



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

EDSON ARAUJO SOARES

ESTUDO DE REDES COMPLEXAS SOB UMA PERSPECTIVA COMBINATÓRIA

FORTALEZA

2021

EDSON ARAUJO SOARES

ESTUDO DE REDES COMPLEXAS SOB UMA PERSPECTIVA COMBINATÓRIA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Física. Área de Concentração: Física da Matéria Condensada.

Orientador: Prof. Dr. André Auto Moreira.

FORTALEZA
2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S653e Soares, Edson Araujo.

Estudo de Redes Complexas sob uma Perspectiva Combinatória / Edson Araujo Soares. – 2021.
57 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Física, Fortaleza, 2021.

Orientação: Prof. Dr. André Auto Moreira.

1. Sistemas Complexos. 2. Redes Complexas. 3. Percolação. 4. Grafo. 5. . I. Título.

CDD 530

EDSON ARAUJO SOARES

ESTUDO DE REDES COMPLEXAS SOB UMA PERSPECTIVA COMBINATÓRIA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Física. Área de Concentração: Física da Matéria Condensada.

Aprovada em 16 / 04 /2021.

BANCA EXAMINADORA

Prof. Dr. André Auto Moreira
Universidade Federal do Ceará (UFC)

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

Prof. Dr. Hênio Henrique Aragão Rêgo
Instituto Federal do Maranhão (IFMA)

AGRADECIMENTOS

A todos os membros do grupo Teoria de Sistemas Complexos, bem como toda a comunidade acadêmica do Departamento de Física da Universidade Federal do Ceará. Agradeço também à Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico pelo suporte financeiro que tornou possível a execução deste trabalho.

RESUMO

Muitos problemas em diversas áreas do conhecimento podem ser abordados por meio da Teoria de Sistemas Complexos, caracterizados como problemas onde um número muito grande de componentes interagem entre si, promovendo a emergência de fenômenos coletivos. A abordagem a tais sistemas são feitas muitas vezes por meio das Redes Complexas, que dentre os seus mais diversos modelos, podem servir no entendimento e tratamento desses sistemas, a partir do mapeamento destes por meio dos chamados grafos. Uma das informações que podem ser acessadas por intermédio da Teoria de Redes são os fenômenos associados à espécies de transições de fase no sistema, isso através da ideia de Percolação, que descreve a transição entre um regime onde existem várias componentes isoladas na estrutura e outro no qual ocorre a formação de uma componente gigante. Estabelecemos uma metodologia com base em argumentos combinatórios que leva a caracterização da distribuição de conectividade e do comportamento da distribuição de tamanho de componentes em função da conectividade média da rede até o limiar de percolação para grafos aleatórios, através da contagem e maximização das formas de construir a estrutura com um dado número de vértices. Usamos ainda as ideias de análise combinatória e arranjos para estabelecer o resultado de enumeração para florestas enraizadas e direcionadas com uma distribuição genérica de conectividade.

Palavras-chave: Sistemas Complexos. Redes Complexas. Percolação. Grafo.

ABSTRACT

Many problems in several areas of knowledge can be approached through Complex Systems Theory, characterized as problems where a very large number of components interact with each other, promoting the emergence of collective phenomena. The approach to such systems are often made through Complex Networks, which among their most diverse models, can serve in the understanding and treatment of these systems, based on the mapping of these systems through the so-called graphs. One of the information that can be accessed through the Network Theory are the phenomena associated with the species of phase transitions in the system, this through the idea of Percolation, which describes the transition between a regime in which there are several isolated components in the structure and another in which the formation of a giant component occurs. We established a methodology based on combinatorial arguments that leads to the characterization of the connectivity distribution and the behavior of the component size distribution as a function of the average network connectivity up to the percolation threshold for random graphs, by counting and maximizing the ways of building the structure with a given number of vertices. We also use the combinatorial analysis and arrangements ideas to establish the enumeration result for rooted and targeted forests with a certain distribution of connectivity.

Keywords: Complex Systems. Complex Networks. Percolation. Graph.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Ilustração esquemática das ilhas de Königsberg | 12 |
| Figura 2 – Grafo associado ao problema das pontes de Königsberg. | 13 |
| Figura 3 – Grafo simples. | 14 |
| Figura 4 – Tipos de grafos | 15 |
| Figura 5 – Caminhos em um Grafo. | 15 |
| Figura 6 – Diâmetro de uma rede | 16 |
| Figura 7 – Distribuição de conectividade em uma rede. | 18 |
| Figura 8 – Modelos de Redes | 21 |
| Figura 9 – Grafo livre de escala | 24 |
| Figura 10 – Comparação da dispersão em redes aleatórias e livres de escala. | 27 |
| Figura 11 – Hubs em redes livre de escala | 28 |
| Figura 12 – Percolação e clusters | 32 |
| Figura 13 – Conectividade do vértice vizinho | 33 |
| Figura 14 – Arranjo por quantidade de elementos | 34 |
| Figura 15 – Tamanhos de componentes em uma floresta | 38 |
| Figura 16 – Exemplo de conjunto | 44 |
| Figura 17 – Iteração entre as etapas 3 e 4 do método de ligação | 46 |
| Figura 18 – Componentes formadas com as escolhas da Figura (17) | 46 |
| Figura 19 – Floresta resultante | 47 |
| Figura 20 – Exemplo de uma floresta contida em outra | 54 |
| Figura 21 – Florestas sequencialmente contidas em outras | 55 |
| Figura 22 – Contagem por corte de arestas | 55 |
| Figura 23 – Contagem por adição de arestas | 56 |

LISTA DE GRÁFICOS

| | |
|---|----|
| Gráfico 1 – Comparação de distribuições Homogêneas e Heterogêneas | 19 |
| Gráfico 2 – Distribuição de conectividade do domínio nd.edu | 23 |
| Gráfico 3 – Dependência de k_{max} com o tamanho da rede | 29 |
| Gráfico 4 – Tamanho da maior componente em função de $\langle k \rangle$ | 37 |
| Gráfico 5 – Previsão analítica da distribuição n_s | 42 |
| Gráfico 6 – Resultado numérico da distribuição n_s excluindo a componente gigante . . . | 43 |

SUMÁRIO

| | | |
|--------------|---|-----------|
| 1 | INTRODUÇÃO | 10 |
| 2 | REDES COMPLEXAS | 12 |
| 2.1 | O problema das Pontes de Königsberg | 12 |
| 2.2 | Grafos e Redes | 13 |
| 2.3 | Caminhos | 15 |
| 2.4 | Conectividade | 17 |
| 2.5 | Distribuição de conectividade | 17 |
| 2.6 | Coefficiente de Agregação | 19 |
| 2.7 | Principais Modelos de Redes | 20 |
| 3 | REDES LIVRES DE ESCALA | 22 |
| 3.1 | Limitações do Modelo Aleatório | 22 |
| 3.2 | Leis de Potência e Redes Livres de escala | 23 |
| 3.3 | Descrição de Distribuição em Lei de Potência | 25 |
| 3.3.1 | Condição de Normalização | 25 |
| 3.3.1.1 | Formalismo Discreto | 25 |
| 3.3.1.2 | Formalismo Contínuo | 26 |
| 3.3.2 | Momentos da Distribuição | 26 |
| 3.4 | Hubs | 28 |
| 3.5 | A função do expoente γ | 29 |
| 4 | PROBLEMA DE PERCOLAÇÃO EM REDES | 31 |
| 4.1 | Solução combinatória de Erdős–Rényi | 32 |
| 4.2 | Distribuição de tamanho de componentes | 38 |
| 5 | ENUMERAÇÃO DE ÁRVORES EM UMA FLORESTA | 44 |
| 6 | CONCLUSÕES E TRABALHOS FUTUROS | 48 |
| | REFERÊNCIAS | 49 |
| | APÊNDICE A – TEOREMA DA INVERSÃO DE LAGRANGE E A SÉRIE DE TAYLOR DA FUNÇÃO $G(X)$ | 52 |
| | ANEXO A – FÓRMULA DE CAYLEY | 54 |

1 INTRODUÇÃO

De acordo com AZEVEDO *et al.* (2009), o cérebro humano é composto por algo em torno de 86 bilhões de neurônios [1], o resultado da interação entre esses tantos neurônios é a emergência de fenômenos coletivos no aspecto macroscópico (cognitivo, emocional, motor, etc.) [2]. Assim como este, podemos caracterizar vários sistemas das mais diversas naturezas por meio da propriedade destes serem constituídos por muitos entes que interagem entre si e que, em decorrência dessa interação, possa ser observado comportamentos macro do sistema que não seriam concebíveis ao tomar um componente isolado, mas somente em função do comportamento coletivo. Sistemas com as características citadas podem ser classificados como *Sistemas Complexos* e são objetos de estudo da *Teoria de Sistemas Complexos*, que, em consonância com THURNER, HANEL e KLINEK (2018), pode ser entendida em termos gerais como a ciência da matéria generalizada interagindo por meio de interações generalizadas [3]. Como exemplos podemos citar ainda como Sistemas Complexos, sistemas físicos, econômicos, sociais e biológicos [4], o que revela o carácter interdisciplinar desse ramo do conhecimento.

Os sistemas complexos têm como uma de suas principais vertentes a *Teoria de Redes Complexas*, que possuem um papel de fundamental importância no estudo e na modelagem de tais sistemas. As redes podem servir para reproduzir as propriedades topológicas que retratam um sistema real por meio da caracterização dos entes que o compõe, bem como estes interagem entre si, isso por meio de estruturas matemáticas chamados *grafos*, no qual vértices e arestas representam constituintes e interações respectivamente [5]. Podemos trazer várias situações de amplo conhecimento na literatura que vem sendo abordada sob a ótica de Redes Complexas, como por exemplo, redes de relações sociais, redes de computadores, redes de comunicação, redes de citação em literatura científica [6] e até redes de malhas aéreas [7], entre tantas outras.

Entender os processos relacionados à evolução e dinâmica de Redes é de fundamental importância para caracterizar fenômenos que ocorrem com os sistemas por ela representados. Por exemplo, os problemas relacionados à *Percolação em Redes*, que representam uma mudança no regime de comportamento no sistema, tal como a transição entre uma fase epidemiológica e não epidemiológica em redes de transmissão de doenças infectocontagiosas [8]. Diante disso, buscamos neste trabalho empregar toda uma metodologia com base em argumentos combinatórios com o intuito de elucidar o problema de Percolação em Redes por meio da predição analítica da região limiar desse fenômeno bem como, ainda com base em nossa proposta combinatória, resultados para problemas de enumeração em teoria dos grafos e, com a finalidade de atestar as previsões analíticas, implementaremos modelos computacionais que

forneçam dados empíricos para tal fim.

A sequência do trabalho desenvolve-se, nessa ordem, com uma introdução acerca dos aspectos primordiais da Teoria de Redes no capítulo 2, uma explanação mais direcionada ao modelo de redes livres de escala no capítulo 3, uma busca no sentido de explorar e solucionar os problemas de percolação aplicados ao modelo aleatório no capítulo 4, a determinação da enumeração para um problema específico em teoria dos grafos no capítulo 5, bem como as considerações e conclusões decorrentes do desenvolvimento do trabalho no capítulo 6.

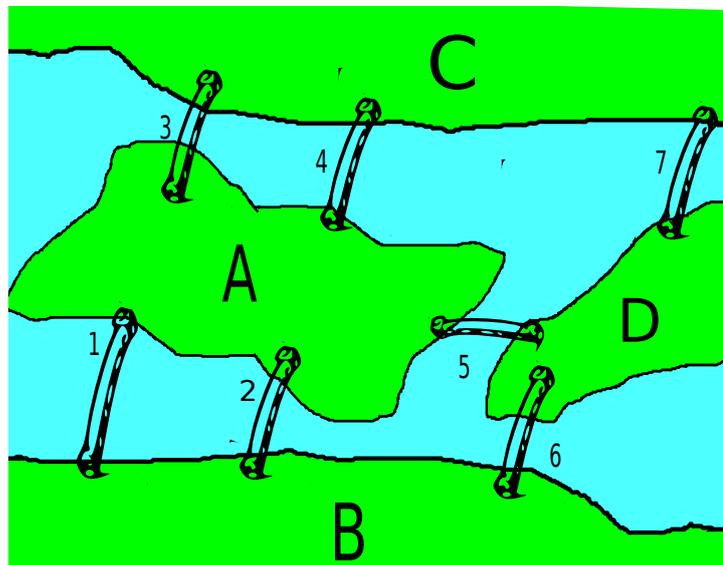
2 REDES COMPLEXAS

De maneira simplificada, uma Rede trata-se da representação de um conjunto de entes que de alguma forma interagem entre si. Estes entes, tais como a maneira de interação, são das mais diversas naturezas [9–12], geralmente tratando de sistemas com um grande número de componentes. A Teoria de Redes Complexas permite a descrição de sistemas altamente complexos bem como suas possíveis formas de evolução e funcionamento. Neste capítulo trataremos de apresentar a maneira como surgiu, os pilares matemáticos e algumas das principais propriedades e características de Redes Complexas.

2.1 O problema das Pontes de Königsberg

Os primórdios da teoria de Redes Complexas remetem ao ano de 1735, quando o matemático suíço Leonhard Euler resolveu o famoso problema das sete pontes de Königsberg [9, 13, 14], já que este problema teria sido primeiro resolvido sob a ideia de grafos, a ferramenta matemática primordial do estudo em Ciência de Redes[9].

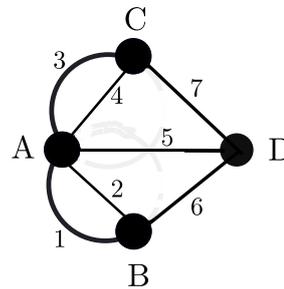
Figura 1 – Ilustração esquemática das ilhas de Königsberg



Fonte: Produzido pelo autor. A Figura representa a configuração das ilhas da cidade de Königsberg, onde o Rio Pergel separa as quatro partes da cidade, conectadas pelas pontes enumeradas de um a sete.

O problema descrito por Euler consistia em determinar a possibilidade ou não de estabelecer-se um percurso entre a ilha de Knepf Hof em Königsberg (no então território da Prússia) e as outras três áreas de terra firme da cidade, separadas pelo rio Pergel (ver Figura 1), por meio das sete pontes que ligavam as ilhas da cidade, de modo a passar apenas uma vez e não menos do que isso sobre cada ponte.

Figura 2 – Grafo associado ao problema das pontes de Königsberg.



Fonte: Produzido pelo autor. Os círculos destacados em preto são os vértices que representam cada uma das ilhas rotuladas como A, B, C e D. As curvas que ligam estes pontos entre si são ditas arestas e representam as sete pontes do problema descrito.

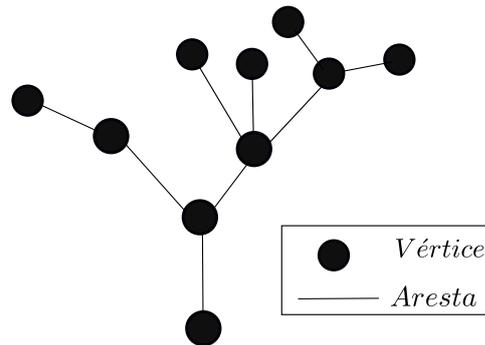
Assim com na Figura 2, Euler representou as ilhas por vértices e as pontes como sendo arestas que ligam cada par de vértices rotulados como A, B, C e D. Como resultado, esse estudo verificou que tal feito só seria possível caso houvessem exatamente zero ou dois vértices com um número ímpar de arestas, uma vez que seria necessário sempre um número par de pontes em cada ilha, uma para entrar e outra para sair, a menos que as ilhas com um número ímpar de pontes fossem o final ou início do trajeto. Assim, Euler concluiu negativamente acerca do problema das pontes e, principalmente, revelou por meio deste estudo a possibilidade de se tratar e buscar resolver problemas por meio de representações acerca da configuração e das características de um dado sistema.

2.2 Grafos e Redes

Um grafo G é definido como sendo um conjunto finito e não vazio $G(V, E)$ de V vértices e E arestas [15–17], trata-se de uma estrutura que representa a topologia da relação de certos objetos. Em Ciência de Redes, os vértices e arestas são comumente referidos, respectivamente, como nós (ou sítios) e links, apesar disso não se tratar de uma distinção rigorosa, dado o fato de que vários autores se referem a tais termos como sinônimos em diversas situações [9, 12, 18]. A Figura 3 exemplifica um grafo simples.

Grafos são estruturas matemáticas abstratas [18], pois a princípio não representam um fenômeno real. A partir do momento em que alguma dessas estruturas possui as propriedades de um sistema, podemos dizer que se trata de uma Rede Complexa, onde cada vértice ou nó passa a representar um componente do sistema em estudo, assim como os links espelham a interação entre estes componentes. É importante salientar para o fato de que, em teoria dos

Figura 3 – Grafo simples.



Fonte: Produzido pelo autor. A Figura mostra um grafo constituído por dez vértices nove arestas.

grafos, o tamanho ou a ordem do grafo é dado pela quantidade presente de arestas, enquanto em Ciência de Redes o tamanho corresponde à quantidade de vértices na rede [19]. Sendo assim, durante a sequência deste trabalho, a quantidade de vértices será considerada o tamanho da rede. A Tabela 1 exemplifica a correspondência entre vértices, links e o que representam em certas redes particulares.

Tabela 1 – Vértices e Links em Redes

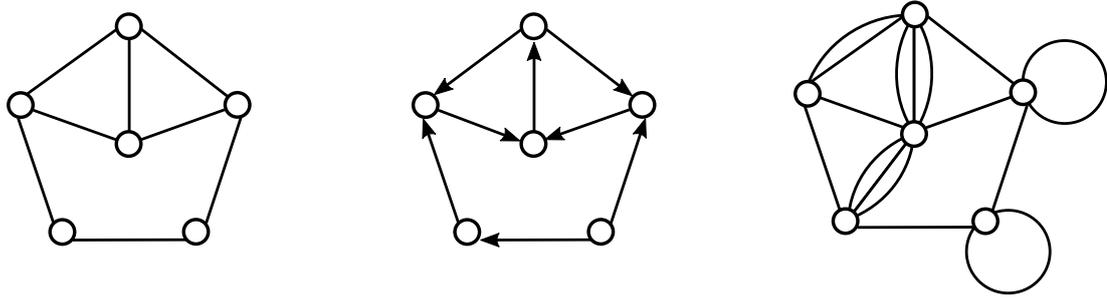
| Rede | Vértice | Link |
|------------------|---------------------------------|--------------------------|
| Internet | Computador ou Roteador | Cabo ou Conexão Wireless |
| Word Wide Web | Página da Web | Hyperlink |
| Rede de Citações | Artigo, Patente | Citação |
| Malha Energética | Estação de Geração ou Substação | Linha de Transmissão |
| Rede de Amigos | Pessoas | Amizade |
| Rede Metabólica | Metabolito | Reação Metabólica |
| Rede Neural | Neurônio | Sinapse |
| Teia Alimentar | Espécies | Predação |

Fonte: Tabela adaptada de Newman: Networks: An Introduction [12].

As arestas em um grafo podem ser direcionadas ou não direcionadas [9, 15]. No primeiro caso, os grafos são denominados grafos direcionados ou dígrafos, significando que as arestas partem de um vértice v_1 em direção a um v_2 . No segundo caso temos os intitulados grafos não direcionados. O mesmo critério vale para classificar redes [9, 12].

Vários sistemas complexos, quando modelados, podem ser representados por grafos que possuem tanto links direcionados quanto não direcionados [9], assim como apresentar vértices ligados a si mesmos (*loop*) em um ou mais vértices, bem como mais de uma ligação entre pares de nós, sendo estes denominados multígrafos [12, 20]. A Figura 4 da exemplos de um grafo direcionado, não direcionado e um multigrafo.

Figura 4: Tipos de grafos



(a) Não direcionado

(b) Direcionado

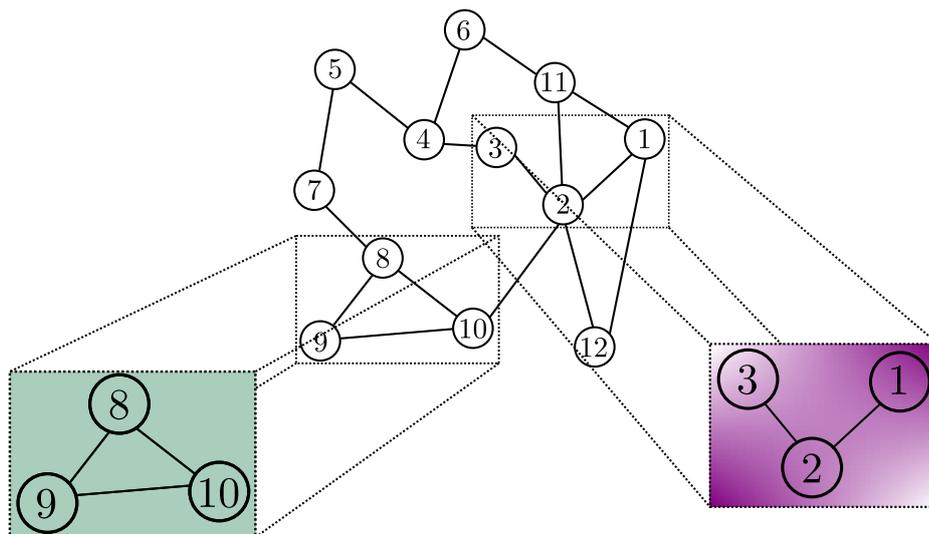
(c) Multigrafo

Fonte: Produzido pelo autor. A Figura (a) mostra um grafo onde as arestas não têm direção, enquanto (b) traz a mesma estrutura exceto pelo fato de todas as arestas agora terem uma direção, apontado de um vértice a outro. A parte (c) da Figura exemplifica um multigrafo que possui *loops* e ligações duplas entre pares de vértices.

2.3 Caminhos

Uma das questões básicas mais relevantes em estudo de grafos é determinar a possibilidade ou não de se estabelecer um caminho entre um vértice i e um j , através de arestas ao longo da estrutura [17, 19]. Tal condição verifica-se sob a existência de um caminho $P_{i,j}$ que conecta tais vértices. Em certas estruturas, é possível também estabelecer caminhos que conectam um vértice a si mesmo, definindo um ciclo na rede. (ver Figura 5).

Figura 5 – Caminhos em um Grafo.



Fonte: Produzido pelo autor. A Figura destaca o menor caminho que parte do vértice 1 até o vértice 3, e destaca também um ciclo formado entre os vértices 8,9 e 10.

Uma rede onde é possível conceber um caminho entre quaisquer vértices é dita uma

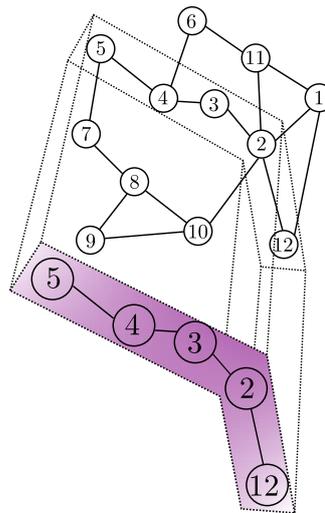
rede conectada [9, 17]. A ausência dessa característica implica a existência de mais de uma componente na rede, ou seja, a existência de subgrafos conectados na rede.

A ideia de caminho leva a três importantes propriedades estatísticas [9, 17–19]:

1. *Menor Caminho* - Consiste no menor caminho possível que conecta um determinado par de nós.
2. *Diâmetro* - Denotado por d_{max} , o diâmetro da rede equivale ao tamanho do menor caminho entre os dois nós mais distantes.
3. *Menor Caminho Médio* - Separação típica entre dois vértices entre cada par possível na rede. Essa grandeza pode ser matematicamente expressa pela equação 2.1.

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j=1,N} d_{i,j}, \quad \text{com } i \neq j \quad (2.1)$$

Figura 6 – Diâmetro de uma rede



Fonte: Produzido pelo autor. O caminho destacado na rede evidencia a menor distância entre os vértices 12 e 5, representado o diâmetro do grafo em questão.

As Figuras 5 e 6 mostram exemplos práticos da identificação destas propriedades em uma rede. Na Figura 5 podemos determinar o menor caminho entre os nós 1 e 3, como sendo igual a dois e o diâmetro da rede, na Figura 6, determinado pela distância entre os vértices 5 e 12, totalizando quatro arestas. Assim como as outras duas propriedades anteriores, é conveniente e, na maioria dos casos, necessário determinar o menor caminho médio por meio da implementação de um algoritmo adequado a cada categoria de rede.

2.4 Conectividade

Uma das principais propriedades de um vértice é o grau de conectividade que este detém [9], ou seja, a quantidade de links que conectam tal vértice diretamente qualquer outro vértice na rede. [17].

Em redes não direcionadas o grau de cada sítio é determinado indistintamente, uma vez que não se trata o fato de as conexões serem de saída ou de entrada [9, 17]. Neste caso, podemos determinar a conectividade total L da rede somando todos os links incidentes em cada vértice e , para carregar a informação de que cada ligação foi contada duas vezes (uma em cada nó de cada par conectado), dividir esta soma por dois, expressado pela equação 2.2.

$$L = \frac{1}{2} \sum_{i=1}^N K_i \quad (2.2)$$

Quando tratamos de redes direcionadas, devemos considerar o grau de vértice que alcançam o vértice (k^{in}), bem como os que partem dele (k^{out}) [9, 17, 18]. A conectividade é dada pela soma entre as ligações, tanto incidentes quanto os de saída. Assim a conectividade total da rede é dada pela equação 2.3.

$$L = \sum_{i=1}^N K_i^{in} = \sum_{i=1}^N K_i^{out} \quad (2.3)$$

Podemos agora expressar uma propriedade estatística especialmente relevante, o grau médio de conectividade de cada nó na rede, sendo dado pela equação 2.4 para redes não direcionadas e a equação 2.5 para redes direcionadas.

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N K_i = \frac{2L}{N} \quad (2.4)$$

$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N K_i^{in} = \langle k^{out} \rangle = \frac{1}{N} \sum_{i=1}^N K_i^{out} = \frac{L}{N} \quad (2.5)$$

Como veremos adiante, os graus de conectividade são distribuídos de maneiras particulares, o que faz com que hajam em muitas redes nós altamente conectados, ao passo que outros sejam de um grau de baixa conexão.

2.5 Distribuição de conectividade

A distribuição de conectividade é uma propriedade crucial na distinção de redes [17]. Tal propriedade, comumente denotada por p_k , é definida como a probabilidade de um sítio, escolhido aleatoriamente em uma rede, ter exatamente um grau k de conectividade [9, 17, 21], ou seja, p_k define a fração de vértices da rede com conectividade k [12]. A distribuição de conectividade é outra propriedade estatística básica em redes, que apesar de geralmente não ser

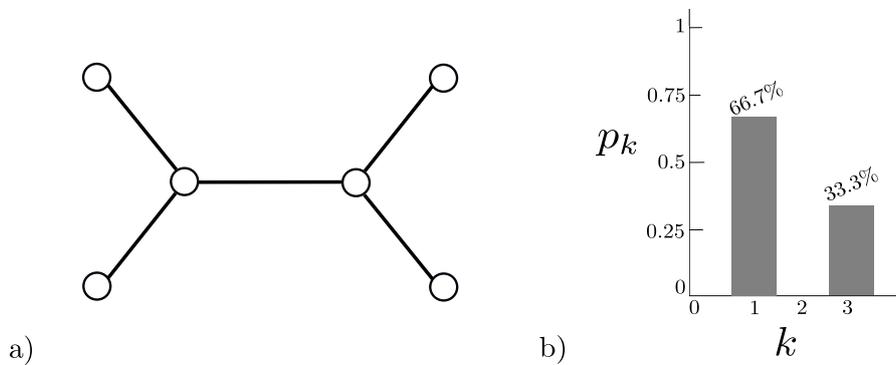
suficiente para caracterizar a estrutura completa de um sistema, em alguns casos esta informação é o bastante para o entendimento de uma rede e seus processos [12, 21].

Como estamos tratando de uma probabilidade, as distribuições de conectividade em cada rede devem satisfazer a condição de normalização, como descrito na equação 2.6 [9].

$$\sum_{k=1}^{\infty} p_k = 1 \quad (2.6)$$

A Figura 7(a) mostra um grafo qualquer no qual podemos exemplificar a manifestação dessa propriedade.

Figura 7 – Distribuição de conectividade em uma rede.



Fonte: Produzido pelo autor. (a) Mostra um grafo contendo seis vértices e cinco arestas, distribuídas de modo que dois nós têm conectividade três e quatro nós têm conectividade dois. (b) Traz a histograma normalizado da distribuição de ligações do grafo em (a)

A forma como as ligações se distribuem ao longo da rede pode ser dada pela razão entre o número de vértices com conectividade k e total de sítios presentes [9], como na equação 2.7.

$$p_k = \frac{N_k}{N} \quad (2.7)$$

De modo que para o grafo da Figura 7(a) podemos determinar os valores para $p_{k=0} \dots p_{k=3}$. O histograma normalizado dessa distribuição é posto ao lado grafo na Figura 7(b).

$$p_0 = 0, \quad p_1 = 4, \quad p_2 = 0, \quad p_3 = 2$$

De posse da informação acerca da distribuição de conectividade, podemos determinar o grau médio de conexão para cada vértice, bem como os momentos superiores da distribuição por meio da equação 3.15, onde $n = 1$ refere-se ao valor médio.

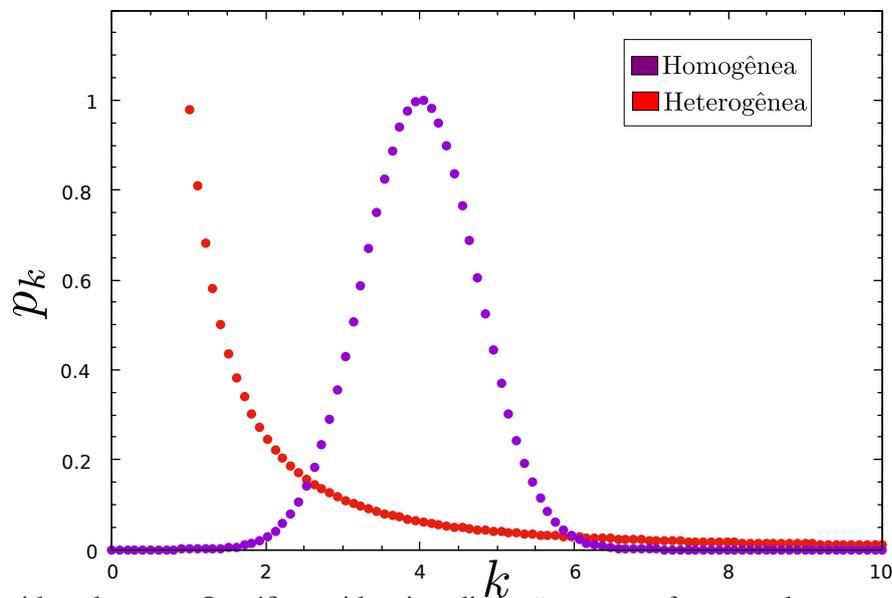
$$\langle k^n \rangle = \sum_n k^n p_k \quad (2.8)$$

Em se tratando de redes direcionadas, podemos aplicar as definições já discutidas,

no entanto, deve-se observar as distribuições como sendo de saída ($P_{k_{out}}$) e de entrada ($P_{k_{in}}$)¹[17].

De acordo com sua distribuição de conectividade, as redes podem ser classificadas como homogêneas e heterogêneas [18, 19, 22]. Redes Homogêneas são caracterizadas por distribuições centradas em torno de um valor médio que decai rapidamente para conectividades que se distanciam da média. Já Redes Heterogêneas possuem um comportamento em forma de lei de potência, apresentando uma curva que decai exponencialmente para grandes valores de conectividade, trataremos em detalhes sobre esta distribuição no capítulo 3. O gráfico 1 mostra as curvas típicas de Redes Homogêneas e Redes Heterogêneas². O grau de heterogeneidade κ

Gráfico 1 – Comparação de distribuições Homogêneas e Heterogêneas



Fonte: Produzido pelo autor. O gráfico evidencia a distinção entre os formatos de curvas de distribuições de Redes Homogêneas e Heterogêneas.

de uma rede é determinado por,

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} \quad (2.9)$$

esta propriedade é especialmente útil no estudo de robustez e ataques a redes [22–24].

2.6 Coeficiente de Agregação

Considere um vértice i em uma rede, conectado a um vértice j e, o mesmo vértice i conectado com um terceiro vértice k . É possível que os vértices j e k também estejam conectados entre si, formando um triângulo na rede. Tal relação, quando verificada, é designada como

¹É possível definir uma distribuição p_{jk} de vértices que possuam simultaneamente j links de entrada e k links de saída, para mais detalhes ver referência [12]

²O gráfico 1 tem fins meramente didáticos, não representando resultados de um estudo real, todos os valores são arbitrários.

transitividade e quantificada por meio do *Coefficiente de Agregação*[9, 12].

Esse parâmetro carrega a informação acerca de o quão conectados são os primeiros vizinhos de um determinado nó e, pode ser formalmente entendido como a probabilidade de dois vizinhos de um nó i estarem conectados [9, 18].

Podemos determinar o coeficiente de agregação para um nó específico i (ou coeficiente de agregação local) com conectividade k_i , por meio da razão entre o número de pares conectados entre os primeiros vizinhos de i (L_i) e número de possíveis pares entre todos os vizinhos $\left(\frac{1}{2}(k_i(k_i - 1))\right)$, como segue na equação 2.10 [9, 12].

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad (2.10)$$

Podemos determinar também o coeficiente de agregação médio da rede, ou seja, o grau de agregação de toda a rede. Para isto basta dividir o coeficiente local de cada vértice pelo tamanho da rede, como na equação 2.11.

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^k C_i \quad (2.11)$$

Tanto C_i quanto $\langle C \rangle$ são números entre 0 e 1, sendo que um coeficiente de agregação igual a 1 representa uma rede perfeitamente transitiva [12] e, quando igual a 0, não há conexões entre vizinhos próximos ou até mesmo não há loops na rede [9, 12].

2.7 Principais Modelos de Redes

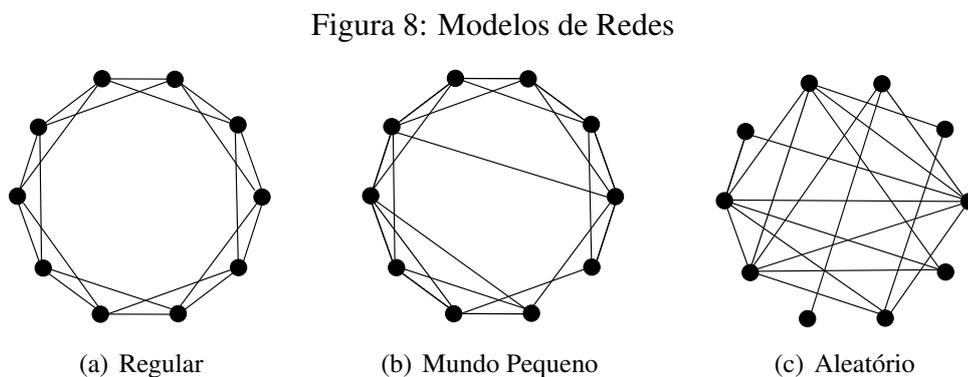
Observando suas particularidades, sistemas complexos são descritos através de redes que traduzam adequadamente suas estruturas, sendo assim, apesar de em certos casos mais de um fenômeno poder ser descrito pela mesma topologia, é de se esperar que existam expressivas variações de redes que possam abrangê-los. A seguir, introduzimos brevemente os principais modelos de redes.

1. *Redes Regulares*- redes caracterizadas por uma estrutura regular (triangular, quadrada, hexagonal, etc.) devido ao seu padrão bem definido de conexões [25]. Esse padrão reflete o fato de cada vértice estar conectado exatamente a k vértices da vizinhança, sendo que, todos os vértices possuem a mesma conectividade [18, 26].
2. *Redes Aleatórias*- em oposição às redes regulares, este modelo apresenta estruturas completamente aleatórias, ou seja, um par qualquer de vértices na rede podem ser aleatoriamente conectado [25–27]. Redes desse tipo seguem distribuições de conectividade do tipo binomial para um número pequeno de vértices na rede e, distribuições de Poisson,

quando se trata de redes com um grande número de vértices [25].

3. *Pequeno Mundo*- é um modelo intermediário, que fica entre o modelo regular e o aleatório. Trata-se, portanto, de uma classe de redes que apresenta uma estrutura parcialmente regular, mas com traços de aleatoriedade [26]. Como a própria nomenclatura sugere, este tipo de rede apresenta a propriedade de *Mundo Pequeno*, esse conceito se traduz no fato de que existe um caminho curto entre a maioria dos pares de nós da rede³ [28].

A maneira mais popular de se conceber redes de mundo pequeno se dá por meio do modelo de *Watts-Strogats (WS)* [25, 27, 28]. Em termos gerais, podemos descrever esse modelo como a construção de uma rede a partir de uma rede regular, onde algumas ligações são reconectadas aleatoriamente ao longo do grafo. A Figura 8 mostra exemplos destes três modelos de rede, nela podemos perceber a possibilidade de, a partir de uma rede regular, construir uma rede de mundo pequeno. Além disso, podemos perceber o fato da penúltima rede listada apresentar uma estrutura intermediária entre os dois extremos (regular e aleatório).



Fonte: Produzido pelo autor.

Outro modelo fundamental, é o de Redes Livres de escala, sob o qual trataremos em maiores detalhes no capítulo seguinte.

³Esse conceito tem sua origem no famoso experimento de Milgran, para mais detalhes ver referencia [10].

3 REDES LIVRES DE ESCALA

O modelo de *Redes Livres de Escala (Scale-Free Networks)* é, dentre todos, o que apresenta maior universalidade, no sentido de que grande parte dos sistemas reais tem suas propriedades melhor mapeadas e reproduzidas por este tipo de rede. Neste capítulo, apresentaremos os principais aspectos e características desse modelo de redes.

3.1 Limitações do Modelo Aleatório

O campo da teoria de grafos, historicamente, é amplamente estudado sob a ótica dos modelos aleatórios, onde se destaca o chamado *Modelo de Erdős–Rényi (ER)*¹ [29]. Como dito no capítulo anterior, redes desse tipo seguem distribuições de conectividade do tipo Poisson, como segue

$$p(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}. \quad (3.1)$$

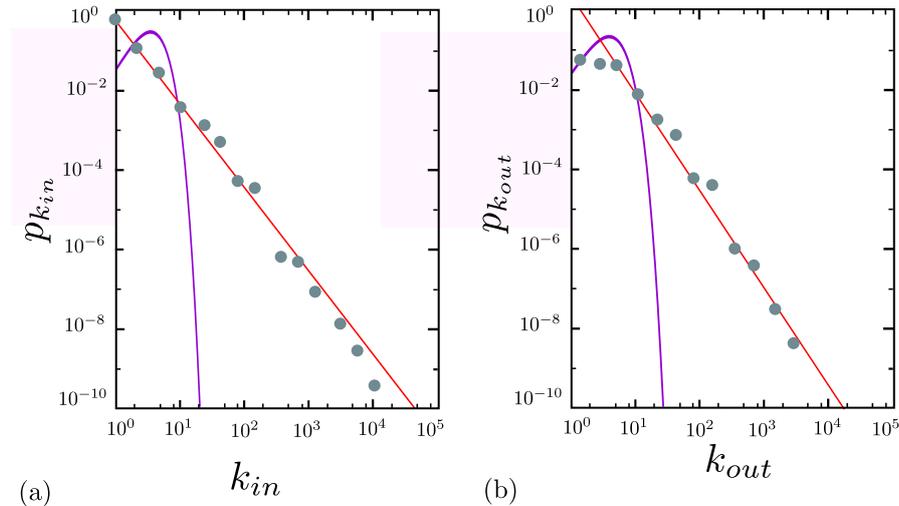
No entanto, vários casos indicam que este padrão de distribuição não reproduz, de fato, a maioria das redes reais, visto que estas possuem propriedades que não são típicas de distribuições normais [9, 11].

Como exemplo, tomemos a WWW (Word Wide Web), uma rede real onde os vértices são documentos e os links são URLs, que permitem acessar um documento a partir de outro através de um clique [9]. Para uma rede dessa natureza, Hawoong Jeong mapeou o domínio nd.edu [9], tal domínio constava de 325.729 documentos e 1.469.680 links [30]. Com esses dados, um estudo publicado por Hawoong, em colaboração com Réka Albert e Albert-László Barabási, intitulado “*The Diameter of the Word Wide Web*” (O Diâmetro da Word Wide Web), em 1999, reproduziu a topologia dessa rede (ver referência [30]). Um dos resultados mostra a distribuição de conectividade de entrada (links que direcionam a um certo documento) e de saída (links que partem de um dado documento) (ver gráfico 2).

Ao se observar a distribuição de conectividade de entrada e de saída (ver gráfico 2), percebe-se que existe uma falha no ajuste da curva referente à distribuição normal, enquanto uma curva em lei de potência consegue reproduzir o comportamento das distribuições de conectividade (entrada e saída) da rede em estudo. Seguindo a ideia de que esta rede poderia ser modelada por uma rede aleatória, não deveriam existir nós com ordem de conectividade elevada (em relação ao valor médio) como revela o gráfico 2. No entanto, tal fato é perfeitamente

¹Trata-se de um modelo para a construção de grafos aleatórios idealizado por Paul Erdős e Alfréd Rényi, em 1959.

Gráfico 2 – Distribuição de conectividade do domínio nd.edu



Fonte: Adaptado de Barabási: *Network Science*. (a) e (b) mostram respectivamente a distribuição de conectividade de links de entrada e de saída em escala logarítmica. Os círculos representam os dados empíricos obtidos por Albert *et al.* [30], as curvas em lilás representam a tentativa de fit por meio de uma distribuição normal e as retas em vermelho em lei de potência. A inclinação das retas determinam o expoente γ como sendo $\gamma_{in} = 2.1$ e $\gamma_{out} = 2.45$, já a curva dada pela distribuição de Poisson tenta ajustar os dados tendo como parâmetros os valores médios da rede $\langle k_{in} \rangle = \langle k_{out} \rangle = 4.60$.

previsível se tomarmos como modelo uma distribuição de conectividade do tipo $p(k) \sim k^{-\gamma}$. Assim, indicamos que redes aleatórias não reproduzem, como evidenciado neste exemplo, a estrutura de certas redes reais.

3.2 Leis de Potência e Redes Livres de escala

O exemplo da WWW, mostrado na seção anterior, revela a emergência de um novo modelo de rede complexa, tendo este uma distribuição de conectividade característica dada por,

$$p_k = Ck^{-\gamma}. \quad (3.2)$$

Como dito anteriormente, esse tipo de expressão é designada como *Lei de Potência*, onde, para o caso de redes complexas, C é uma constante a ser determinada pela condição de normalização e γ é o expoente característico da distribuição chamado *parâmetro de escala*. Note que, ao tomarmos o logaritmo da equação 3.2, temos

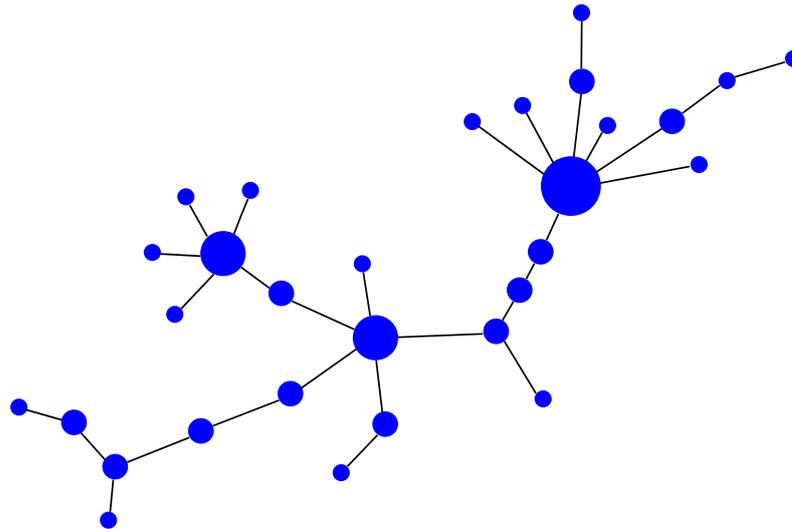
$$\ln p_k = -\gamma \ln k + C_0. \quad (3.3)$$

O que significa que o expoente γ representa a inclinação da reta em escala logarítmica (ver gráfico 2).

Redes que obedecem distribuições de conectividade da natureza da equação 3.2, são chamadas de *Redes Livres de Escala (Scale Free Network -SF)* [9, 28]. O termo *livre de escala*

refere-se ao fato de que, como veremos na seção 3.5, para um regime específico do expoente γ , as flutuações em torno do valor da conectividade média serem arbitrariamente grandes ou mesmo tender ao infinito no limite onde o número de vértices da rede tende a valores muito altos ($N \rightarrow \infty$). A Figura 9 mostra um exemplo de estrutura de uma rede livre de escala, representada, neste caso, por um tipo específico de grafo chamado *Árvore*².

Figura 9 – Grafo livre de escala .



Fonte: Produzido pelo autor. Exemplo clássico de uma estrutura em lei de potência ($\gamma \cong 2,5$ para este caso), uma árvore exibe características específicas, como por exemplo, o fato de haver apenas um único caminho possível entre qualquer par de vértices. Redes livres de escala, no entanto, podem exibir estruturas bem mais complexas do que essa, por exemplo, com a presença de múltiplos ciclos.

Vários são os exemplos de sistemas reais cujas estruturas podem ser modeladas por redes com distribuição tipo leis de potência, a saber, a colaboração entre atores em filmes, redes de malhas energéticas, redes de citações em artigos científicos, etc [32]. Tal dado, revela a universalidade das redes livre de escala. De fato, a maioria dos sistemas reais de que se tem informação na literatura são melhor representadas por redes desse tipo, tipicamente com expoentes na faixa $2 \leq \gamma \leq 3$, apesar de valores acima desse intervalo serem também possíveis [12].

Apesar dos bons resultados, na prática, poucas distribuições de dados empíricos são exatamente ou perfeitamente leis de potência em toda sua extensão. Ela se aplica na maioria dos casos para valores a partir de um certo valor mínimo ou mesmo para a “cauda da curva”[12,33]. Assim como pode haver alguma divergência em regimes de grandes valores de k como, por exemplo, um valor máximo possível de conectividade para os vértices mais conectados [12,34].

²Grafo completamente conectado, não direcionado e acíclico. Um conjunto de árvores desconectadas é chamado *Floresta* (para mais detalhes ver referência [31])

3.3 Descrição de Distribuição em Lei de Potência

Para um melhor entendimento de redes livres de escala, faz-se necessário uma discussão acerca das principais propriedades estatísticas, como segue.

3.3.1 Condição de Normalização

Assim como qualquer outra distribuição de probabilidade, as distribuições de conectividade em forma de lei de potência devem obedecer à condição de normalização. Observamos essa condição tanto na forma discreta da distribuição, quanto na forma contínua.

3.3.1.1 Formalismo Discreto

A condição de normalização de 3.2 considerando o aspecto discreto, é dado por

$$\sum_{k=1}^{\infty} p_k = 1. \quad (3.4)$$

Assim, observando 3.2 e 3.4 temos que,

$$\begin{aligned} \sum_{k=1}^{\infty} Ck^{-\gamma} &= 1, \\ C &= \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}}. \end{aligned} \quad (3.5)$$

O termo $\sum_{k=1}^{\infty} k^{-\gamma}$ é a conhecida forma da *Função Zeta de Riemann*, denotada por $\zeta(\gamma)$. Dessa forma a constante de normalização C é determinada por,

$$C = \frac{1}{\zeta(\gamma)}. \quad (3.6)$$

Podemos então concluir que, para o regime discreto,

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}. \quad (3.7)$$

Note que não consideramos a possibilidade de existência de vértices com grau $k = 0$. Isso pois, para uma distribuição em lei de potência pura, a existência de vértices isolados implicaria uma probabilidade infinita ($p_{k=0} \rightarrow \infty$), provocando uma grave inconsistência, já que a probabilidade é sempre algo no intervalo entre 0 e 1.

Como dito no capítulo 2, a equação 3.7 pode ser interpretada como a probabilidade de que um nó aleatoriamente escolhido tenha exatamente conectividade k .

3.3.1.2 Formalismo Contínuo

De maneira análoga ao caso discreto, podemos estabelecer,

$$p(k) = Ck^{-\gamma}. \quad (3.8)$$

E impor novamente a condição de normalização,

$$\int_{k_{min}}^{\infty} p(k)dk = 1. \quad (3.9)$$

Combinando as equações 3.8 e 3.9, temos,

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma}} \quad (3.10)$$

Um processo de integração simples leva ao valor da constante que normaliza a distribuição, como segue,

$$C = (\gamma - 1)k_{min}^{(\gamma-1)}. \quad (3.11)$$

Dessa forma podemos determinar a distribuição normalizada como sendo

$$p(k) = (\gamma - 1)k_{min}^{\gamma-1}k^{-\gamma}. \quad (3.12)$$

onde k_{min} é o menor valor de conexão a partir do qual a lei de potência se aplica.

Para o caso contínuo, a interpretação estatística tem uma pequena variação, pois agora o fator

$$\int_{k_1}^{k_2} p(k)dk, \quad (3.13)$$

representa a probabilidade de que um vértice aleatoriamente escolhido tenha conectividade no intervalo entre k_1 e k_2 [9].

Como dito também anteriormente, é importante atentar ao fato de estas informações se referirem ao modelo na sua forma pura. A aplicação em dados empíricos geralmente requer adequações. Ainda assim, essas propriedades servem como ponto de partida para tais [12].

3.3.2 Momentos da Distribuição

Podemos determinar o n -ésimo momento de uma distribuição como segue,

$$\langle k^n \rangle = \sum_{k_{min}}^{\infty} k^n p_k \approx \int_{k_{min}}^{\infty} k^n p_k dk. \quad (3.14)$$

Pra o caso de redes livre de escala, temos

$$\langle k^n \rangle = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n - \gamma + 1}, \quad (3.15)$$

onde

$k_{min} \sim$ é um valor fixo

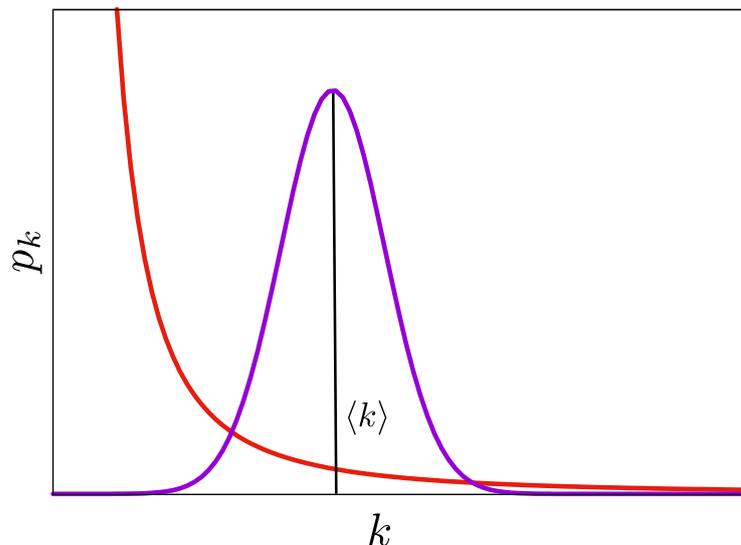
$k_{max} \sim$ é o grau do nó mais conectado

No limite em que $N \rightarrow \infty$ tem-se que, $k_{max} \rightarrow \infty$. Nesta situação, se $n \leq \gamma - 1$ o n -ésimo momento é finito, no entanto, se $n \geq \gamma - 1$ a equação 3.15 diverge, portanto, os momentos nessa condição também divergem.

O momento referente a $n = 1$ é o valor médio da distribuição $\langle k \rangle$. O segundo momento ($\langle k^2 \rangle$) é também uma propriedade estatística relevante, pois é necessária para obtenção da variância (σ_k) na forma,

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2. \quad (3.16)$$

Figura 10 – Comparação da dispersão em redes aleatórias e livres de escala.



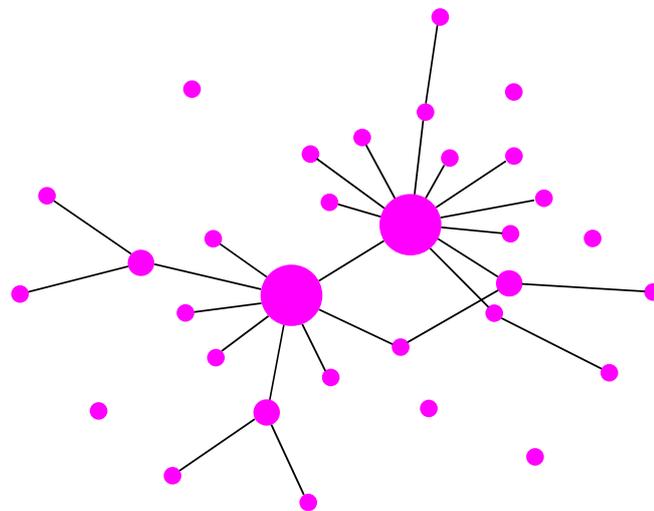
Fonte: Adaptado de Barabási: Network Science. A Figura mostra que, para uma distribuição do tipo normal, a probabilidade de encontramos vértices que distanciam do valor médio vai a zero rapidamente, revelando o fato de que nesse tipo de distribuição σ_k é pequeno, enquanto para a distribuição em lei de potência esse valor diverge ($\sigma_k \rightarrow \infty$).

A variância carrega a informação acerca de o quão dispersos são valores da distribuição em relação ao valor médio, ou seja, os valores devem ser distribuídos ao longo do intervalo $\langle k \rangle \pm \sigma_k$. Se ocorrer de o segundo momento da distribuição divergir, a variância também diverge, o que significa que as flutuações em torno do valor médio são muito grandes ou mesmo tendem ao infinito. Esse é justamente o fator que caracteriza a propriedade livre de escala. A Figura 10 mostra que esse intervalo é muito maior para redes em lei de potência em comparação a redes aleatórias.

3.4 Hubs

Uma das principais características de redes livres de escala, em detrimento à redes aleatórias, é a existência de vértices altamente conectados, em comparação ao valor médio. Verifica-se nesse tipo de rede a existência de muitos nós “pobres” em links e uma pequena quantidade de nós com um número muito grande de conexões. Estes nós são chamados de *Hubs* e despenham um papel fundamental na estrutura da rede. A Figura 11 mostra um exemplo pictórico de rede livre de escala onde é possível identificar os hubs.

Figura 11 – Hubs em redes livre de escala



Fonte: Produzido pelo autor. Os vértices mais conectados são realçados por um pelo seu tamanho, evidenciando a existência de dois vértices altamente conectados na rede.

A Figura 10 serve no entendimento da manifestação dessa propriedade, pois evidencia o fato de que, para redes livre de escala, existe uma probabilidade alta (p_k) de serem encontrados vértices com valores baixos de k e uma probabilidade pequena, embora existente, de serem encontrados vértices com altos valores de k . Este ultimo caso é expressamente proibido para redes aleatórias, uma vez que a probabilidade p_k vai a zero rapidamente para valores distantes do valor médio ($\langle k \rangle$). De maneira mais enfática, gráfico 2, referente a uma subrede da WWW, também é útil na verificação desse fato. Note que a reta em vermelho (plot em escala logarítmica de uma lei de potência) permite ou prediz a existência vértices ordens de grandeza maior do que o esperado para uma distribuição do tipo aleatória (curva lilás, também em escala logarítmica), mesmo as duas possuindo o mesmo valor médio.

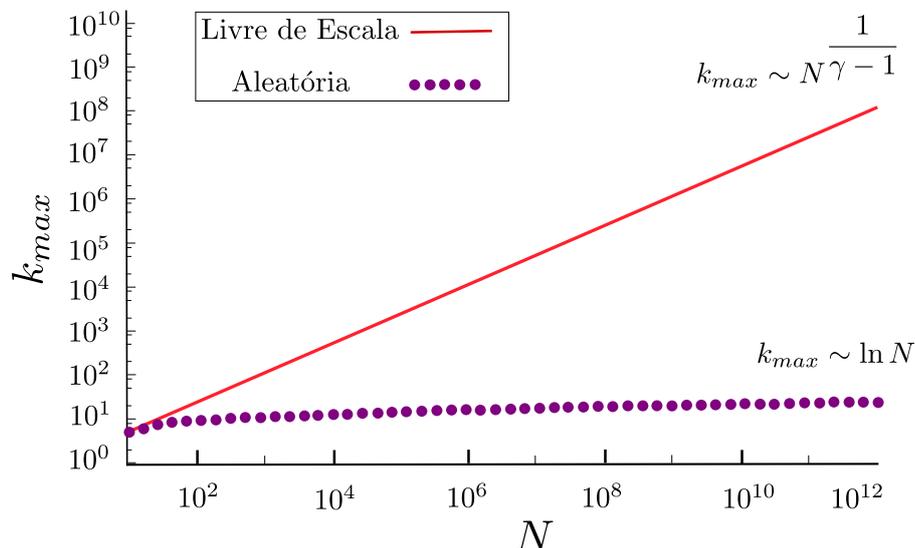
A existência dos hubs confere às redes livres de escala uma alta resistência à falhas (ou ataques) aleatórios, em contrapartida, estas são altamente vulneráveis à remoções direcionadas aos vértices mais conectadas [35, 36].

Como todas as redes reais são finitas, uma questão importante é conhecer para cada rede o tamanho ou a capacidade máxima esperada de conectividade dos vértices mais conectados. Esta propriedade é chamada de *natural cutoff* (corte natural) (k_{max}) e amplamente estudada na literatura, bem como suas implicações [34,37]. BARABÁSI (2016) determina a dependência do valor de k_{max} para redes livres de escala como sendo

$$k_{max} = k_{min} N^{\frac{1}{\gamma-1}}. \quad (3.17)$$

Onde k_{min} é o grau do nó menos conectado. A dependência de natureza polinomial, revela que

Gráfico 3 – Dependência de k_{max} com o tamanho da rede



Fonte: Adaptado de Brabási: Network Science. O gráfico mostra o comportamento de k_{max} em função do tamanho da rede N para uma rede em lei de potência e outra aleatória, ambas contendo o mesmo $\langle k \rangle = 3$. Para a rede em livre de escala, $\gamma = 2.5$.

a medida que o tamanho da rede cresce, os vértices mais conectados tendem a ser ordens de grandeza maiores do que os nós de conectividade mínima. Em comparação à redes aleatórias, onde $k_{max} \sim \ln N$ (ver referência [9]), esse comportamento é significativamente distinto, como evidenciado pelo gráfico 3.

3.5 A função do expoente γ

Uma informação relevante acerca de sistemas complexos modelados por redes livres de escala consiste na oferecida por meio do expoente da lei de potência (γ). Este fator traduz, de acordo com seu valor, características estatísticas e estruturais fundamentais de tais redes. Estas propriedades podem ser identificadas e agrupadas em regimes bem definidos, como listado

abaixo:

1. ($\gamma \leq 2$) - Usando os critérios da seção 3.3.2, podemos notar que para estes valores de γ o valor médio da distribuição diverge ($\langle k \rangle \rightarrow \infty$), assim como todos os momentos de ordem superior. O caminho médio da rede ($\langle d \rangle$) não depende do tamanho (N) [38], revelando um comportamento anômalo para este regime. Por estes e outros motivos, redes com esse regime de expoentes não podem existir (ver referência [9]).
2. ($2 < \gamma < 3$) - Novamente usando os critérios da seção 3.3.2, constatamos que os momentos de ordens superiores a $n = 2$ divergem. Isso confere, de fato, a propriedade livre de escala, indicando a existência de hubs altamente conectados. Redes no regime livres de escala exibem, devido à presença dos seus hubs “gigantes” a propriedade de *Ultra Mundo Pequeno*, onde o caminho mede assume a forma $\ln(\ln N)$ [38].
3. ($\gamma > 3$) - Neste regime são recuperadas propriedades de redes aleatórias, como o fato de o primeiro e o segundo momento serem finitos. Redes desse tipo exibem a propriedade de mundo pequeno, sendo $\langle d \rangle \sim \ln N$ [38]. Para grandes valores de N , essas redes são indistinguíveis de redes aleatórias.

4 PROBLEMA DE PERCOLAÇÃO EM REDES

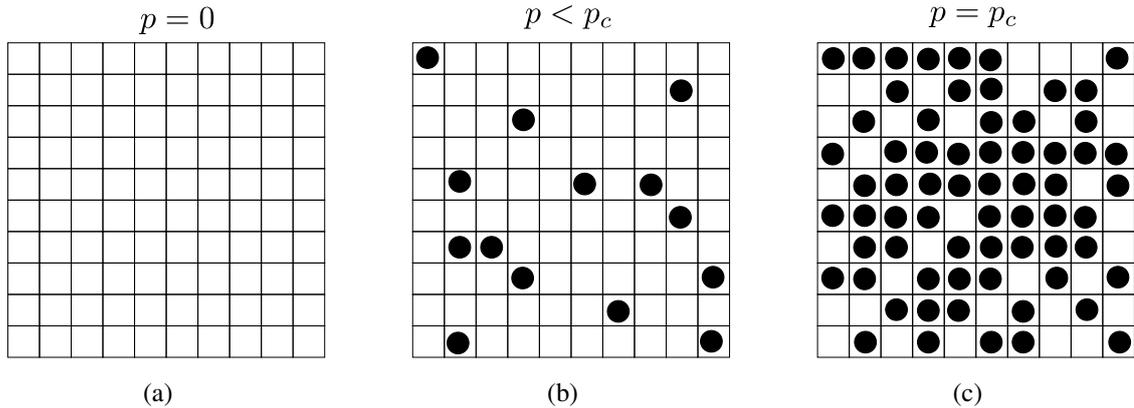
Um problema de especial relevância e amplamente estudado em Ciência de Redes reside em responder a seguinte pergunta: o quão resistente é uma rede, em relação à remoção categórica de vértices e/ou arestas? Ou seja, de que forma a estrutura se comporta sob esse tipo de ação até se desintegrar em pequenos fragmentos e suas implicações para cada sistema [7, 36, 39, 40]. O problema inverso também é de grande interesse: quando um conjunto fragmentado de nós torna-se uma rede quase totalmente conectada? A *Teoria da Percolação* é quem dispõe das ferramentas necessárias ao estudo desse fenômeno. Esse campo de estudo foi fundado por *Broadbent* e *Flantnansley* [41] em um artigo publicado no ano de 1957 e intitulado “*Percolation processes: I. Crystals and mazes*” (Processos de percolação: I. Cristais e desordens), no texto os autores introduzem a ideia de percolação por diversos exemplos práticos [42].

Um bom exemplo para o entendimento acerca dessa teoria consiste na seguinte situação [43, 44]: em uma folha, constituída de várias pequenas caixas, pontos são colocados aleatoriamente dentro de cada uma com probabilidade p , de maneira independente (ver Figura 12). Dois pontos em caixas vizinhas são ditos conectados, e formam um *cluster* de tamanho dois ou mais, dependendo de quantos pontos estão conectados em sequência. É de se esperar que o tamanho do maior cluster cresça à medida que a probabilidade p se aproxime de 1. No entanto, um fato não trivial, é que existe um limiar de probabilidade p_c a partir do qual praticamente todos os sítios da folha farão parte de um único cluster chamado de *cluster infinito* [43]. Essa mudança brusca que ocorre para a probabilidade p_c pode ser comparado a uma transição de fase¹ e, é a manifestação do fenômeno de percolação. Assim, podemos classificar a percolação como um fenômeno crítico e a probabilidade p_c como o ponto crítico na qual as propriedades do sistema são alteradas [43].

Para as mais distintas topologias de rede, a percolação se manifesta, com seus respectivos pontos de criticidade, no surgimento de componentes gigantes, ou seja, uma componente na qual a probabilidade p_∞ de um nó aleatoriamente escolhido ser encontrado é praticamente 1, como veremos adiante.

¹Transições de fase tratam-se de processos físicos no qual algumas das propriedades dos materiais são abruptamente alteradas, como por exemplo a transição entre os estados físicos da água.

Figura 12: Percolação e clusters



Fonte: Produzido pelo autor. (a) mostra uma folha onde não há probabilidade de nenhuma caixa ser ocupada por um ponto ($p = 0$). (b) Mostra o surgimento de alguns pontos na folha com probabilidade abaixo da probabilidade crítica, pares caixas vizinhas e povoadas formam pequenos clusters. (c) Mostra uma configuração ilustrativa do sistema na probabilidade crítica, onde quase todos ou todos os pontos formam um único cluster gigante. Este exemplo se aproxima á percolação de uma rede quadrada.

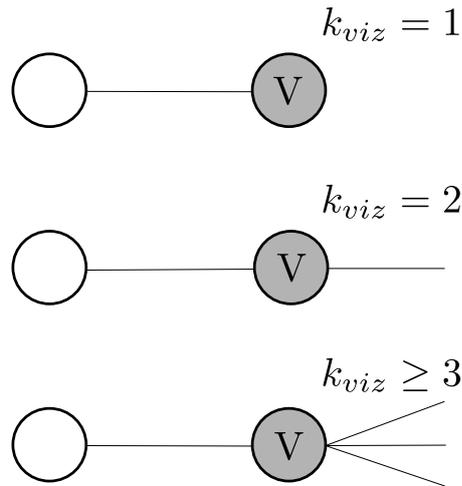
4.1 Solução combinatória de Erdős–Rényi

Como mencionado no capítulo 3, o modelo de Erdős–Rényi (ER) é usado para gerar grafos aleatórios. Vamos abordar agora sob um aspecto combinatório a solução do problema de percolação em grafos acíclicos aleatórios.

Existem varias soluções do problema de Erdős–Rényi. Uma simples segue o seguinte raciocínio. Seguindo uma conexão encontramos um vértice com grau k_{viz} . Como seguimos uma conexão aleatória, existe a possibilidade de encontrar um vértice de grau 1, que chamamos folha, nesse caso o ramo dessa conexão é um terminal. Existe a possibilidade de encontramos um vértice de grau 3 ou mais, nesse caso o ramo se bifurca. Finalmente, se encontramos um vértice de grau 2 o ramo segue sem bifurcar (ver Figura 13). Se seguimos um grande número de conexões, o número esperado de conexões depois dos vértices vizinhos será alterado de um fator $\langle k_{viz} \rangle - 1$. Essa redução de -1 da conta da conexão que foi seguida para atingir o vértice vizinho. Ou seja, se $\langle k_{viz} \rangle > 2$ o número de vértices encontrados deve crescer rapidamente a cada passo, atingindo uma fração considerável da rede. Caso contrário, se $\langle k_{viz} \rangle < 2$ esse número provavelmente vai diminuir e vai se atingir apenas um pequeno número de vértices quando comparado com a rede toda. O caso $\langle k_{viz} \rangle = 2$ é o ponto crítico. Note que não consideramos que algumas dessas conexões podem ser dirigidas aos mesmos vértices. Essa hipótese pode ser justificada em algumas situações, e o caso de Erdős–Rényi é um desses. Vamos voltar a esse ponto mais adiante.

O ponto crítico depende então do valor de $\langle k_{viz} \rangle$. Note que esse valor difere do grau médio $\langle k \rangle = 2|E|/|N|$, lembrando que $|E|$ é o número de conexões e $|N|$ é número de vértices.

Figura 13: Conectividade do vértice vizinho



Os vértices destacados à direita exemplificam as situações onde um ramo chega um terminal ($k_{viz} = 1$), segue sem se bifurcar ($k_{viz} = 2$) ou se ramifica ($k_{viz} \geq 3$).

Obviamente seguindo uma conexão não se pode encontrar um vértice de grau nulo. Da mesma forma, um vértice de grau alto é mais provável de ser o vizinho encontrado aleatoriamente do que um de grau 1, por exemplo. De forma geral, é razoável que a probabilidade $P_{viz}(k_{viz})$ de seguindo uma conexão encontrar um vértice de grau k_{viz} é proporcional ao grau do vértice, $P_{viz} = P(k)k_{viz}/\langle k \rangle$, onde $P(k)$ é a distribuição de graus.

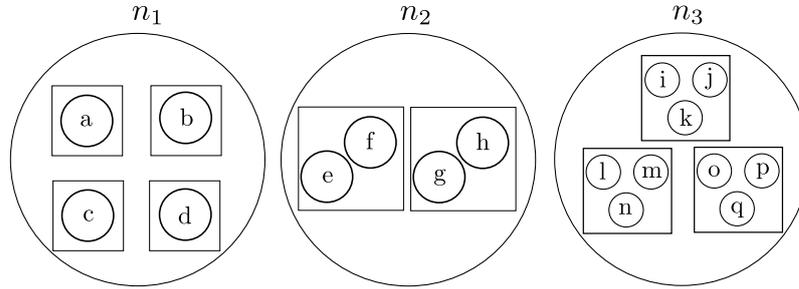
Temos então que determinar a distribuição de graus $P(k)$ no modelo de Erdős-Rényi. Para isso definimos $n_k = NP(k)$ o número esperado de vértices com grau k . Vamos obter n_k no caso aleatório pelo método da maximização da enumeração. Vamos imaginar cada vértice como uma caixa. O grau de um vértice pode ser representado pelo número de elementos em uma caixa. Como cada conexão aumenta o grau de dois vértices, temos um total de $2|E|$ elementos para serem distribuídos entre $|N|$ caixas. Temos de determinar o número $\Omega_1[n_k]$ de formas de distribuir os elementos, dado que n_k caixas terão k elementos, como na Figura 14, onde os elementos são representados por caracteres do alfabeto e as caixas por retângulos, bem como circunferências que representam grupos com a mesma ocupação. Para decidir que elemento é colocado em qual caixa colocamos os elementos e as caixas em ordens aleatórias. Temos então um fator $(|N|!.2.|E|!)$. Se trocamos a ordem de dois elementos dentro de uma mesma caixa não muda a distribuição dos elementos. Temos então que dividir nossa enumeração por $\prod k!^{n_k}$. Se trocamos duas caixas com o mesmo número de elementos também não muda a distribuição e temos que dividir também por $\prod n_k!$. Finalmente chegamos a

$$\Omega_1[n_k] = \frac{2|E|!N!}{\prod n_k!k!^{n_k}}. \quad (4.1)$$

Como as redes são amostradas aleatoriamente a distribuição n_k mais provável é a que maximiza

$\Omega[n_k]$ com os vínculos

Figura 14 – Arranjo por quantidade de elementos



Fonte: Produzido pelo autor. Os elementos foram ordenadas em ordem alfabética e ordem crescente de ocupação, sendo que o conjunto ou n_1 tem quatro caixas de tamanho 1 ($n_{S=1} = 4$), n_2 duas caixas de tamanho 2 ($n_{S=2} = 2$) e, n_3 , três caixas de tamanho 3 ($n_{S=3} = 3$).

$$\sum n_k = N \quad \text{e} \quad \sum kn_k = 2E. \quad (4.2)$$

Para encontrar tal distribuição, primeiramente calculamos $\log \Omega_1[n_k]$, considerando que $N \gg 1$.

$$\ln \Omega_1[n_k] = \ln \left(\frac{2E!N!}{\prod_k (k!)^{n_k} \prod_k n_k!} \right)$$

$$\ln \Omega_1[n_k] = \ln 2E!N! - \sum_k (n_k \ln k! + \ln n_k!) \quad (4.3)$$

usando a *aproximação de Stirling*:

$$\ln x! \sim x(\ln x - 1) \quad \text{desde que } x \gg 1,$$

na equação 4.3, temos

$$\ln \Omega_1[n_k] = \ln 2E!N! - \sum_k n_k (\ln k! + \ln n_k - 1) \quad (4.4)$$

O ponto de máximo da equação 4.4 pode ser encontrado por meio da diferenciação de $\ln \Omega_1$ em relação a n_k , como segue.

$$\begin{aligned} \delta \ln \Omega_1[n_k] &= -\delta \sum_k n_k (\ln k! + \ln n_k - 1) \\ \delta \ln \Omega_1[n_k] &= -\sum_k \delta n_k (\ln k! + \ln n_k - 1 + n_k \frac{1}{n_k}) \\ \delta \ln \Omega_1[n_k] &= -\sum_k \delta n_k (\ln k! + \ln n_k) \end{aligned}$$

com os vínculos dados (equação 4.2), fazemos

$$\begin{aligned}
\delta f(\ln \Omega_1[n_k], \lambda_1, \lambda_2) &= 0 \\
\delta(\ln \Omega_1[n_k] - \lambda_1 \sum_k n_k - \lambda_2 \sum_k kn_k) &= 0 \\
-\sum_k \delta n_k (\ln k! + \ln n_k) - \lambda_1 \sum_k \delta n_k - \lambda_2 \sum_k \delta kn_k &= 0 \\
-\sum_k \delta n_k (\ln k! + \ln n_k + \lambda_1 + \lambda_2 k) &= 0
\end{aligned}$$

Com a inclusão dos multiplicadores de Lagrange, podemos considerar os δn_k independentes e

$$\begin{aligned}
\ln n_k &= -\lambda_1 - \lambda_2 k - \ln k! \\
n_k &= \frac{e^{-\lambda_1} e^{-\lambda_2 k}}{k!}.
\end{aligned} \tag{4.5}$$

Vamos chamar $e^{-\lambda_1} = C_1$ e $e^{-\lambda_2} = C_2$, assim a equação 4.5 fica

$$n_k = \frac{C_1 C_2^k}{k!}. \tag{4.6}$$

dos vínculos sabemos que,

$$\begin{aligned}
N = \sum_k n_k &= \sum_{k=0}^{\infty} C_1 \frac{C_2^k}{k!} \\
N &= C_1 e^{C_2},
\end{aligned} \tag{4.7}$$

e

$$\begin{aligned}
2E &= \sum_k kn_k = C_1 \sum_{k=0}^{\infty} \frac{C_2^k k}{k!} \\
2E &= C_1 \sum_{k=0}^{\infty} C_2 \frac{d}{dC_2} \frac{C_2^k}{k!} = C_1 C_2 \sum_{k=0}^{\infty} \frac{d}{dC_2} \frac{C_2^k}{k!} \\
2E &= C_1 C_2 \frac{d}{dC_2} e^{C_2} = C_1 C_2 e^{C_2}.
\end{aligned} \tag{4.8}$$

Juntando as equações 4.7 e 4.8, temos

$$\begin{aligned}
2E &= NC_2 \\
C_2 &= \frac{2E}{N} = \mu,
\end{aligned} \tag{4.9}$$

portanto, μ é justamente o valor médio de conectividade. Substituindo a equação

4.9 na equação 4.7, temos

$$\begin{aligned} N &= C_1 e^\mu \\ C_1 &= \frac{N}{e^\mu} \end{aligned} \quad (4.10)$$

De posse da informação acerca das constantes C_1 e C_2 (equações 4.9 e 4.10), podemos finalmente encontrar a distribuição n_k , substituindo estes resultados na equação 4.6.

$$n_k = \frac{N\mu^k}{k!e^\mu} \quad (4.11)$$

Podemos então concluir que a distribuição de grau no modelo de Erdős-Rényi é a distribuição de Poisson

$$P(k) = e^{-\mu} \mu^k / k!. \quad (4.12)$$

Os momentos da distribuição podem ser determinados via

$$\langle k^n \rangle = \sum_{k=0}^{\infty} k^n p(k), \quad (4.13)$$

onde $n = 1$ é o valor médio da distribuição calculado como

$$\begin{aligned} \langle k \rangle &= \sum_{k=0}^{\infty} k p(k) \\ \langle k \rangle &= \sum_{k=0}^{\infty} k e^{-\mu} \mu^k / k! \\ \langle k \rangle &= e^{-\mu} \sum_{k=0}^{\infty} k \mu^k / k! \\ \langle k \rangle &= e^{-\mu} \left(\mu + \frac{2\mu^2}{2!} + \frac{3\mu^3}{3!} + \frac{4\mu^4}{4!} \dots \right) \\ \langle k \rangle &= e^{-\mu} \mu \left(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} \dots \right) \\ \langle k \rangle &= e^{-\mu} \mu \\ \langle k \rangle &= \mu. \end{aligned} \quad (4.14)$$

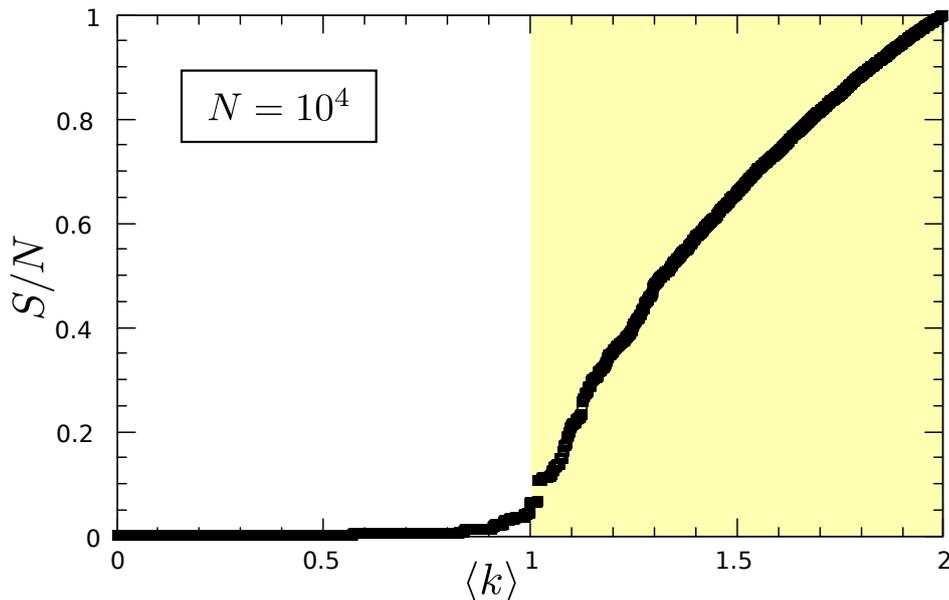
Bem como para o segundo momento ($n = 2$),

$$\begin{aligned} \langle k^2 \rangle &= \sum_{k=0}^{\infty} k^2 e^{-\mu} \mu^k / k! \\ \langle k^2 \rangle &= \sum_{k=0}^{\infty} (k(k-1) + k) e^{-\mu} \mu^k / k! \\ \langle k^2 \rangle &= \sum_{k=0}^{\infty} k(k-1) e^{-\mu} \mu^k / k! + \sum_{k=0}^{\infty} k e^{-\mu} \mu^k / k! \end{aligned}$$

$$\begin{aligned}
\langle k^2 \rangle &= e^{-\mu} \sum_{k=0}^{\infty} k(k-1)\mu^k/k! + \langle k \rangle \\
\langle k^2 \rangle &= e^{-\mu} \left(\frac{2\mu^2}{2!} + \frac{6\mu^3}{3!} + \frac{12\mu^4}{4!} + \frac{20\mu^5}{5!} + \dots \right) + \mu \\
\langle k^2 \rangle &= e^{-\mu} \mu^2 \left(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots \right) + \mu \\
\langle k^2 \rangle &= e^{-\mu} \mu^2 e^{\mu} + \mu \\
\langle k^2 \rangle &= \mu(\mu + 1).
\end{aligned} \tag{4.15}$$

Chegando, dessa forma, a $\langle k \rangle = \mu$ e $\langle k^2 \rangle = \mu(1 + \mu)$, o que leva a $\langle k_{viz} \rangle = 1 + \mu$. O ponto crítico de Erdős-Rényi acontece quando o grau médio é $\mu = 1$. Esse resultado verifica-se numericamente, como no gráfico 4, onde a simulação feita para uma rede de 10^4 vértices mostra a evolução do tamanho da maior componente da rede em função da conectividade média, evidenciando a mudança no regime próximo da região onde $\langle k \rangle = 1$, tal comportamento sugere de fato a existência de uma transição de fase.

Gráfico 4 – Tamanho da maior componente em função de $\langle k \rangle$



Fonte: Produzido pelo autor. Para uma rede de $N = 10^4$ vértices, o gráfico mostra a formação de uma componente gigante que cresce abruptamente a partir de $\langle k \rangle$ próximo a 1.

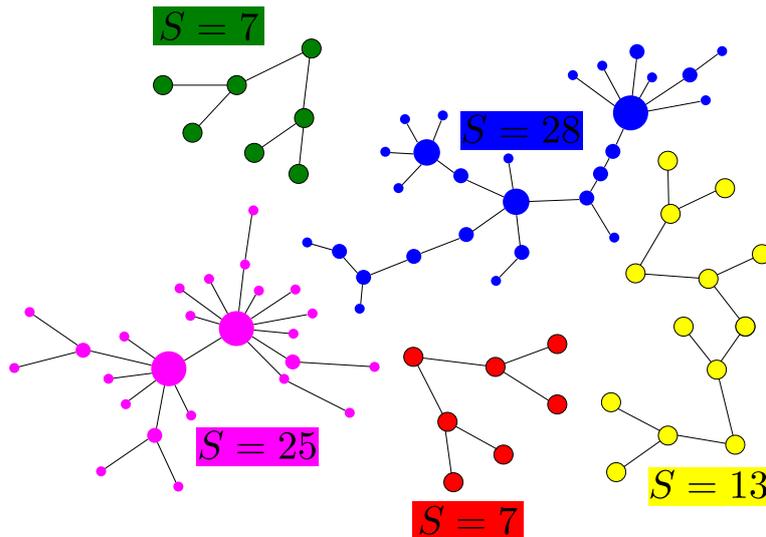
4.2 Distribuição de tamanho de componentes

Utilizando agora o método da maximização para determinar a distribuição de tamanhos de agregados no modelo de Erdős-Rényi, ou seja, a enumeração das formas de ligar N vértices rotulados em um grafo acíclico de m componentes de tal forma que n_1 componentes tenham um vértice, n_2 tenha dois vértices e assim por diante (ver Figura ??), que denotaremos por $\Omega_2[n_S]$. Assim como antes, devemos escolher quais dos N vértices entram em cada componente, sendo irrelevante a ordem, isso dá $\frac{N!}{\prod_{S=1} (S!)^{n_S}}$ formas distintas, como já visto. Agora devemos ligar os vértices dentro de cada componente. Um importante resultado em teoria dos grafos, conhecido como *Fórmula de Cayley*, diz que existem N^{N-2} diferentes grafos acíclicos conectados, ou seja, árvores, com N vértices rotulados, como provado no anexo A [45, 46]. Sendo assim, a fórmula de Cayley enumera todas as possíveis formas de ligação em cada componente. Como temos n_S componentes de mesmo tamanho, para todos os tamanhos teremos $\prod_{S=1} S^{(S-2)n_S}$ formas de arranjos, mas sempre que permutamos duas ou mais componentes de mesmo tamanho não teremos um novo estado, dessa forma devemos dividir esse fator pelas $\prod_{S=1} n_S!$ repetições e, finalmente, as configurações possíveis são

$$\Omega_2[n_S] = N! \prod_{S=1} \left[\left(\frac{S^{S-2}}{S!} \right)^{n_S} \frac{1}{n_S!} \right]. \quad (4.16)$$

A distribuição n_S é encontrada, assim como antes, pelo procedimento de maximização de

Figura 15: Tamanhos de componentes em uma floresta



A distribuição para este exemplo é : $n_{S=7} = 2$, $n_{S=13} = 1$, $n_{S=25} = 1$ e $n_{S=28} = 1$.

$\Omega_2[n_S]$. Vamos considerar que m e N são grandes e que são obedecidos aos vínculos

$$\sum n_S = m \quad \text{e} \quad \sum S n_S = N. \quad (4.17)$$

Tomando o logaritmo da equação 4.16, temos

$$\begin{aligned} \ln \Omega_2[n_S] &= \ln N! + \sum_S \left((\ln S^{S-2})^{n_S} - \ln(S!)^{n_S} - \ln n_S! \right) \\ \ln \Omega_2[n_S] &= \ln N! + \sum_S \left(n_S (\ln S^{S-2} - \ln S!) - \ln n_S! \right). \end{aligned} \quad (4.18)$$

Usando a aproximação de Stirling para $\ln n_S!$ na equação 4.18, vem

$$\ln \Omega_2[n_S] = \ln N! + \sum_S n_S \left((S-2) \ln S - \ln S! - \ln n_S + 1 \right). \quad (4.19)$$

Novamente, tomamos a diferenciação de $\Omega_2[n_S]$ em relação a n_S .

$$\begin{aligned} \delta \ln \Omega_2[n_S] &= \sum_S \delta n_S \left((S-2) \ln S - \ln S! - \ln n_S + 1 \right) + n_S \frac{-1}{n_S} \\ \delta \ln \Omega_2[n_S] &= \sum_S \delta n_S \left((S-2) \ln S - \ln S! - \ln n_S \right) \end{aligned} \quad (4.20)$$

Com a inclusão dos multiplicadores de Lagrange e obedecendo aos vínculos das equações 4.17, fazemos

$$\begin{aligned} \delta (\ln \Omega_2[n_S] - \lambda_1 \sum n_S - \lambda_2 \sum S n_S) &= 0 \\ \sum_S \delta n_S \left((S-2) \ln S - \ln S! - \ln n_S - \lambda_1 - \lambda_2 S \right) &= 0 \\ (S-2) \ln S - \ln S! - \ln n_S - \lambda_1 - \lambda_2 S &= 0 \\ \ln n_S &= (S-2) \ln S - \ln S! - \lambda_1 - \lambda_2 S. \end{aligned} \quad (4.21)$$

Resolvendo a equação 4.21 para n_S , temos

$$n_S = \frac{S^{S-2} e^{-\lambda_1} e^{-S\lambda_2}}{S!} \quad (4.22)$$

Chamando $y = e^{-\lambda_1}$ e $x = e^{-\lambda_2}$ em 4.22, temos:

$$n_S = \frac{S^{S-2} y x^S}{S!} \quad (4.23)$$

Prosseguimos com a determinação das constantes x e y da equação 4.23, por meio dos vínculos 4.17.

$$\begin{aligned}
N &= \sum_S S n_S = \sum_S \frac{S y x^S S^{S-2}}{S!} \\
N &= \sum_S \frac{y x^S S^{S-1}}{S!} \\
N &= y \sum_S \frac{x^S S^{S-1}}{S!}
\end{aligned} \tag{4.24}$$

e

$$\begin{aligned}
m &= \sum_S n_S = \sum_S \frac{y x^S S^{S-2}}{S!} \\
m &= y \sum_S \frac{x^S S^{S-2}}{S!} = y \sum_S \frac{x^S S^{S-1}}{S S!} \\
m &= y \sum_S \int_0^x \frac{x^{S-1} S^{S-1}}{S!} \\
m &= y \sum_S \int_0^x \frac{x^S S^{S-1}}{x S!}
\end{aligned} \tag{4.25}$$

O somatório presente na equação 4.24 é a forma contraída da expansão em série de Taylor da função implícita $g(x)$ definida como $g(x)e^{-g(x)} = x$, calculada por meio do *Teorema da Inversão de Lagrange* [47]. Este e outros resultados usados adiante são provados no apêndice A. Ao identificar esse somatório nas equações 4.24 e 4.25, podemos substituí-los por $g(x)$ e, dessa forma, as reescrevemos respectivamente como

$$N = y g(x) \tag{4.26}$$

e

$$m = y \int_0^x \frac{g(x)}{x}. \tag{4.27}$$

Para um conjunto de grafos acíclicos construído a partir de N vértices e distribuídos em m componentes, existem $N - m$ arestas (E). O número médio de conexões por vértice μ é, portanto

$$\mu = \frac{2E}{N} = \frac{2(N - m)}{N}. \tag{4.28}$$

Isso pois, cada aresta conecta um par de vértices.

Substituindo os resultados para N e m das equações 4.26 e 4.27, respectivamente, na equação 4.28, temos

$$\mu = \frac{2 \left(y g(x) - y \int_0^x \frac{g(x)}{x} \right)}{y g(x)}. \tag{4.29}$$

No apêndice A é provado a relação $\frac{d}{dx} \left(g(x) \left(1 - \frac{g(x)}{2} \right) \right) = \frac{g(x)}{x}$, o que significa que a antiderivada presente na equação 4.29 é igual ao derivando da expressão proposta e, por esse motivo, podemos substituir a integral por $g(x) \left(1 - \frac{g(x)}{2} \right)$, como segue

$$\begin{aligned} \mu &= \frac{2 \left(yg(x) - yg(x) \left(1 - \frac{g(x)}{2} \right) \right)}{yg(x)} \\ \mu &= 2 \left(1 - 1 + \frac{g(x)}{2} \right) \\ \mu &= g(x). \end{aligned} \quad (4.30)$$

Descobrimos, portanto, que a função $g(x)$ se trata na verdade do número médio de ligações μ . Com base nessa informação podemos determinar as constantes x e y da distribuição na equação 4.23. Substituindo a equação 4.30 na 4.26, temos

$$y = \frac{N}{\mu} \quad (4.31)$$

e que, diretamente com a substituição da equação 4.30 na definição da função $g(x)$, determinamos

$$x = \mu e^{-\mu}. \quad (4.32)$$

Enfim, determinamos a distribuição n_S substituindo os resultados das equações 4.31 e 4.32 na equação 4.23

$$n_S = \frac{S^{S-2} (\mu e^{-\mu})^S}{S!} \left(\frac{N}{\mu} \right). \quad (4.33)$$

Para proceder com a análise, vamos reescrever e desenvolver 4.33 da seguinte forma

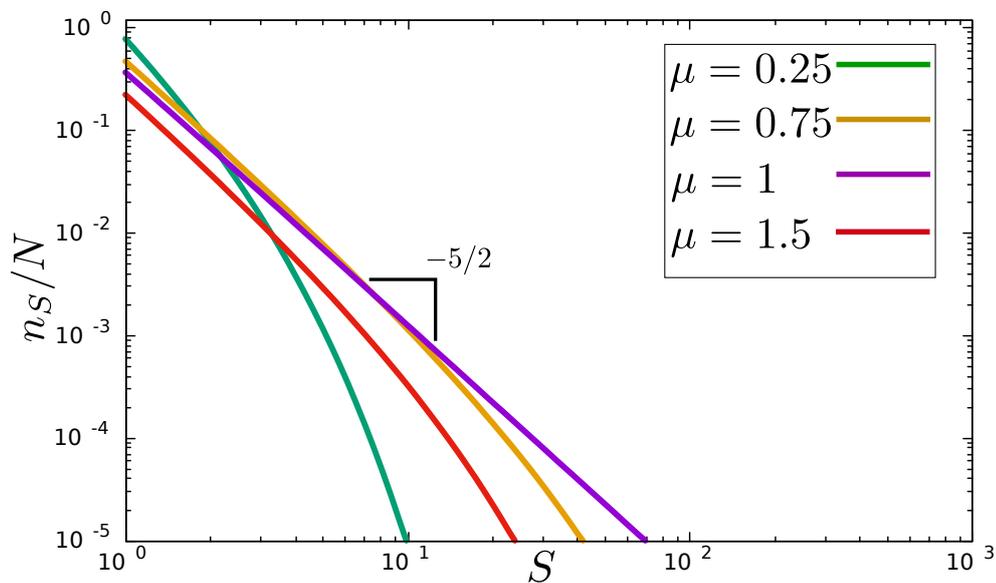
$$\begin{aligned} n_S &= \frac{NS^{S-2} \mu^S e^{-\mu S}}{S! \mu} \\ n_S &= \frac{NS^{S-2} \mu^{S-1} e^{-\mu S}}{S!} \end{aligned} \quad (4.34)$$

Para $S \gg 1$ na equação 4.34, podemos utilizar a *Fórmula de Stirling* aproximando $S! \sim \sqrt{2\pi S} \left(\frac{S}{e} \right)^S$. Além disso, vamos escrever também $\mu^{S-1} = e^{(S-1)\ln \mu}$, dessa forma, temos

$$\begin{aligned} n_S &\sim \frac{N}{\sqrt{2\pi}} \frac{S^{S-2} e^{(S-1)\ln \mu} e^{-\mu S}}{S^{1/2} \left(\frac{S}{e} \right)^S} \\ n_S &\sim S^{-5/2} e^{(S-1)\ln \mu + (-\mu S) + S} \\ n_S &\sim S^{-5/2} e^{-(\mu-1)S + (S-1)\ln \mu}. \end{aligned} \quad (4.35)$$

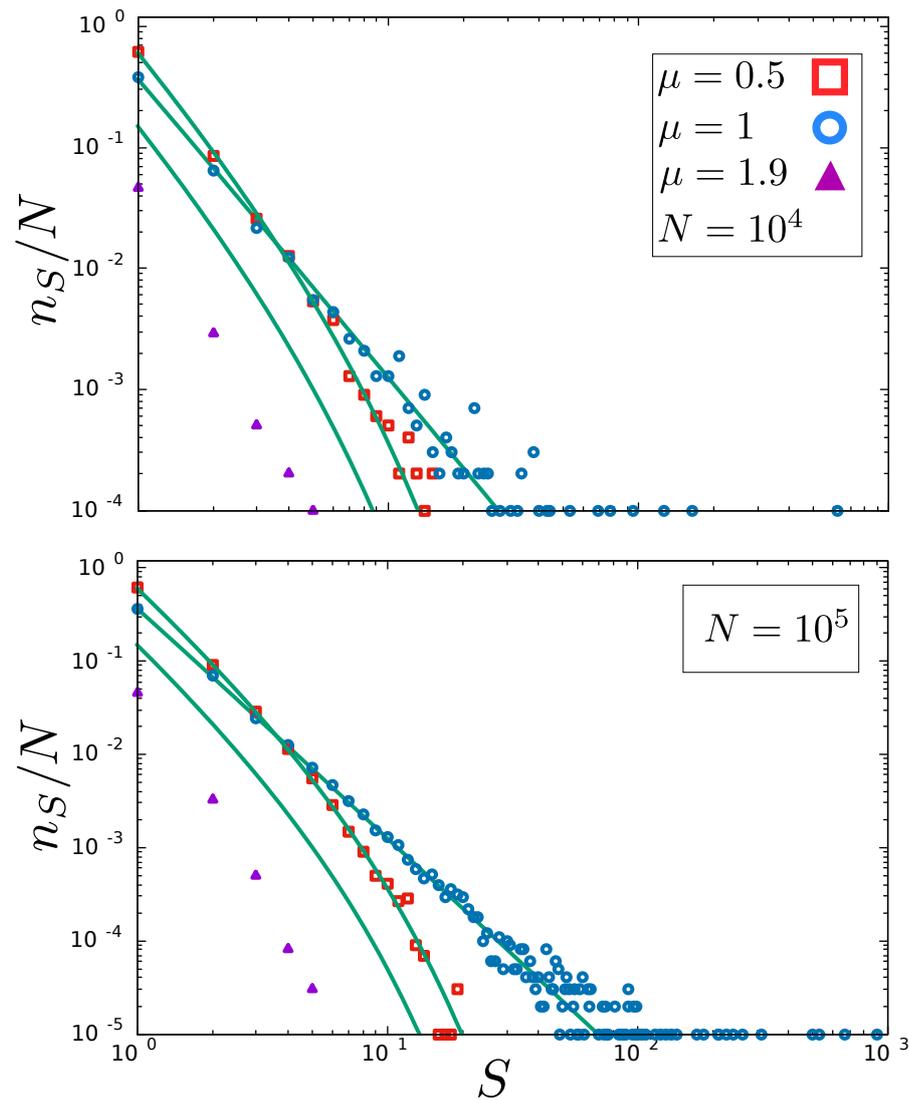
A equação 4.35 revela o que a distribuição de componentes n_s tem, predominantemente, um comportamento exponencial que decresce rapidamente para S grande. No entanto, quando $\mu = 1$ o termo exponencial desaparece e então teremos somente a contribuição em lei de potência devido ao termo $S^{-5/2}$, que decai muito menos rapidamente, em comparação à contribuição exponencial. Essa mudança de comportamento é típica de fenômenos de transição de fase e, nesse caso, a conectividade média $\mu = 1$, em concordância ao que foi visto na seção anterior, representa o ponto crítico, a partir do qual a floresta percola, e passa a compor praticamente apenas uma componente gigante.

Gráfico 5 – Previsão analítica da distribuição n_s



Fonte: Produzido pelo autor. Distribuição n_s para vários valores de μ , de acordo com a equação 4.34

O gráfico 5 mostra o comportamento da distribuição n_s para vários valores de μ , podemos observar que no ponto crítico são permitidas componentes para altos valores de S , o que é o indicativo do fato de que a partir desse momento, novas conexões certamente irão ligar grandes árvores, proporcionando o surgimento de uma componente que contém grande parte dos vértices, em acordo com as características de processos de transição de fase. Esse resultado é sustentado por simulações computacionais que, numericamente, calculam a distribuição de tamanho das componentes em grafos do tipo floresta, a partir da criação de grafos aleatórios para vários tamanhos de rede e para certos valores de conectividade média (μ), como mostrado no gráfico 6. O Resultado numérico mostra ainda que, excluindo a componente gigante, a distribuição se aproxima do previsto na equação 4.34 à medida que N cresce.

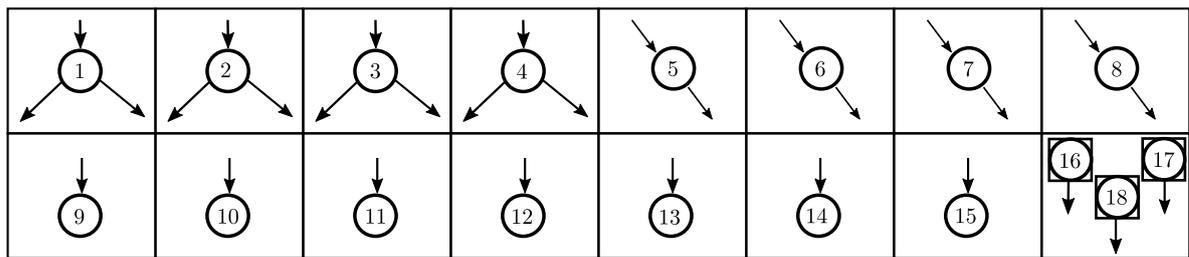
Gráfico 6 – Resultado numérico da distribuição n_s excluindo a componente gigante

Fonte: Produzido pelo autor. A simulação feita para redes de 10^4 e 10^5 revelam uma expressiva acurácia para a solução proposta da distribuição de tamanho de componentes antes do ponto crítico.

5 ENUMERAÇÃO DE ÁRVORES EM UMA FLORESTA

A exemplo dos mais diversos resultados de enumeração em teoria dos Grafos, buscamos enumerar agora, a partir de nossos argumentos combinatoriais, as Ω_0 formas de construir florestas enraizadas, direcionadas e rotuladas em T componentes, dado um conjunto de vértices onde N_k vértices tenham exatamente grau k de conectividade. Impomos a condição de que a raiz de cada árvore trata-se de uma folha, ou seja, um vértice de conectividade 1, e selecionamos dentre os vértices de grau 1 um grupo especial de T vértices para serem raízes, o que significa que estes vértices não farão parte aos pares em nenhuma componente, ou seja, estes vértices nunca podem estar na mesma componente. Nossos argumentos são construídos de maneira intuitiva vamos, portanto, estabelecer um exemplo prático de grande valor didático, tomemos o conjunto de vértices exemplificado na figura (16).

Figura 16 – Exemplo de conjunto



Fonte: Produzido pelo autor. Conjunto de vértices com distribuição de graus definida. Os vértices de rótulo 16,17 e 18 são os vértices especiais que serão raízes

Dado um conjunto de vértices com uma distribuição discreta de conectividade, o número de componentes será dada por

$$T = \sum_k \left(1 - \frac{k}{2}\right) N_k. \quad (5.1)$$

Dessa forma, para o conjunto da figura (16), podemos dividir os vértices do conjunto em exatamente três árvores. Assim, escolhamos raízes fixas os vértices destacados no último quadro (16,17,18).

Observemos os seguintes detalhes provenientes do fato das árvores serem direcionadas:

- As raízes apontam para um vértice e nunca são apontadas;
- Todo vértice que não é raiz deve ser apontado uma única vez;
- Os vértices com grau $k > 1$ devem apontar a outros $(k - 1)$ vértices;

- Os vértices de grau 1 que não são raízes nunca apontam a outro vértice.

Com base nessas observações, construímos um método simples e intuitivo de conectar estes vértices e consequentemente formar as componentes, descrito nas seguintes etapas:

1. Criamos uma lista L_1 de vértices “apontáveis”, que contém todos os vértices que não são raízes. O tamanho dessa lista será $(N - T)$, onde N é o número total de vértices.
2. Uma segunda lista L_2 contém os vértices que possuem grau ($k > 1$), ou seja, vértices que apontam e não são raízes. Cada um dos vértices com essa característica entra $(k - 1)$ vezes na lista, ou seja, o número de vezes que o vértice aponta na estrutura, ao final, essa lista terá $(N - 2T)$ elementos.
3. Escolhemos um dos elementos da lista L_2 a ser conectado com um elemento da lista L_1 . Em L_2 , podemos tomar qualquer um dos $(N - 2T)$ elementos, o que reflete um problema de arranjo com repetição, uma vez alguns desses vértices entram na lista mais de uma vez e todos serão selecionados. Assim, o primeiro fator da contagem carrega a informação acerca das formas de selecionar sequencialmente elementos de L_2 , na forma

$$\frac{(N - 2T)!}{\prod_k \left[(k - 1)! \right]^{N_k}}. \quad (5.2)$$

4. Uma vez selecionado o elemento de L_2 , devemos conectá-lo com um dos elementos da lista de apontáveis, sendo que, no primeiro passo temos $(N - T - 1)$ opções na lista L_1 , isso, pois um dos termos não é permitido. Ligar um vértice a ele mesmo levaria a um loop no grafo. No passo seguinte temos $(N - T - 2)$ opções, isso porque um dos elementos foi apontado e outro leva a um loop ou um ciclo. No j -ésimo passo, j opções não estão disponíveis, $j - 1$ por já terem sido apontados e 1 por levar à um loop ou um ciclo. Para o último elemento da lista L_2 haverá exatamente T opções disponíveis em L_1 . Portanto, podemos escrever nosso segundo fator da contagem como segue

$$(N - T - 1) \times (N - T - 2) \times (N - T - 3) \times \dots \times T = \frac{(N - T - 1)!}{(T - 1)!}. \quad (5.3)$$

A figura (17) exemplifica a iteração entre as etapas 3 e 4 para o conjunto da figura (16).

O resultado para as escolhas exemplificadas na figura (17) é a formação das componentes da figura (18).

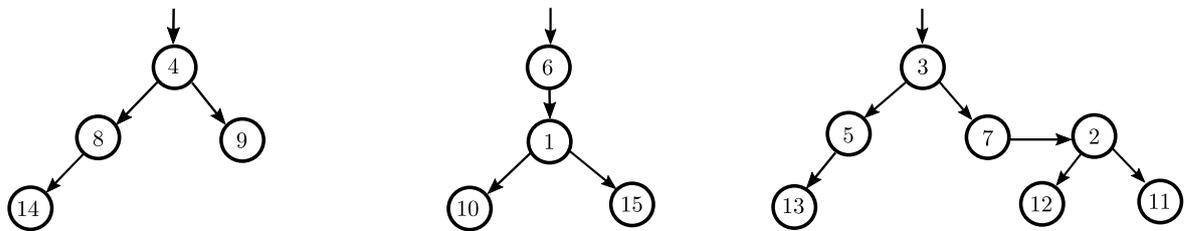
5. Resta agora enumerar todas as formas de conectar as T raízes nas componentes formadas, ou seja, conectar as raízes nos vértices apontáveis da lista L_1 que ainda não tenham sido apontados. Trivialmente podemos fazer isso de $T!$ formas distintas, portanto, esse é o

Figura 17 – Iteração entre as etapas 3 e 4 do método de ligação

| | | | | | | | | | | | | | | | |
|---------|----|----|----|----|---|----|-----|------|----|---|----|-----|---|---|---|
| Passo | I | II | II | IV | V | VI | VII | VIII | IX | X | XI | XII | | | |
| L_1 : | 10 | 15 | 12 | 11 | 5 | 7 | 8 | 9 | 13 | 1 | 2 | 14 | 3 | 6 | 4 |
| L_2 : | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 6 | 7 | 8 | | | |

Fonte: Produzido pelo autor. Cada vértice da lista L_2 é conectado a um da lista L_1 a cada passo como descrito nas etapas 3 e 4.

Figura 18 – Componentes formadas com as escolhas da Figura (17)

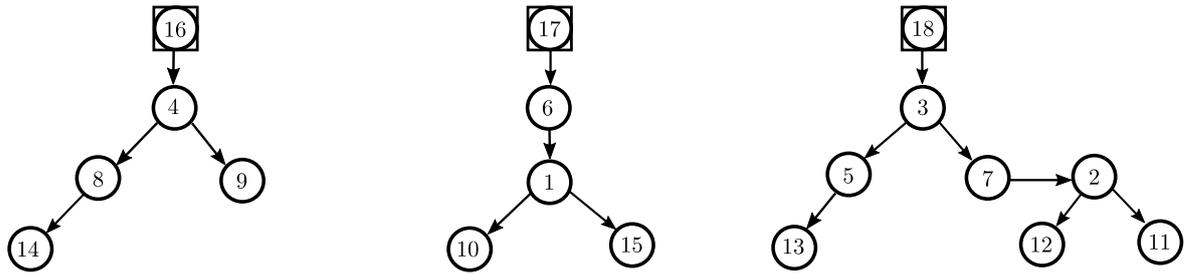


Fonte: Produzido pelo autor. A escolhas feitas no exemplo dado levam, dentre todas as possíveis, a essa configuração, resta apenas escolher das raízes estarão em cada componente.

terceiro fator da contagem. Podemos então finalizar uma das florestas possíveis para o exemplo de conjunto da figura (16), mostrada na figura (19).

6. O quarto e último fator da enumeração dá conta da quantidade de vezes que a formação de pares entre as duas listas leva à mesma configuração, decorrente do fato de conectar o mesmo conjunto de pares, apenas em uma ordem diferente. Esse fator é igual à quantidade de formas de tomar uma sequência de $(N - 2T)$ pares, dada por $(N - 2T)!$

Figura 19 – Floresta resultante



Fonte: Produzido pelo autor. Após a escolha das raízes dentre as combinações possíveis, a configuração final é estabelecida.

Juntando todos os fatores encontrados, encontramos finalmente

$$\Omega_0 = T \frac{(N - T - 1)!}{\prod_k \left[[(k - 1)!]^{N_k} \right]}. \quad (5.4)$$

A equação (5.4) enumera todo o *ensemble* de estados acessíveis a um sistema que possa descrito nos moldes apresentados na caracterização deste problema de contagem, assim, além de estabelecer um resultado para enumeração em teoria dos grafos, essa fórmula pode servir, em potencial, para estudos de sistemas físicos ou de outras áreas dos sistemas complexos a partir de um tratamento dela com base nos métodos da mecânica Estatística e da Termodinâmica. A verificar as possibilidades.

6 CONCLUSÕES E TRABALHOS FUTUROS

Com este trabalho, estabelecemos um estudo das características fundamentais de redes complexas, por meio da exploração do aspecto matemático primordial e das propriedades fenomenológicas, que vem servindo durante anos como modelo para estudos de sistemas reais.

Para um dos mais relevantes dos problemas envolvendo redes, o de percolação, propomos aqui uma solução com base em uma análise combinatória atrelada a maximização das configurações possíveis de organização para esse sistema. Utilizamos essa metodologia para encontrar a distribuição de graus do modelo aleatório e resolver o problema de percolação para grafos acíclicos. Obtivemos ainda como resultado uma expressão analítica para distribuição de tamanho de componentes como função da conectividade média da rede. A validade destes resultados adquire respaldo ao concordar com outros amplamente difundidos na literatura e também na sustentação por resultados numéricos, obtidos no decorrer do trabalho, através de simulações computacionais, realizadas através da construção de redes do referido tipo. Estes resultados apontam sempre uma transição de fase nos sistemas quando a conectividade média é maior que 1, sendo este o ponto crítico associado à percolação.

Ainda fazendo uso da metodologia proposta, usando ideias oriundas da análise combinatórias e arranjos, conseguimos estabelecer uma enumeração de todas as formas possíveis de construir uma floresta direcionada e enraizada a partir de um dado conjunto de vértices com uma certa distribuição de graus pré-estabelecida onde um certo número de folhas sejam tidos com um grupo exclusivo de raízes. Por sua vez, esse conjunto de estados acessíveis revela uma potencialidade de no tratamento de sistemas que de alguma forma possam se adequar às condições descritas, por meio dos conhecidos métodos da mecânica estatística, bem como os empregados na determinação para os resultados da distribuição de conectividade e da distribuição de tamanhos de componentes presentes no capítulo 4.

Com perspectivas ao desenvolvimento do trabalho, em relação aos resultados do capítulo 4, pretendemos estender o método combinatório proposto neste estudo à redes que permitam a formação de ciclos e loops entre seus vértices e mesmo ao modelo de Redes Livres de Escala. Em relação aos resultados obtidos capítulo 5, vislumbramos a possibilidade de estabelecer uma conexão entre sistemas físicos (ou mesmo de outros campos) de interesse e o modelo já descrito, com o intuito de buscar situações de interesse prático que possam ser estudados sob os conceitos e técnicas da mecânica estatística a partir da enumeração de estados acessíveis aqui desenvolvida.

REFERÊNCIAS

- [1] AZEVEDO, F. A. et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, Wiley Online Library, v. 513, n. 5, p. 532–541, 2009.
- [2] CHIALVO, D. R. Emergent complex neural dynamics. *Nature physics*, Nature Publishing Group, v. 6, n. 10, p. 744–750, 2010.
- [3] THURNER, S.; HANEL, R.; KLIMEK, P. *Introduction to the theory of complex systems*. [S.l.]: Oxford University Press, 2018.
- [4] GOLDENFELD, N.; KADANOFF, L. P. Simple lessons from complexity. *science*, American Association for the Advancement of Science, v. 284, n. 5411, p. 87–89, 1999.
- [5] GOH, K.-I.; KAHNG, B.; KIM, D. Universal behavior of load distribution in scale-free networks. *Physical review letters*, APS, v. 87, n. 27, p. 278701, 2001.
- [6] DOROGOVTSEV, S. N.; MENDES, J. F. Scaling properties of scale-free evolving networks: Continuous approach. *Physical Review E*, APS, v. 63, n. 5, p. 056125, 2001.
- [7] CALLAWAY, D. S. et al. Network robustness and fragility: Percolation on random graphs. *Physical review letters*, APS, v. 85, n. 25, p. 5468, 2000.
- [8] MEYERS, L. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, v. 44, n. 1, p. 63–86, 2007.
- [9] BARABÁSI, A.-L. et al. *Network science*. [S.l.]: Cambridge university press, 2016.
- [10] WATTS, D. J. *Six degrees: The science of a connected age*. [S.l.]: WW Norton & Company, 2004.
- [11] ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*, APS, v. 74, n. 1, p. 47, 2002.
- [12] NEWMAN, M. *Networks: An Introduction*. [S.l.]: Oxford University Press, 2010.
- [13] EULER, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, p. 128–140, 1741.
- [14] LOPES, F. J. A.; TáBOAS, P. Z. Euler e as pontes de königsberg. *Revista Brasileira de História da Matemática*, p. 23–32, 2015.
- [15] HARARY, F. *Graph theory*. [S.l.]: Addison-Wesley, 1969.
- [16] NETO, P. O. B. Grafos: teorias, modelos e algoritmos. 4 edição. *E. Blücher. São Paulo*, 2006.
- [17] BARRAT, A.; BARTHELEMY, M.; VESPIGNANI, A. *Dynamical processes on complex networks*. [S.l.]: Cambridge university press, 2008.

- [18] AGUIAR, S. G. de. *Contribuição ao estudo de Redes Complexas: Modelo de Afinidade com Métrica*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2012.
- [19] MATA, A. S. D. *Transições de Fase em um Processo de Reação-Difusão em metapopulações Heterogêneas*. Dissertação (Mestrado) — Universidade Federal de Viçosa, 2011.
- [20] COHEN, R.; HAVLIN, S. *Complex networks: structure, robustness and function*. [S.l.]: Cambridge university press, 2010.
- [21] DOROGOVTSSEV, S. N. *Lectures on complex networks*. [S.l.]: Oxford University Press New York, 2010. v. 24.
- [22] TANAKA, G.; MORINO, K.; AIHARA, K. Dynamical robustness in complex networks: the crucial role of low-degree nodes. *Scientific reports*, Nature Publishing Group, v. 2, p. 232, 2012.
- [23] HOLME, P. et al. Attack vulnerability of complex networks. *Physical review E*, APS, v. 65, n. 5, p. 056109, 2002.
- [24] SHAO, S. et al. Percolation of localized attack on complex networks. *New Journal of Physics*, IOP Publishing, v. 17, n. 2, p. 023049, 2015.
- [25] LEWIS, T. G. *Network science: Theory and applications*. Wiley Publishing, 2009.
- [26] STAM, C. J.; REIJNEVELD, J. C. Graph theoretical analysis of complex networks in the brain. *Nonlinear biomedical physics*, Springer, v. 1, n. 1, p. 3, 2007.
- [27] WANG, X.; CHEN, G. Complex networks: Small-world, scale-free and beyond. *Circuits and Systems Magazine, IEEE*, v. 3, p. 6 – 20, 02 2003.
- [28] COSTA, L. da F. et al. Characterization of complex networks: A survey of measurements. *Advances in Physics*, Taylor Francis, v. 56, n. 1, p. 167–242, 2007. Disponível em: <https://doi.org/10.1080/00018730601170527>.
- [29] NEWMAN, M. E.; STROGATZ, S. H.; WATTS, D. J. Random graphs with arbitrary degree distributions and their applications. *Physical review E*, APS, v. 64, n. 2, p. 026118, 2001.
- [30] ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Diameter of the world-wide web. *nature*, Nature Publishing Group, v. 401, n. 6749, p. 130–131, 1999.
- [31] LIU, C. L. *Elements of discrete mathematics*. [S.l.]: McGraw-Hill Computer Science Series, 1986. v. 2.
- [32] BARABÁSI, A. L., & albert, r.(1999). emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 509512.
- [33] CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. Power-law distributions in empirical data. *SIAM review*, SIAM, v. 51, n. 4, p. 661–703, 2009.
- [34] BOGUNÁ, M.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Cut-offs and finite size effects in scale-free networks. *The European Physical Journal B*, Springer, v. 38, n. 2, p. 205–209, 2004.

- [35] BARABÁSI, A.-L.; BONABEAU, E. Scale-free networks. *Scientific american*, JSTOR, v. 288, n. 5, p. 60–69, 2003.
- [36] ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Error and attack tolerance of complex networks. *nature*, Nature Publishing Group, v. 406, n. 6794, p. 378–382, 2000.
- [37] ZHOU, S.; MONDRAGÓN, R. J. Structural constraints in complex networks. *New Journal of Physics*, IOP Publishing, v. 9, n. 6, p. 173, 2007.
- [38] COHEN, R.; HAVLIN, S. Scale-free networks are ultrasmall. *Physical review letters*, APS, v. 90, n. 5, p. 058701, 2003.
- [39] COHEN, R. et al. Breakdown of the internet under intentional attack. *Physical review letters*, APS, v. 86, n. 16, p. 3682, 2001.
- [40] COHEN, R.; BEN-AVRAHAM, D.; HAVLIN, S. Percolation critical exponents in scale-free networks. *Physical Review E*, APS, v. 66, n. 3, p. 036113, 2002.
- [41] BOLLOBÁS, B. et al. *Percolation*. [S.l.]: Cambridge University Press, 2006.
- [42] BROADBENT, S. R.; HAMMERSLEY, J. M. Percolation processes: I. crystals and mazes. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society*. [S.l.], 1957. v. 53, n. 3, p. 629–641.
- [43] BUNDE, A.; HAVLIN, S. *Fractals and disordered systems*. [S.l.]: Springer Science & Business Media, 2012.
- [44] STAUFFER, D.; AHARONY, A. *Introduction to percolation theory*. [S.l.]: CRC press, 2018.
- [45] PITMAN, J. Coalescent random forests. *Journal of Combinatorial Theory, Series A*, Elsevier, v. 85, n. 2, p. 165–193, 1999.
- [46] AIGNER, M.; ZIEGLER, G. M. *Proofs from the Book*. [S.l.]: Springer, 2004.
- [47] ABRAMOWITZ, M.; STEGUN, I. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. 1972*. [S.l.]: Dover, New York. 14 p.

**APÊNDICE A - TEOREMA DA INVERSÃO DE LAGRANGE E A SÉRIE DE
TAYLOR DA FUNÇÃO $G(X)$**

O teorema da inversão de Lagrange diz: Dado uma função $f(g)$, que obedece $f(0) = 0$, além de $\frac{df}{dg}|_{g=0} \neq 0$, se

$$x = f(g),$$

então

$$g(x) = \sum_{k=1}^{\infty} \left[\frac{d^{k-1}}{dg^{k-1}} \left(\frac{g'}{f(g')} \right) \right]_{g'=0} \frac{x^k}{k!}.$$

Podemos usar isso para encontrar a série de Taylor de $g(x)$, onde

$$g(x)e^{-g(x)} = x. \tag{A.1}$$

Sabendo que $f(a=0) = 0 = x_0$, escrevemos:

$$g(x) = a + \sum_{k=1}^{\infty} g_k \frac{(x-x_0)^k}{k!} = \sum_{k=1}^{\infty} g_k \frac{(x)^k}{k!}$$

onde:

$$g_k = \lim_{g \rightarrow 0} \left[\frac{d^{k-1}}{dg^{k-1}} \left(\frac{g-a}{f(g)-f(a)} \right)^k \right] = \lim_{g \rightarrow 0} \left[\frac{d^{k-1}}{dg^{k-1}} \left(\frac{g}{f(g)} \right)^k \right]$$

Dessa forma, temos,

$$\begin{aligned} g(x) &= \sum_{k=1}^{\infty} \lim_{g \rightarrow 0} \left[\frac{d^{k-1}}{dg^{k-1}} \left(\frac{g}{f(g)} \right)^k \right] \frac{(x)^k}{k!} \\ g(x) &= \sum_{k=1}^{\infty} \lim_{g \rightarrow 0} \left[\frac{d^{k-1}}{dg^{k-1}} \left(\frac{g}{ge^{-g}} \right)^k \right] \frac{(x)^k}{k!} \\ g(x) &= \sum_{k=1}^{\infty} \lim_{g \rightarrow 0} \left[\frac{d^{k-1}}{dg^{k-1}} (e^{gk}) \right] \frac{(x)^k}{k!} \\ g(x) &= \sum_{k=1}^{\infty} \lim_{g \rightarrow 0} \left[(k)^{k-1} e^{gk} \right] \frac{(x)^k}{k!} \\ g(x) &= \sum_{k=1}^{\infty} \left[(k)^{k-1} \right] \frac{(x)^k}{k!} \\ g(x) &= -x + x^2 - \frac{3}{2}x^3 + \frac{8}{3}x^4 \dots \end{aligned} \tag{A.2}$$

Podemos mostrar ainda que vale a relação $\frac{dg(x)}{dx} = \frac{e^{g(x)}}{1-g(x)}$.

$$\begin{aligned}
 g(x)e^{-g(x)} &= x \\
 \frac{d}{dx}(g(x)e^{-g(x)}) &= \frac{d}{dx}x \\
 g'(x)e^{-g(x)} - g(x)g'(x)e^{-g(x)} &= 1 \\
 g'(x)\left(e^{-g(x)} - g(x)e^{-g(x)}\right) &= 1
 \end{aligned} \tag{A.3}$$

Resolvendo a equação A.3 para $g'(x)$, temos

$$\begin{aligned}
 g'(x) &= \frac{1}{e^{-g(x)} - g(x)e^{-g(x)}} \\
 g'(x) &= \frac{1}{e^{-g(x)}(1-g(x))} \\
 g'(x) &= \frac{e^{g(x)}}{(1-g(x))}.
 \end{aligned} \tag{A.4}$$

O resultado A.4 serve para mostrar que $\frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) = \frac{g(x)}{x}$, como segue

$$\begin{aligned}
 \frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) &= g'(x)(1-g(x)/2) + g(x)(-g'(x)/2) \\
 \frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) &= g'(x) - g'(x)\frac{g(x)}{2} - g(x)\frac{g'(x)}{2} \\
 \frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) &= g'(x) - g(x)g'(x) \\
 \frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) &= g'(x)(1-g(x)).
 \end{aligned} \tag{A.5}$$

Usando o resultado de A.4 em A.5, temos

$$\begin{aligned}
 \frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) &= \frac{e^{g(x)}}{(1-g(x))}(1-g(x)) \\
 \frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) &= e^{g(x)}.
 \end{aligned} \tag{A.6}$$

Da definição da função $g(x)$ (A.1) decorre diretamente o fato de que $e^{-g(x)} = x/g(x)$. Substituindo essa informação na equação A.5, temos finalmente que

$$\frac{d}{dx}\left(g(x)\left(1 - \frac{g(x)}{2}\right)\right) = \frac{g(x)}{x}. \tag{A.7}$$

ANEXO A – FÓRMULA DE CAYLEY

Vamos determinar a enumeração das formas de ligar n vértices distinguíveis em um grafo acíclico com uma só componente, esta demonstração foi inicialmente proposta por *Jim Pitman* [45] e encontra-se igualmente detalhada na referência [46]. Denotemos por:

$f_{n,k} \sim$ Conjunto de todas as florestas enraizadas com n vértices em k componentes,

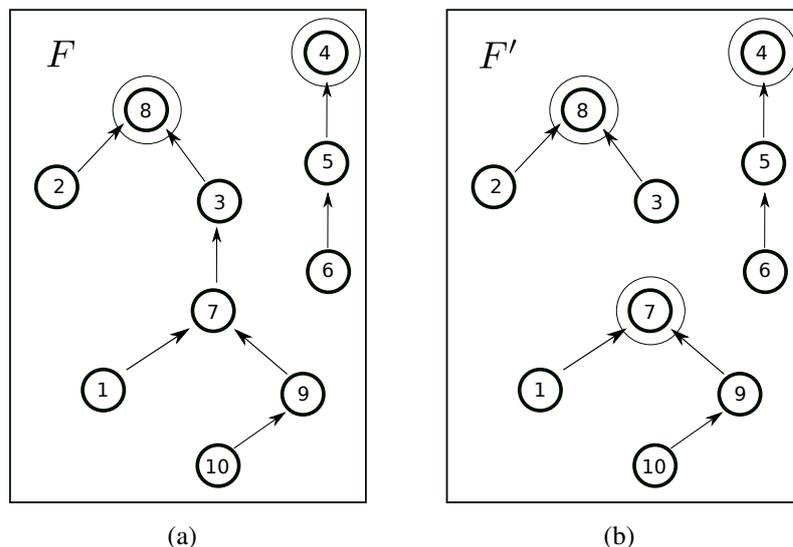
$F_{n,1} \sim$ Conjunto de todas as árvores enraizadas de n vértices e uma componente e,

$T_n \sim$ Conjunto de árvores não enraizadas de n vértices em uma componente.

Veja que $F_{n,1} = nT_n$, isso porque estamos tratando de árvores enraizadas e qualquer vértice pode ser raiz, gerando nT_n diferentes árvores.

Considere agora que a floresta $F_{n,k} \in f_{n,k}$ como um grafo direcionado, considere ainda que a floresta F' está contida na floresta $F_{n,k}$ (ou simplesmente F), ou seja, todos os grafos direcionados de F' estão presentes também em F , como na Figura 20, onde os grafos de F' podem ser facilmente identificados em F . Note que se F contém F' , F' tem mais componentes do que F .

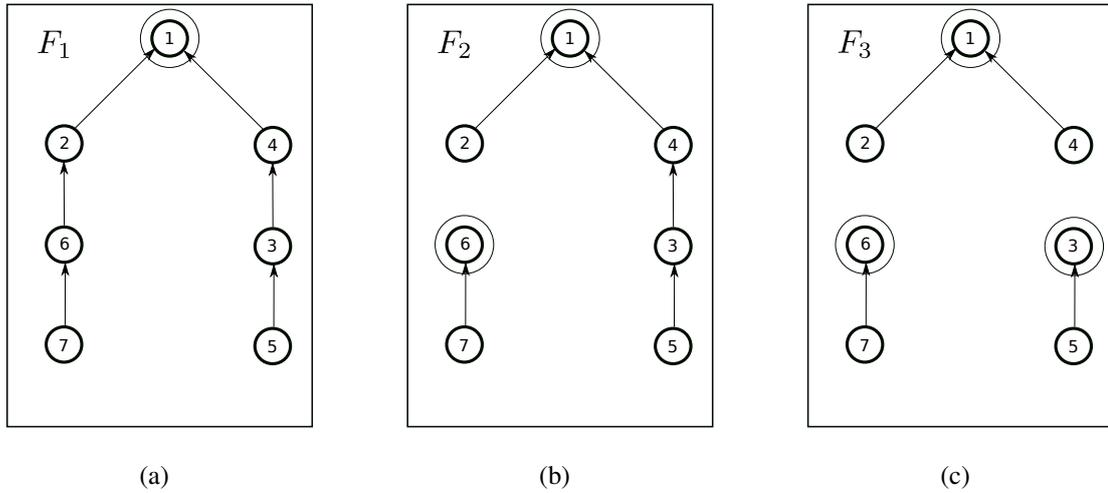
Figura 20: Exemplo de uma floresta contida em outra



Fonte: Produzido pelo autor. (a) representa a floresta F que consta de duas componentes. As raízes dos grafos são destacadas por um círculo em torno do vértice. (b) é a floresta F' que possui três componentes, podemos ver que as três árvores de F' que podem ser encontrados em F por meio da exclusão da aresta que liga os vértices 7 e 3.

Tomemos a sequência de florestas $F_{n,1}, F_{n,2}, F_{n,3}, \dots, F_{n,k}$, dizemos que essa é uma sequência de refinamento se $F_{n,i} \in F_{n,i}$, e $F_{n,i}$ contém $F_{n,i+1}$ para todo i (ver Figura 21). Agora

Figura 21: Florestas sequencialmente contidas em outras



Fonte: Produzido pelo autor. As árvores da floresta F_3 (c) estão contidas na floresta F_2 assim como F_3 está contida em F_1 (a). Dessa forma se configura a sequência F_1, F_2, F_3 .

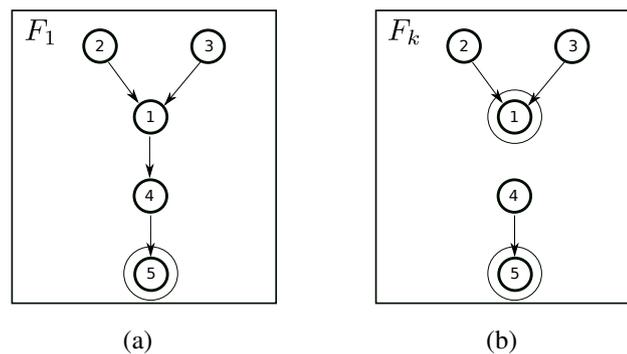
tomamos uma floresta fixa $F_{n,k}$ em $f_{n,k}$, por simplicidade vamos omitir adiante o índice n já que ele é constante, e denotamos:

$N(F_k) \sim$ número de árvores enraizadas que contem F_k ,

$N^*(F_k) \sim$ número de seqüências de refinamento terminadas em F_k ($F_1, F_2, F_3 \dots F_k$).

Vamos contar $N(F_k)$ de duas formas: iniciando em F_1 e iniciando em F_k . Suponhamos que $F_1 \in F_{n,1}$ e contém F_k , como na Figura 22. Note que, na Figura 22, para sair de F_1 e

Figura 22: Contagem por corte de arestas



Fonte: Produzido pelo autor. O procedimento aqui consiste em, partindo da floresta F_1 , excluir a cada passo uma aresta, até que a última floresta F_k ser construída. Como no exemplo a exclusão da ligação entre os vértices 1 e 4 em F_1 gerando F_k (b).

construir F_k , o fazemos cortando uma ligação a cada passo até chegar em F_k , assim o total de cortes será:

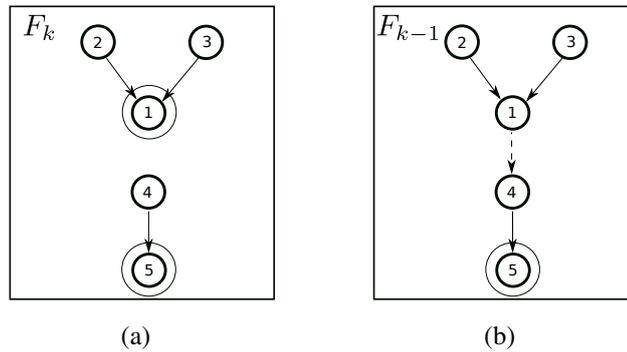
$$(n-1) - (n-k) = k-1. \quad (\text{A.1})$$

Sendo que podemos escolher qualquer ordem de arestas para cortar $(k-1)!$, assim temos:

$$N^*(F_k) = N(F_k)(k-1)! \quad (\text{A.2})$$

Agora fazemos o contrário, partindo de F_k vamos construir a sequência $(F_k, F_{k-1}, F_{k-2}, \dots, F_1)$ com $k \geq j \geq 2$. Para fazer F_k virar F_{k-1} devemos adicionar uma aresta entre duas componentes, apontando um vértice a para uma raiz b que não seja da mesma componente, como na Figura 23. Assim há a possibilidade de escolher quaisquer dos n vértices e

Figura 23: Contagem por adição de arestas



Fonte: Produzido pelo autor. Partindo da floresta F_k , adicionar a cada passo uma aresta entre vértices de componentes distintas, até que a última floresta seja construída. Como no exemplo a adição da ligação entre os vértices 1 e 4 em F_k gerando F_{k-1} (b).

ligar com qualquer umas das $k-1$ raízes disponíveis. Dessa forma, temos que, sucessivamente

$$\begin{aligned} N^*(F_k) &= n(k-1)n(k-2)n(k-3)\dots n(1) \\ N^*(F_k) &= n^{k-1}(k-1)! \end{aligned} \quad (\text{A.3})$$

Equacionado as expressões A.2 e A.3, temos

$$\begin{aligned} N(F_k)(k-1)! &= n^{k-1}(k-1)! \\ N(F_k) &= n^{k-1}. \end{aligned} \quad (\text{A.4})$$

Agora fazemos $n = k$ na equação A.4, assim F_k consiste em n vértices e n componentes ($F_{n,k=n}$), ou seja, um conjunto de vértices isolados. Portanto, visto que essa é a ramificação máxima e que, por consequência, juntar todos leva à construção de um grafo com uma só componente, $N(F_{k=n}) = F_n$, 1 conta o número de todas as possíveis árvores enraizadas de uma só componente

$$F_{n,1} = N(F_{k=n}) = n^{n-1}. \quad (\text{A.5})$$

Desa forma, o número de árvores não enraizadas T_n sobre n vértices e uma componente, ou

mais especificamente, as formas de ligar n vértices rotulados em um grafo acíclico de uma componente, é dado por

$$\begin{aligned} nT_n &= F_{n,1} = n^{n-1} \\ T_n &= n^{n-2} \end{aligned} \tag{A.6}$$

Que é a famosa fórmula de Cayley!