Ricardo Fabiano Samila

Modelos de Grafos Aleatórios e Aplicações

Monografia apresentada como Trabalho de Graduação no curso de Bacharelado em Ciência da Computação, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Jair Donadelli Jr.

Curitiba

2005

Resumo

Muitas redes complexas do mundo real podem ser modeladas através de grafos que denotam as relações existentes entre seus elementos. Para compreender os princípios de organização de tais redes, são propostos modelos de grafos aleatórios que codificam em sua estrutura propriedades existentes nas redes complexas. Um intenso estudo tem sido feito nos últimos anos sobre modelagem de redes complexas (em especial a Internet e a *World-Wide-Web*), uma vez que a existência de modelos próximos da realidade é extremamente útil para o avanço tecnológico, pois possibilita o estudo analítico da topologia das redes, a otimização de protocolos de comunicação e a resolução de problemas em geral. O objetivo deste trabalho é entender os fundamentos e as aplicações da modelagem de redes complexas como grafos aleatórios. Nele são apresentados exemplos de redes reais que podem ser modeladas através de grafos, bem como os modelos tradicional e generalizado de grafos aleatórios que codificam características de sistemas complexos. Também são descritas ferramentas úteis no estudo das redes, como o método probabilístico, funções geradoras e análise espectral. Por último é feito um estudo da tolerância a erros em redes, conforme aspectos topológicos de sua estrutura.

Satan is the only one who seems to understand.

Cake

Sumário

1 Redes do mundo real ..... p. 1

1.1 Introdução ..... p. 1

1.2 Conceitos de redes complexas ..... p. 2

1.2.1 Mundos pequenos (*Small worlds*) ..... p. 2

1.2.2 Agrupamentos (*Clustering*) ..... p. 3

1.2.3 Distribuição de grau ..... p. 3

1.3 Topologia de redes reais ..... p. 4

2 Grafos aleatórios – modelo tradicional ..... p. 8

2.1 O modelo de Erdős-Rényi ..... p. 8

2.2 Propriedades de quase todos os grafos ..... p. 12

2.3 Evolução do grafo ..... p. 13

2.4 Funções limiares ..... p. 15

2.5 O método probabilístico ..... p. 17

2.6 Distribuição de grau ..... p. 20

3 Generalização de grafos aleatórios ..... p. 23

3.1 Grafos aleatórios com distribuição de grau arbitrária ..... p. 24

3.2 Funções geradoras de probabilidade ..... p. 26

3.3 Tamanhos de componentes e transições de fase ..... p. 29

3.4 Distância média ..... p. 32

4 Análise espectral de grafos ..... p. 36

4.1 Espectro de grafos . . . . . p. 36
4.2 Aplicação de técnicas espectrais em grafos da Internet . . . . . p. 39
4.3 Autovalores de grafos com distribuição de grau *power-law* . . . . . p. 44

5 Tolerância a erros em redes . . . . . p. 49
5.1 Redes aleatórias (modelo binomial) . . . . . p. 51
5.2 Redes livres de escala . . . . . p. 52

Referências . . . . . p. 54

1 Redes do mundo real

1.1 Introdução

Muitos sistemas do mundo real têm sua estrutura formada por uma rede complexa que pode ser modelada através de um grafo no qual os vértices correspondem aos componentes elementares da rede e as arestas entre os vértices denotam as relações existentes entre tais elementos.

Devido ao alto grau de complexidade e dinâmica dessas redes, geralmente é inviável (ou até mesmo impossível) descrever a sua topologia para entender determinadas propriedades e comportamentos. Para isso existem várias abordagens de modelagem das redes reais através de grafos aleatórios, o que permite uma aproximação das redes para facilitar a compreensão de suas propriedades e a análise de certas medidas quantitativas, que podem ser calculadas a partir do modelo criado.

O primeiro modelo de grafo aleatório foi criado pelos matemáticos húngaros Paul Erdős e Alfréd Rényi [8]. De acordo com tal modelo, o grafo é construído iniciando-se com n vértices e cada par de vértices forma uma aresta com probabilidade p, o que resulta em um grafo com aproximadamente $pn(n-1)/2$ arestas distribuídas aleatoriamente. Embora esse modelo tenha sido utilizado durante décadas para modelar redes complexas, o crescente interesse na área fez com que cientistas reconsiderassem esse paradigma, uma vez que experimentos demonstraram que as redes reais por trás dos sistemas não são fundamentalmente aleatórias, no sentido do modelo tradicional de Erdős-Rényi, também conhecido como modelo *binomial* de grafo aleatório. De acordo com experimentos realizados, as

redes reais expunham alguns princípios de organização que deveriam ser codificados também em sua topologia. Assim, para incluir tais princípios na modelagem das redes reais, são propostas ferramentas e medidas que permitem a captura em termos quantitativos desses princípios de organização que desviam dos grafos aleatórios binomiais.

Exemplos de redes complexas do mundo real vão desde sistemas minúsculos, como a célula (que pode ser descrita como uma rede onde elementos químicos são conectados por reações químicas) até grandes sistemas, como a *World-Wide-Web* (onde as páginas são conectadas por *hyperlinks*). A topologia desses e de outros sistemas será melhor descrita na seção 1.3.

1.2 Conceitos de redes complexas

No estudo das redes reais e das interações entre seus componentes, há vários conceitos de propriedades e medidas obtidas através da modelagem com a utilização de grafos. Três desses conceitos são definidos a seguir, pela sua importância e estudo aprofundado realizado nos últimos anos.

1.2.1 Mundos pequenos (*Small worlds*)

O conceito de mundos pequenos descreve o fato de que, apesar de o tamanho das redes ser geralmente grande, na maioria delas existe um caminho relativamente curto entre quaisquer dois vértices. A distância entre dois vértices de um grafo corresponde ao número de arestas ao longo do caminho mais curto entre eles.

Uma conhecida manifestação do fenômeno de "mundos pequenos" é o conceito de "seis graus de separação", descoberto pelo psicólogo social Stanley Milgram (1967), que concluiu que há um caminho de relacionamentos com comprimento típico de seis entre a maioria das pessoas nos Estados Unidos.

A propriedade de mundos pequenos caracteriza a maioria das redes complexas: os atores em Hollywood estão na média dentro de três coatuações de distância entre um e

outro, ou os elementos químicos em uma célula são separados tipicamente por três reações.

Embora intrigante, o conceito de mundo pequeno não indica um princípio particular de organização, Erdős e Rényi demonstraram que a distância típica entre quaisquer dois vértices de um grafo aleatório é proporcional ao logaritmo do tamanho do grafo. Portanto grafos aleatórios também apresentam a propriedade de mundos pequenos.

1.2.2 Agrupamentos (*Clustering*)

Uma propriedade comum de redes sociais é a formação de grupos que representam círculos de amigos ou conhecimentos em que todo membro conhece outro membro. Essa tendência ao agrupamento é quantificada pelo coeficiente de agrupamento (definido por Watts e Strogatz, 1998): seja i um vértice selecionado na rede com $d(i)$ arestas que unem ele a $d(i)$ outros vértices. A relação entre o número E_i de arestas que existem entre os $d(i)$ vértices e o número total de pares de vértices $d(i) \cdot (d(i) - 1)/2$ dá o valor do coeficiente de agrupamento do vértice i:

$$C_i = \frac{2E_i}{d(i) \cdot (d(i) - 1)}. \tag{1.1}$$

O coeficiente de agrupamento da rede inteira é a média dos C_i's de cada vértice. Como a distribuição das arestas em um grafo aleatório (de Erdős-Rényi) é binomial com probabilidade p, o coeficiente de agrupamento é p. Contudo, foi demonstrado que na maioria das redes complexas – senão em todas – o coeficiente de agrupamento é normalmente maior do que em um grafo aleatório binomial com mesmo número de vértices e arestas.

1.2.3 Distribuição de grau

Nem todos os vértices de um grafo pertencem a um mesmo número de arestas. O grau de um vértice, ou o seu número de arestas, é caracterizado pela função de distribuição $P(k)$, que corresponde ao número de vértices com grau k. A distribuição de grau de um

grafo aleatório binomial é uma distribuição de Poisson[1]

$$P(k) \sim \frac{e^{\lambda}\lambda^{k}}{k!}, \tag{1.2}$$

onde

$$\lambda = n\binom{n-1}{k}p^{k}q^{n-1-k} \tag{1.3}$$

e $q = 1 - p$. Como em um grafo aleatório binomial as arestas são colocadas independentemente, a maioria dos vértices tem aproximadamente o mesmo grau, próximo do grau médio $\overline{k} = (n-1)p$ do grafo, portanto a distribuição de grau possui um pico em $P(\overline{k})$.

Entretanto, resultados recentes mostram que para a maioria das grandes redes complexas a distribuição de grau desvia significativamente de uma distribuição de Poisson. Em particular, para um grande número de redes, incluindo a *World-Wide-Web* e a Internet, a distribuição de grau tem uma cauda *power-law*[2]

$$P(k) \sim k^{-\gamma}. \tag{1.4}$$

Tais redes com essa distribuição são chamadas livres de escala (*scale-free*). Enquanto algumas redes expõem uma cauda exponencial, freqüentemente a forma funcional de $P(k)$ ainda desvia significativamente de uma distribuição de Poisson esperada para um grafo aleatório binomial.

1.3 Topologia de redes reais

1. **World-Wide-Web**

A *World-Wide-Web* (WWW) é a maior rede para a qual estão disponíveis informações topológicas atualmente, embora sua estrutura seja altamente dinâmica. Os

[1] A distribuição de Poisson, usada freqüentemente para modelar dados de contagem, possui um parâmetro λ e determina a probabilidade de obter exatamente k indivíduos, conforme a equação (1.2). A variância de uma Poisson é igual à sua média, λ.

[2] A distribuição *power-law* (lei de potência) é uma curva logarítmica que decresce abruptamente a níveis próximos de um mínimo e mantém-se assim, construindo uma longa "cauda". Tal distribuição implica que uma abundância de vértices possui apenas algumas arestas e uma minoria pequena, mas significativa, tem a grande maioria de arestas.

vértices do grafo que representa a WWW são as páginas *Web* e as arestas são os *hyperlinks* (URLs) que ligam uma página a outra. O tamanho desse grafo estava próximo de 1 bilhão de vértices no fim de 1999 [5].

O interesse na *Web* como um grafo aleatório tem prosperado depois de ter sido descoberto que a distribuição de grau das páginas *Web* segue leis de potência. Como as arestas da WWW são dirigidas, a rede é caracterizada por duas distribuições de grau: a distribuição de arestas que saem de um vértice, $P_{\text{out}}(k)$, que é a probabilidade de que uma página tenha k *hyperlinks* para outras páginas e a distribuição de arestas que entram em um vértice, $P_{\text{in}}(k)$, que é a probabilidade de que k *hyperlinks* apontem para uma certa página. Vários estudos têm estabelecido que ambos $P_{\text{out}}(k)$ e $P_{\text{in}}(k)$ seguem leis de potência:

$$P_{\text{out}}(k) \sim k^{-\gamma_{\text{out}}} \quad \text{e} \quad P_{\text{in}}(k) \sim k^{-\gamma_{\text{in}}}. \tag{1.5}$$

Apesar do elevado número de vértices, resultados de experimentos mostraram que a *Web* apresenta a propriedade "mundo pequeno", tendo a distância média entre dois vértices relativamente curta. Foi também constatado que a *Web* apresenta um coeficiente de agrupamento superior ao de um grafo aleatório binomial.

2. **Internet**

A Internet é a rede das ligações físicas existentes entre os computadores e outros equipamentos de telecomunicação. A topologia da Internet é estudada em dois diferentes níveis: no nível dos roteadores, os vértices são os roteadores e as arestas são as conexões físicas entre eles; já no nível interdomínio (sistema autônomo), cada domínio (composto por centenas de roteadores e computadores) é representado por um vértice simples, e uma aresta é traçada entre dois domínios se existe pelo menos um roteador que os conecta. Estudos feitos com a Internet em ambos os níveis concluíram que em cada caso a distribuição de grau segue leis de potência. Resultados de experimentos realizados com essa e outras redes são apresentados por Albert e Barabási em [4].

A Internet como uma rede também apresenta alto índice de agrupamento e distância média relativamente curta. A partir de estudos realizados com a Internet no nível de domínio entre 1997 e 1999 foi encontrado coeficiente de agrupamento variando entre $0,18$ e $0,3$, maior que o coeficiente de um grafo aleatório binomial com parâmetros similares, que é $C_{\text{rand}} \simeq 0,001$. A distância média entre vértices da Internet no nível de domínio variou entre $3,70$ e $3,77$, e no nível de roteador esteve em torno de 9, o que indica sua característica de mundo pequeno.

3. **Rede de colaboração entre atores de filmes**

Uma base de dados muito estudada é a rede de colaboração entre atores de filmes, baseada no *Internet Movie Database* que contém todos os filmes e seus elencos desde 1890. Nessa rede os vértices são os atores e dois vértices têm uma aresta em comum se os atores correspondentes atuaram em algum filme juntos.

A distância média entre vértices da rede de atores é próxima à de um grafo aleatório binomial com mesmos tamanho e grau médio, mas seu coeficiente de agrupamento é mais de 100 vezes maior que o de um grafo aleatório binomial. Esse grafo (assim como o de colaboração científica, visto a seguir) apresenta uma propriedade particular: sua estrutura é bipartida.

4. **Grafo de colaboração científica**

Uma rede de colaboração similar à dos atores de filmes pode ser construída para cientistas, onde os vértices são os cientistas e dois vértices são conectados se os dois cientistas correspondentes escreveram um artigo juntos. Redes de colaboração científica estudadas mostraram distância média pequena coeficiente de agrupamento alto. A distribuição de grau de todas as redes de colaboração estudadas segue leis de potência.

5. **Redes celulares**

Uma modelagem realizada descreve a célula como um grafo no qual os vértices são os substratos e as arestas representam as reações químicas (predominantemente dirigidas)

das quais esses substratos podem participar. A distribuição das arestas tem seguido leis de potência para todos os organismos estudados, o coeficiente de agrupamento é indeterminado (devido à natureza dirigida da rede), e a distância média encontrada foi aproximadamente a mesma em todos os organismos, com um valor de 3,3.

Uma outra rede importante caracterizando a célula descreve interações entre proteínas, onde os vértices correspondem às proteínas e são conectados se foi experimentalmente demonstrado que estas podem se manter unidas. Um estudo dessas interações físicas mostra que a distribuição de grau também segue leis de potência.

6. **Redes ecológicas**

Cadeias alimentares são freqüentemente usadas para quantificar a interação entre várias espécies de um ambiente. No grafo de uma cadeia alimentar, os vértices são espécies e as arestas representam relações predador-presa entre eles.

Em um estudo feito com as sete maiores cadeias alimentares, foi constatado que o grau médio do grafo de diferentes cadeias alimentares difere amplamente, que espécies de um determinado habitat estão há três ou menos arestas uma da outra, e que as cadeias alimentares são também altamente agrupadas.

7. **Redes de citação**

Uma rede bastante complexa é formada pelas citações em publicações científicas, com os vértices sendo os artigos publicados e uma aresta dirigida representando uma referência para um artigo publicado previamente.

Como resultados encontrados a partir de estudos realizados com redes de citação, a probabilidade de um artigo ser citado k vezes segue uma lei de potência (distribuição de grau "entrante"). Já a distribuição de grau das arestas que partem dos vértices apresenta uma cauda exponencial.

2 Grafos aleatórios – modelo tradicional

Grafos são as estruturas matemáticas que melhor representam redes reais. Uma subárea da teoria dos grafos diz respeito aos grafos aleatórios, em que as arestas são distribuídas aleatoriamente. Como existem redes que apresentam topologia complexa e princípios de organização desconhecidos que parecem aleatórios, a teoria dos grafos aleatórios é freqüentemente usada no estudo de tais redes.

O estudo dos grafos aleatórios iniciou-se com o trabalho de Paul Erdős e Alfréd Rényi (1959) [8], após Erdős ter descoberto que métodos probabilísticos são muito úteis na solução de problemas da teoria dos grafos. A seguir são descritas as principais características do modelo de grafo aleatório de Erdős e Rényi.

2.1 O modelo de Erdős-Rényi

Em seu artigo pioneiro de 1959, Erdős definiu um grafo aleatório com n vértices fixos e m arestas selecionadas aleatoriamente das $M = \binom{n}{2}$ arestas possíveis. No total existem $\binom{M}{m}$ grafos com n vértices e m arestas, formando um espaço de probabilidade em que todo grafo com m arestas possui a mesma probabilidade de ocorrer.

Fixamos o conjunto de vértices $V = \{0, \ldots, n-1\}$, contendo n elementos. O objetivo é colocar o conjunto $\mathcal{G} = \mathcal{G}(n)$ de todos os grafos em V dentro de um espaço de probabilidade, para então tentar responder questões que tipicamente são feitas sobre objetos aleatórios: Qual é a probabilidade de que um grafo $G \in \mathcal{G}$ tenha esta ou aquela propriedade? Qual é o valor esperado de um dado invariante de G, como sua cintura ou

número cromático esperados?

Um grafo aleatório G pode ser gerado através dos passos seguintes. Para cada par de vértices e no conjunto de todos os pares $\binom{V}{2}$, deve ser decidido por algum experimento aleatório se e deve ou não deve ser uma aresta de G, como por exemplo o lançamento de uma moeda, considerando que a moeda apresenta a probabilidade p de resultar "cara" em um lançamento. O grafo inicialmente não contém arestas, isto é, $G = (V, \emptyset)$ e, para cada par de vértices $\{i, j\} \in \binom{V}{2}$, a moeda é lançada. Se o resultado for "cara", a aresta $\{i, j\}$ é colocada no grafo, isto é, $\{i, j\} \in E(G)$; caso contrário ("coroa"), $\{i, j\} \notin E(G)$. Um grafo genérico assim definido é denotado por $G_{n,p}$ e o chamamos de *grafo aleatório* com n vértices e probabilidade de aresta p.

Os experimentos aleatórios são representados independentemente, e para cada um deles a probabilidade de sucesso – isto é, de aceitar e como uma aresta de G – é igual a algum número fixo[1] $p \in [0, 1]$. Então se G_0 é algum grafo fixo em V, com m arestas, por exemplo, o evento elementar $\{G_0\}$ tem uma probabilidade de $p^m q^{\binom{n}{2}-m}$ (onde $q = 1 - p$): com essa probabilidade, o grafo G gerado aleatoriamente é esse particular grafo G_0. Quando a probabilidade $p = \frac{1}{2}$, todo grafo com n arestas pode ocorrer com a mesma probabilidade, como mencionado no primeiro parágrafo desta seção.

Para construir a medida de probabilidade em $\mathcal{G}$ formalmente, é definido inicialmente para todo par de vértices $e \in \binom{V}{2}$ seu próprio espaço de probabilidade $\Omega_e = \{0_e, 1_e\}$, e escolhido $P_e(\{1_e\}) = p$ e $P_e(\{0_e\}) = q$ como as probabilidades de seus dois eventos elementares. Equivalentemente, para o espaço de probabilidade $\mathcal{G} = \mathcal{G}(n, p)$ é utilizado o espaço produto

$$\Omega = \prod_{e \in [V]^2} \Omega_e. \tag{2.1}$$

[1] Freqüentemente, o valor de p irá depender da cardinalidade n do conjunto V no qual os grafos aleatórios são gerados; assim, p irá ter o valor $p = p(n)$ de alguma função $n \mapsto p(n)$. Contudo, V (e então n) é fixo pela definição de $\mathcal{G}$: para cada n separadamente, é construído um espaço de probabilidade dos grafos G em $V = \{0, \ldots, n-1\}$, e dentro de cada espaço a probabilidade de que $e \in [V]^2$ seja uma aresta de G tem o mesmo valor para todo e.

Assim, formalmente, um elemento de Ω é um mapa ω que atribui para todo $e \in \binom{V}{2}$ ou 0_e ou 1_e, e a medida de probabilidade P em Ω é o produtório de todas as medidas P_e. Na prática, ω corresponde ao grafo G em V cujo conjunto de arestas é

$$E(G) = \{e|\omega(e) = 1_e\}, \tag{2.2}$$

e G é chamado de *grafo aleatório binomial* em V com probabilidade de aresta p.

Qualquer conjunto de grafos em V é considerado um *evento* em $\mathcal{G}(n,p)$. Em particular, para todo $e \in [V]^2$ o conjunto

$$A_e = \{\omega|\omega(e) = 1_e\} \tag{2.3}$$

de todos os grafos G em V com $e \in E(G)$ é um evento: o evento em que e é uma aresta de G. Os eventos A_e são independentes e ocorrem com probabilidade p. [1]

A primeira propriedade de grafos aleatórios estudada por Erdős e Rényi foi o surgimento de subgrafos.[2] Considerando o evento de que o grafo aleatório G contém um subgrafo H; seja $|V(H)| = k$ e $|E(H)| = \ell$. A probabilidade desse evento $H \subseteq G$ é o produto das probabilidades A_e sobre todas as arestas $e \in H$, então $P[H \subseteq G] = p^\ell$. Por outro lado, a probabilidade de H ser um subgrafo induzido[3] de G é $p^\ell q^{\binom{k}{2}-\ell}$: agora as arestas ausentes em H devem estar ausentes em G também, e elas o fazem independentemente com probabilidade $q = 1 - p$.

Já a probabilidade P_H de que G tenha um subgrafo induzido *isomorfo* a H é relativamente mais difícil de calcular: como as instâncias possíveis de H em subconjuntos de V se sobrepõem, os eventos de que elas ocorrem em G não são independentes. Contudo, a soma (sobre todos os k-conjuntos $U \subseteq V$) das probabilidades $P[H \simeq G[U]]$ é sempre um limite superior para P_H, uma vez que P_H é a medida da união de todos esses eventos.

No contexto de grafos aleatórios, cada uma das propriedades familiares de um

[2]Um grafo G_1 é um subgrafo de G se $V(G_1) \subset V(G)$, $E(G_1) \subset E(G)$ e cada $e \in E(G_1)$ é um par $e \in [V(G_1)]^2$.

[3]Um subgrafo de G induzido por $u \subseteq V(G)$ é $G[u] = \left(u, E(G) \cap \binom{u}{2}\right)$.

grafo (como grau médio, cintura, número cromático, etc) é uma variável aleatória não-negativa em $\mathcal{G}(n,p)$, isto é, uma função

$$X : \mathcal{G}(n,p) \to [0,\infty). \tag{2.4}$$

O cálculo do valor médio de uma variável aleatória X pode ser uma forma eficiente de estabelecer a existência de um grafo G tal que $X(G) < \alpha$ para algum $\alpha > 0$ e, além disso, se G tem alguma propriedade desejada $\mathcal{P}$, desde que $\mathcal{P}$ dependa do valor da variável X. Se o valor médio de X é pequeno, então $X(G)$ não pode ser grande para mais do que alguns grafos em $\mathcal{G}(n,p)$, porque $X(G) \geq 0$ para todos os grafos $G \in \mathcal{G}(n,p)$. Portanto X deve ser pequeno para muitos grafos em $\mathcal{G}(n,p)$, e é esperado que exista algum grafo com a propriedade $\mathcal{P}$.

Essa idéia simples é o princípio de muitas provas de existência através de grafos aleatórios e, quantificada, corresponde à desigualdade de Markov:

Teorema 1 (Desigualdade de Markov). *Seja $X \geq 0$ uma variável aleatória em $\mathcal{G}(n,p)$ e $\alpha > 0$, então*

$$P[(X-\mu)^2 \geq \alpha^2] \leq \frac{\mathbb{E}((X-\mu)^2)}{\alpha^2}.$$

Demonstração.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{G \in \mathcal{G}(n,p)} P(\{G\}) \cdot X(G) \\
&\geq \sum_{\substack{G \in \mathcal{G}(n,p) \\ X(G) \geq \alpha}} P(\{G\}) \cdot X(G) \\
&\geq \sum_{\substack{G \in \mathcal{G}(n,p) \\ X(G) \geq \alpha}} P(\{G\}) \cdot \alpha \\
&= P[X \geq \alpha] \cdot \alpha.
\end{aligned}$$

□

2.2 Propriedades de quase todos os grafos

A teoria dos grafos aleatórios estuda as propriedades do espaço de probabilidade associado a grafos com n vértices quando $n \to \infty$. Muitas propriedades desses grafos aleatórios podem ser determinadas através de argumentos probabilísticos. Nesse aspecto, Erdős e Rényi usaram a definição de que quase todo grafo tem uma propriedade Q se a probabilidade de ter Q tende a 1 quando $n \to \infty$.

Uma propriedade de grafo (*graph property*) é uma classe de grafos fechada sob isomorfismo. Se $p = p(n)$ é uma função fixa e $\mathcal{P}$ é uma propriedade de grafo, é possível identificar o comportamento da probabilidade $P[G \in \mathcal{P}]$ para $G \in \mathcal{G}(n,p)$ quando $n \to \infty$. Se essa probabilidade tende a 1 quando $n \to \infty$, diz-se que $G \in \mathcal{P}$ para *quase todo* $G \in \mathcal{G}(n,p)$, ou que *quase certamente* $G \in \mathcal{P}$; analogamente, se a probabilidade tende a 0, diz-se que *quase nenhum* $G \in \mathcal{G}(n,p)$ tem a propriedade $\mathcal{P}$.

O conceito é ilustrado através da demonstração da seguinte proposição:

Proposição 2. *Para toda constante $p \in (0,1)$ e todo grafo abstrato*[4] *H, quase todo $G \in \mathcal{G}(n,p)$ contém uma cópia induzida de H.*

Demonstração. Seja H um grafo dado, e $k = |H|$. Se $n \geq k$ e $U \subseteq \{0,\ldots,n-1\}$ é um conjunto fixo de k vértices de G, então $G[U]$ é isomorfo a H com probabilidade $r > 0$. Essa probabilidade r depende de p, mas não de n. Agora G contém uma coleção de $\lfloor \frac{n}{k} \rfloor$ conjuntos U disjuntos. A probabilidade de que nenhum dos grafos correspondentes $G[U]$ seja isomorfo a H é $(1-r)^{\lfloor \frac{n}{k} \rfloor}$, pois esses eventos são independentes pela disjunção dos conjuntos de arestas $[U]^2$. Portanto:

$$P[H \not\subseteq G \text{ induzido}] \leq (1-r)^{\lfloor \frac{n}{k} \rfloor} \underset{n\to\infty}{\longrightarrow} 0.$$

□

[4] Só o tipo isomórfico de H é relevante, não seus atuais vértices e conjuntos de arestas. Portanto, nesse contexto, a palavra 'subgrafo' tem o sentido de 'isomorfo a um subgrafo'.

Uma interessante característica da maioria dos resultados que mostram que quase todo (ou quase nenhum) grafo possui certa propriedade é que os valores de p não assumem qualquer papel: se quase todo grafo em $\mathcal{G}(n, \frac{1}{2})$ tem uma propriedade, então o mesmo vale para quase todo grafo em $\mathcal{G}(n, \frac{1}{1000})$. Isso ocorre porque para a maioria das propriedades, a ordem de grandeza de p em torno da qual a propriedade considerada irá ocorrer fica bem abaixo de qualquer valor constante, geralmente é uma função de n tendendo a zero quando $n \to \infty$.

2.3 Evolução do grafo

Um grafo aleatório é construído através de adições sucessivas de arestas aleatórias, a partir de um conjunto de n vértices isolados inicialmente. Tal processo é denominado *evolução* do grafo aleatório. A cada estágio da evolução, os grafos obtidos possuem maior probabilidade de aresta p, eventualmente obtendo um grafo completo[5] para $p \to 1$.

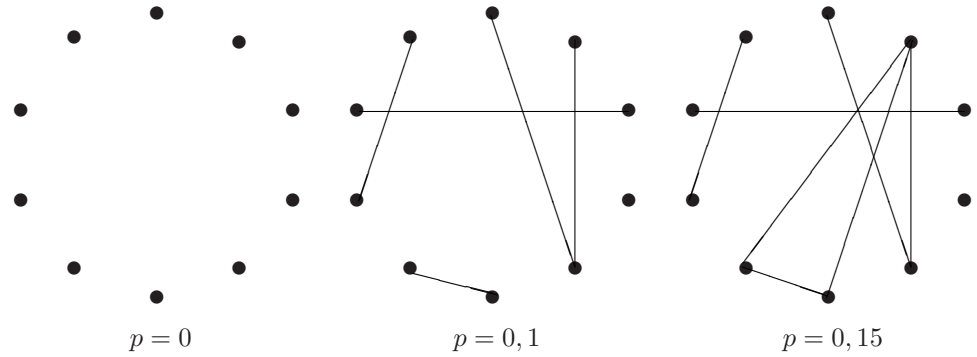


Figura 1: Processo de evolução de um grafo aleatório, iniciando-se com 10 vértices e conectando-os com probabilidades $p = 0,1$ e $p = 0,15$.

O principal problema na teoria dos grafos aleatórios é determinar em qual probabilidade de aresta p o grafo irá possuir uma propriedade específica. A maior descoberta de Erdős e Rényi foi que muitas propriedades importantes de grafos aleatórios aparecem repentinamente: em uma dada probabilidade, ou quase todo grafo tem a propriedade Q, ou quase nenhum grafo tem a propriedade Q. Portanto a transição de uma propriedade ser muito improvável para ser muito provável é usualmente rápida.

[5]Um grafo completo possui todas as arestas possíveis: $G = \left(V, \binom{V}{2}\right)$.

Se a probabilidade de aresta p variar com o número n de vértices, para valores de p cuja ordem de grandeza fica abaixo de n^{-2}, um grafo aleatório $G \in \mathcal{G}(n,p)$ quase certamente não possui absolutamente nenhuma aresta. À medida que p aumenta, G adquire mais estrutura: a partir de $p = \sqrt{n}n^{-2}$, quase certamente existe um componente com mais de dois vértices, os quais crescem em árvores, e em torno de $p = n^{-1}$ surgem os primeiros ciclos, depois, o grafo torna-se conexo, e assim por diante.

Considerando um grafo aleatório com n vértices e considerando que a probabilidade de aresta p é proporcional a n^z, onde z é um parâmetro ajustável que pode assumir qualquer valor entre $-\infty$ e 0. Para $z < -\frac{3}{2}$ quase todo grafo contém apenas vértices e arestas isolados. Quando z passa por $-\frac{3}{2}$, árvores de ordem 3 aparecem, quando z alcança $-\frac{4}{3}$, árvores de ordem 4 aparecem, e conforme z se aproxima de -1, o grafo contém árvores de ordem cada vez maior. Contudo, contanto que $z < -1$, tal que o grau médio do grafo $\overline{k} = pn \to 0$ quando $n \to \infty$, o grafo é uma união de árvores disjuntas, e não contém ciclos. Exatamente quando z passa por -1, a probabilidade de ciclos de todas as ordens (constantes) passa de 0 para 1. Ciclos de ordem 3 também podem ser vistos como subgrafos completos de ordem 3. Subgrafos completos de ordem 4 aparecem quando $z = -\frac{2}{3}$, e conforme z aumenta, surgem os de ordem maior. Finalmente, quando z aproxima-se de 0, quase todo grafo está próximo do grafo completo de n vértices. Tudo isso é resumido na figura 2.

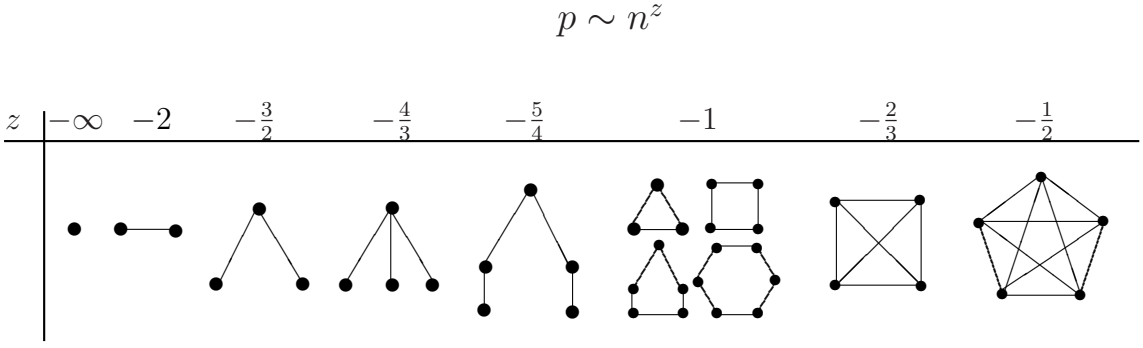


Figura 2: As probabilidades críticas em que diferentes subgrafos surgem em um grafo aleatório.

É comum analisar o desenvolvimento de grafos aleatórios quando p cresce: para cada $p = p(n)$, consideram-se as propriedades compartilhadas por quase todos os grafos

em $\mathcal{G}(n,p)$ como propriedades 'do' típico grafo aleatório $G \in \mathcal{G}(n,p)$, e estuda-se como G tem suas características alteradas com a taxa de crescimento de p. A evolução de grafos aleatórios ocorre em saltos relativamente repentinos: existem probabilidades de aresta *críticas* t que são limiares em que quase nenhum grafo em $\mathcal{G}(n,p)$ tem a propriedade considerada quando $p \ll t$ e quase todo grafo em $\mathcal{G}(n,p)$ tem a propriedade quando $p \gg t$. A probabilidade limiar será detalhada na seção seguinte.

2.4 Funções limiares

Uma função real $t = t(n)$ com $t(n) \neq 0, \forall n$ é uma *função limiar* para uma propriedade $\mathcal{P}$ de um grafo se o seguinte vale para todo $p = p(n)$ e $G \in \mathcal{G}(n,p)$:

$$\lim_{n\to\infty} P[G \in \mathcal{P}] = \begin{cases} 0 & \text{se } \frac{p}{t} \to 0 \text{ quando } n \to \infty \\ 1 & \text{se } \frac{p}{t} \to \infty \text{ quando } n \to \infty. \end{cases} \tag{2.5}$$

Se $\mathcal{P}$ tem uma função limiar t, então qualquer múltiplo positivo ct de t é também uma função limiar para $\mathcal{P}$. Assim, funções limiares nesse sentido são únicas, independentes de uma constante multiplicativa (embora para algumas propriedades pode-se definir limiares restritos em que o fator constante é crucial).

Propriedades ditas *monótonas* são propriedades fechadas sob a adição ou remoção de arestas (como a conectividade de um grafo, por exemplo). Todas as propriedades monótonas (com algumas exceções triviais) possuem funções limiares.

Considerando uma propriedade da forma

$$\mathcal{P} = \{G \in \mathcal{G}(n) | X(G) > x\}, \tag{2.6}$$

onde $X \geq x$ é um evento em $\mathcal{G}(n,p)$. Inúmeras propriedades podem ser expressas dessa forma, por exemplo: se X denota o número de árvores geradoras, então $\mathcal{P}$ corresponde à propriedade de ser conexo, para $x = 0$.

A prova de que $\mathcal{P}$ tem uma função limiar t consiste de duas partes: a primeira deve mostrar que quase nenhum grafo $G \in \mathcal{G}(n,p)$ tem $\mathcal{P}$ quando p é pequeno comparado a t, e a segunda que quase todo grafo G tem $\mathcal{P}$ quando p é grande.

Para a primeira parte da prova, basta encontrar um limite superior para o valor esperado $\mathbb{E}(X)$ e usar a desigualdade de Markov (teorema 1): se $\mathbb{E}(X)$ é pequeno então $X(G)$ pode ser positivo apenas para alguns $G \in \mathcal{G}(n,p)$.

Na segunda parte da prova, para mostrar que $P[X > 0]$ é grande, não é suficiente limitar $\mathbb{E}(X)$ inferiormente: como X não está limitado superiormente, $\mathbb{E}(X)$ pode ser grande simplesmente porque X é muito grande em apenas alguns grafos G, então X pode ainda ser zero para a maioria dos $G \in \mathcal{G}(n,p)$. Para provar que $P[X > 0] \to 1$, deve-se então mostrar que X não desvia muito de sua média, o que pode ser feito com a utilização da *variância*:

$$\sigma^2 = \mathbb{E}((X - \mu)^2), \tag{2.7}$$

onde $\mu = \mathbb{E}(X)$ e $\sigma > 0$.

A variância σ^2 (ou *segundo momento*) de X é uma medida quadrática que indica quanto X desvia de sua média. Como $\mathbb{E}$ é linear, a definição é equivalente a:

$$\sigma^2 = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - \mu^2. \tag{2.8}$$

As medidas μ e σ^2 sempre se referem a uma variável aleatória em um espaço de probabilidade fixo. No caso em que são considerados os espaços $\mathcal{G}(n,p)$, ambas são funções de n.

Teorema 3 (Desigualdade de Chebyshev). *Da desigualdade de Markov (teorema 1), temos:*

$$P[|X - \mu| \geq t] \leq \frac{\mathbb{E}(X - \mu)^2}{t^2}.$$

A desigualdade de Chebyshev [1] afirma que X não pode desviar muito de sua média, e também pode ser usada para provar que $X(G) > 0$ para quase todo grafo em $\mathcal{G}(n,p)$:

Afirmação 4. *Se $\mu > 0$ para n grande, e $\sigma^2/\mu^2 \to 0$ quando $n \to \infty$, então $X(G) > 0$ para quase todo $G \in \mathcal{G}(n,p)$.*

Demonstração. Qualquer grafo G com $X(G) = 0$ satisfaz $|X(G) - \mu| = \mu$. Daí a desigualdade de Chebyshev (com $\lambda = \mu$) implica que:

$$P[X = 0] \leq P[|X - \mu| \geq \mu] \leq \sigma^2/\mu^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Como $X \geq 0$, então quase certamente $X > 0$, isto é, $X(G) > 0$ para quase todo $G \in \mathcal{G}(n,p)$. □

Um exemplo interessante de função limiar é um corolário de um teorema de Erdős e Rényi:

Corolário 1. *Se $k \geq 3$, então $t(n) = n^{-1}$ é uma função limiar para a propriedade de conter um k-ciclo.*

Essa função limiar é independente do tamanho fixo k do ciclo considerado: na evolução de grafos aleatórios, ciclos de todos os tamanhos (constantes) surgem praticamente ao mesmo tempo, como visto na seção sobre evolução de grafos aleatórios (2.3). Um fenômeno similar ocorre para árvores (em que a função limiar não depende de sua forma, apenas de sua ordem), e também para subgrafos completos.

2.5 O método probabilístico

O método probabilístico é uma ferramenta poderosa da matemática discreta utilizada na prova de existência. Para provar a existência de um objeto com alguma propriedade desejada, define-se um espaço de probabilidade em alguma classe de objetos

grande (e certamente não vazia), e então mostra-se que um elemento desse espaço possui a propriedade desejada com probabilidade positiva (maior do que zero).

Como exemplo, a seguir é apresentada a aplicação do método probabilístico para o teorema de Erdős: *Dado qualquer inteiro positivo k, existe um grafo G com cintura $g(G) > k$ e número cromático $\chi(G) > k$.*[6] A demonstração é apresentada no teorema 5.

Considerando *curtos* os ciclos de tamanho no máximo k, e *grandes* os conjuntos com $\frac{|G|}{k}$ ou mais vértices. O teorema de Erdős afirma que existem grafos com cintura arbitrariamente grande e número cromático arbitrariamente alto. Para prová-lo seria suficiente achar um grafo G sem ciclos curtos e sem conjuntos independentes[7] grandes de vértices: então as classes de cor em quaisquer colorações dos vértices de G são *pequenas* (não-grandes), portanto é necessário mais do que k cores para colorir G, isto é, $\chi\alpha \geq n$.

Como esse grafo G pode ser encontrado? Se for escolhido p suficientemente pequeno, então um grafo aleatório $\mathcal{G}(n, p)$ provavelmente não teria quaisquer ciclos (curtos). Se for escolhido p suficientemente grande, então G provavelmente não teria conjuntos independentes grandes de vértices. A questão é: as duas faixas de p se sobrepõem, ou seja, é possível escolher p tal que, para algum n_0, p seja tanto suficientemente pequeno para que $P[g > k] \geq \frac{1}{2}$ quanto suficientemente grande para que $P[\alpha < n/k] \geq \frac{1}{2}$, para todo $n > n_0$? Se sim, $\mathcal{G}(n, p)$ irá conter pelo menos um grafo sem quaisquer ciclos curtos ou conjuntos independentes grandes.

Porém, tal escolha de p é impossível, pois as duas faixas de p não se sobrepõem. Deve-se manter p abaixo de n^{-1} para fazer a ocorrência de ciclos curtos em G improvável – mas para qualquer p dessa grandeza, G provavelmente não terá ciclos, então será bipartido e portanto existirão pelo menos $n/2$ vértices independentes. Isso será melhor detalhado na seção sobre funções limiares (2.4).

Para fazer grandes conjuntos independentes improváveis, deve-se fixar p acima

[6] A cintura de um grafo é o número de arestas contidas em seu ciclo mais curto (ou ∞ caso o grafo seja acíclico). O número cromático de um grafo G é o valor mínimo de k, tal que $V(G)$ pode ser particionado em k classes e $E(G)$ contém arestas conectando apenas vértices de diferentes classes.

[7] Um conjunto independente em um grafo é um conjunto de vértices não adjacentes entre si.

de n^{-1}, em $n^{\epsilon-1}$ para algum $\epsilon > 0$. Se ϵ for pequeno o suficiente então aparecerão apenas *alguns* ciclos curtos em G, mesmo comparado a n. Se for excluído um vértice de cada um desses ciclos, então o grafo H obtido não irá ter ciclos curtos, e seu número de independência $\alpha(H)$ será no máximo o de G. Como H não é muito menor que G, seu número cromático ainda será grande, então terá sido encontrado um grafo com ambos cintura e número cromático grandes.

Para a prova formal do teorema de Erdős, primeiramente é demonstrado que uma probabilidade de aresta $p = n^{\epsilon-1}$ é sempre grande o suficiente para assegurar que $G \in \mathcal{G}(n,p)$ 'quase certamente' não possui conjuntos independentes grandes de vértices. Para isso, a seguinte afirmação deve ser provada[8]: *Seja $k > 0$ um inteiro, e seja $p = p(n)$ uma função de n tal que $p \geq (6k \ln n)n^{-1}$. Então*

$$\lim_{n\to\infty} P\left[\alpha \geq \frac{1}{2}n/k\right] = 0. \tag{2.9}$$

Teorema 5. *Para todo inteiro k existe um grafo H com cintura $g(H) > k$ e número cromático $\chi(H) > k$.*

Demonstração. Seja $k \geq 3$, $0 < \epsilon < \frac{1}{k}$, e $p = n^{\epsilon-1}$. Seja $X(G)$ o número de ciclos curtos em um grafo aleatório $G \in \mathcal{G}(n,p)$, ou seja, o número de ciclos de tamanho no máximo k. O valor esperado de ciclos de tamanho $\leq k$ é:

$$E(X) = \sum_{i=3}^{k} \frac{(n)_i}{2i} p^i \leq \sum_{i=3}^{k} n^i p^i \leq \frac{1}{2}(k-2)n^k p^k;$$

como $(np)^i \leq (np)^k$ (pois $np = n^\epsilon \geq 1$), tem-se:

$$\begin{aligned} P\left[X \geq \frac{n}{2}\right] &\leq \frac{E(X)}{n/2} \\ &\leq (k-2)n^{k-1}p^k \\ &= (k-2)n^{k-1}n^{(\epsilon-1)k} \\ &= (k-2)n^{k\epsilon-1}. \end{aligned}$$

[8] A demonstração da afirmação (2.9) não é apresentada aqui mas pode ser encontrada em [1].

Como $k\epsilon - 1 < 0$ devido à escolha de ϵ, isso implica que

$$\lim_{n\to\infty} P\left[X \geq \frac{n}{2}\right] = 0.$$

Seja n suficientemente grande tal que $P[X \geq \frac{n}{2}] < \frac{1}{2}$ e $P[\alpha \geq \frac{1}{2}n/k] < \frac{1}{2}$; isso é possível pela escolha de p. Então existe um grafo $G \in \mathcal{G}(n,p)$ com menos que $\frac{n}{2}$ ciclos curtos e $\alpha(G) < \frac{1}{2}n/k$. Seja H o grafo obtido pela remoção de um vértice de cada um desses ciclos. Então $|H| \geq \frac{n}{2}$ e H não possui ciclos curtos, portanto $g(H) > k$. Por definição de G,

$$\chi(H) \geq \frac{|H|}{\alpha(H)} \geq \frac{n/2}{\alpha(G)} > k.$$

□

2.6 Distribuição de grau

Em um grafo aleatório com probabilidade p, o grau $k_i = d(i)$ de um vértice i segue uma distribuição binomial com parâmetros $n-1$ e p:

$$P(k_i = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}. \tag{2.10}$$

Essa probabilidade representa o número de formas em que k arestas podem ser traçadas de um determinado vértice: a probabilidade de k arestas é p^k, a probabilidade da ausência de arestas adicionais é $(1-p)^{n-1-k}$, e há $\binom{n-1}{k}$ formas equivalentes de selecionar os k vértices finais para essas arestas. Além do mais, se i e j são vértices diferentes, $P(k_i = k)$ e $P(k_j = k)$ são próximos a variáveis aleatórias independentes. Para encontrar a distribuição de grau do grafo, é preciso estudar o número de vértices com grau k, X_k. O objetivo é determinar a probabilidade de X_k assumir um determinado valor, $P(X_k = r)$.

Conforme (2.10), o valor esperado do número de vértices com grau k é

$$E(X_k) = nP(k_i = k) = \lambda_k, \tag{2.11}$$

onde

$$\lambda_k = n\binom{n-1}{k}p^k(1-p)^{n-1-k}. \tag{2.12}$$

A distribuição dos valores de X_k, $P(X_k = r)$, é próxima de uma distribuição de Poisson

$$P(X_k = r) = e^{-\lambda_k}\frac{\lambda_k^r}{r!}. \tag{2.13}$$

Portanto o número de vértices com grau k segue uma distribuição de Poisson com valor médio λ_k. O valor esperado da distribuição (2.13) não é uma constante, mas sim a função λ_k (2.12). A distribuição de Poisson decresce rapidamente para valores elevados de r, com o desvio padrão $\sigma_k = \sqrt{\lambda_k}$. Isto quer dizer que X_k não diverge muito do resultado aproximado $X_k = nP(k_i = k)$, válido caso os vértices sejam independentes. Então com aproximação razoável a distribuição de grau de um grafo aleatório é uma distribuição binomial

$$P(k) = \binom{n-1}{k}p^k(a-p)^{n-1-k}, \tag{2.14}$$

que para n grande pode ser substituída por uma distribuição de Poisson

$$P(k) \simeq e^{-pn}\frac{(pn)^k}{k!} - e^{-\overline{k}}\frac{\overline{k}^k}{k!}. \tag{2.15}$$

Uma simulação que realizamos com um grafo aleatório de tamanho $n = 13000$ e probabilidade de aresta $p = 0,0013$ comprovou a proximidade entre a distribuição de graus do grafo aleatório e a distribuição de Poisson. O valor esperado para uma distribuição de Poisson com esses parâmetros é $pn = 16,9$. O grau médio obtido com a simulação foi $\overline{k} = 16,8986$. A comparação é ilustrada na figura 3.

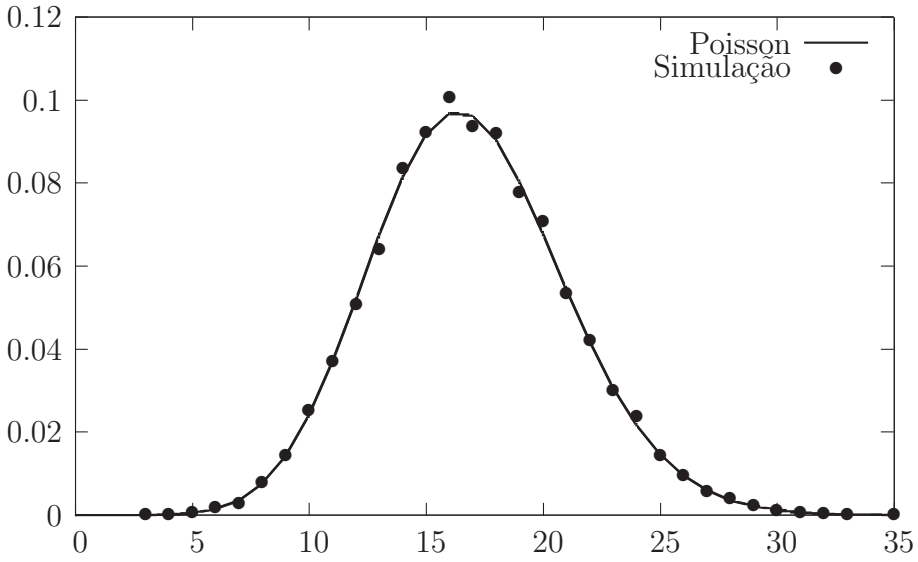


Figura 3: Distribuição de grau da simulação de um grafo aleatório. O eixo x do gráfico corresponde ao grau k, e o eixo y é a fração de vértices do grafo que possuem grau k. O gráfico compara a distribuição de graus do grafo simulado com o valor esperado da distribuição de Poisson.

Resultados de estudos a respeito da existência e exclusividade dos graus mínimo e máximo de um grafo aleatório indicam que, para um grande intervalo de valores p, ambos os graus mínimo e máximo são determinados e finitos. Por exemplo, se $p(n) \sim n^{-1-1/k}$ (o grafo é um conjunto de árvores isoladas de ordem $\leq k+1$), quase nenhum grafo possui vértices com grau maior que k. Em outro extremo, se $p = [\ln(n) + k\ln(\ln(n)) + c]/n$, quase todo grafo aleatório tem grau mínimo de pelo menos k. Além disso, para um p suficientemente alto ($pn/\ln(n) \to \infty$), o grau máximo de quase todos os grafos aleatórios tem a mesma ordem de grandeza que o grau médio.

Portanto, apesar de o posicionamento das arestas ser aleatório, um grafo aleatório típico é bastante homogêneo: a maioria dos seus vértices pertence a um mesmo número de arestas.

3 Generalização de grafos aleatórios

Como visto na seção sobre topologia de redes reais (1.3), tais redes diferem do modelo tradicional de grafo aleatório em sua distribuição de grau. O grafo aleatório de Erdős-Rényi possui distribuição de grau próxima de uma Poisson, já na maioria das redes reais estudadas, a distribuição de grau segue uma lei de potência onde o número de vértices com grau k é aproximadamente $P(k) \sim k^{-\gamma}$. Como leis de potência não possuem uma escala característica, essas redes são chamadas *redes livres de escala*.

Outra diferença significativa é que redes reais costumam apresentar um alto coeficiente de agrupamento, ou seja, se dois vértices possuem um vizinho em comum, então a probabilidade de conexão entre esses dois vértices é mais alta. O coeficiente de agrupamento (1.1) representa a probabilidade de dois vizinhos de um dado vértice serem também vizinhos um do outro. Em muitas redes do mundo real estudadas, o coeficiente de agrupamento encontrado obteve um valor muito alto, comparado ao grafo aleatório de Erdős-Rényi, no qual as probabilidades de aresta são independentes por definição, então não há maior probabilidade de dois vértices serem conectados se eles têm ou não um vizinho em comum.

Devido a essas características das redes reais que não são capturadas pelo modelo de grafo aleatório binomial, foram propostos modelos diferentes para descrevê-las. Uma abordagem comum é construir um modelo generalizado de grafo aleatório que tenha a distribuição de grau como parâmetro de entrada, mas que seja aleatório em todos os outros aspectos, ou seja, as arestas são formadas por vértices aleatórios, mas com a condição de manter uma distribuição de grau específica.

Essa generalização revela características das redes que não podem ser obtidas a partir de grafos aleatórios binomiais, como a forma de desenvolvimento do tamanho e da topologia dos agrupamentos, ou a existência de um limiar no qual o agrupamento gigante[1] surge, ou no qual o grafo torna-se conexo. Para chegar a esses valores de limiar, é preciso identificar o parâmetro relevante na evolução das arestas do grafo. No caso do grafo aleatório binomial, o parâmetro é a probabilidade de aresta. Já no grafo generalizado, como a restrição é que a distribuição de grau siga uma lei de potência, o parâmetro de controle é o expoente γ da distribuição. Assim, as redes aleatórias livres de escala são analisadas através da variação de γ para a verificação da existência de um valor limiar de γ no qual propriedades importantes dos grafos são alteradas subitamente.

3.1 Grafos aleatórios com distribuição de grau arbitrária

Construir um grafo aleatório com distribuição de grau diferente da Poisson é relativamente simples. Uma maneira é construí-lo a partir de uma *seqüência* de graus, um conjunto específico $\{k_i\}$ dos graus $d(i)$ dos vértices $i = 1, \ldots, n$ do grafo. A seqüência pode ser gerada diretamente a partir da distribuição p_k desejada, em que p_k estabelece a probabilidade de que um vértice tenha grau k.

O método para a construção do grafo tendo a seqüência de graus é o seguinte: inicia-se com n vértices isolados, e a cada vértice i é atribuído um número k_i de "ganchos" ("pontas de aresta saindo do vértice") e então escolhem-se pares desses ganchos aleatoriamente, os quais são unidos criando uma aresta completa. Quando todos os ganchos forem usados, o grafo aleatório resultante apresenta a seqüência de graus desejada.[2]

Esse método cria um grafo com a distribuição de grau desejada, mas não permite a especificação de um coeficiente de agrupamento, que é uma propriedade importante. Por exemplo, o número médio de vizinhos de um vértice v em um grafo com distribuição

[1] Um agrupamento gigante é uma componente conexa de um grafo G que cresce linearmente com o tamanho de G, isto é, seu tamanho é de ordem $\sim \varepsilon n$, onde $n = |V(G)|$ e $0 < \varepsilon \leq 1$.

[2] Como cada aresta liga dois vértices, o número total de ganchos deve ser par. Portanto a restrição para a seqüência de graus $\{k_i\}$ é que $\sum_i k_i$ seja par.

de grau p_k é $z = \overline{k} = \sum_{k\geq 0} kp_k$.

Supondo que se deseja calcular o número médio de segundos vizinhos (vizinhos dos vizinhos) de v. Em uma rede com alto coeficiente de agrupamento, muitos dos segundos vizinhos de um vértice são também primeiros vizinhos, e esse efeito deveria ser tratado para evitar a contagem repetida do número de segundos vizinhos. Contudo, esse problema pode ser desconsiderado para n grande, pois a probabilidade de que um dos segundos vizinhos de v seja também primeiro vizinho tende a n^{-1} no grafo aleatório, qualquer que seja a distribuição de grau.

No cálculo do número de segundos vizinhos também deve-se considerar o seguinte: como um vértice com grau alto possui mais arestas conectadas a ele, então há uma probabilidade maior de que qualquer aresta no grafo seja conectada a ele, proporcionalmente ao grau do vértice. Assim, a distribuição de probabilidade do grau de um vértice para o qual uma aresta conduz é proporcional a kp_k, e não apenas p_k. Para o cálculo do número dos segundos vizinhos de v, é preciso saber o número de arestas que saem de cada vértice vizinho de v, com exceção da aresta que liga v ao seu vizinho, ou seja, o grau total k do vizinho menos um. Portanto a distribuição normalizada é:

$$q_{k-1} = \frac{kp_k}{\sum_j jp_j} \implies q_k = \frac{(k+1)p_{k+1}}{\sum_j jp_j}, \tag{3.1}$$

onde q_k é a distribuição do número de segundos vizinhos do vértice. O grau médio desse vértice vizinho de v é então:

$$\sum_{k\geq 0} kq_k = \frac{\sum_{k=0}^{\infty} k(k+1)p_{k+1}}{\sum_j jp_j} = \frac{\sum_{k=0}^{\infty}(k-1)kp_k}{\sum_j jp_j} = \frac{\overline{k}^2 - \overline{k}}{\overline{k}}, \tag{3.2}$$

que é o número médio de vértices a uma distância 2 do vértice v alcançados por meio de um de seus vizinhos. Multiplicando isso pelo grau médio de v, que é $\overline{k}$, tem-se o número médio de segundos vizinhos de um vértice:

$$z_2 = \overline{k}^2 - \overline{k}. \tag{3.3}$$

O cálculo pode ser expandido para vizinhos a qualquer distância m de um vértice v, apenas conhecendo o número médio de primeiros e segundos vizinhos, através da iteração da equação:

$$z_m = \frac{\overline{k}^2 - \overline{k}}{\overline{k}} z_{m-1} = \frac{z_2}{z_1} z_{m-1} \quad \Longrightarrow \quad z_m = \left[\frac{z_2}{z_1}\right]^{m-1} z_1. \tag{3.4}$$

3.2 Funções geradoras de probabilidade

Uma abordagem para a análise de propriedades de grafos aleatórios com distribuição de grau arbitrária é a utilização de funções geradoras de probabilidade, conforme formalismo desenvolvido em [7]. Aqui são apresentados os conceitos e propriedades de funções geradoras, e em seguida são demonstradas as aplicações das funções na análise de grafos aleatórios.

Uma *função geradora* tem a forma:

$$G(z) = g_0 + g_1 z + g_2 z^2 + \ldots = \sum_{n \geq 0} g_n z^n, \tag{3.5}$$

e diz-se que $G(z)$ (ou apenas G) é a função geradora para a seqüência $\langle g_0, g_1, g_2, \ldots \rangle$.

Uma *função geradora de probabilidade* é uma representação alternativa de uma distribuição de probabilidade. Por exemplo a distribuição p_k, que é a distribuição dos graus do vértice de um grafo, tem a seguinte função geradora correspondente:

$$G_0(x) = \sum_{k \geq 0} p_k x^k, \tag{3.6}$$

onde p_k é a probabilidade de que um vértice do grafo tenha grau k. Como p_k é uma distribuição de probabilidade, então

$$G_0(1) = \sum_{k \geq 0} p_k = 1. \tag{3.7}$$

Como a distribuição de probabilidade é positiva, $G_0(x)$ é convergente para todo $|x| \leq 1$. Os cálculos a seguir são limitados à região $|x| \leq 1$ e irão mostrar algumas

propriedades importantes das funções geradoras.

A função $G_0(x)$ encapsula toda informação contida na distribuição de probabilidade p_k, uma vez que a distribuição p_k é dada pela k-ésima derivada de G_0:

$$p_k = \frac{1}{k!} \left. \frac{d^k G_0}{dx^k} \right|_{x=0}, \tag{3.8}$$

o que pode ser demonstrado através de uma série de Taylor:

$$\begin{aligned} F(x) &= F(0) + \frac{F^{(1)}(0)x}{1!} + \frac{F^{(2)}(0)x^2}{2!} + \dots \\ &= \sum_{k \geq 0} F^{(k)}(0) \frac{x^k}{k!} \\ &= \sum_{k \geq 0} \left(\frac{d}{dx} \right)^k F(0) \frac{x^k}{x!} \end{aligned}$$

onde $F^{(n)}$ denota a n-ésima derivada de F. Portanto:

$$F^{(k)} = \frac{\left(\frac{d}{dx}\right)^k F(0)}{k!}. \tag{3.9}$$

A média da distribuição de probabilidade gerada por uma função geradora – o grau médio $\overline{k}$ de um vértice no caso de $G_0(x)$ – pode ser calculada diretamente por derivação:

$$G_0'(1) = \sum_k k p_k = \overline{k}. \tag{3.10}$$

Assim, a partir de uma função geradora, é possível calcular a média da distribuição de probabilidade que ela gera. Outros momentos da distribuição podem ser calculados através de derivadas mais altas, em geral, tem-se:

$$\overline{k}^n = \sum_k k^n p_k = \left[\left(x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1}. \tag{3.11}$$

Se uma função geradora gera a distribuição de probabilidade de alguma medida quantitativa k de um objeto, como o grau de um vértice, então a distribuição do total de k somado em n realizações independentes do objeto é gerada pelo coeficiente de x^n, a n-ésima potência da função geradora. Por exemplo, se forem escolhidos n vértices de

um grafo, então a distribuição da soma dos graus desses vértices é gerada por $[G_0(x)]^n$. Para ilustrar essa propriedade, considerando o caso de apenas dois vértices, o quadrado $[G_0(x)]^2$ da função geradora para um único vértice pode ser expandida como:

$$\begin{aligned}
[G_0(x)]^2 &= \left[\sum_k p_k x^k\right]^2 \\
&= \sum_{jk} p_j p_k x^{j+k} \\
&= p_0 p_0 x^0 + (p_0 p_1 + p_1 p_0)x^1 + (p_0 p_2 + p_1 p_1 + p_2 p_0)x^2 \\
&\quad + (p_0 p_3 + p_1 p_2 + p_2 p_1 + p_3 p_0)x^3 + \ldots
\end{aligned}$$

O coeficiente da potência x^n nessa expressão é exatamente a soma de todos os produtos $p_j p_k$ tal que $j+k=n$, o que dá corretamente a probabilidade de que a soma dos graus dos dois vértices seja n. Essa propriedade vale para todas as potências da função geradora.

Outra distribuição importante na análise dos grafos aleatórios é a distribuição dos graus dos vértices alcançados a partir de uma aresta aleatória. Uma aresta alcança um vértice com probabilidade proporcional ao grau daquele vértice, o qual tem, portanto, uma distribuição de probabilidade do grau proporcional a kp_k. A distribuição normalizada é gerada por:

$$\frac{\sum_k k p_k x^k}{\sum_k k p_k} = x\frac{G_0'(x)}{G_0'(1)}, \tag{3.12}$$

pois, conforme (3.10), $G_0'(1) = \sum_k k p_k$, e

$$\begin{aligned}
\sum_k k p_k x^k &= x \sum_k k p_k x^{k-1} \\
&= x\left[\sum_k p_k x^k\right]' \\
&= x G_0'(1).
\end{aligned}$$

Ao escolher um vértice aleatório e seguir cada uma das arestas desse vértice para alcançar os k vizinhos mais próximos, então os vértices alcançados têm a distribuição das arestas restantes gerada por essa função, menos uma potência de x, para descontar

a aresta pela qual o vértice foi alcançado. Assim, a distribuição de tais arestas é gerada pela função

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)} = \frac{G_0'(x)}{z}, \tag{3.13}$$

onde z é o grau médio ($\overline{k}$). A probabilidade de que cada uma dessas arestas esteja conectada ao vértice inicial, ou a qualquer um de seus vizinhos imediatos, tende a n^{-1} (onde n é o tamanho do grafo) e portanto pode ser desconsiderada quando $n \to \infty$. Assim, usando a propriedade de potências de funções geradoras, a função geradora para a distribuição de probabilidade do número de *segundos* vizinhos do vértice inicial pode ser expressa como

$$\sum_k p_k [G_1(x)]^k = G_0(G_1(x)). \tag{3.14}$$

E o número médio z_2 de segundos vizinhos é

$$z_2 = \left[\frac{d}{dx} G_0(G_1(x))\right]_{x=1} = G_0'(1)G_1'(1) = G_0''(1). \tag{3.15}$$

Nas próximas seções serão apresentados cálculos de algumas propriedades de interesse dos grafos aleatórios com distribuição de grau específica, utilizando funções geradoras de probabilidade.

3.3 Tamanhos de componentes e transições de fase

Um estudo feito a partir da generalização de grafos aleatórios é relacionado à transição de fase, na qual surge um componente gigante no grafo, explicada a seguir.

A partir da equação (3.4), que calcula o número médio de vizinhos dos vértices v de um grafo a uma distância m, está claro que se $z_2 > z_1$, então a expressão irá divergir exponencialmente quando $m \to \infty$, portanto o número médio de vizinhos do vértice v a qualquer distância é infinito, e deve existir um componente gigante no grafo. Já se $z_2 < z_1$, o número médio de vértices é finito, e portanto o grafo não deve possuir um componente gigante.

Isto quer dizer que a transição de fase em que o componente gigante aparece é no ponto onde $z_1 = z_2$. Equivalentemente, como $z_1 = \overline{k}$ e com a igualdade (3.3):

$$\begin{aligned}
\overline{k}^2 - \overline{k} &= \overline{k} \\
\overline{k}^2 - 2\overline{k} &= 0 \\
\sum_k k^2 p_k - 2\sum_k k p_k &= 0 \\
\sum_k (k^2 - 2k) p_k &= 0 \\
\sum_k k(k-2) p_k &= 0.
\end{aligned}$$

Portanto, a condição para o surgimento do componente gigante é:

$$\sum_k^{\infty} k(k-2) p_k = 0. \tag{3.16}$$

Uma característica interessante dessa condição (3.16) é que os vértices de grau 0 e grau 2 não contribuem com o somatório, devido ao fator $k(k-2)$, e portanto não afetam a posição da transição de fase da existência do componente gigante. A explicação para isso é que os vértices de grau 0 claramente não afetam a existência do componente gigante, podem ser adicionados ou removidos do grafo sem influenciar o componente gigante possivelmente existente. Já os vértices de grau 2 também não influem, uma vez que sua remoção não altera a estrutura topológica do grafo, pois todos os vértices de grau 2 ficam no meio de arestas entre outros pares de vértices, por exemplo, considerando 3 vértices "–A–B–C–" e removendo o vértice B (de grau 2), obtém-se "–A–C–".

A distribuição de tamanhos das componentes conexas do grafo pode também ser analisada com o auxílio de funções geradoras de probabilidade, como descrito a seguir.

Uma componente conexa pode ser determinada com um algoritmo de busca em largura, partindo de um vértice aleatório, e seguindo suas arestas para alcançar os vizinhos, e em seguida as arestas dos vizinhos, repetidamente até que não sejam encontrados novos vértices. O conjunto dos vértices alcançados nessa busca corresponde a uma componente

conexa. Esse algoritmo é implicitamente descrito por uma função geradora.

Seja $H_1(x)$ a função geradora para a distribuição dos tamanhos de componentes alcançados por uma aresta aleatória e seguindo um de seus vértices. H_1 satisfaz a equação iterativa

$$H_1(x) = \frac{\sum_k kp_k[H_1(x)]^k}{\sum_k kp_k} = xG_1(H_1(x)). \tag{3.17}$$

Na equação acima, kp_k corresponde à probabilidade de uma aresta aleatória alcançar um vértice com grau k, e $[H_1(x)]^k$ representa as k maneiras em que o componente pode ser continuado recursivamente (através dos vizinhos de um vértice alcançado previamente). Iniciando de um vértice aleatório, tem-se um componente ao fim de cada aresta deixando o vértice, e portanto a função geradora para o tamanho da componente conexa inteira é

$$H_0(x) = x\sum_k p_k[H_1(x)]^k = xG_0(H_1(x)). \tag{3.18}$$

O tamanho médio do componente é dado pela primeira derivada da função geradora:

$$\overline{s} = H_0'(1) = 1 + \frac{G_0'(1)}{1 - G_1'(1)}, \tag{3.19}$$

que diverge quando $G_1'(1) = 1$, uma condição para o surgimento do componente gigante, conforme (3.13) e as equações abaixo:

$$\begin{aligned}
G_1(x) &= \frac{\sum_k kp_k x^{k-1}}{\sum_k kp_k} \\
G_1'(x) &= \frac{\sum_k k(k-1)p_k x^{k-2}}{\sum_k kp_k} \\
G_1'(1) &= \frac{\sum_k k(k-1)p_k}{\sum_k kp_k} = 1 \\
\sum_k k(k-1)p_k &= \sum_k kp_k \\
\sum_k k(k-1)p_k - \sum_k kp_k &= 0 \\
\sum_k (k(k-1)-k)p_k &= 0 \\
\sum_k k(k-2)p_k &= 0,
\end{aligned}$$

que é exatamente a condição da equação (3.16). Para qualquer distribuição de grau para a qual o somatório acima é negativo, o grafo não possui componente gigante, mas distribuições de grau que resultam em um somatório positivo levam ao surgimento de um componente gigante.

Quando o grafo não possui um componente gigante (abaixo da transição de fase), a função geradora $H_0(x)$ gera a distribuição de probabilidade para todos os componentes do grafo. Quando existe um componente gigante, $H_0(x)$ gera a distribuição de probabilidade apenas para os componentes finitos. Portanto $H_0(1)$ não assume o valor 1, mas sim $1 - S$, onde S é a fração de vértices contidos no componente gigante. O resultado $S = 1 - H_0(1)$ permite o cálculo do tamanho S do componente gigante:

$$S = 1 - G_0(u), \tag{3.20}$$

onde $u \equiv H_1(1)$ é a menor solução real não-negativa da equação $u = G_1(u)$, pois $H_0(1) = G_0(H_1(1)) = G_0(G_1(H_1(1)))$, conforme as equações (3.17) e (3.18).

3.4 Distância média

Outra medida estudada é a distância média entre dois vértices do grafo (comprimento médio de menor caminho). Se não há componente gigante no grafo, portanto abaixo da transição de fase conforme a equação (3.16), então a maioria dos pares de vértices não estão conectados um ao outro, nesse caso não faz muito sentido o cálculo da distância entre vértices.

Por outro lado, acima da transição de fase, quando existe um componente gigante no grafo, todos os vértices contidos no componente estão conectados por algum caminho a todos os outros. A equação (3.4) pode ser usada para o cálculo da distância média em grafos aleatórios, pois ela determina o número médio de vértices a uma distância m de um dado vértice v do componente gigante.

Quando o número total de vértices a uma distância m é igual ao tamanho n do

grafo, m é igual ao "raio" r do grafo em torno do vértice v. Como $\frac{z_2}{z_1} \gg 1$ acima da transição, o número de vértices a uma distância m cresce rapidamente com m, o que significa que a distância entre v e a maioria dos vértices do grafo é em torno de r, e r é aproximadamente igual à distância média ℓ entre vértices. Portanto, acima da transição, o comprimento médio de caminho ℓ é dado aproximadamente por $z_\ell \simeq n$, ou:

$$\ell = \frac{\log(n/z_1)}{\log(z_2/z_1)} + 1. \tag{3.21}$$

No caso do grafo aleatório de Erdős-Rényi, onde $z_1 = z$ e $z_2 = z^2$, o comprimento médio de caminho é dado por $\ell = \frac{\log n}{\log z}$.

Conforme a equação (3.21), a distância média entre vértices aumenta logaritmicamente com o tamanho n do grafo. Mesmo para grafos muito grandes a distância esperada entre dois vértices é relativamente curta. Em redes sociais, esse fenômeno é conhecido como "mundo pequeno", como visto na seção 1.2.1.

O cálculo do número médio de m-ésimos vizinhos, conforme a equação (3.4), pode ser expresso em termo de funções geradoras como

$$z_m = [G_1'(1)]^{m-1} G_0'(1) = \left[\frac{z_2}{z_1}\right]^{m-1} z_1, \tag{3.22}$$

onde z_1 e z_2 são o número de primeiros e segundos vizinhos, respectivamente. A equação acima é obtida a partir das distribuições de probabilidade para os primeiros e segundos vizinhos, $G_0(x)$ e $G_0(G_1(x))$. A distribuição dos m-ésimos vizinhos é gerada por $G_0(G_1(\ldots G_1(x)\ldots))$, com $m-1$ iterações da função G_1 sobre ela mesma. Definindo $G^{(m)}(x)$ como a função geradora para os m-ésimos vizinhos, tem-se

$$G^{(m)}(x) = \begin{cases} G_0(x) & \text{para } m = 1 \\ G^{(m-1)}(G_1(x)) & \text{para } m \geq 2. \end{cases} \tag{3.23}$$

Então o número médio z_m de m-ésimos vizinhos é

$$z_m = \left.\frac{dG^{(m)}}{dx}\right|_{x=1} = G_1'(1) G^{(m-1)\prime}(1) - G_1'(1) z_{m-1}. \tag{3.24}$$

Como $z_1 = z = G_0'(1)$, verifica-se a igualdade da equação (3.22).

A partir disso é possível estimar a distância média ℓ entre dois vértices do grafo. Assumindo que todos os vértices do grafo podem ser alcançados em ℓ passos, tem-se

$$1 + \sum_{m=1}^{\ell} z_m = n, \tag{3.25}$$

onde n é o tamanho do grafo. A equação (3.22) então pode ser usada para obter o valor estimado de ℓ em função do número médio de primeiros e segundos vizinhos:

$$\begin{aligned} 1 + \sum_{m=1}^{\ell} \left(\frac{z_2}{z_1}\right)^{m-1} z_1 &= n \\ 1 + z_1 \sum_{m=1}^{\ell} \left(\frac{z_2}{z_1}\right)^{m-1} &= n \\ \sum_{m=1}^{\ell} \left(\frac{z_2}{z_1}\right)^{m-1} &= \frac{n-1}{z_1}. \end{aligned}$$

Considerando $r = \frac{z_2}{z_1}$ e $r > 1$,

$$\begin{aligned} \sum_{m=1}^{\ell} r^{m-1} &= \frac{n-1}{z_1} \\ r^0 + r^1 + r^2 + \ldots + r^{\ell} &= \frac{r^{\ell} - 1}{r - 1} \\ \frac{r^{\ell} - 1}{r - 1} &= \frac{n-1}{z_1} \\ r^{\ell} &= \frac{(r-1)(n-1)}{z_1} + 1 \\ \ell &= \log_r \left(\frac{(r-1)(n-1) + z_1}{z_1}\right) \\ &= \log_r((r-1)(n-1) + z_1) - \log_r z_1. \end{aligned}$$

Substituindo r por $\frac{z_2}{z_1}$,

$$\begin{aligned} \ell &= \log_r \left(\left(\frac{z_2}{z_1} - 1\right)(n-1) + z_1\right) - \log_r z_1 \\ &= \log_r \left(\left(\frac{z_2 - z_1}{z_1}\right)(n-1) + z_1\right) - \log_r z_1 \\ &= \log_r \left(\frac{(z_2 - z_1)(n-1) + z_1^2}{z_1}\right) - \log_r z_1 \end{aligned}$$

$$
\begin{aligned}
&= \log_r((z_2 - z_1)(n-1) + z_1^2) - 2\log_r z_1 \\
&= \frac{1}{\ln r}\left(\ln((z_2 - z_1)(n-1) + z_1^2) - 2\ln z_1\right) \\
&= \frac{\ln[(z_2 - z_1)(n-1) + z_1^2] - \ln z_1^2}{\ln(z_2/z_1)}.
\end{aligned}
$$

Como para a maioria dos grafos $n \gg z_1$ e $z_2 \gg z_1$, a equação se reduz a

$$\ell = \frac{\ln(n/z_1)}{\ln(z_2/z_1)} + 1. \tag{3.26}$$

A partir desse resultado da distância média ℓ, pode-se concluir que ℓ escala logaritmicamente com n para todo grafo aleatório, qualquer que seja a distribuição de grau. Outra propriedade interessante da distância média é que embora ℓ seja uma medida global, pode ser calculada em função de medidas locais (como o grau médio de um vértice), e apenas o número de primeiros e segundos vizinhos são importantes para o cálculo de ℓ, assim dois grafos com diferentes distribuições de grau, mas mesmos valores de z_1 e z_2, terão a mesma distância média.

4 Análise espectral de grafos

Nesta seção será descrita uma ferramenta bastante útil na implementação de algoritmos em grafos, a análise espectral, baseada nos polinômios característicos, autovalores e autovetores da matriz de adjacências dos grafos. A partir da aplicação de técnicas espectrais pode-se obter importantes características topológicas de grafos, bem como encontrar soluções aproximadas de problemas para os quais não existe algoritmo de tempo polinomial, o que torna a análise espectral uma eficiente ferramenta de aproximação.

4.1 Espectro de grafos

Todo grafo G sobre os vértices $\{1, \ldots, n\}$ pode ser representado por uma matriz quadrada $A_{n\times n} = (a_{ij})$, onde $a_{ij} = 1$ se $\{i, j\} \in E(G)$ e $a_{ij} = 0$ caso contrário, em outras palavras, uma posição da matriz assume o valor 1 se os vértices correspondentes à linha e à coluna formam uma aresta. Tal matriz é denominada *matriz de adjacências* do grafo, e percebe-se claramente que a matriz é simétrica no caso de grafos não-dirigidos.

Um *autovalor* de uma matriz $A_{n\times n}$ é o número $\lambda \in \mathbb{C}$ se existe um vetor $\mathbf{v} \neq 0$ tal que

$$A\mathbf{v} = \lambda\mathbf{v}. \tag{4.1}$$

O vetor $\mathbf{v}$ é denominado *autovetor* associado ao autovalor λ. Um autovalor de uma matriz não é único, tampouco os autovetores associados a um autovalor.

Teorema 6. *Seja A uma matriz $n \times n$. As seguintes propriedades são equivalentes:*

- *A é invertível;*[1]
- *$A\mathbf{v} = 0$ tem apenas a solução trivial;*
- *os vetores coluna de A são linearmente independentes;*
- $\det(A) \neq 0$.

A equação (4.1) pode ser escrita na forma $(A - \lambda\mathrm{Id})\mathbf{v} = 0$, onde Id é a matriz identidade de ordem n. Pelo Teorema 6, para que sejam admitidas soluções não-nulas, a matriz denotada pelo termo $A - \lambda\mathrm{Id}$ deve ser não-invertível, ou seja, $\det(A - \lambda\mathrm{Id}) = 0$. O determinante $\det(A - \lambda\mathrm{Id})$ é um polinômio em λ denominado *polinômio característico* de A:

$$p_A(x) = \det(A - x\mathrm{Id}), \tag{4.2}$$

cujo grau corresponde à ordem da matriz A (ou seja, n). Portanto, a condição para a existência de autovetores associados a um autovalor λ pode ser expressa por

$$p_A(\lambda) = 0 \Leftrightarrow \exists \mathbf{v}, A\mathbf{v} = \lambda\mathbf{v}. \tag{4.3}$$

Pelo Teorema Fundamental da Álgebra, toda matriz possui autovalor $\lambda \in \mathbb{C}$. Se A é real e λ é real, então A possui autovetores reais. A *multiplicidade algébrica* de um autovalor λ, $\mathrm{ma}(\lambda)$, é a multiplicidade da raiz λ de $p_A(x) = 0$, ou seja, o número de fatores $(x - \lambda)$ no polinômio característico da matriz.

O espectro de um grafo é o conjunto dos autovalores de sua matriz de adjacências, normalmente denotado por

$$\mathrm{Sp}(G) = (\lambda_1, \lambda_2, \ldots, \lambda_n), \tag{4.4}$$

que corresponde à seqüência dos autovalores em ordem não-crescente, isto é, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

[1] Uma matriz A é invertível se existe uma matriz B tal que $AB = BA = \mathrm{Id}$ (matriz identidade).

O espectro de um grafo não determina o grafo unicamente, isto é, existem grafos não isomorfos que compartilham o mesmo espectro. Contudo, ele determina o número de vértices, arestas, e muitas outras propriedades topológicas do grafo.

A Teoria Espectral dos Grafos, utilizada nas análises e aplicações de técnicas descritas no decorrer desta seção, é o estudo das propriedades de um grafo relacionadas aos polinômios característicos, autovalores e autovetores de sua matriz de adjacências (ou a matriz com alguma forma de normalização, como a Laplaciana).

Abaixo são listados alguns fatos para ilustrar o relacionamento entre as propriedades estruturais de um grafo e os autovalores e autovetores da matriz de adjacências (os fatos não são provados aqui, mas maiores detalhes podem ser encontrados em [11]):

1. O maior autovalor λ_1 de um grafo d-regular[2] é d e o autovetor correspondente é constante.

2. Todos os autovalores λ_i, $2 \leq i \leq n$, de um grafo d-regular uniformemente aleatório (isto é, todo grafo d-regular sobre n vértices é igualmente possível) são $|\lambda_i| = O(\sqrt{d})$, quase certamente.

3. Os autovalores λ_i, $1 \leq i \leq n$, de um grafo com m arestas e grau máximo d são limitados por $|\lambda_i| \leq \min\{\sqrt{m}, d\}$.

4. O espectro da união de grafos disjuntos é a união de seus espectros.

5. Se A e B são as matrizes de adjacências de grafos não necessariamente disjuntos com autovalores $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ e $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_n$, então os autovalores de sua união $C = A + B$ são $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_n$ com $\alpha_i - \beta_n \leq \gamma_i \leq \alpha_i + \beta_1$, $1 \leq i \leq n$.

[2]Um grafo é dito d-regular quando todos os seus vértices possuem grau d.

4.2 Aplicação de técnicas espectrais em grafos da Internet

Como existem algoritmos eficientes para o cálculo dos autovalores e autovetores de matrizes relacionadas a grafos, muitos problemas são resolvidos com a utilização de técnicas espectrais, baseadas no espectro dos grafos, o que permite a obtenção de informações sobre as propriedades topológicas dos grafos, bem como a aproximação de valores para os quais o cálculo não pode ser feito em tempo polinomial.

Alon [10] descreve a aplicação de técnicas espectrais para a resolução de três problemas NP-difíceis: bissecção, coloração e cálculo do número de independência. É possível desenvolver algoritmos eficientes que resolvem tais problemas para quase todos os grafos em classes definidas apropriadamente, o que é desejável, pois estes problemas surgem freqüentemente na prática e os grafos de entrada não são necessariamente os piores casos.

Quando a análise espectral não é feita diretamente na matriz de adjacências do grafo, são realizadas normalizações na matriz antes da aplicação das técnicas. A mais utilizada é a matriz Laplaciana, definida a seguir.

A matriz Laplaciana de um grafo G é definida por

$$Q = D - A, \tag{4.5}$$

onde $D = (d_{u,v})_{u,v \in V}$ é a matriz diagonal em que $d_{u,u}$ é o grau de u em G, $d(u)$, e $d_{u,v} = 0$ para todo $u \neq v$. Percebe-se que todas as linhas e colunas da matrizes somam zero. Tanto a matriz de adjacências de um grafo quanto sua matriz Laplaciana são simétricas e, portanto, possuem autovalores reais e uma base ortonormal de autovetores.

Há uma estreita relação entre os autovalores das matrizes de adjacências (A) e Laplaciana (Q) e várias propriedades estruturais do grafo, o que pode ser útil na implementação de algoritmos eficientes, visto que os autovalores e autovetores das matrizes podem ser calculados eficientemente.

A seguir são apresentadas análises espectrais realizadas em grafos da Internet, no nível de sistema autônomo (AS), onde cada vértice representa uma rede sob um específico controle administrativo, e dois vértices são conectados se existe uma conexão física entre duas estações das redes correspondentes aos vértices.

Em [11] a análise espectral é realizada examinando-se os autovetores correspondentes aos maiores autovalores de matrizes relacionadas à matriz de adjacências do grafo. O método de filtragem espectral executado sugere agrupamentos de sistemas autônomos com proximidade semântica natural (seja por localização geográfica ou interesses comuns), e tais propriedades de agrupamento podem insinuar padrões de tráfego, o que permite a previsão de congestionamento em determinados pontos da rede. Além disso são apresentados outros resultados a partir da análise espectral, como a classificação hierárquica de sistemas autônomos, a persistência de certas propriedades do grafo no decorrer do tempo, etc.

O método de *filtragem espectral*, utilizado por [11], é baseado nos maiores autovalores de matrizes relacionadas à matriz de adjacências, e procura agrupamentos nos autovetores associados a estes autovalores. A primeira referência aos maiores autovalores da matriz de adjacências é a "lei de potência de autovalores", que foi relatada em [13] e será descrita na seção 4.3. Há diversas aplicações para a relação entre a filtragem espectral e a conectividade do grafo, especialmente na recuperação de informações e mineração de dados, onde agrupamentos representam grupos de dados com proximidade semântica.

Os autovetores correspondentes aos maiores autovalores tendem a capturar características globais do grafo, tais como grupos de vértices $S \in V$ para os quais a razão

$$\frac{\text{arestas em } S}{\text{arestas incidentes a } S} = \frac{|\{\{i,j\} \in E : i \in S, j \in S\}|}{|\{\{i,j\} \in E : i \in S, j \in V\}|} \tag{4.6}$$

é pequena, indicando agrupamentos de conectividade relativamente alta e, assim, grande proximidade semântica. Encontrar um conjunto S que minimiza a razão (4.6) acima é um problema NP-difícil, e portanto o método espectral é um eficiente algoritmo de

aproximação. Os autovetores correspondentes aos menores autovalores tendem a capturar ruído, ou características locais que são explícitas ou podem ser facilmente computadas a partir do grafo.

Em linhas gerais, o método de filtragem espectral aplicado por [11], para uma matriz simétrica $A_{n\times n}$, é dividido em dois passos:
PASSO 1: Calcule os k maiores autovalores de A e seus respectivos autovetores. O parâmetro k depende da aplicação e da instância do grafo, mas é sempre uma ou duas ordens de grandeza menor do que n.
PASSO 2: Para cada i, $1 \leq i \leq k$, seja $\mathbf{e}$ o autovetor associado a λ_i. Ordene os vértices de acordo com o peso atribuído por $\mathbf{e}$. Corte na direção dos extremos mais positivos e mais negativos. Os grupos resultantes são candidatos para agrupamento e/ou proximidade semântica.

A análise espectral realizada por [11] foi executada sobre topologias AS da Internet. Os dados topológicos foram obtidos do NLANR[3]. Os fatos mencionados no final da seção 4.1 dizem respeito basicamente a grafos regulares. Porém, na Internet isso não se aplica, pois os graus das topologias são relativamente distorcidos, e o método da filtragem espectral se deteriora rapidamente quando há grande variação na freqüência dos graus. Para contornar esse problema, a análise da topologia não-dirigida foi realizada sobre a normalização estocástica $N(A)$ da matriz de adjacências A do grafo, definida abaixo.

A normalização estocástica, ou normalização de freqüência inversa, é uma abordagem utilizada para restaurar a filtragem espectral em grafos não-regulares. Basicamente, a normalização divide cada entrada a_{ij} da matriz pela soma $\sum_j a_{ij}$ das entradas da linha correspondente, obtendo uma matriz onde todas as linhas somam 1. A matriz estocástica resultante da normalização, assim como matrizes simétricas, também possui um espectro completo de autovalores e autovetores reais, e o intervalo dos autovalores é normalizado no intervalo $(-1, 1)$. Se todas as entradas diagonais forem $a_{ii} = 1/2$ e as demais entradas forem multiplicadas por $1/2$, o intervalo dos autovalores é alterado para $(0, 1)$. Em suma,

[3]National Laboratory for Applied Network Research, http://www.nlanr.net/

a normalização estocástica $N(A) = (na_{ij})$ de uma matriz $A = (a_{ij})$ é denotada por

$$N(A) = \begin{cases} na_{ij} = 1/2 & \text{para } i = j \\ na_{ij} = \frac{1}{2}\frac{a_{ij}}{\sum_j a_{ij}} & \text{para } i \neq j \end{cases} . \tag{4.7}$$

Como resultado da análise espectral, foi observado que os autovetores correspondentes aos maiores autovalores da matriz normalizada indicam agrupamentos de vértices na topologia AS da Internet, que possuem alguma proximidade semântica (seja por localização geográfica ou por interesses comuns). Além disso, análises realizadas com dados topológicos obtidos em diferentes períodos mostraram que o comportamento espectral da topologia da Internet é consistente no decorrer do tempo, o que confirma a idéia de que o espectro é uma característica robusta de uma topologia.

Outra análise interessante foi realizada com o estudo das propriedades topológicas do grafo em conexão com a funcionalidade da rede, onde os autovetores correspondentes aos maiores autovalores são relacionados com o desempenho da rede, para prever possíveis locais de congestionamento. Para uma topologia de rede não-dirigida, supondo que seja enviada uma unidade de tráfego através do caminho mais curto, de cada vértice para todo outro vértice da rede. Isso induz uma *tensão*[4] para cada ligação, definida pelo número total de caminhos que passam pela ligação. A tensão máxima pode ser considerada como um indicador de congestionamento na rede. Observou-se que topologias com autovalores mais altos tendem a exibir pior comportamento referente à tensão nas ligações, portanto conclui-se que, devido ao forte agrupamento, a topologia AS da Internet é mais vulnerável a congestionamentos do que se fosse mais homogênea e menos agrupada.

Ainda um outro estudo realizado a partir da filtragem espectral foi a classificação dos sistemas autônomos pelo primeiro autovetor. Foi proposta uma heurística que atribui "importância" e níveis hierárquicos aos sistemas autônomos, com uma adaptação do recurso de *pagerank* do Google, usado para inferir a qualidade de páginas *Web*. Na WWW, uma ligação apontando de uma página i para uma página j indica uma "transferência"

[4] *link stress*, no artigo original.

de importância de i para j. De forma análoga, na Internet, uma aresta apontando de um cliente i para um provedor j (no caso de topologia dirigida) pode ser considerada como uma transferência de importância similar. Maiores detalhes sobre a classificação a partir dos autovetores correspondentes aos maiores autovalores podem ser obtidos em [11].

Uma análise espectral similar da topologia da Internet no nível de sistemas autônomos foi realizada por D. Vukadinović *et al.* [12], mas sobre o espectro Laplaciano normalizado de grafos AS, conforme definição a seguir. Para a análise foram utilizados grafos da topologia AS obtidos entre 1997 e 2001, e com a aplicação de técnicas espectrais foram identificadas propriedades estruturais dos grafos que não são capturadas pelos métodos convencionais, e também foi observado que certas propriedades permanecem estáveis no decorrer do tempo, apesar do crescimento do número de vértices e arestas.

Estudos anteriores foram baseados nos maiores autovalores das matrizes dos grafos, mas na análise de [12] é observado todo o conjunto de autovalores do grafo, no caso o espectro Laplaciano normalizado, pois tem a vantagem de todos os autovalores pertencerem ao intervalo $[0, 2]$, o que torna fácil a comparação do espectro de grafos com diferentes tamanhos. Seja $d(v)$ o grau do vértice v. A matriz Laplaciana normalizada utilizada na análise é definida por

$$\mathcal{L}(G)(u,v) = \begin{cases} 1 & \text{se } u = v \text{ e } d(v) \neq 0, \\ -\frac{1}{\sqrt{d(u)d(v)}} & \text{se } u \text{ e } v \text{ são adjacentes,} \\ 0 & \text{caso contrário.} \end{cases} \tag{4.8}$$

Se A é a matriz de adjacências do grafo G e D é uma matriz diagonal tendo $d_{ii} = d(v_i)$, então $\mathcal{L}(G) = D^{-\frac{1}{2}}(D - A)D^{\frac{1}{2}}$. O espectro Laplaciano normalizado (*nls*) é o conjunto de autovalores de $\mathcal{L}(G)$, isto é, todos os valores λ tais que $\mathcal{L}(G)\mathbf{u} = \lambda\mathbf{u}$ para algum $\mathbf{u} \in \mathbb{R}^n, \mathbf{u} \neq 0$.

O espectro Laplaciano normalizado $\mathcal{L}(G)$, como dito acima, para qualquer G, está contido no intervalo fechado $[0, 2]$ e o menor autovalor é sempre 0. A multiplicidade do autovalor 0 é igual ao número de componentes conexas de G. O maior autovalor é menor

ou igual a 2, onde a igualdade indica uma componente conexa bipartida. O espectro de um grafo é a união do espectro de suas componentes conexas. Além dessas e outras propriedades interessantes do espectro Laplaciano normalizado, [12] apresenta um limite inferior para a multiplicidade do segundo maior autovalor, baseando-se em características topológicas do grafo, além de observações sobre a persistência do espectro através do tempo, e outros resultados e interpretações físicas.

4.3 Autovalores de grafos com distribuição de grau *power-law*

As leis de potência da topologia da Internet foram relatadas inicialmente por Faloutsos [13], onde são observadas três relações *power-law* a partir da análise de propriedades do grafo que demonstram certa regularidade, apesar da aleatoriedade aparente da Internet. Foi observado que todas as três leis de potência são consistentes através do tempo, apresentando variação insignificante na análise de grafos obtidos em diferentes épocas. As análises foram realizadas com grafos da Internet inter-domínio (sistemas autônomos) e as leis de potência são descritas a seguir.

A primeira lei de potência é o *expoente de classificação* $\mathcal{R}$. Tomando-se os graus $d(v)$ dos vértices ordenados em ordem decrescente, e sendo r_v a "classificação" de um vértice, seu índice na ordem decrescente dos graus, a relação entre a classificação r_v e o grau $d(v)$ segue uma lei de potência. Portanto, o grau $d(v)$ de um vértice v é proporcional à classificação do vértice, r_v, elevada à potência de uma constante $\mathcal{R}$:[5]

$$d(v) \sim r_v^{\mathcal{R}}. \tag{4.9}$$

[5]O símbolo "$\sim$" nas equações quer dizer "proporcional a".

A segunda lei de potência é o *expoente de grau* $\mathcal{O}$. Seja f_d a freqüência de um grau d, ou seja, o número de vértices no grafo cujo grau é d. A relação entre a freqüência f_d e o grau d segue uma lei de potência. Portanto, a freqüência f_d de um grau d é proporcional ao grau d elevado à potência de uma constante $\mathcal{O}$:

$$f_d \sim d^{\mathcal{O}}. \tag{4.10}$$

A terceira e última lei de potência relatada por [13] – a mais importante para o estudo dessa seção – é o *auto-expoente* ε, identificada a partir de propriedades dos autovalores dos grafos da Internet. Como visto anteriormente, os autovalores de um grafo estão estreitamente relacionados com muitas propriedades topológicas básicas tais como o diâmetro, o número de arestas, árvores geradoras, componentes conexas, etc. Tomando-se os autovalores da matriz de adjacências do grafo em ordem não-crescente, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, a relação entre o autovalor λ_i e sua ordem i segue uma lei de potência. Portanto, os autovalores λ_i de um grafo são proporcionais à sua ordem i, elevada a potência de uma constante ε:

$$\lambda_i \sim i^{\varepsilon}. \tag{4.11}$$

O auto-expoente ε é definido como a inclinação do gráfico dos autovalores λ_i ordenados contra sua ordem i, em escala $\log \times \log$. Foi observado que os auto-expoentes de grafos AS da Internet analisados em diferentes épocas são praticamente iguais, o que significa que o auto-expoente captura a propriedade da Internet que caracteriza todas as instâncias, apesar do incremento no tamanho. Por outro lado, o auto-expoente de grafos analisados no nível de roteador variou significativamente, o que mostra que o auto-expoente pode fazer distinção entre famílias diferentes de grafos.

Como observado por [11], a lei de potência de autovalores de Faloutsos é uma conseqüência direta da lei de potência de graus (descrita no capítulo 3), o que pode ser observado pelos seguintes passos:

PASSO 1: Decomponha uma topologia AS A em $A = F + E$. Inicialmente F é o conjunto de vértices que têm os k maiores graus, $d_1, d_2, \ldots, d_k$, e F não contém arestas. Seja E o grafo da topologia AS inteira. Algumas arestas serão removidas de E e adicionadas a F, para criar k estrelas[6] disjuntas em F, pelo seguinte processo: Para cada vértice v que não está em F, se v é incidente a k_v vértices de F, escolha um destes vértices u com probabilidade proporcional ao grau de u no grafo inteiro, faça a aresta $\{v, u\}$ incidente ao vértice $u \in F$ e remova a aresta $\{v, u\}$ de E. Note que F é agora um conjunto de estrelas de vértices disjuntos com graus $d'_1, d'_2, \ldots, d'_k$, e E é a topologia AS inicial onde todas as arestas pertencentes às estrelas foram removidas.

PASSO 2: Note que os autovalores de uma estrela de grau d são $\pm\sqrt{d-1}$ e 0 com multiplicidade $d-1$. Portanto os maiores autovalores de F são $\sqrt{d'_1}, \sqrt{d'_2}, \ldots, \sqrt{d'_k}$.

PASSO 3: Para topologias AS típicas, observa-se que o procedimento acima, para $k = 100$, dá $d'_i \sim d_i$, $1 \leq i \leq k$, daí os maiores autovalores de F são próximos de $\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_k}$, e os maiores autovalores de E são, no pior caso, estritamente menores que $\sqrt{d_1}$ e, na média, $\sqrt{d_1}/5$. Os maiores autovalores de $A = F + E$ podem ser considerados próximos de $\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_k}$. Daí, para grafos onde os maiores graus seguem Zipf[7], como [13] relata da Internet AS, os maiores autovalores seguem uma lei de potência com expoente próximo de 0,5, também como [13] relata para a Internet AS.

Como visto em (4.10), grafos da topologia da Internet exibem o comportamento *power-law* (lei de potência), em que o número de vértices de grau k é proporcional a k^β, com $\beta > 0$. Fan Chung *et al.* [14] mostraram que se $\beta > 2,5$, o maior autovalor de um grafo *power-law* aleatório é quase certamente $(1 + o(1))\sqrt{\Delta}$,[8] onde Δ é o grau máximo do grafo. Também foi mostrado que os k maiores autovalores de um grafo *power-law* aleatório com expoente β têm distribuição *power-law* com expoente $2\beta - 1$ se o grau máximo é suficientemente grande, onde k é uma função que depende de β, Δ e d, o grau médio. Quando $2 < \beta < 2,5$, o maior autovalor é próximo de $c\Delta^{3-\beta}$ para alguma

[6] Uma estrela é um grafo bipartite completo da forma $K_{1,s}$, onde um dos vértices possui grau $s = n-1$, e todos os demais possuem grau 1.

[7] Zipf é uma lei de potência $P_i \sim 1/i^a$, com o expoente a próximo de 1.

[8] Uma função $f(n)$ é $o(g(n))$ se $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

constante c dependente de β e do grau médio d.

Os resultados apresentados por [14] para grafos *power-law* são uma conseqüência imediata de um resultado genérico para autovalores de grafos aleatórios com distribuição de grau arbitrária. O modelo utilizado foi uma generalização do grafo aleatório de Erdős-Rényi com a seqüência de graus esperados $(w_1, w_2, \ldots, w_n)$, conforme descrito na seção 3.1. O maior autovalor λ_1 da matriz de adjacências de um grafo aleatório nesse modelo depende do grau máximo Δ e do grau médio de segunda ordem $\tilde{d}$, definido por:

$$\tilde{d} = \frac{\sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} w_i}. \tag{4.12}$$

Observou-se que λ_1 é quase certamente $(1 + o(1))\sqrt{\Delta}$ se $\sqrt{\Delta}$ é maior que $\tilde{d}$ por um fator de $\log^2 n$ e λ_1 é quase certamente $(1 + o(1))\tilde{d}$ se $\sqrt{\Delta}$ é menor que $\tilde{d}$ por um fator de $\log n$. Em outras palavras, λ é, assintoticamente, o máximo de $\sqrt{\Delta}$ e $\tilde{d}$ se os dois valores de $\sqrt{\Delta}$ e d são muito distantes. Além do mais, se o k-ésimo maior grau esperado m_k é maior que $\tilde{d}$ por um fator de $\log n$, então os k maiores autovalores são $(1+o(1))\sqrt{m_k}$.

Para grafos *power-law* aleatórios, onde a distribuição de grau segue lei de potência com expoente β (número de vértices com grau k proporcional a $k^{-\beta}$), foi demonstrado o seguinte teorema, onde Δ é o grau máximo e d é o grau médio do grafo.

Teorema 7. *Para um grafo G com distribuição de grau* power-law *com expoente β, tem-se:*

- *Para $\beta \geq 3$ e $m > d^2 \log^3 n$, o maior autovalor do grafo aleatório* power-law *é quase certamente $(1 + o(1))\sqrt{m}$.*
- *Para $3 > \beta > 2,5$ e $m > d^{\frac{\beta-2}{\beta-2,5}} \log^{\frac{3}{\beta-2,5}} n$, o maior autovalor é quase certamente $(1 + o(1))\sqrt{m}$.*
- *Para $2 < \beta < 2,5$ e $m > \log^{\frac{3}{2,5-\beta}} n$, o maior autovalor é quase certamente $(1+o(1))\tilde{d}$.*

- *Para* $k < n(\frac{d}{m \log n})^{\beta-1}$ *e* $\beta > 2,5$*, quase certamente os* k *maiores autovalores do grafo* power-law *aleatório com expoente* β *têm distribuição* power-law *com expoente* $2\beta - 1$*, sendo* m *suficientemente grande.*

A prova desse teorema e maiores detalhes sobre os resultados alcançados para autovalores de grafos com distribuição de grau *power-law* podem ser encontrados em [14].

5 Tolerância a erros em redes

Redes de comunicação têm a necessidade de oferecer confiabilidade para que o tráfego de informações entre as estações conectadas ocorra de maneira adequada, com mecanismos que garantam a continuidade de seu funcionamento após a ocorrência de eventuais erros (queda de conexão, remoção de estação, entre outros). Redes grandes e complexas costumam ser altamente robustas nesse sentido: a ocorrência de falhas locais em pontos específicos raramente acarreta em problemas no funcionamento global da rede.

A robustez que faz com que a estabilidade de uma rede seja mantida após a ocorrência de erros é normalmente atribuída à existência de ligações redundantes em sua estrutura física. Entretanto, além da redundância, a topologia da rede também assume um importante papel na tolerância a tais erros nos sistemas complexos. Albert e Barabási [4] apresentaram um estudo sobre robustez e tolerância a erros causados após alterações em aspectos topológicos da rede, isto é, após a remoção de vértices e/ou arestas.

Se for removida uma fração p das arestas de um grafo G arbitrário com n vértices, a probabilidade de que o grafo resultante permaneça conexo depende da probabilidade de remoção p. Para uma grande classe de grafos G existe uma probabilidade limiar $p_c(n)$ tal que se $p < p_c(n)$ o subgrafo resultante da remoção continua conexo, mas se $p > p_c(n)$ o subgrafo é desconexo. Assim como na evolução do grafo aleatório (seção 2.3), a probabilidade crítica depende do número de vértices no grafo.

Uma vez que a remoção de um vértice implica na perda de todas as arestas às quais o vértice pertence, a remoção de vértices provoca mais danos do que a remoção de arestas de um grafo. A seguir são apresentados resultados que indicam uma estreita

relação entre topologia e robustez da rede, convencionando que uma rede é denominada robusta (tolerante a erros) se contém um agrupamento gigante composto pela maioria dos vértices, mesmo após uma fração de seus vértices ter sido removida.

Foi realizada uma comparação da tolerância a erros entre o modelo de grafo aleatório de Erdős-Rényi e o modelo livre de escala, analisando-se grafos com o mesmo número de vértices e arestas, diferindo apenas em sua distribuição de grau. Em ambos os grafos o estudo foi feito com a utilização de dois critérios de remoção de vértices: o primeiro consiste em remover vértices selecionados aleatoriamente (todos com a mesma probabilidade), para simular erros aleatórios em pontos locais da rede; o segundo critério é a remoção do vértice de maior grau, a cada passo, simulando um ataque intencional contra a rede, já que a ausência dos vértices altamente conectados tende a comprometer mais a integridade do sistema.

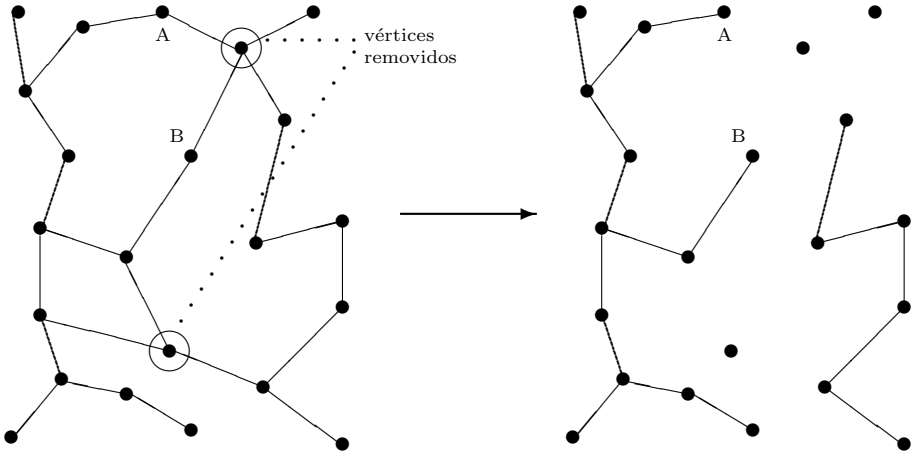


Figura 4: Ilustração dos efeitos da remoção de um vértice em uma rede inicialmente conectada. No estado inicial a distância entre os vértices A e B é 2, mas após a remoção de dois vértices ela aumenta para 6 e a rede é dividida em 5 agrupamentos isolados.

Partindo de uma rede inicialmente conexa, a cada passo é removido um vértice (o que implica na remoção de todas as arestas às quais o vértice pertencia, rompendo alguns caminhos entre os vértices restantes). Uma forma de monitorar o rompimento de uma rede inicialmente conexa é estudar o tamanho relativo do maior agrupamento que permanece conexo, S, e a distância média ℓ desse agrupamento, em função da fração f dos vértices removidos do sistema. Após uma remoção, é esperado que o tamanho do

maior agrupamento diminua e que a distância média ℓ aumente conforme os vértices são removidos da rede.

5.1 Redes aleatórias (modelo binomial)

No modelo de grafo aleatório binomial de Erdős-Rényi, a maioria dos vértices apresentam os maiores graus, e o maior grau do grafo é próximo do grau médio, pn. Foram observadas em uma rede aleatória binomial as mudanças no tamanho relativo do maior agrupamento S, ou seja, a fração de vértices contidos no maior agrupamento da rede, e também sua distância média ℓ, à medida que vértices aleatórios são removidos.

Inicialmente a rede é conexa, portanto o tamanho relativo do maior agrupamento é $S = 1$. Certamente o tamanho de S diminui quando a fração f de vértices removidos aumenta. Mais do que isso, existe uma fração crítica f_c em que S cai para 0, o que indica que a rede é dividida em minúsculos agrupamentos isolados. O comprimento médio de caminho ℓ apresenta comportamento semelhante: aumenta junto com a fração f quando são rompidos caminhos na rede, e tem pico em f_c. Porém, quando a rede se divide em agrupamentos isolados, ℓ também diminui, pois nesse caso o tamanho do maior agrupamento é muito menor.

No caso de um ataque intencional, quando os vértices com o maior número de arestas são atingidos, a rede é rompida muito antes que no caso da remoção aleatória de vértices e, portanto, a fração crítica f_c é muito mais baixa que no caso aleatório. Isso é compreensível, uma vez que a cada passo o maior número possível de arestas é removido da rede.

5.2 Redes livres de escala

Em se tratando de redes livres de escala (cuja distribuição de grau segue uma lei de potência), o resultado é bastante diferente do modelo aleatório binomial, quando uma fração de seus vértices é removida aleatoriamente. Simulações indicam que enquanto S – o tamanho do maior agrupamento – diminui, ele alcança 0 em uma fração f mais alta. Ao mesmo tempo, o comprimento médio de caminho ℓ aumenta mais lentamente que no caso aleatório binomial.

Simplificando, redes livres de escala apresentam uma robustez excepcional contra falhas aleatórias de vértices. Isso ocorre porque a grande maioria dos vértices de um grafo livre de escala possui grau muito baixo. Com isso, a probabilidade de um vértice aleatório pertencer a poucas arestas é mais alta.

Entretanto, quando é realizada a remoção preferencial de vértices (simulando um ataque intencional), em que os vértices de maior grau são removidos a cada passo, o resultado é ainda pior que no caso da rede aleatória binomial, pois uma minoria dos vértices possuem os graus mais elevados, e a remoção de tais vértices compromete rapidamente o funcionamento da rede.

Em suma, simulações indicam que redes livres de escala apresentam uma robustez topológica contra falhas aleatórias de vértices. A origem dessa tolerância a erros está em sua topologia heterogênea: vértices de grau baixo são muito mais abundantes que vértices com alto grau, desse modo a seleção aleatória irá, com maior probabilidade, afetar os vértices que assumem um papel marginal na topologia total da rede. Porém, a mesma heterogeneidade torna as redes livres de escala frágeis perante ataques intencionais, pois a remoção dos vértices altamente conectados tem um efeito dramático sobre a rede.

A tolerância a erros em redes estudada aqui é estática, pois diz respeito apenas ao sentido topológico da rede, na forma que um vértice removido afeta outros. Por outro lado, em muitas redes existe um aspecto dinâmico para tolerância a erros: a remoção de

um vértice pode afetar a funcionalidade de outros vértices também, em cascata. Albert e Barabási [4] mencionam sobre um estudo de tais falhas em cascata e apresentam análises da tolerância a erros em redes reais (incluindo a Internet e a *World-Wide-Web*).

Após a análise da tolerância a erros a partir de alterações na topologia da rede, conclui-se que redes livres de escala são mais robustas do que redes aleatórias (no modelo binomial de Erdős-Rényi) contra falhas de vértices aleatórios (simulando erros reais), porém são mais vulneráveis quando os vértices mais conectados (de maior grau) são tomados como alvo (simulando ataques intencionais).

Referências

[1] R. Diestel. *Graph Theory.* Springer-Verlag, 2000.

[2] R. Graham, D. Knuth, O. Patashnik. *Concrete Mathematics.* Addison-Wesley, 1999.

[3] J. A. Bondy e U. S. R. Murty. *Graph Theory With Applications.* Macmillan Press, 1977.

[4] R. Albert e A. Barabási. "Statistical Mechanics of Complex Networks". Reviews of Modern Physics. 2001.

[5] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal. "The Web as a Graph". Symp. Principles of Database Systems. 1999.

[6] M. E. J. Newman. "Random Graphs as Models of Networks". Handbook of Graphs and Networks. 2002.

[7] M. E. J. Newman, S. H. Strogatz, D. J. Watts. "Random Graphs With Arbitrary Degree Distributions and Their Applications". Physical Review. 2000.

[8] P. Erdős, A. Rényi. "On Random Graphs". Publicationes Mathematicae Debrecen. 1959.

[9] M. Molloy, B. Reed. "A Critical Point for Random Graphs With a Given Degree Sequence". Random Structures & Algorithms. 1995.

[10] N. Alon. "Spectral Techniques in Graph Algorithms". Theoretical Informatics, Third Latin American Symposium. 1998.

[11] C. Gkantsidis, M. Mihail, E. Zegura. "Spectral Analysis of Internet Topologies". In Proceedings of Infocom. 2003.

[12] D. Vukadinović, P. Huang, T. Erlebach. "A Spectral Analysis of the Internet Topology". Technical report, ETH Zurich. 2001.

[13] M. Faloutsos, P. Faloutsos, C. Faloutsos. "On Power-Law Relationships of the Internet Topology". ACM SIGCOMM. 1999.

[14] F. Chung, L. Lu, V. Vu. "Eigenvalues of Random Power law Graphs". Annals of Combinatorics. 2003.

[15] A. Medina, A. Lakhina, I. Matta, J. Byers. BRITE: An Approach to Universal Topology Generation. 2001. *http://www.cs.bu.edu/brite/.*