vamos nos concentrar na *estrutura* das redes, onde serão apresentadas as definições e representação das mesmas, bem como a caracterização através de medidas, e alguns modelos teóricos propostos, de maneira a reproduzir características estruturais observadas em sistemas reais. No Capítulo 3 a seguir introduziremos o conceito de dinâmicas, apresentando alguns modelos que serão empregados nesta tese.

2.1 Definições básicas e representação

Redes complexas são representações de sistemas sob a forma de *grafos*. Uma *rede complexa* $\mathcal{G}(\mathcal{V}, \mathcal{E})$ é formada por um conjunto $\mathcal{V} = \{1, \dots, N\}$ de N *nós* ou *vértices* – que constituem os elementos individuais do sistema – e um conjunto $\mathcal{E} = \{\{i, j\}; i, j \in \mathcal{V}, i \neq j\}$ de ℓ *arestas* ou *conexões*. A aresta $\{i, j\}$ indica que há uma interação entre os nós i e j , que pode ser recíproca ou não. Caso tal interação não seja recíproca, ou seja, a conectividade se faz presente apenas do nó i para o nó j , substituímos a expressão $\{i, j\}$ pelo par ordenado (i, j) , e dizemos que a rede é *direcionada* ou *dirigida*. Note que nesta situação a reciprocidade somente ocorre quando ambas as conexões (i, j) e (j, i) estão definidas em \mathcal{E} . É possível representar uma rede complexa através de uma matriz binária $N \times N$ $\mathbf{A}(\mathcal{G})$, denominada *matriz de adjacência* de \mathcal{G} , onde $a_{i,j} = 1$ se uma aresta conecta o nó i ao nó j , e $a_{i,j} = 0$ caso contrário. Tal matriz é simétrica para redes não-direcionadas. A rede também pode ter grandezas numéricas associadas às suas conexões, denominadas na literatura de grafos de *custos* ou *pesos*. Neste caso, podemos definir uma matriz $\mathbf{W}(\mathcal{G})$ como a *matriz de pesos* de \mathcal{G} , inteira ou real, onde $w_{i,j} = w(i, j)$, peso associado à aresta (i, j) . A Figura 2.1 ilustra os conceitos de direcionamento, peso e representação matricial de redes complexas.

2.2 Caracterização através de medidas

A *caracterização topológica* de redes complexas através de diversas *medidas* – algumas importadas da teoria de grafos, outras propostas pela própria linha de pesquisa em redes – permite uma melhor compreensão dos sistemas modelados pelas mesmas. A cardinalidade N do conjunto de nós \mathcal{V} , e a cardinalidade do conjunto de arestas \mathcal{E} – que denotaremos por ℓ – ou seja, respectivamente o número de nós e arestas de uma rede $\mathcal{G}(\mathcal{V}, \mathcal{E})$ – constituem dois exemplos de tais medidas. Porém, dada a diversidade de aspectos locais e globais que tais sistemas podem incorporar, tais grandezas são insuficientes para uma descrição detalhada da estrutura dos mesmos. De fato, ao longo dos últimos anos, uma infinidade de medidas foi proposta, de maneira a classificar, analisar e explicar a topologia de diversos sistemas. Uma revisão de muitas destas medidas pode ser encontrada em (4). Algumas importantes medidas de rede, que serão consideradas neste texto, serão apresentadas a

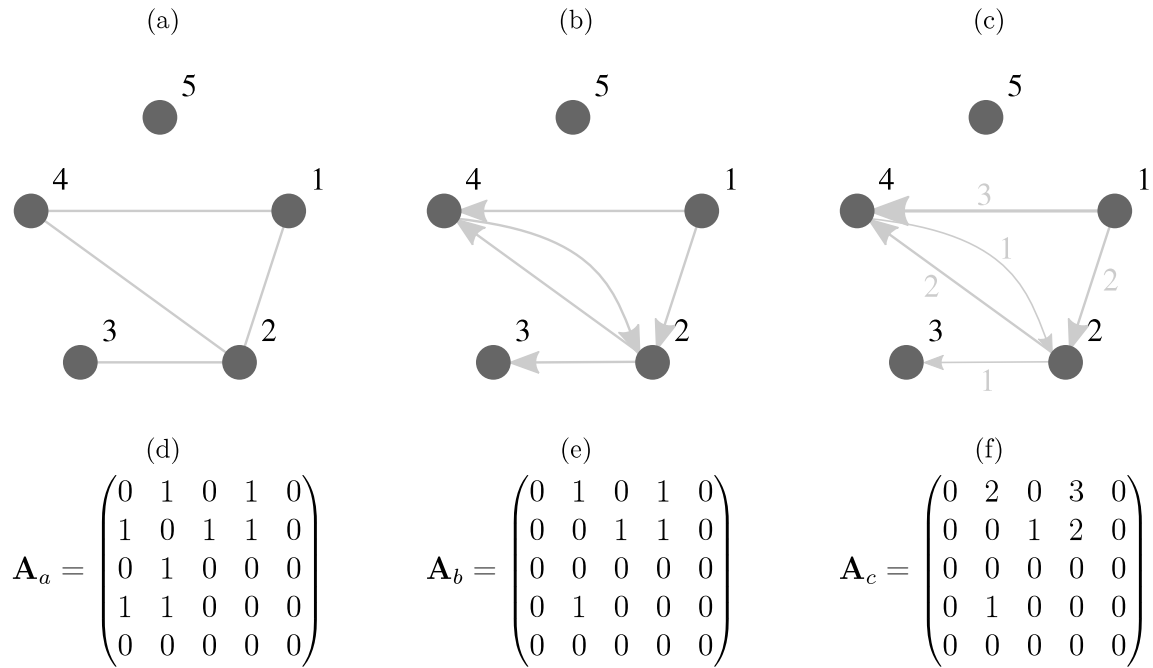


Figura 2.1 – Exemplos de redes complexas: uma rede não-direcionada e sem peso (a), uma rede direcionada e sem peso (b), e uma rede direcionada e com peso – representado pela espessura das arestas – (c), com suas respectivas matrizes de adjacência \mathbf{A}_a (d), \mathbf{A}_b (e); e de peso \mathbf{A}_c (f). \mathbf{A}_a é simétrica, por representar a conectividade de uma rede não-direcionada. O mesmo não ocorre em relação a \mathbf{A}_b , onde a reciprocidade de conexões ocorre somente entre os nós 2 e 4. Note pela matriz de pesos \mathbf{A}_c que tal reciprocidade pode não ocorrer também em relação aos pesos.

seguir.

2.2.1 Grau dos nós e grau médio da rede

Consideremos um nó específico i de uma dada rede *não-direcionada* $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Ao definir $\mathcal{N}(i)$ como o conjunto dos nós diretamente conectados a i – denominado *vizinhança* do nó i – temos que o valor $k(i) \equiv |\mathcal{N}(i)|$ é o **grau** do nó i . O grau do nó também pode ser dado pela matriz de adjacência $\mathbf{A}(\mathcal{G})$ por meio da expressão $k(i) = \sum_j a_{i,j}$. Tal nó é dito *isolado* se $k(i) = 0$ ($\mathcal{N}(i) = \emptyset$). Na Figura 2.1(a), por exemplo, o nó 5 é isolado ($k(5) = 0$), ao passo que os graus dos nós 1 a 4 são dados por 2, 3, 1 e 2, respectivamente. O **grau médio da rede** nada mais é do que o valor médio do grau de todos os nós que compõem a mesma:

$$\langle k \rangle = \frac{1}{N} \sum_i k(i).$$

Determinar o grau médio através da expressão acima é custoso, especialmente para valores elevados de N . Porém, o valor $\langle k \rangle$ pode ser diretamente obtido a partir de N e ℓ através da equação

$$\langle k \rangle = 2 \frac{\ell}{N}, \quad (2.1)$$

onde o valor 2 na expressão decorre do fato de que cada aresta está conectada a dois nós. Sendo N o número de nós da rede, o maior número possível de arestas é $\ell_{max} = N(N-1)/2$.

No caso de *redes direcionadas*, podemos dividir a vizinhança $\mathcal{N}(i)$ em dois subconjuntos disjuntos $\mathcal{P}(i)$ e $\mathcal{S}(i)$, que contêm respectivamente os *predecessores* e *sucessores* de i . Os predecessores de i são os nós de \mathcal{G} que possuem arestas direcionadas a i , enquanto os sucessores de i são os nós onde chegam arestas direcionadas que partem de i . Definimos $k_{in}(i) = |\mathcal{P}(i)|$ e $k_{out}(i) = |\mathcal{S}(i)|$ como respectivamente o **grau de entrada** e o **grau de saída** do nó i . Por exemplo, na Fig. 2.1, itens (b) e (c), o grau de entrada do nó 4 é 2, ao passo que o grau de saída do mesmo é 1. Para o grau 1 o grau de entrada é 0, ao passo que para o nó 3 é o grau de saída que se anula. De modo análogo, podemos estimar o grau de entrada médio e o grau de saída médio da rede,

$$\langle k_{in} \rangle = \langle k_{out} \rangle = \frac{\ell}{N}. \quad (2.2)$$

Note a diferença entre a expressão acima e a Eq. 2.1: na anterior, o máximo número possível de arestas é $N(N-1)/2$ ao passo que para redes direcionadas, tal valor é o dobro, devido à possibilidade de definição de duas arestas distintas envolvendo o mesmo par de nós, sendo uma em cada direção. Ainda, $\langle k_{in} \rangle$ e $\langle k_{out} \rangle$ assumem o mesmo valor, dado que o mesmo número de nós e conexões estão envolvidos.

Densidade de arestas

A **densidade de arestas** é uma medida alternativa ao grau médio da rede, dada pela normalização do mesmo,

$$\delta = 2 \frac{\ell}{N(N-1)}, \quad (2.3)$$

isto é, $\delta = \ell/\ell_{max} \in [0, 1]$. A densidade é máxima quando cada nó da rede está conectado a todos os demais. A densidade de arestas da expressão anterior é válida para redes não-dirigidas, ao passo que para redes dirigidas, o valor é dado por $\delta = \ell/[N(N-1)]$.

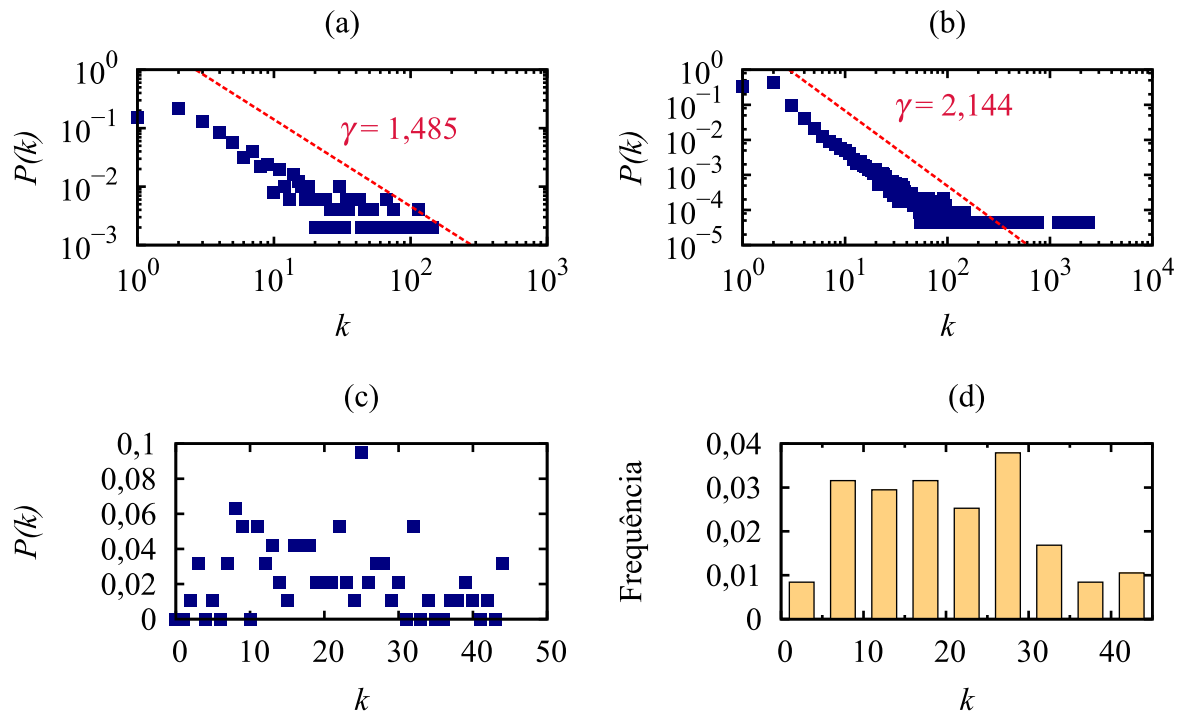


Figura 2.2 – Distribuição de grau para algumas redes reais: (a) Rede de aeroportos (40, 41) – $N = 500$, $\langle k \rangle = 11,92$ – com os 500 aeroportos mais movimentados dos Estados Unidos, e respectivas conexões; (b) Estrutura da Internet em nível de sistemas autônomos (42, 43) – $N = 22963$, $\langle k \rangle = 4,219$; (c) e (d) Sistema tálamo-cortical do gato (28, 32, 33, 44) – $N = 95$, $\langle k \rangle = 20,126$. Nos casos (a) e (b), notamos que a distribuição de grau segue uma *lei de potências*, caracterizada pelo decaimento linear de $P(k)$ em função de k observado nos gráficos logarítmicos: $P(k) \sim k^{-\gamma}$, onde o valor γ é dado nos gráficos. Tal comportamento será revisto a seguir, na Seção 2.3.3. Para o sistema tálamo-cortical do gato, vemos que o comportamento é diferente, entretanto – item (c). $P(k)$ não segue um padrão ao longo dos valores de k , e uma representação alternativa que ilustra com mais clareza o comportamento da distribuição na rede envolve o uso de histograma – item (d) – onde as barras do mesmo estão normalizadas de maneira que a área total seja 1.

2.2.2 Distribuição de graus

A distribuição de graus $P(k)$ de uma rede \mathcal{G} nos fornece o panorama de sua interconectividade, o que não pode ser obtido através de medidas escalares como o grau médio ou a densidade de arestas. Pode ser expressa na forma de uma função ou de um conjunto discreto de valores (histograma de frequências). Adiante, apresentaremos a distribuição para modelos teóricos de rede. A Fig. 2.2 ilustra exemplos de distribuições para diferentes sistemas reais, incluindo o sistema tálamo-cortical do gato, que será tratado na Seção 4.1. A definição pode ser estendida para redes dirigidas, onde podemos observar o

comportamento tanto da distribuição de graus de entrada quanto de graus de saída.

2.2.3 Entropia de grau

Uma maneira de complementar a informação do grau médio da rede é calculando a *entropia de grau* (45–47) da mesma, dada pela expressão

$$h = - \sum_{k=0}^{N-1} p(k) \log [p(k)] , \quad (2.4)$$

onde $p(k) \equiv |\{i \in \mathcal{V}; k(i) = k\}|/N$ é a probabilidade de ocorrência do grau k , e é tal que $\sum_k p(k) = 1$. A entropia é nula em *redes regulares*, onde todos os nós tem o mesmo grau, ao passo que atinge seu valor máximo $h = \log(N)$ para redes com distribuição uniforme dos nós ($p(k) = 1/N, \forall k$). A entropia é uma medida global da heterogeneidade presente na conectividade do sistema, e está associada à resistência da rede a falhas aleatórias em seus elementos (47).

2.2.4 Correlações de grau e assortatividade

Vimos que a entropia de grau fornece uma medida de heterogeneidade da rede, ao passo que a distribuição de grau nos permite visualizar tal heterogeneidade. Podemos ir além e observar *localmente* como a conectividade se comporta. Por exemplo: para um dado sistema, será que um determinado nó se conectará a outros aleatoriamente, isto é, desprezando características dos mesmos? Embora pareça plausível que um nó possa se conectar a outro independente de características do segundo, isto não ocorre na maioria dos casos reais: é comum encontrar em tais sistemas exemplos de *padrões de conectividade* que se estabelecem pela preferência de nós de se conectarem com outros nós de características similares ou mesmo opostas. Várias características dos nós podem ser consideradas (48), entretanto as atenções são voltadas ao grau dos mesmos (49): é observado em redes reais a tendência de nós com alto ou baixo grau de estabelecerem conexões com nós de grau similar (fenômeno chamado de *assortatividade*) ou então o caso contrário, caracterizado pela alta

frequência de conexões ligando nós com alta e baixa conectividade (*disassortatividade*).

Newman (50) propõe uma maneira de estimar quantitativamente o quão uma rede é (dis)assortativa, através do coeficiente de correlação

$$r = \frac{\ell^{-1} \sum_{(i,j)} k_i k_j - \left[\ell^{-1} \sum_{(i,j)} \left(\frac{k_i + k_j}{2} \right) \right]^2}{\ell^{-1} \sum_{(i,j)} \left(\frac{k_i^2 + k_j^2}{2} \right) - \left[\ell^{-1} \sum_{(i,j)} \left(\frac{k_i + k_j}{2} \right) \right]^2}, \quad (2.5)$$

onde k_i e k_j denotam o grau dos nós relativos a uma conexão $\{i, j\}$ definida na rede. Em redes *assortativas* ($r > 0$), nós com grau de mesma ordem conectam-se entre si, ao passo que em redes *disassortativas* ($r < 0$), nós com grau baixo estabelecem conexões com nós que possuem alta conectividade. Sabe-se da literatura (48, 50) que redes sociais de maneira geral são assortativas, ao passo que diversos sistemas tecnológicos e redes biológicas, como a rede neuronal do nemátodo *C. elegans* (51), por exemplo, são disassortativas. O coeficiente r é estimado em termos do comportamento observado nas conexões da rede. Uma maneira alternativa de analisar a assortatividade é através do *grau médio dos vizinhos mais próximos*

$$k_{nn}(i) = \frac{1}{k(i)} \sum_{j \in \mathcal{N}(i)} k(j).$$

Ao se calcular o valor médio de $k_{nn}(i)$ para todos os N_k nós com grau k da rede, temos

$$\begin{aligned} K_{nn}(k) &= \frac{1}{N_k} \sum_{i; k(i)=k} k_{nn}(i) \\ &\equiv \sum_{k'} k' P(k'/k), \end{aligned} \quad (2.6)$$

onde a expressão inferior relaciona o valor de $K_{nn}(k)$ com as correlações existentes entre os graus de nós conectados. Na ausência de correlações, $P(k'/k)$ somente depende de k' e portanto $K_{nn}(k)$ é constante. Em relação ao grau médio dos vizinhos, a rede é assortativa se $K_{nn}(k)$ é função crescente de k , e disassortativa se a função decresce. A Fig. 2.3 ilustra exemplos de redes assortativas e disassortativas.

Os conceitos aqui apresentados se estendem para redes dirigidas, onde se pode considerar o grau de entrada e/ou de saída para analisar os padrões de conectividade.

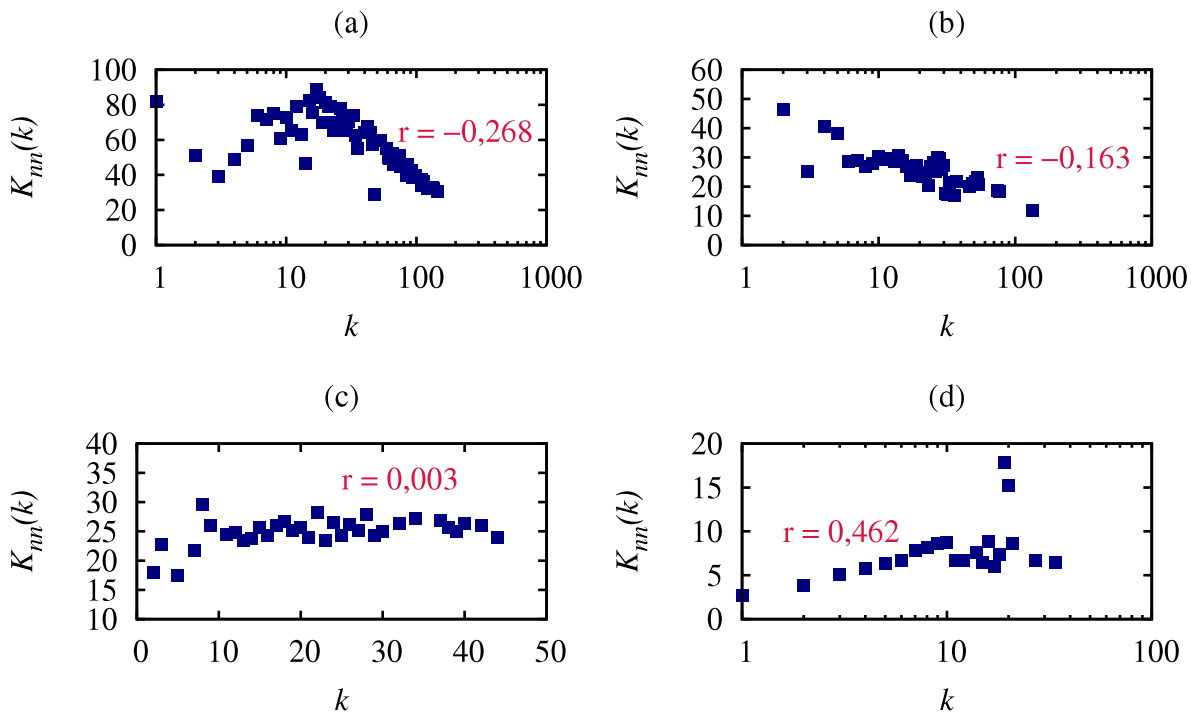


Figura 2.3 – Correlações de grau e assortatividade em redes reais: (a) rede de aeroportos (40, 41); (b) sistema neuronal do nemátodo *C. elegans* (51), $N = 297$, $\langle k \rangle = 14,465$; (c) sistema tálamo-cortical do gato (28); e (d) Rede de coautores em artigos em redes complexas (52), $N = 1589$, $\langle k \rangle = 3,451$. r corresponde ao valor do coeficiente de assortatividade dado pela Eq. 2.5 – que *não* se iguala à correlação dos dados nos gráficos, sendo de (a) a (d) iguais a $-0,786$, $-0,487$, $0,445$ e $0,397$, respectivamente. Ao passo que a rede de colaborações (d) é assortativa, caso típico de redes sociais, as demais – com exceção da rede cortical do gato, exemplo de rede aparentemente livre de correlações – são disassortativas como geralmente ocorre em redes biológicas e tecnológicas (48).

2.2.5 Simetria

Em redes dirigidas, pode ser interessante analisar qual o grau de simetria das conexões que definem a mesma. Embora uma simples medida da simetria da rede (53) seja dada por

$$S = \frac{\ell^{\leftrightarrow}}{\ell},$$

isto é, a razão de arestas bidirecionais ℓ^{\leftrightarrow} sobre o número total de conexões ℓ . Em uma rede dirigida puramente unidirecional, como é o caso da rede de citações de artigos (2) – onde um artigo só pode citar outros que já tenham sido publicados – temos que $S = 0$, ao passo que no outro extremo, em redes dirigidas puramente bidirecionais (como na Internet), $S = 1$. Sistemas reais, entretanto, apresentam quase que exclusivamente

valores intermediários de S . A medida S apresenta certas limitações (54), como a falta de comparação com o valor S_{rand} esperado em uma rede aleatória do mesmo tamanho e com o mesmo grau médio. S_{rand} será maior em redes aleatórias com uma densidade de arestas maior, visto que uma maior quantidade de arestas recíprocas será formada ao acaso, o que impede o uso de S na comparação de duas redes com diferentes níveis de conectividade. Garlaschelli e Loffredo (54) propuseram uma medida alternativa,

$$g_S = \frac{\sum_{i \neq j} (a_{ij} - \delta)(a_{ji} - \delta)}{\sum_{i \neq j} (a_{ij} - \delta)^2} \equiv \frac{S - \delta}{1 - \delta}, \quad (2.7)$$

onde $\delta = \sum_{i \neq j} a_{ij} / [N(N - 1)] = \ell / [N(N - 1)]$ é a densidade de arestas para redes dirigidas, conforme visto na Seção 2.2.1. A proposição é baseada no fato de que $S_{rand} \equiv \delta$, porque em uma rede aleatória e livre de correlações de grau a chance de se encontrar uma aresta bidirecional entre dois nós é a mesma de se encontrar uma aresta dirigida entre dois vértices quaisquer, δ . A medida g_S , daqui em diante referida como *índice de simetria de Garlaschelli e Loffredo*, corresponde a um coeficiente de correlação, a exemplo do coeficiente de assortatividade visto na seção anterior: $g_S > 0$ (< 0) denotam redes *recíprocas* (*antirrecíprocas*), onde arestas bidirecionais ocorrem com frequência maior (menor) que a esperada no caso aleatório, onde $g_S = 0$. Na Seção 4.3, veremos que redes formadas a partir de neurônios com formas diferentes geram padrões distintos na reciprocidade nas conexões.

2.2.6 Força

Em redes com peso, podemos definir uma outra medida, similar ao grau, denominada **força** (*strength*) do nó i , dada pela soma dos pesos das arestas adjacentes ao mesmo, isto é,

$$s(i) = \sum_j w_{i,j}. \quad (2.8)$$

De maneira análoga ao visto para a medida de grau, é possível definir, para redes direcionadas, as forças de entrada e saída, $s_{in}(i)$ e $s_{out}(i)$, que não serão objeto de estudo neste texto.

2.2.7 Coeficiente de agrupamento

Dado um nó qualquer i de uma rede $\mathcal{G}(\mathcal{V}, \mathcal{E})$, e sua vizinhança $\mathcal{N}(i)$, é possível definir um *subgrafo* $\mathcal{H}(\mathcal{N}(i), \mathcal{E}_i)$ de \mathcal{G} , formado pelos vizinhos de i e as conexões entre os mesmos – representadas pelo conjunto \mathcal{E}_i . Sendo $\ell_i \equiv |\mathcal{E}_i|$ o número de arestas de tal subgrafo, temos que o *coeficiente de agrupamento* do nó i é dado pela *densidade de arestas* – cf. Eq. (2.3) – do mesmo,

$$\begin{aligned} C(i) &= 2 \frac{\ell_i}{k(i)[k(i) - 1]} \\ &\equiv \frac{(\mathbf{A}^3)_{i,i}}{k(i)[k(i) - 1]}. \end{aligned}$$

O termo na segunda linha é a definição equivalente do coeficiente, obtida através da matriz de adjacência $\mathbf{A}(\mathcal{G})$: $(\mathbf{A}^3)_{i,i} = \sum_{j \neq i} \sum_{h \neq \{i,j\}} a_{ij} a_{ih} a_{jh}$ é o i -ésimo elemento da diagonal do produto \mathbf{A}^3 (55). Dado que $C(i)$ fornece a densidade de arestas $\{j, h\}$ entre vizinhos $j \neq h$ de i , quanto maior o valor de $C(i)$, maior a presença de *triângulos* na rede \mathcal{G} tendo i como um de seus vértices. Triângulos são ciclos de tamanho 3 da forma $\{i, j\}, \{i, h\}, \{j, h\}$.

O *coeficiente de agrupamento da rede* é dado pela média aritmética dos coeficientes de agrupamento individuais:

$$C = \frac{1}{N} \sum_i C(i). \quad (2.9)$$

Note que $C(i)$ (e portanto C) varia entre 0 e 1. O coeficiente de agrupamento é uma *propriedade local* da rede, uma vez que indica a interconectividade local média entre cada um dos nós e sua vizinhança. O coeficiente de agrupamento está associado ao *paradigma de pequeno-mundo*, conforme ilustrado por (37): uma rede classificada como “pequeno-mundo” é definida por uma combinação de alto coeficiente de agrupamento e baixo comprimento médio dos menores caminhos, ou seja, embora existe uma forte interconectividade local, é possível, a partir de um nó, alcançar qualquer outro percorrendo-se poucas conexões. Em redes direcionadas, adotamos a extensão proposta por Fagiolo (55), onde o coeficiente do nó i é dado por

$$C(i) = \frac{(\mathbf{A} + \mathbf{A}^T)_{i,i}^3}{2 \{k_{tot}(i) [k_{tot}(i) - 1] - 2 \overleftrightarrow{k}(i)\}}, \quad (2.10)$$

onde $(\mathbf{A} + \mathbf{A}^T)_{i,i}^3 = \sum_{j \neq i} \sum_{h \neq \{i,j\}} (a_{ij} + a_{ji})(a_{ih} + a_{hi})(a_{jh} + a_{hj})$, e $k_{tot}(i) = k_{in}(i) + k_{out}(i) = \sum_j (a_{ij} + a_{ji})$ e $\overleftrightarrow{k}(i) = \sum_j a_{ij} a_{ji} = (\mathbf{A}^2)_{i,i}$ são respectivamente o *grau total* do nó i e o número de conexões bidirecionais (recíprocas) envolvendo o mesmo.

2.2.8 Comprimento dos menores caminhos e eficiência

Um *caminho* de comprimento H entre dois nós $i \neq j$ é uma sequência de H arestas distintas $\{i, v_1\}, \{v_1, v_2\}, \dots, \{v_{H-2}, v_{H-1}\}, \{v_{H-1}, j\}$ que *interconecta* os mesmos. O(s) *menor(es) caminho(s)* ou *caminho(s) geodésico(s)* entre tais nós é(são), dentre todos os caminhos possíveis entre i e j , o(s) de menor comprimento, cujo valor denotaremos como d_{ij} . O *comprimento médio dos menores caminhos* em uma rede é a média de tal valor sobre todos os caminhos geodésicos possíveis. Porém tal medida é sensível com relação ao número de *componentes conexas* da rede \mathcal{G} , isto é, subconjuntos $\mathcal{G}_r(\mathcal{V}_r, \mathcal{E}_r)$, $r = 1, 2, \dots$ de \mathcal{G} tais que $\mathcal{V}_r \cap \mathcal{V}_s = \emptyset$, $\mathcal{E}_r \cap \mathcal{E}_s = \emptyset$, $\forall r \neq s$: se $i \in \mathcal{G}_i$ e $j \in \mathcal{G}_j$, $i \neq j$, não há caminhos interconectando i e j , situação denotada pela expressão $d_{ij} = \infty$. Tais ocorrências não são computadas no cálculo do comprimento médio, e para contornar tal situação, (56) propôs uma medida dada por

$$E = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} \frac{1}{d_{ij}}, \quad (2.11)$$

denominada **eficiência global** da rede \mathcal{G} . Tal medida é assim denominada por estar associada à *eficiência* da rede em transportar informações entre diferentes nós, inversamente proporcional à distância entre os mesmos. Note que $E \in [0, 1]$, e o quociente de normalização $1/[N(N-1)]$ se deve à existência de no máximo $N(N-1)$ caminhos distintos entre os nós da rede, embora para redes não-direcionadas necessariamente $d_{ij} = d_{ji}$ e podemos reescrever a Eq. (2.11) como $E = 2/[N(N-1)] \sum_i \sum_{j>i} \frac{1}{d_{ij}}$.

2.2.9 Medidas de centralidade

A *centralidade* de um nó fornece uma estimativa de sua *importância* dentro da rede, em termos da *comunicabilidade* da mesma, ou seja, da transmissão de informação pelo sistema, através da topologia (nós e arestas) do mesmo. Diversas medidas de centralidade foram propostas (4), contudo apresentaremos duas medidas, a *centralidade de intermediação* e o *índice k -shell*.

Centralidade de intermediação

Seja σ_{jk} o número de caminhos geodésicos – veja Seção 2.2.8 – entre os nós j e k (é possível que haja mais de um caminho entre j e k com comprimento mínimo d_{jk}). Sendo $\sigma_{jk}(i)$ o(s) caminho(s) geodésico(s) que passam pelo nó i , definimos a *centralidade de intermediação* (*betweenness centrality*) (57) do mesmo como o valor médio, considerando-se todos os pares j, k , da fração $\sigma_{jk}(i)/\sigma_{jk}$:

$$C_B(i) = \frac{2}{(N-1)(N-2)} \sum_{j \neq i} \sum_{k \neq \{i,j\}} \frac{\sigma_{jk}(i)}{\sigma_{jk}}. \quad (2.12)$$

A fração à esquerda dos somatórios normaliza o valor de $C_B(i)$ em redes não-dirigidas. O caso dirigido não será nosso objeto de estudo.

Índice k -shell

O índice *k -shell*, por sua vez, é uma medida de centralidade que envolve o grau dos nós (58), e sua estimativa é baseada num método de *decomposição*, da seguinte maneira:

1. Seja m o *grau mínimo* da rede, isto é, $k(i) \geq m, \forall i \in \mathcal{G}$. Para todo nó j com grau m , atribua índice de *k -shell* m : $k_S(j) = m$.
2. “Remova” da decomposição os nós cujo valor *k -shell* foi atribuído e suas respectivas arestas.
3. Após a etapa 2, é possível o surgimento de novos nós com grau *menor ou igual a m* – nós que estavam em contato com outros que foram removidos. Repita os itens 1 e 2, atribuindo índice m a tais nós, até que apenas restem nós com grau superior a m .
4. Repita o processo para $k = m + 1, m + 2, \dots$ até que todo nó i possua um índice *k -shell* $k_S(i)$. Seu valor é sempre menor ou igual ao grau do mesmo, $k_S(i) \leq k(i)$.

Ao final da decomposição, a rede está subdividida em diversas “casas”, cada uma definida por um valor do índice. As casas cujo índice correspondente é maior definem a região mais central ou o “núcleo” da rede, ao passo que as restantes envolvem os nós mais à “periferia” da mesma. Kitsak e colaboradores (59) mostraram que em determinadas redes sociais existe uma alta correlação entre o índice *k-shell* dos nós e sua importância na propagação de processos epidêmicos. Nesta tese comparamos, para alguns modelos de rede (veja Seção 2.3) e a rede de aeroportos mencionada na Fig. 2.2 as correlações entre a propagação de epidemias e diversas características dos nós, incluindo as medidas de centralidades aqui apresentadas e a *acessibilidade* dos nós, que será apresentada a seguir. A Figura 2.4 ilustra um exemplo de rede onde os nós são classificados de acordo com o índice.

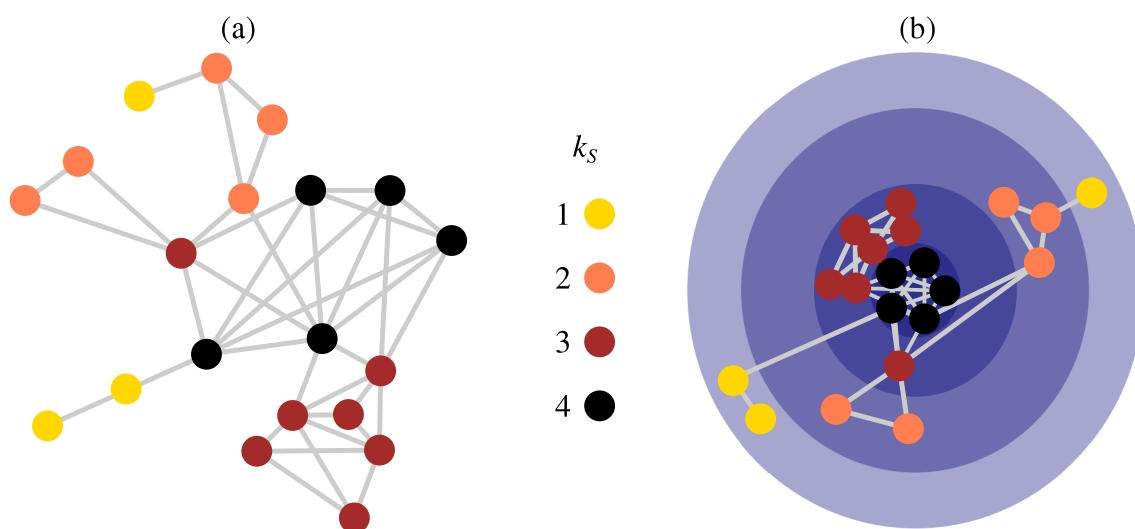


Figura 2.4 – Índice *k-shell*. (a) grafo contendo nós classificados de acordo com o índice. (b) Representação alternativa do mesmo, destacando as “casas” respectivas de cada valor – veja texto para detalhes.

2.2.10 Acessibilidade

A *acessibilidade* de um nó i (35) corresponde a uma estimativa de quantos nós são efetivamente alcançados ou *acessados*, por meio de *caminhadas aleatórias* (60–62) de comprimento H partindo do mesmo. É dada pela expressão

$$A_H(i) = \exp \left\{ - \sum_{j=1}^N P_H(i, j) \log [P_H(i, j)] \right\}. \quad (2.13)$$

Cada termo $P_H(i, j)$ corresponde à probabilidade de um nó j ser visitado no passo H da caminhada, de modo que $\sum_j P_H(i, j) = 1$. Em uma caminhada aleatória, se no instante t o agente ou caminhante encontra-se em um nó i , a probabilidade de que o mesmo visite o nó j no instante $t + 1$ é $P_1(i, j) = 1/k(i) = a_{ij}/\sum_h a_{ih}$, que denominamos *probabilidade de transição* de i para j . É possível, a partir da matriz de adjacência \mathbf{A} da rede, definir a *matriz de transição* \mathcal{T} tal que $\mathcal{T}_{ij} = P_1(i, j)$. As definições aqui são validas tanto para redes não-dirigidas como para redes dirigidas. Podemos estender o conceito de transição para redes com pesos, onde $P_1(i, j) = w_{ij}/\sum_h w_{ih} = w_{ij}/s(i)$. $s(i)$ é a soma dos pesos das conexões adjacentes a i , denominada *força (strength)* do nó i . Os valores $P_H(i, j)$ são obtidos a partir da matriz de transição, sendo dados pelos respectivos coeficientes de \mathcal{T}^H , isto é, $P_H(i, j) = (\mathcal{T}^H)_{i,j}$.

O termo negativo no interior da exponencial é denominado *diversidade* de i , e corresponde à entropia das probabilidades $P_H(i, j)$. Na hipótese de que $N_H \leq N$ nós sejam acessados a partir de i com a mesma probabilidade, $P_H(i, j) = 1/N_H$, então a diversidade é $\log(N_H)$, o que implica $A_H(i) = N_H$. Em outras palavras, N_H nós são efetivamente acessados a partir de i , em H etapas. A diversidade é reduzida quando a probabilidade de se atingir determinados nós é maior do que outros, de modo que $A_H(i) < N_H$ – veja Fig. 2.5, onde o conceito de acessibilidade é explorado para redes não-direcionadas e sem pesos. Note que se apenas um nó j pode ser acessado em caminhadas de H etapas a partir de i , então $A_H(i) = 1$, pois a diversidade será nula. Logo, $0 \leq A_H(i) \leq N$, indicando que a acessibilidade pode ser vista como uma medida do *grau efetivo* do nó i . Para cada nó, infinitos valores de acessibilidade podem ser estimados, para $H = 1, 2, \dots$. Nesta tese, entretanto, tomamos H como o arredondamento para inteiro do comprimento médio dos menores caminhos definido na Seção 2.2.8.

2.3 Modelos

Além das medidas topológicas de redes, diversos modelos foram propostos, visando tentar reproduzir características topológicas de sistemas reais. Destacamos alguns destes modelos a seguir, que serão utilizados nesta tese.

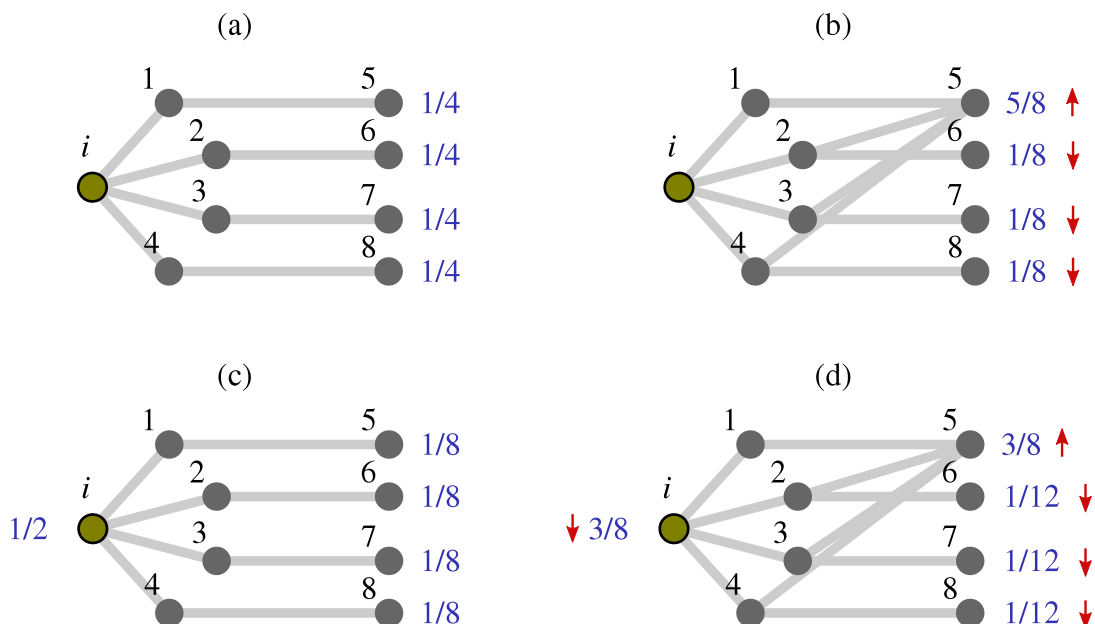


Figura 2.5 – Caminhadas aleatórias e acessibilidade: duas maneiras de se obter o valor da acessibilidade do nó i : a partir de caminhadas *self-avoiding* – itens (a) e (b) – e caminhadas simples – itens (c) e (d). Suponha que desejamos estimar o valor $A_2(i)$, ou seja, a acessibilidade do nó i para $H = 2$. No caso de caminhadas *self-avoiding* – veja itens (a) e (b) – nós e arestas podem ser visitados somente uma vez, de modo que caminhadas de comprimento $H = 2$ iniciadas em i podem visitar somente os nós 5 a 8. No item (a), a probabilidade de visita é a mesma para cada um destes nós, $1/4$. Desta forma, $A_2(i) = 4$ – veja Eq. (2.13). No item (b), novas conexões são adicionadas, aumentando a probabilidade de acesso ao nó 5 e conseqüentemente diminuindo a probabilidade de acesso aos demais, conforme indicado pelas setas. Desta forma, a entropia das probabilidades, $P_2(i, j)$, $j = 5, \dots, 8$, diminui em comparação com o cenário do item (a), e neste caso, $A_2(i) = 1,594 \ll 4$. Os itens (c) e (d) correspondem à situação análoga a dos itens anteriores, porém envolvendo caminhadas simples – conforme descrito no texto. Tais caminhadas não possuem as mesmas restrições de não-repetição do caso *self-avoiding*, de modo que agora o nó i pode ser também visitado no passo $H = 2$. De maneira análoga ao visto para os itens (a) e (b), a inserção na rede de mais arestas que envolvam o nó de número 5 aumenta a chance de o mesmo ser visitado no segundo passo das caminhadas, reduzindo a chance de visita aos demais, porém agora a acessibilidade do nó i assume os valores $A_2(i) = 4$ e $A_2(i) = 3,884$, para os itens (c) e (d), respectivamente. Embora caminhadas *self-avoiding* sejam úteis para a compreensão do conceito de acessibilidade, seu cálculo é altamente custoso, de modo que nesta tese apenas utilizaremos a acessibilidade que envolva caminhadas simples – as quais o cálculo das probabilidades $P_H(i, j)$ pode ser dado pela matriz de transição da rede, conforme explicado no texto.

2.3.1 Rede aleatória de Erdős-Rényi (modelo ER)

Criado em meados da década de 1960, pelos autores Paul Erdős e Alfred Rényi (63), o modelo consiste de uma rede formada por n nós, onde interconectam-se pares dos mesmos com probabilidade p_{ER} , de tal forma que o número esperado de arestas é $\langle \ell \rangle = p_{ER}n(n-1)/2$, e o grau médio, $\langle k \rangle = p_{ER}(n-1)$. É possível gerar uma rede de grau médio K definindo-se $p_{ER} \equiv K/(n-1)$. A distribuição de graus $P(k)$ tende a uma distribuição de Poisson no limite termodinâmico, conforme podemos ver na Fig. 2.6.

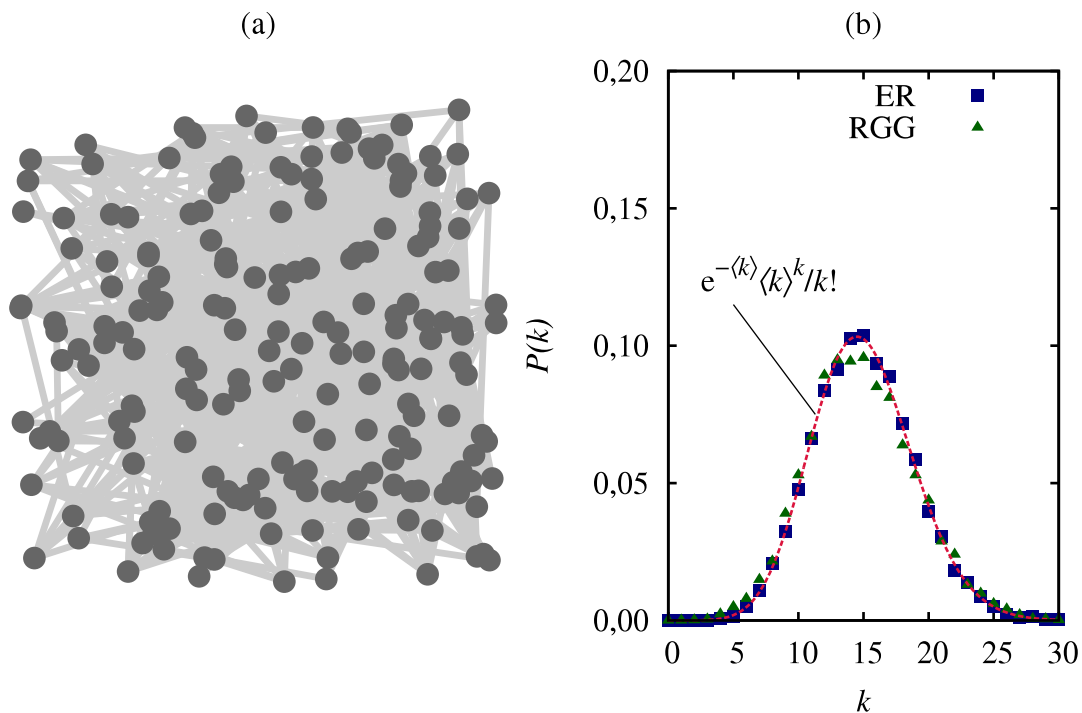


Figura 2.6 – Modelo ER. (a) Exemplo de rede modelada, de tamanho $N = 200$ e com grau médio $\langle k \rangle \approx 6$. (b) Distribuição de graus de uma rede ER (quadrados) e de uma rede geográfica RGG (triângulos) – ver Seção 2.3.4, ambas com $N = 10000$ nós e grau médio $\langle k \rangle = 14,998$ e $\langle k \rangle = 14,791$, respectivamente. Como podemos ver, $k \sim \text{Pois}(\langle k \rangle) = \exp(-\langle k \rangle) \langle k \rangle^k / k!$.

2.3.2 Rede pequeno-mundo de Watts-Strogatz (modelo WS)

Modelo proposto por Duncan Watts e Steven Strogatz (37), que visa explicar a propriedade de “pequeno-mundo” observada em muitos sistemas reais, como redes sociais (ex.: (39)), mas que não é capturada pelo modelo clássico de Erdős e Rényi: quando comparada a uma rede ER de mesmo tamanho e mesmo grau médio, a rede pequeno-mundo apresenta maior coeficiente de agrupamento; por outro lado, se comparada a uma rede regular (onde todos os nós possuem o mesmo grau) de mesmo tamanho e conectividade, a rede pequeno-mundo apresentam um valor mais baixo para o comprimento médio dos caminhos mínimos. Uma rede pequeno-mundo é gerada a partir de uma rede regular onde cada um dos n nós se conecta a um valor fixo k de vizinhos – $\ell = nk/2$. Então, percorrem-se as arestas da rede, trocando-se, com uma probabilidade p_{WS} , um dos nós envolvidos na conexão por outro aleatoriamente escolhido, tomando-se a precaução de se evitar conexões duplicadas, de forma que o número de arestas se mantém – veja Fig.

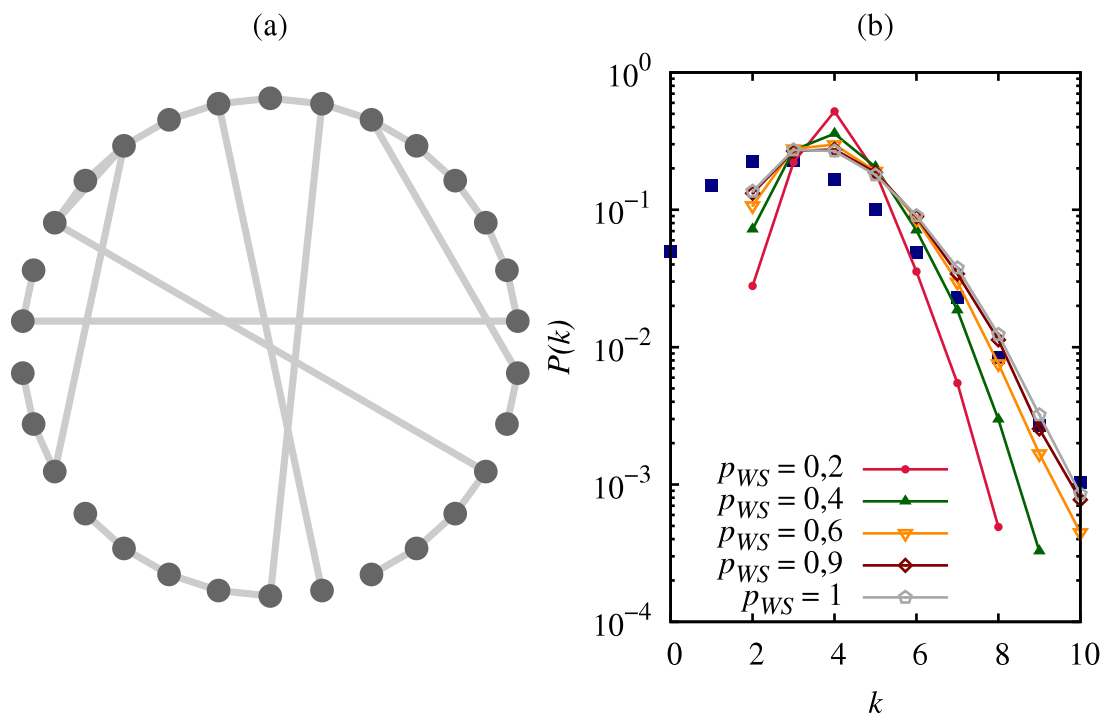


Figura 2.7 – Modelo WS. (a) Rede “pequeno-mundo” é gerada a partir de uma estrutura regular, onde cada nó se conecta aos vizinhos mais próximos. Então, para cada conexão, o nó de uma das extremidades é trocado com probabilidade p_{WS} . No exemplo, $N = 30$, $\langle k \rangle = 2$, $p_{WS} = 1/5$. (b) Distribuições de grau $P(k)$ para redes geradas a partir do modelo WS de tamanho $N = 1000$ com $\langle k \rangle = 4$, para diversos valores p_{WS} . Os quadrados correspondem a uma rede aleatória ER de mesmo tamanho e conectividade equivalente.

2.7(a). Para $p_{WS} = 0$ não há alterações, e a rede é regular, caracterizada por um alto coeficiente de agrupamento, e baixa eficiência (valor elevado do comprimento médio dos menores caminhos). Já no outro extremo, quando $p_{WS} = 1$, a rede se comporta como o modelo de Erdős-Rényi: eficiência aprimorada, porém ao custo de baixa conectividade local (coeficiente de agrupamento baixo) – Fig. 2.7(b). Os autores mostram em seu trabalho que para valores baixos de p_{WS} , é possível uma situação intermediária, com grande redução do comprimento médio dos caminhos entre os nós, com manutenção de alta conectividade local, o que caracteriza estruturas do tipo “pequeno-mundo”. O grau médio da rede no modelo é k , independente do valor p_{WS} .

2.3.3 Rede livre-de-escala de Barabási-Albert (modelo BA)

Modelo proposto (38) pelos autores para explicar a lei-de-potência observada na distribuição de grau em estruturas construídas pelo homem, como a rede da Internet (13).

Diferentemente de outros modelos, nesta rede o número de nós não é fixado ao início da construção do mesmo: de imediato, uma rede de m_0 nós é criada (instante $t_0 = 0$). A seguir, para cada passo de tempo, um nó é inserido na rede e ligado a $m \leq m_0$ nós presentes, seguindo o princípio de *conexão preferencial*: a probabilidade de um novo nó se conectar a outro nó i já existente na rede é proporcional ao grau deste, dada por $\Pi(i) = k_i / (\sum_j k_j)$, $j = 1, \dots, m_0 + t - 1$. A operação é repetida de modo que no instante t , a rede é formada por $m_0 + t$ nós e mt arestas. Quando $n = m_0 + t \rightarrow \infty$, $\langle k \rangle \rightarrow 2m$, e o parâmetro m corresponde ao *grau mínimo* da rede. No modelo, a distribuição de grau segue uma *lei de potências* – Figura 2.8(b) – ou seja, $k \sim k^{-\gamma}$, onde no limite termodinâmico, $\gamma = 3$. A principal diferença em relação aos demais modelos é que a conectividade não é portanto limitada ($\langle k^2 \rangle \rightarrow \infty$), e redes BA são caracterizadas pela presença de nós com alto grau k , os chamados *hubs*.

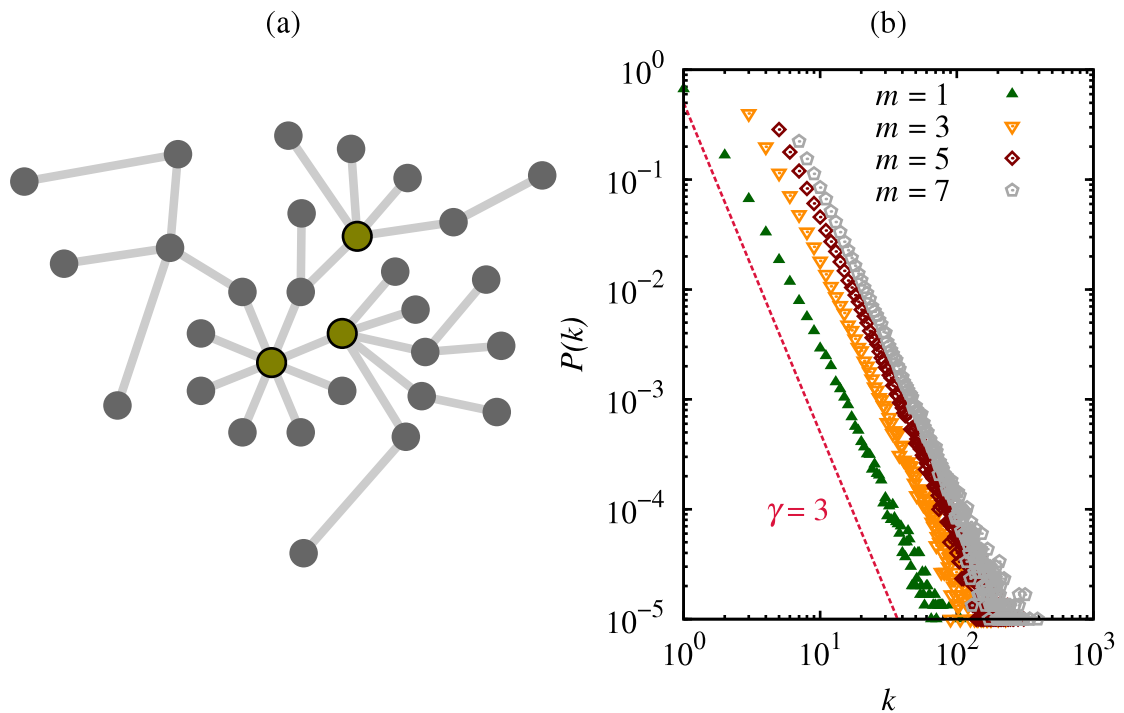


Figura 2.8 – Modelo BA. (a) Exemplo de rede com $N = 30$ nós e $\langle k \rangle \approx 2$. Em destaque, os *hubs* (nós com alta conectividade), cuja presença é característica típica do modelo. (b) Distribuições $P(k)$ de redes de tamanho $N = 300000$ geradas a partir do modelo BA, para diversos valores de m . A distribuição segue uma lei de potências, de modo que o decaimento de $P(k)$ com k das redes é similar ao da curva tracejada, isto é, $k \sim k^{-\gamma}$, com $\gamma = 3$. Diversas redes reais também exibem este comportamento, como podemos ver na Fig. 2.2.

2.3.4 Modelos geográficos

É comum em diversos sistemas reais que os elementos que os constituem estejam *mapeados* em algum ponto específico dos mesmos. Um exemplo clássico são as redes de transporte, onde cada ponto no sistema pode representar um porto, um aeroporto ou uma cidade, e as conexões entre os mesmos são definidas por conexões aéreas, rodovias, etc. Em tais sistemas, diversos aspectos podem determinar e (in)viabilizar a formação de tais conexões, como custo, demanda, distância etc. A *distância* em específico é determinante na conectividade, uma vez que quanto mais separados dois nós em um sistema, menor é a probabilidade de que os mesmos se conectem. Tais redes são denominadas *espaciais* ou *geográficas*. Uma revisão recente e abrangente sobre o tópico é encontrada em (64). Diversas modelagens de redes geográficas existem na literatura. Um simples modelo consiste em espalhar nós em posições aleatórias do espaço e conectar cada par de nós que esteja separado a uma distância menor ou igual a um certo valor R , conforme descrito a

seguir.

Grafo geométrico aleatório (modelo RGG)

Em um grafo geométrico aleatório (RGG, abreviatura do inglês) (65) envolvendo N nós, cada nó i ocupa uma posição $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$ no espaço d -dimensional, onde cada componente $0 \leq x_{ij} < 1$, $j = 1, \dots, d$ é aleatoriamente amostrada de uma distribuição uniforme. Em outras palavras, nós são distribuídos dentro do cubo unitário d -dimensional $[0, 1]^d$, cujo volume é 1, independente da dimensão d . Dois nós i e j são conectados sempre que $|\mathbf{x}_i - \mathbf{x}_j| \leq R$, onde R é um valor positivo real que corresponde à *distância de corte* do modelo. Uma interpretação de como as conexões se formam é associar a cada nó i uma esfera d -dimensional de raio $r \equiv R/2$ com centro em \mathbf{x}_i , sendo que dois nós i e j são interconectados caso as esferas com centros em \mathbf{x}_i e \mathbf{x}_j se sobreponham, como na Fig. 2.9 para o caso $d = 2$.

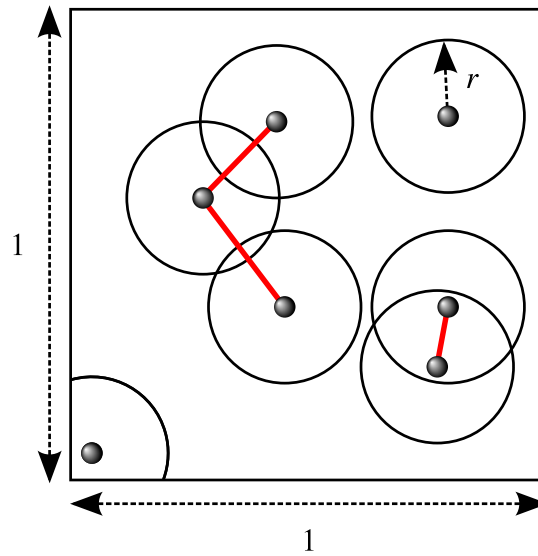


Figura 2.9 – Conexões em um grafo geométrico aleatório (RGG) bidimensional ($d = 2$). Associa-se a cada nó uma esfera de raio $r \equiv R/2$ com centro em suas coordenadas, e conexões entre dois nós são criadas sempre que ocorre sobreposição entre suas respectivas esferas.

De acordo com Dall e Christensen (65), a probabilidade de dois nós aleatoriamente escolhidos i e j estarem conectados é igual ao volume da esfera d -dimensional de raio R , uma vez que nós estão mapeados dentro do cubo d -dimensional de volume unitário. Ou seja,

$$p(0 \leq |\mathbf{x}_i - \mathbf{x}_j| \leq R) \equiv V_d(R) = p_R = \frac{\pi^{d/2} R^d}{\Gamma\left(\frac{d+2}{2}\right)}, \quad (2.14)$$

onde $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$ é a função gama, que estende o conceito de fatorial para números reais e complexos.

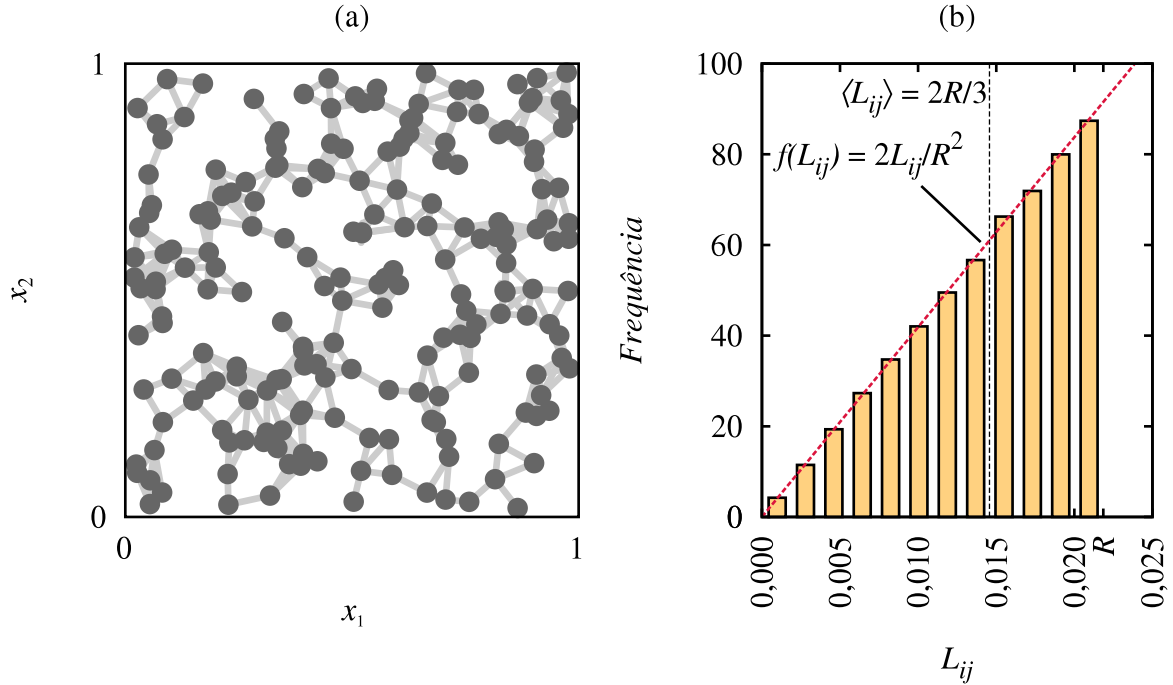


Figura 2.10 – Modelo geográfico RGG. (a) Exemplo de rede, com tamanho $N = 200$ e grau médio $\langle k \rangle = 5,19$ – confronto com a rede ER dada na Fig. 2.6(a) – Note que a posição dos nós é definida no espaço, e conexões só são formadas entre nós próximos uns aos outros, quando a distância entre os mesmos é menor que uma distância de corte R . (b) Distribuição do tamanho das conexões, L_{ij} , para uma rede de tamanho $N = 10000$ e grau médio $\langle k \rangle = 14,791$. Neste caso, a distância de corte é $R \approx 0,0219$. A curva que melhor ajusta a distribuição corresponde à derivada da Eq. 2.17. A distribuição de graus desta rede é dada no item (b) da Fig. 2.6.

Em uma rede com N nós cuja conectividade média é $\langle k \rangle$, temos que $p_R = \langle k \rangle / (N - 1)$. Logo, um RGG com grau médio $\langle k \rangle$ terá distância de corte

$$R(N, d, \langle k \rangle) = \frac{1}{\sqrt{\pi}} \left[\frac{\langle k \rangle}{N - 1} \Gamma \left(\frac{d + 2}{2} \right) \right]^{\frac{1}{d}} \quad (2.15)$$

Nesta tese, nos limitamos ao caso $d = 2$, de modo que as equações 2.14 e 2.15 são reescritas como

$$p_R = \pi R^2, \quad R(N, \langle k \rangle) = \sqrt{\frac{\langle k \rangle}{\pi(N - 1)}}. \quad (2.16)$$

Ainda, em duas dimensões, o *tamanho* L_{ij} de uma conexão aleatoriamente escolhida $\{i, j\}$ é tal que

$$P(0 \leq L_{ij} \leq L) = \left(\frac{L}{R}\right)^2, \quad (2.17)$$

de modo que o *tamanho médio das arestas* será dado por

$$\langle L_{ij} \rangle = \frac{2R}{3}.$$

A Figura 2.10 mostra a distribuição dos tamanhos das conexões L_{ij} para uma rede RGG quando $d = 2$.