

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

**Medidas de Centralidade em Grafos
e Aplicações em redes de dados**

por

Elizandro Max Borba

Dissertação submetida como requisito parcial
para a obtenção do grau de
Mestre em Matemática Aplicada

Prof. Dr. Vilmar Trevisan
Orientador

Porto Alegre, 4 de Outubro de 2013.

CIP – CATÁLOGO NA PUBLICAÇÃO

Borba, Elizandro Max

Medidas de Centralidade em Grafos e Aplicações em redes de dados / Elizandro Max Borba. Porto Alegre: PPGMAp da UFRGS, 2013.

61 p.: il.

Dissertação (mestrado) Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Matemática Aplicada, Porto Alegre, 2013.

Orientador: Trevisan, Vilmar

Dissertação: Matemática Aplicada

Grafos, Análise de Redes, Centralidade, Vulnerabilidade.

Medidas de Centralidade em Grafos e Aplicações em redes de dados

por

Elizandro Max Borba

Dissertação submetida ao Programa de Pós-Graduação em Matemática Aplicada do Instituto de Matemática da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de

Mestre em Matemática Aplicada

Linha de Pesquisa: Análise Numérica e Computação Científica

Orientador: Prof. Dr. Vilmar Trevisan

Banca Examinadora:

Prof. Dr. Leonardo Silva de Lima
CEFET/RJ

Profa. Dra. Luciana Salete Buriol
Instituto de Informática-UFRGS

Prof. Dr. Carlos Hoppen
Instituto de Matemática-UFRGS

Porto Alegre, 4 de Outubro de 2013.

Sumário

Lista de Figuras.....	vi
Lista de Abreviaturas.....	viii
Resumo	x
Abstract	xi
1. Introdução.....	1
2. Conceitos Básicos.....	6
2.1 Introdução	6
2.2 Teoria dos Grafos	6
2.3 Álgebra Linear e Teoria Espectral dos Grafos	9
3. Análise de Redes.....	11
3.1 Introdução	11
3.2 Histórico	12
3.3 Modelos de rede.....	16
3.3.1 Grafos Aleatórios	17
3.3.2 Redes com conexão preferencial.....	18
3.4 Ferramentas da Análise de redes	19
3.4.1 Distribuição de grau.....	20
3.4.2 Diâmetro e distância	20
3.4.3 Agrupamento e estrutura de comunidades	21
3.4.4 Motifs.....	22
3.4.5 Mediação e Centralidade.....	24
3.4.6 Processos.....	25
4. Medidas de Centralidade.....	26
4.1 Introdução	26

4.2	As três medidas básicas: grau, proximidade e intermediação	28
4.2.1	Centralidade de grau.....	29
4.2.2	Centralidade de proximidade	30
4.2.3	Centralidade de intermediação.....	30
4.3	Outras medidas de centralidade	32
4.3.1	Medidas similares à de grau	32
4.3.2	Medidas similares à de proximidade	36
4.3.3	Medidas similares à de intermediação.....	36
4.3.4	Medidas delta	37
5.	Aplicações em redes de dados	39
5.1	Introdução	39
5.2	Detecção de comunidades.....	40
5.2.1	O algoritmo de Girván-Newman (GN)	43
5.2.2	Modularidade	44
5.2.3	Comunidades em uma rede de dados.....	45
5.3	Pontos de vulnerabilidade	50
5.3.1	Falhas e ataques	50
5.3.2	Análise da vulnerabilidade de uma rede de dados.....	51
6.	Conclusão.....	59
	Bibliografia	61
	Índice Remissivo	65

Lista de Figuras

Figura 1: Um exemplo de um grafo.....	1
Figura 2: Centralidades calculadas no grafo da Figura 1: grau, proximidade, intermediação e autovetor.....	3
Figura 3: Blogosfera da eleição presidencial americana de 2004.....	12
Figura 4: Grafo de coautoria de papers tendo Paul Erdős como centro e contando com nomes como Hilbert, Gauss e Einstein.....	13
Figura 5: Versão atualizada e ampliada do grafo contendo os autores com EN 1 e 2.	14
Figura 6: Os “seis graus de separação”.....	15
Figura 7: Exemplo de distribuição de grafos aleatórios e com conexão preferencial.	19
Figura 8: Comunidades em uma rede de roteadores.....	22
Figura 9: Os 13 motifs possíveis de 3 nós.....	23
Figura 10: Redes classificadas em “superfamílias” de acordo com os escores-Z de seus motifs.....	24
Figura 11: Grafo estrela S_6	27
Figura 12: Neste grafo, o nó em destaque tem papel central evidente, ainda que seu grau seja mínimo.	27
Figura 13: uma rede com sua estrutura de comunidades evidenciada.....	40
Figura 14: Um exemplo de árvore hierárquica.....	42
Figura 15: Uma pequena rede antes e depois do algoritmo GN executado pelo software <i>yEd</i>	45
Figura 16: Modularidade das partições.....	46
Figura 17: Rede de dados não-particionada.....	48
Figura 18: A mesma rede, após a aplicação do algoritmo GN.....	49
Figura 19: Rede com 94 nós, que serão "alvo" de simulações de ataque.....	52
Figura 20: Centralidades calculadas na rede de exemplo: grau, proximidade, intermediação e autovetor.....	53

Figura 21: Tamanho da maior componente conexa da rede em função do número de iterações do algoritmo 5.2.....	56
Figura 22: Comparação do resultado, para a centralidade de intermediação, dos algoritmos 5.2 e 5.3.....	58

Lista de Abreviaturas

k_i	grau do vértice v_i
$A(G)$	matriz de adjacência do grafo G
a_{ij}	elemento na i -ésima linha e j -ésima coluna da matriz A
$\vec{1}$	vetor com todas as entradas iguais a 1
e_i	vetor com a i -ésima entrada igual a 1 e demais entradas iguais a zero
$d(a, b)$	distância entre os vértices a e b
K_n	grafo completo com n vértices
S_n	grafo estrela com $n + 1$ vértices
$\rho(A)$	maior autovalor em módulo de A
EN	Número de Erdős
GN	algoritmo de Girvan-Newman
IP	Internet Protocol
WWW	World Wide Web
$c_X(v)$	centralidade (absoluta) do vértice v ; o ndice X  substituído por um identificador da centralidade especfica
$c'_X(v)$	centralidade (relativa) do vértice v
$\mathbf{c}_X(G)$	vetor contendo as centralidade de todos os vértices de G
g_{ab}	nmero de geodsicas entre os vértices a e b

- g_{avb} número de geodésicas entre os vértices a e b
que passam pelo vértice v
- $P(G)$ coesão do grafo G
- $\mathcal{E}(G)$ eficiência do grafo G
- $c_{BL}(l)$ centralidade de intermediação do link l
- $Q(G^*)$ modularidade da partição G^*

Resumo

A Análise de Redes trata do estudo da estrutura de uma rede a fim de obter informações importantes sobre seus elementos e suas interações. Um aspecto relevante da análise de uma rede é decidir quais são os elementos mais importantes ou centrais de uma rede, através do uso das medidas de centralidade. Neste trabalho, apresentamos um panorama sobre a área da Análise de Redes, e faremos um *survey* sobre as principais medidas de centralidade, mostrando suas motivações e definições. Em seguida, apresentaremos duas aplicações das centralidades às redes de dados: a obtenção da estrutura de comunidades de uma rede de roteadores e a avaliação dos pontos de vulnerabilidade de uma rede.

Abstract

Network Analysis is the field that studies the structure of a network in order to retrieve important information about its elements and interactions between them. Deciding which are the most important or central elements of a network is a relevant aspect in the analysis of a network; this can be achieved through the use of centrality measures. In this work, we present an overview of the area of Network Analysis, making a survey of the main centrality measures, along with their motivations and definitions. We also present two applications of centrality measures to data networks: retrieving the community structure of a network of routers, and assessing the vulnerability spots of a network.

1. Introdução

Um grafo é um par ordenado de conjuntos $G = (V, E)$ onde $E = \{ij = \{i, j\}\}$, sendo que i e j devem pertencer a V . Os elementos $v \in V$ são chamados de *vértice* ou *nós* de G , e os elementos $e \in E$ são chamados de *arestas* ou *links*. Grafos são representados graficamente por pontos ligados por linhas, como mostra a Figura 1.

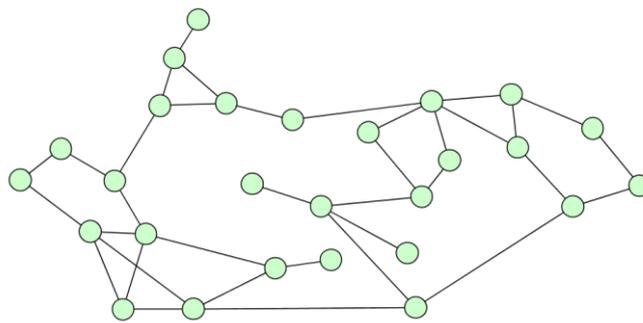


Figura 1: Um exemplo de um grafo.

Outras formas de representação de grafos incluem listas e matrizes. Um exemplo importante de matriz associada a um grafo é a sua matriz de adjacência $A_{n \times n}$, definida por $a_{ij} = 1$ se existe aresta entre os nós i e j , e por zero nas demais posições.

Diversos sistemas no mundo real podem ser representados através de redes. Dessa forma, grafos são uma forma natural de representar matematicamente esses sistemas. A *Análise de Redes* é a área do conhecimento que investiga a estrutura de uma rede a fim de obter informações importantes sobre seus elementos e suas interações.

Nas últimas décadas, o interesse nessa área tem crescido. Surgiram aplicações em áreas tão diversas quanto a biologia e a linguística [1], porém as redes tecnológicas têm tido um lugar de destaque, especialmente com o advento e a popularização da Internet. As *redes de dados*, cujos elementos mais importantes

são os roteadores que direcionam o tráfego de dados na Internet e em redes domésticas e corporativas, são de especial interesse.

Outro evento que vem motivando a pesquisa na área é o crescimento das redes sociais online, que geram grandes conjunto de dados. Também surgiram diversos softwares específicos, incorporando diversos algoritmos e permitindo uma visualização gráfica de alta qualidade. A análise de uma rede utiliza várias ferramentas, algumas das quais introduziremos neste trabalho.

Um aspecto relevante dessa análise é decidir quais são os elementos mais importantes ou centrais de uma rede. As *medidas de centralidade* são uma forma de quantificar essa importância. Uma medida de centralidade de nó é uma função $c_X: V \rightarrow \mathbb{R}$ construída de modo que a relação de ordem entre $c_X(v_i)$ e $c_X(v_j)$ reflita a percepção de que v_i é *mais central* do que v_j , em algum sentido. Medidas de centralidade de aresta $c_Y: E \rightarrow \mathbb{R}$ são constituídas de maneira análoga. Em diversas aplicações, a centralidade de um elemento é associada à importância desse elemento na estrutura da rede.

Desde a década de 1950, várias centralidades surgiram na literatura; Freeman [2] fez um dos primeiros esforços para organizar essas medidas. De acordo com Freeman, as medidas de centralidade de nó básicas são:

- *centralidade de grau*, dada por $c_D(v_i) = k_i$, onde k_i é o grau no nó v_i .
- *centralidade de proximidade*, definida como o recíproco da soma das distâncias de v aos demais nós, ou seja, $c_C(v) = (\sum_{a \in V} d(v, a))^{-1}$ onde $d(v, a)$ é a distância (tamanho do caminho de comprimento mínimo) entre os nós v e a .
- *centralidade de intermediação*, dada pela proporção das geodésicas (caminhos de tamanho mínimo) entre dois nós quaisquer que passam pelo nó em questão, ou seja, $c_B(v) = \sum_{a, b \neq v} \frac{g_{avb}}{g_{ab}}$, onde g_{ab} é número de geodésicas entre a e b , e g_{avb} é o número dessas que passam por v .

Existem também medidas espectrais; sendo λ_1 o maior autovalor em módulo da matriz de adjacência e \mathbf{v}_1 o autovetor associado, a *centralidade de autovetor* do vértice v_i é definida como a i -ésima posição de \mathbf{v}_1 .

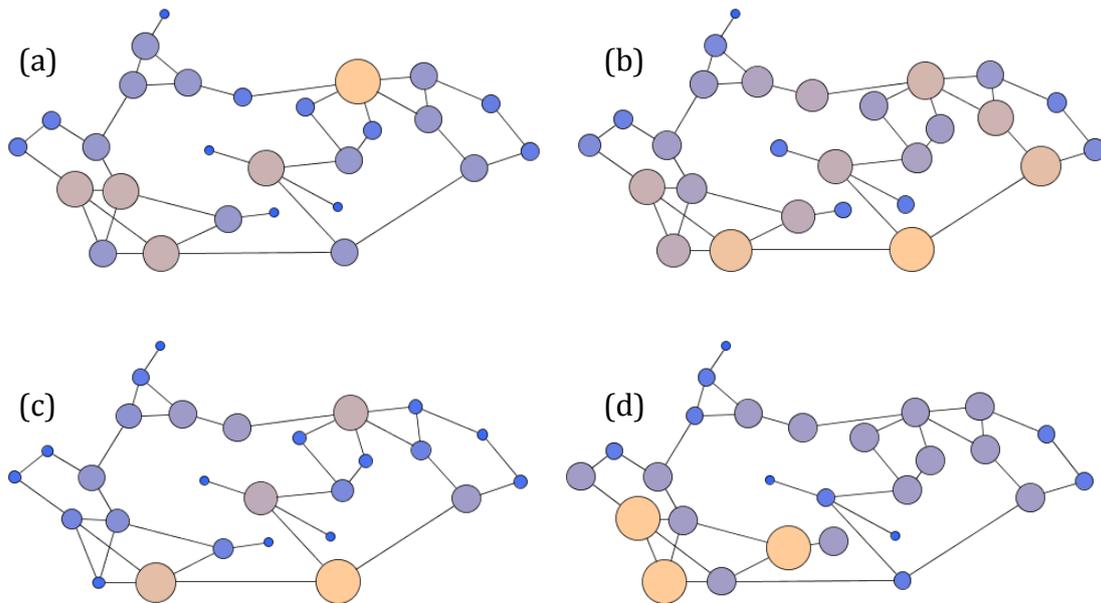


Figura 2.; Centralidades calculadas no grafo da Figura 1: grau (a), proximidade (b), intermediação (c) e autovetor (d). Nós maiores e mais claros são mais centrais.

Várias aplicações têm surgido para as centralidades. Mecanismos de buscas como o Google organizam a relevância de seus resultados baseados em uma variante da centralidade de autovetor [3]. Girván e Newman utilizaram uma versão da centralidade de intermediação adaptada aos links da rede a fim de detectar a estrutura de comunidades de uma rede [4].

Entre os problemas que podem causar impacto em uma rede, estão *falhas* nos links e *ataques* aos nós. Exemplos de falhas são rompimentos de cabos e perda de sincronismo de transmissão. Já ataques são fruto de ação intencional e frequentemente maliciosa. Os chamados *cyberataques*, que exploram os protocolos de comunicação de dados a fim de indisponibilizar um servidor ou roteador (ou seja, um nó da rede), são algo comum hoje em dia.

Silva [5] obteve uma relação entre as medidas de centralidade e confiabilidade (resistência a falhas). No caso da ampliação de uma rede, buscava-se a inserção de um link de modo a aumentar a confiabilidade da rede. Os melhores resultados foram obtidos quando inserido um link entre o nó menos central e o mais distante dele. Quanto à questão da vulnerabilidade (susceptibilidade a ataques), o presente trabalho apresenta uma aplicação das medidas de centralidade que trata da questão da identificação dos pontos mais vulneráveis de uma rede. Para tal, tomamos o lugar do atacante, a fim de remover da rede os nós de modo a causar maior fragmentação, ou seja, de modo a “quebrar” a rede em diversas componentes conexas. Quanto menor o tamanho da maior componente conexa, mais bem sucedido é o ataque.

O objetivo deste trabalho é apresentar algumas aplicações das centralidades no contexto das redes de dados. As redes usadas nos exemplos são adaptadas de redes de roteadores de uma empresa de telecomunicações. Usaremos primariamente os softwares *yEd* e *Gephi* para analisar os grafos gerados. Esses programas estão munidos de algoritmos para o cálculo das centralidades e outros parâmetros, e dispõem de ferramentas para a manipulação dos grafos associados, como ordenamento de nós por um dado parâmetro e remoção de nós, que serão utilizadas nessas aplicações.

O presente trabalho divide-se em quatro capítulos. No capítulo 2, apresentamos os conceitos fundamentais de Teoria dos Grafos, para fins de referência; também citamos alguns resultados de Álgebra Linear que serão utilizados no restante do trabalho.

O capítulo 3 oferece um amplo panorama da área da Análise de Redes, incluindo um histórico desde seus primórdios, até a intensificação na pesquisa, especialmente nas últimas décadas.

A seguir, no capítulo 4, é aprofundado o estudo das medidas de centralidade, apresentando as principais medidas encontradas na literatura e sua classificação.

Finalmente, no capítulo 5, apresentamos duas aplicações das medidas de centralidade no âmbito específico das redes de dados. Inicialmente, utilizamos o algoritmo de Girvan-Newman para obter a estrutura de comunidades de uma rede, a fim de sugerir uma distribuio de grupos de roteadores entre equipes que as administraro. A seguir, faremos uma simulao de ataque que tem o objetivo de identificar pontos de vulnerabilidade de uma rede especfica de roteadores.

2. Conceitos Básicos

Neste capítulo, apresentaremos conceitos básicos e terminologias de Teoria dos Grafos, além de alguns resultados de Álgebra Linear que serão utilizados nos capítulos posteriores.

2.1 Introdução

Informalmente, um grafo é algo que pode ser representado por elementos ligados por linhas; esta também é a representação mais intuitiva. Um resumo dos conceitos básicos é apresentado na seção 2.2. Mais detalhes podem ser encontrados no livro de Diestel [6].

A seção 2.3 contém algumas definições e resultados de Álgebra Linear e da Teoria Espectral dos Grafos, que serão usados em capítulos posteriores. Os detalhes encontram-se no livro de Datta [7].

2.2 Teoria dos Grafos

Um grafo é um par ordenado de conjuntos $G = (V, E)$ onde $E = \{ij = \{i, j\}\}$, sendo que i e j devem pertencer a V . Os elementos $v \in V$ são chamados de *vértices* de G , e os elementos $e \in E$ são chamados de *arestas*. Dois vértices são ditos adjacentes ou vizinhos se existe uma aresta ligando-os. Um *subgrafo* de G é um grafo $G' = (V', E')$ tal que $V' \subset V$. Um *laço* é uma aresta do tipo ii , ou seja, que liga um vértice a si próprio.

Ao longo do texto, alguns desses termos podem ser substituídos por sinônimos adequados ao contexto. É comum usar-se o termo *rede* para referir-se

ao grafo que representa essa mesma rede. Em redes de comunicação, o termo *link* é mais adequado do que *aresta*; o mesmo vale para *amizade* na análise de redes sociais, assim como o termo *amigos* é um substituto óbvio para *vizinhos* nesse mesmo contexto. Alguns sinônimos usuais de *vértices* são: *nós*, *pontos*, *atores*.

Um grafo G é dito *não-orientado* se as relações entre os nós são simétricas, ou seja, se os pares ab não forem ordenados. Por exemplo, uma relação de amizade costuma ser recíproca: tanto faz dizer que a é amigo de b ou b é amigo de a . Se os pares forem ordenados, G é dito *orientado*. Um exemplo são as ruas de uma cidade, que podem ter apenas um sentido. G é dito *simples* se não contém loops e não há mais de uma aresta entre dois vértices quaisquer. Caso contrário, G é chamado de *multigrafo*.

Uma representação comum de um grafo é a sua matriz de adjacência $A_{n \times n}$ onde $a_{ij} = 1$ se existe a aresta ij , e zero nas demais posições. Para grafos não-orientados, a matriz é simétrica.

O número k_i de vizinhos de um nó v_i é seu *grau*. A partir de A , para um grafo não-orientado, temos que, $k_i = \sum_{j=1}^n a_{ij} = e_i^T A \vec{1}$. Para grafos orientados, definimos o *grau de saída* como $k_i^{out} = \sum_j a_{ij} = e_i^T A \vec{1}$ e o *grau de entrada* como $k_i^{in} = \sum_j a_{ji} = e_i^T A^T \vec{1}$.

Dois vértices são ditos *adjacentes* se existe uma aresta os conectando (ou incidindo neles). Duas arestas são ditas adjacentes se incidem em um mesmo vértice.

Às arestas de um grafo pode-se associar uma função peso $w: E \rightarrow \mathbb{R}$. Normalmente, $w(e)$ recebe um valor relevante à aplicação. Um exemplo típico: se V representa cidades e E a malha rodoviária, w pode ser o comprimento do segmento de rodovia representado.

Um *passeio* entre dois vértices a e b é uma sequência de arestas adjacentes que determina uma sequência de vértices adjacentes partindo de a e chegando em b . Para grafos orientados, cada vértice i do passeio deve estar ligado ao próximo

vértice j na sequência por uma aresta orientada no sentido ij . Um *caminho* é um passeio com vértices distintos, e um *ciclo* é um caminho que começa e termina no mesmo nó. O *comprimento* de um passeio é dado pelo número de arestas deste ou, em grafos com peso, pela soma dos pesos. Uma *geodésica* é um caminho de tamanho mínimo entre dois nós. O comprimento de qualquer geodésica entre a e b é a *distância* entre a e b , que é denotada por $d(a, b)$. Se não há caminho entre a e b , convencionou-se que $d(a, b) = \infty$.

O *diâmetro* de G é a distância máxima entre dois nós, ou seja, $\text{diam}(G) = \max_{a,b \in V} d(a, b)$. A *excentricidade* de um nó traduz a ideia de o quanto um nó está distante dos demais. Define-se como $e(v_i) = \max_{v \in V} d(v_i, v)$. O *raio* do grafo é definido por $r(G) = \min_{v \in V} e(v)$, e o *centro* do grafo é $c(G) = \{v \in V : e(v) = r(G)\}$.

Um grafo de n nós é dito *completo* se todo par de vértices está conectado por uma aresta. Denotamos um grafo assim por K_n . Uma *clique* é um subgrafo completo.

Um grafo G é dito *conexo* se sempre existe um caminho entre dois vértices quaisquer; caso contrário, G é dito *desconexo*. Se G é orientado com caminhos de “ida” e “volta” entre dois vértices quaisquer, G é dito *fortemente conexo*. Cada subgrafo maximal conexo de G é uma *componente* de G . Um conjunto de nós V' (ou arestas E') é dito um *conjunto de corte* de G se $G - V'$ (ou $G - E'$) é desconexo. Um conjunto de corte é chamado de *ponto de articulação* de G se contiver apenas um nó, ou de *ponte* se contiver apenas uma aresta. Um *bloco* é um subgrafo conexo sem ponto de articulação.

Grafos também são usualmente representados e armazenados na forma de listas, sendo que as mais utilizadas são:

- *lista de adjacência*: $L_I = \{\{v_i, v_j\} : v_i, v_j \in V\}$; os pares são considerados ordenados se o grafo for orientado, com o primeiro e segundo elementos sendo respectivamente a origem e o destino da seta.

- lista de incidência: $L_A = \{e_i, v_j\}$ tal que e_i incide em v_j .

Além da matriz de adjacência, diversas matrizes podem ser associadas a um grafo; consideremos que $|V| = n$ e $|E| = m$.

- matriz de incidência: $B_{n \times m}$ onde $b_{ij} = 1$ se e_j incide em v_i , e zero nas demais posições.

- matriz geodésica: $S_{n \times n}$ onde $s_{ik} = d(v_i, v_k)$.

- matriz diagonal: $D_{n \times n}$ onde $d_{ii} = \deg(v_i)$, e zero nas demais posições.

- matriz Laplaciana: $L = D - A$.

2.3 Álgebra Linear e Teoria Espectral dos Grafos

A *Teoria Espectral dos Grafos* trata da análise dos grafos a partir de suas matrizes associadas, usando diversas ferramentas da Álgebra Linear, com ênfase especial nos autovalores e autovetores dessas matrizes.

Uma matriz $A = [a_{ij}]$ é dita *não-negativa* se $a_{ij} \geq 0$ para qualquer i, j . Se A é quadrada de ordem n , é dita *irredutível* se não existe nenhuma decomposição da forma $N_1 \cup N_2 = \{1, 2, \dots, n\}$, $N_1 \neq \emptyset \neq N_2$, $N_1 \cap N_2 = \emptyset$ tal que $a_{ij} = 0$ para $i \in N_1$ e $j \in N_2$.

Teorema 2.1 (Perron-Frobenius): Seja $A_{n \times n}$ é não-negativa e irredutível, e $\rho(A)$ o seu maior autovalor em módulo. Então,

- a) $\rho(A) > 0$
- b) Existe um autovetor positivo \mathbf{v}_1 associado a $\rho(A)$.
- c) $\rho(A)$ é um autovalor algebricamente simples.

A seguir, alguns resultados e algumas definições de Álgebra Linear que serão usados mais tarde no texto.

A p -norma de um vetor $x \in \mathbb{R}^n$ é definida por

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad p \geq 1.$$

Para uma matriz A , uma *norma matricial subordinada* a uma dada p -norma é definida por

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

Seja $\|\cdot\|$ uma norma matricial subordinada qualquer. Se para uma sequência de matrizes $\{A_k\}_{k=1}^{\infty}$ existe uma matriz A tal que $\|A_n - A\| \rightarrow 0$ quando $n \rightarrow \infty$, dizemos que a sequência *converge* para a matriz A , que é então dita *convergente*.

Teorema 2.2: Seja a sequência $\{A^n\}_{n=1}^{\infty}$ de potências da matriz A . Então, $A^n \rightarrow 0$ se e somente se $|\lambda_i| < 1$ para qualquer autovalor λ_i de A .

Teorema 2.3: Para qualquer autovalor λ_i de A ,

$$|\lambda_i| \leq \|A\|, \tag{2.1}$$

onde $\|\cdot\|$ é qualquer norma matricial subordinada. Em particular, se $\rho(A) = \max |\lambda_i|$,

$$\rho(A) \leq \|A\|. \tag{2.2}$$

Corolário 2.4: Uma matriz A é convergente se $\|A\| < 1$, onde $\|\cdot\|$ é uma norma matricial subordinada.

Teorema 2.5: Se A é convergente, então

$$I + A + A^2 + \dots = \sum_{k=0}^{\infty} A^k \rightarrow (I - A)^{-1}. \tag{2.3}$$

3. Análise de Redes

Nesse capítulo, exibiremos um panorama da área de Análise de Redes. Partimos de uma revisão histórica; passamos pelos principais modelos matemáticos de redes, e por fim serão apresentadas algumas das principais ferramentas usadas nessas análises. Essa apresentação das ferramentas tem como objetivo dar uma visão ampla (embora não profunda) da área.

3.1 Introdução

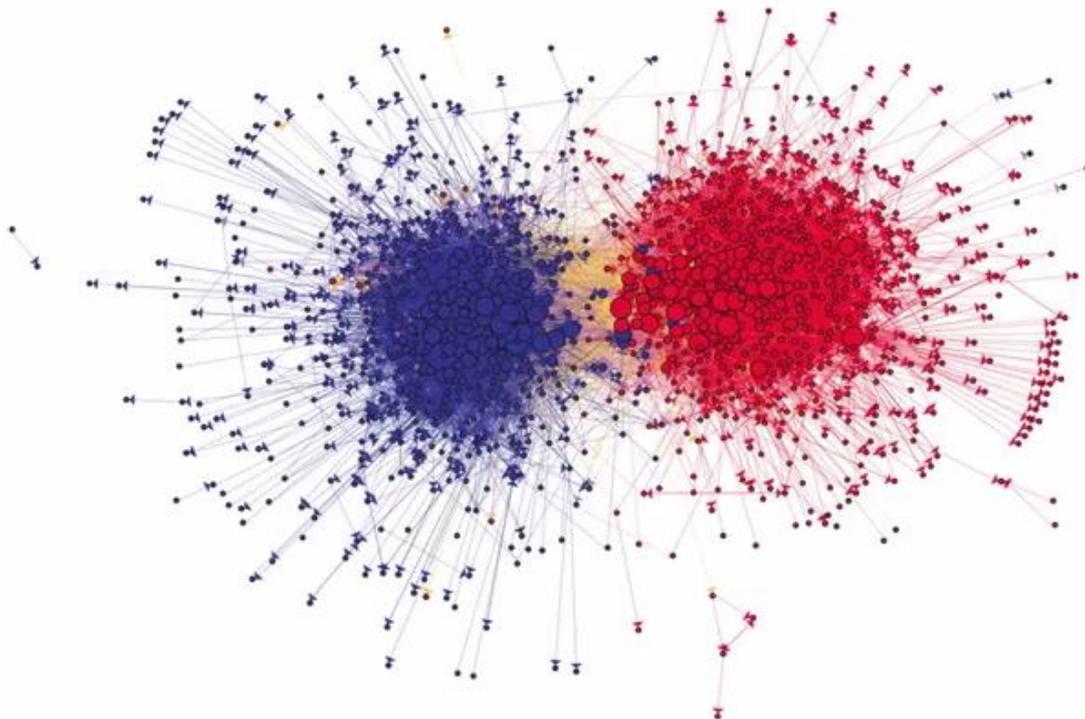
Diversos sistemas no mundo real podem ser representados através de redes e, portanto, modeladas como grafos. Analisando a estrutura do grafo que representa a rede, podemos obter informações importantes sobre seus elementos e suas interações. Esse é o papel da *Análise de Redes*, também chamada de *Teoria de Redes* ou *Ciência das Redes*.

Hoje em dia, programas como NetLogo, Gephi, Pajek e yEd, entre outros, permitem fazer diversos tipos de análise automaticamente, e estão munidos de algoritmos de layout que permitem uma boa visualização da estrutura da rede, como no exemplo da Figura 3.

Iniciamos, na seção 3.2, com um breve histórico da Análise de Redes, mostrando como a importância dessa área vem crescendo, especialmente nas últimas décadas.

A seguir, na seção 3.3, veremos dois dos principais modelos matemáticos de redes: grafos aleatórios e redes com conexão preferencial, bem como suas motivações e principais propriedades.

Finalmente, a seção 3.4 apresenta algumas das principais ferramentas usadas nessas análises: parâmetros calculados a partir da rede ou de seus elementos, e padrões a serem reconhecidos.



**Figura 3: Blogosfera da eleição presidencial americana de 2004 [8].
Neste grafo, cada ponto é um blog democrata (azul) ou republicano (vermelho).**

3.2 Histórico

A resolução do *Problema das Sete Pontes de Königsberg* por Euler , em 1736, é tido como o marco inicial da Teoria dos Grafos, e pode ser também considerada um exemplo de Análise de Redes. Entretanto, é apenas no século 20 que o estudo se intensifica.

Em 1926, Boruvka publica um algoritmo para encontrar a árvore geradora mínima em um grafo, almejando a construção de uma rede elétrica eficiente na região da Morávia (República Checa) [9]. Na década de 30, Moreno [10] usou grafos (“sociogramas”) para representar relações sociais; essa foi uma das primeiras tentativas de análise sistemática de redes sociais.

Em 1959, Erdős e Rényi [11,12] criaram modelos para criação de grafos aleatórios, hoje conhecidos como *grafos ER*, apresentando várias propriedades dos

grafos gerados por esses modelos. Porém, algumas dessas propriedades tornam os grafos ER pouco apropriados para o estudo de redes reais, como veremos mais adiante. Porém, os trabalhos de Erdős e Rényi estabeleceram um marco na área, permanecendo como ponto de partida para novos modelos.

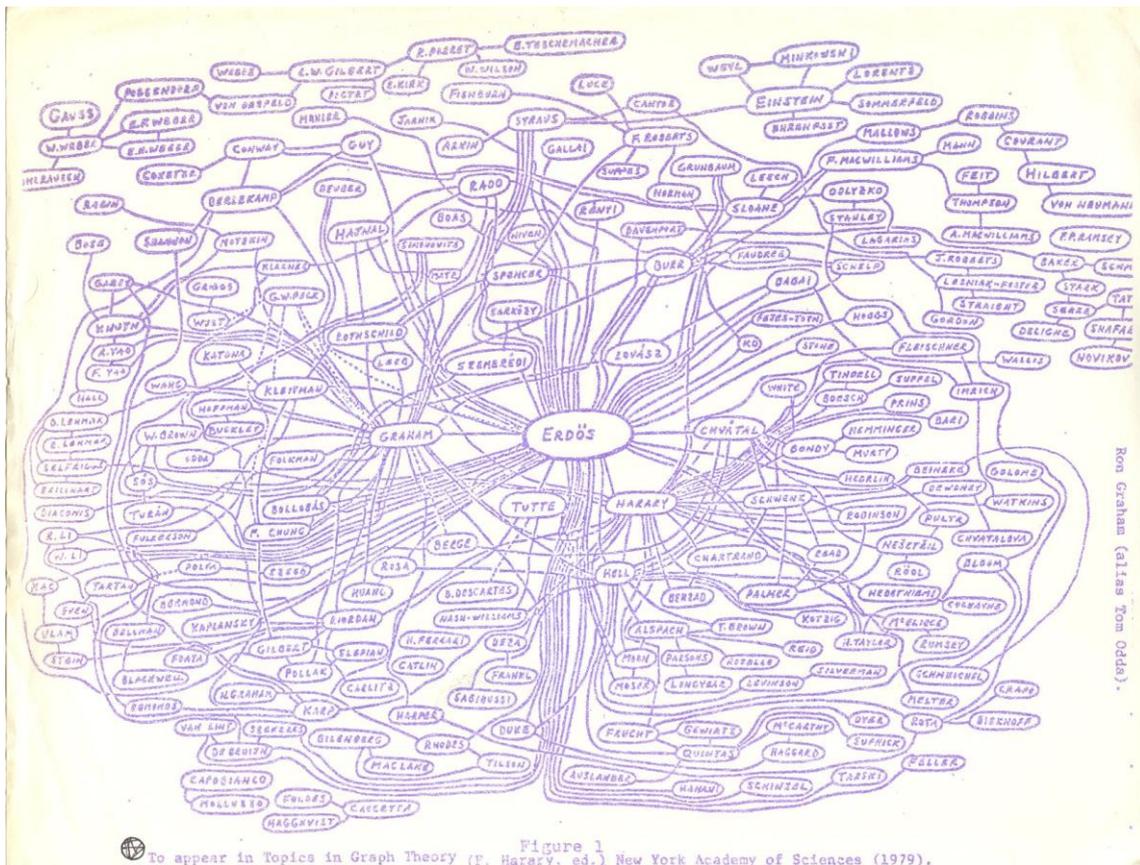


Figura 4: Grafo de coautoria de artigos tendo Paul Erdős como centro [13] e contando com nomes como Hilbert, Gauss e Einstein. O desenho foi feito à mão em 1979.

A figura de Erdős como presumivelmente o matemático mais prolífico da história inspirou um interessante exercício de Análise de Redes: a determinação do “número de Erdős” (que abreviaremos por EN, *Erdős Number*) de um matemático. O EN do próprio Erdős é 0; aqueles que colaboraram em papers com ele recebem o EN 1; aos colaboradores dos colaboradores (que não tenham colaborado com o próprio Erdős) é atribuído um EN 2, e assim sucessivamente. Ou seja, se cada matemático for um nó em um grafo e as arestas representarem coautoria em

papers, o EN de cada nó é a distância deste ao nó que representa o próprio Erdős. A Figura 4 mostra um esboço manual desse grafo, e a Figura 5 é uma versão atualizada gerada por software. É interessante notar que normalmente o EN é surpreendentemente baixo. Num levantamento [13] contendo quase 270 mil matemáticos, a média foi 4,65 e o desvio padrão 1,21, sendo que o EN máximo foi 13 e em apenas 2% dos casos o EN ultrapassou 8.

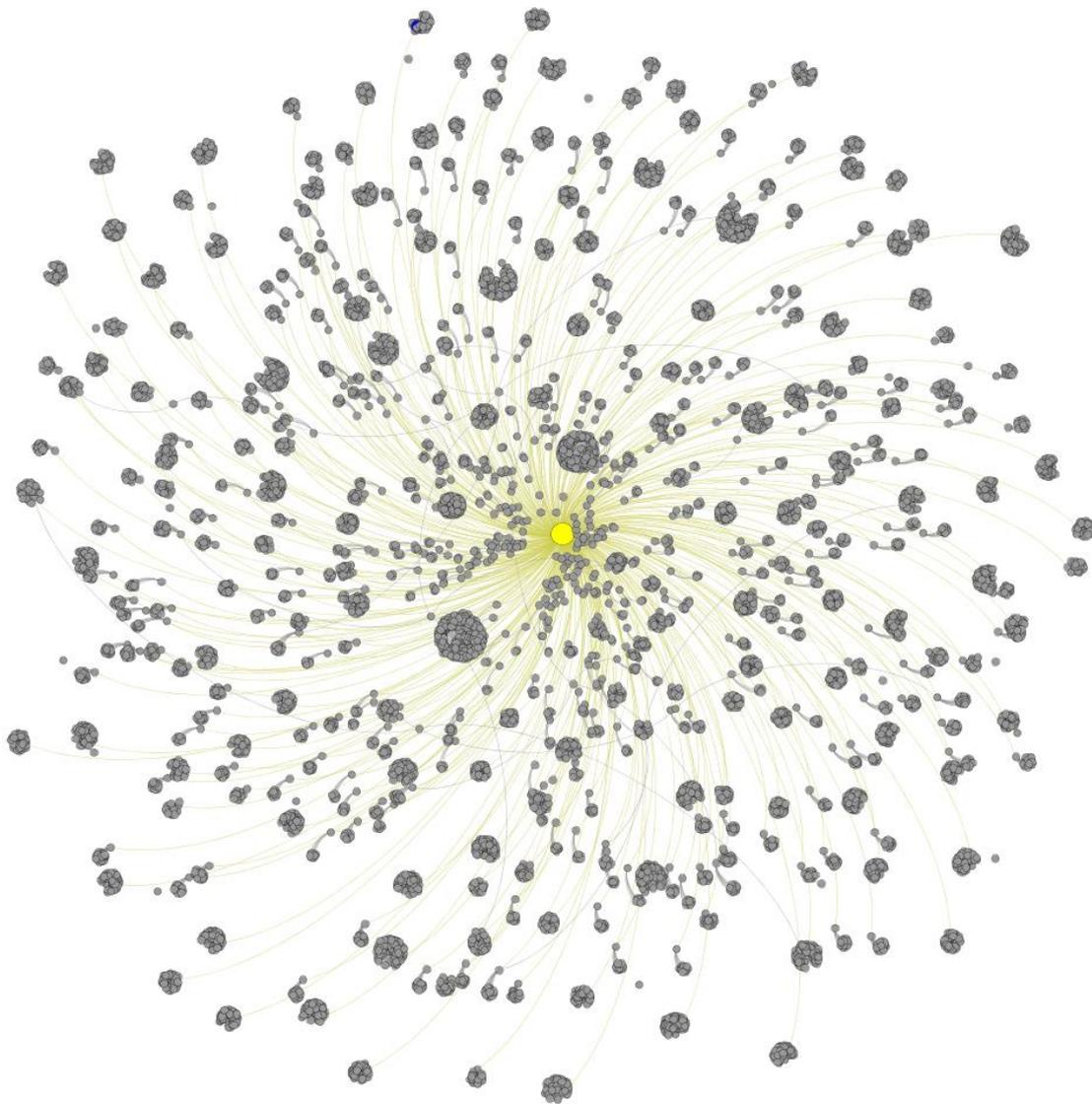


Figura 5: Versão atualizada e ampliada do grafo contendo os autores com EN 1 e 2. O grafo foi gerado no software *Gephi* pelo autor deste trabalho a partir de dados encontrados em [13]. Erdős é o ponto amarelo no centro, e o ponto azul na parte superior corresponde ao orientador deste trabalho.

Em 1967, o psicólogo americano Stanley Milgram conduziu experimentos projetados para demonstrar o “Fenômeno do Mundo Pequeno” [14]. Várias cartas endereçadas a pessoas em Boston foram dadas a pessoas em lugares

comparativamente remotos, como Nebraska e Iowa. Se o portador não conhecesse o destinatário, a carta deveria ser entregue a algum conhecido do portador que se julgasse que tivesse maior probabilidade de conhecer o destinatário. Entre as cartas que chegaram, a média de passos (arestas) entre o portador inicial e o destinatário ficou por volta de 6; daí se popularizou a noção dos “Seis Graus de Separação”, em que duas pessoas quaisquer estão ligadas por uma cadeia de relativamente poucos conhecidos.

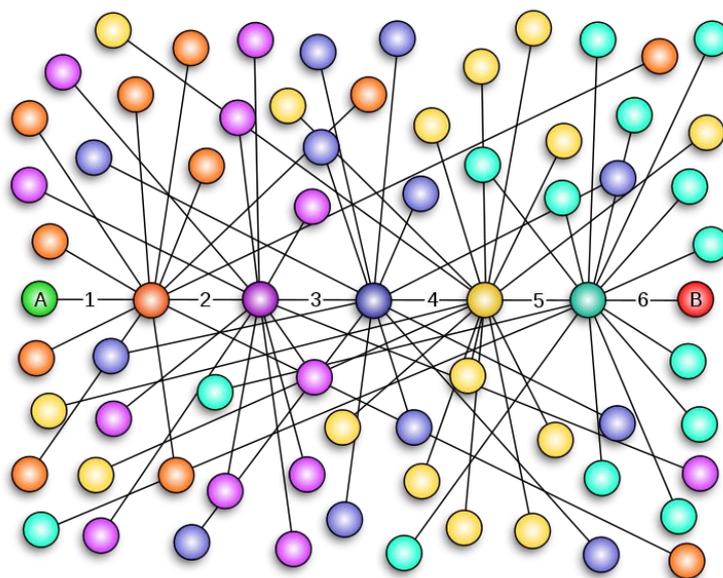


Figura 6: Os “seis graus de separação”.

Na década de 70, a ideia de centralidade começa a ganhar mais foco. Freeman [2] foi o primeiro a tentar organizar as várias medidas já existentes na literatura.

Em 1998, foi publicado na revista *Nature* o artigo *Collective dynamics of 'small-world' networks*, de Watts e Strogatz [15], retomando a ideia do “mundo pequeno” de Milgram. Os autores propõem algumas modificações nos modelos ER, refletindo melhor o fenômeno do mundo pequeno.

Nesse mesmo ano, surge o Google, sistema de busca na internet cujo cerne é um algoritmo que usa o algoritmo *PageRank* para identificar as páginas mais

relevantes para os termos buscados. Veremos mais sobre a motivação do *PageRank* na seção 4.3.1.

Na década de 2000, temos relatos de usos da análise de redes sociais aplicadas a redes terroristas. Em 2002, Krebs [16] publica um artigo analisando as conexões entre os envolvidos nos ataques às torres do World Trade Center, em 11 de setembro de 2001. Em 2003, Hougham [17] relata que análises similares levaram à captura do ditador iraquiano Saddam Hussein. Como reflexo desses e outros episódios, em 2005 um comitê do governo americano considerou a Ciência de Redes como uma área estratégica para a segurança nacional [18].

O surgimento das redes sociais online, em especial o Orkut e o Facebook, ambos surgidos em 2004, gerou grandes conjuntos de dados para a análise de redes, e essas redes e subconjuntos delas são objeto de intensa pesquisa até hoje. Assim, os grafos analisados hoje em dia têm tamanho massivo, muitas vezes da ordem de milhões de nós ou mais.

3.3 Modelos de rede

A existência de semelhanças entre diversas redes levou naturalmente à criação de modelos matemáticos descrevendo a criação e evolução de uma rede. Podemos, então, comparar redes reais com redes criadas por esses modelos; as previsões de um modelo podem então ser confirmadas, ou então o modelo pode ser corrigido e aprimorado. Por exemplo, citamos anteriormente que o modelo de grafos aleatórios prevê o fenômeno do “mundo pequeno”, porém não é adequado para modelar redes com *hubs*, o que motivou a criação de novos modelos.

Veremos nessa seção dois dos principais tipos de modelos: *grafos aleatórios* e *redes com conexão preferencial*, apresentando suas motivações e a concordância destes com redes reais. Especialmente os grafos aleatórios, ainda que mostrem pouca concordância com redes reais, servem como pano de fundo para

comparação de resultados, como veremos, por exemplo, no caso dos *motifs* da seção 3.2.4.

3.3.1 Grafos Aleatórios

Uma ideia inicial é considerar simplesmente que as conexões de uma rede formam-se de maneira aleatória. Foram dois os modelos desse tipo criados por Erdős e Rényi; os grafos gerados são então conhecidos como *grafos ER*. Consideremos o conjunto de todos os $\binom{N}{K}$ possíveis grafos com N vértices e K arestas, assumindo que cada um destes grafos tem a mesma probabilidade de ocorrer. Um grafo escolhido aleatoriamente desse conjunto é chamado de *grafo aleatório*. Nesse contexto, diz-se que se uma propriedade A , com probabilidade $P(A)$, é tal que $\lim_{n \rightarrow \infty} P(A) = 1$, dizemos que A vale para *quase todo grafo*.

Para gerar grafos conforme a função de probabilidade do primeiro modelo [11], inicia-se com N nós desconectados e selecionam-se pares de nós aleatoriamente, gerando um link entre eles¹, até que haja K links. Denotamos por $G_{N,K}^{ER}$ um grafo ER com N nós e K links.

Um resultado obtido a partir desse modelo é que, para um grafo de grau médio $\langle k \rangle$, para N suficientemente grande, a geodésica média L tende a $\log_{\langle k \rangle} N$. Isso significa que o fenômeno do mundo pequeno encontra justificativa nesse modelo. Por exemplo, estimando a população mundial em $N = 7 \cdot 10^9$ e assumindo uma média de conexões por pessoa (família, amigos, conhecidos), de $\langle k \rangle = 40$, obtemos $L \cong 6$.

No segundo modelo [12], também se inicia com N nós desconectados, e cada um dos $\binom{N}{2}$ links tem uma probabilidade p de existir. Denotamos esse grafo por $G_{N,p}^{ER}$. A probabilidade de $G_{N,p}^{ER}$ ter K links é $p^K(1-p)^{\binom{N}{2}-K}$. Existe uma probabilidade crítica $p_c = \frac{1}{N}$, correspondente a um grau médio crítico $\langle k \rangle_c = 1$. Conforme os casos para p , tem-se que [19]:

¹ Conexões múltiplas são proibidas.

- $p < p_c \Rightarrow$ toda componente de $G_{n,p}^{\text{ER}}$ quase certamente tem tamanho $\leq O(\ln N)$ e possui no máximo um ciclo.
- $p = p_c \Rightarrow$ quase certamente a maior componente tem tamanho $O(N^{2/3})$.
- $p > p_c \Rightarrow$ há uma componente de tamanho $O(N)$ com $O(N)$ ciclos, e as demais têm tamanho $\leq O(\ln N)$ e possuem no máximo um ciclo.

Seguem algumas propriedades importantes de grafos ER [20]. Para N grande e $\langle k \rangle$ fixo, a distribuição de graus $P(k)$ é aproximadamente Poisson. Se $p \geq \ln N / N$, então $G_{N,p}^{\text{ER}}$ é quase certamente conexo. $\text{Diam}(G)$ e L variam pouco ao redor de $\ln N / \ln \langle k \rangle$. O coeficiente de aglomeração é dado por $C = p = \frac{\langle k \rangle}{N}$, logo é evanescente.

Existem também na literatura extensões do modelo ER partindo de distribuições $P(k)$ arbitrárias. Um exemplo é o *Modelo da Configuração* [21], resumido a seguir. Uma configuração desejada da distribuição é dada por uma sequência $D = \{k_1, \dots, k_N\}$ tal que $\sum k_i = 2K$. Um grafo de N vértices gerado por esse modelo a partir da sequência D é denotado por $G_{N,D}^{\text{conf}}$. Para cada nó i são atribuídos k_i “meios-links”, e a seguir pareiam-se todos os “meios-links” aleatoriamente. O modelo pode ser ajustado para permitir ou não multigrafos e loops.

3.3.2 Redes com conexão preferencial

Vimos que a distribuição de graus $P(k)$ para grafos aleatórios é aproximadamente Poisson. Ao estudar redes reais, entretanto, descobriu-se que a maioria delas exibe uma distribuição de graus dos nós seguindo uma lei de potência $P(k) = \alpha k^{-\gamma}$, com $2 < \gamma < 3$ (a Figura 7 compara as distribuições de $P(k)$ observadas nos dois modelos). Isso motivou a criação dos *modelos de redes com conexão preferencial* (em inglês, *preferential attachment*). A ideia é que, quando um nó entra na rede, ele não se conecta aleatoriamente, mas “prefere” conectar-se com nós de alto grau (ou seja, a probabilidade de um novo nó u se

conectar a um nó existente v é proporcional ao grau de v). Por exemplo, numa rede de citações de papers, é mais provável que um novo paper cite papers que já foram citados várias vezes, em vez de outros menos citados.

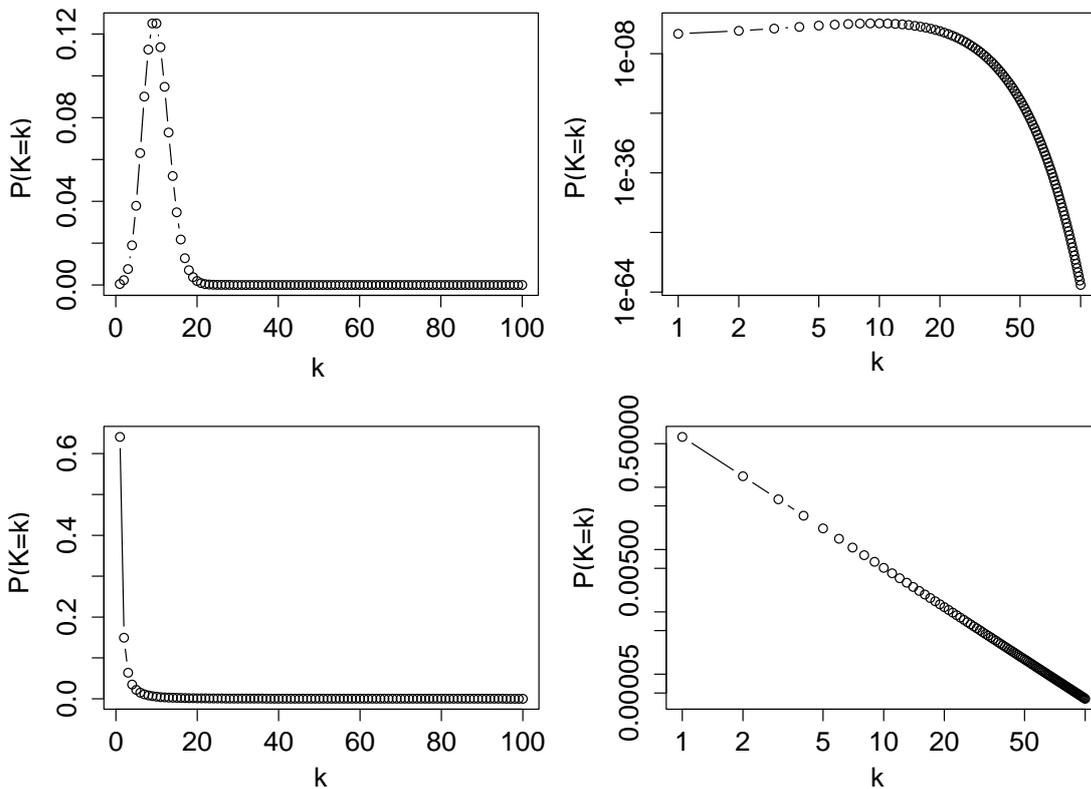


Figura 7: Exemplo de distribuição de grafos aleatórios (acima) e com conexão preferencial (abaixo). Os gráficos à direita estão em escala logarítmica.

Um dos modelos dessa classe é o modelo de Barabási e Albert [22], que emula a evolução da WWW. A partir de um subgrafo inicial, cada novo nó i (uma nova página da web) entra na rede com m links, sendo que a probabilidade de i se conectar a j é proporcional a k_j , gerando uma distribuição $P(k) = 2m^2k^{-3}$.

3.4 Ferramentas de Análise de redes

Ao longo das últimas décadas, diversas ferramentas têm sido estudadas a fim de melhor caracterizar a estrutura de redes e seus componentes. Estas incluem o cálculo de parâmetros e a detecção de padrões, tanto globais (considerando a rede como um todo) quanto locais (considerando apenas um nó ou link, ou um

pequeno grupo), bem como a análise de processos em uma rede em evolução. A seguir apresentaremos uma lista com alguns exemplos dessas ferramentas.

- *Parâmetros globais*: distribuição de graus, comprimento de caminho, centralidade.

- *Parâmetros locais*: mediação, grau.

- *Padrões locais*: aglomeração, transitividade, *motifs*.

- *Processos*: difusão, coordenação, falhas.

Nessa seção introduziremos brevemente algumas dessas ferramentas.

3.4.1 Distribuição de grau

Um parâmetro global básico é a distribuição do grau k dos nós de uma rede; denotaremos por $P(k)$ essa distribuição, e por $\langle k \rangle$ o grau médio. Num grafo aleatório, essa distribuição tende a ser normal ao redor de $\langle k \rangle \approx \log n$. Em muitas redes reais, entretanto, a maioria dos nós tem grau baixo e alguns poucos nós têm grau alto, tendo um papel de “concentradores” (*hubs*), caracterizando uma distribuição seguindo uma lei de potência $P(k) \sim k^{-\gamma}$. Isso motivou o desenvolvimento de outros modelos, como o de conexão preferencial, que vimos anteriormente, entre outros.

3.4.2 Diâmetro e distância

Métricas de diâmetro e distância média também são importantes para tipificar uma rede. Experimentos como o de Milgram, citado na seção 3.2, mostraram que várias redes grandes podem ter distâncias médias surpreendentemente curtas (fato que, como vimos, é confirmado por alguns modelos).

O problema de encontrar geodésicas em redes tem importantes aplicações práticas; dois exemplos importantes são os algoritmos de roteamento em redes IP, como o OSPF (*Open shortest path first*) que é baseado no algoritmo de Dijkstra, que fornece, a partir de um ponto de origem, o caminho mais curto aos demais nós da rede. Outra aplicação importante da avaliação de geodésicas é em serviços de roteamento; um exemplo é o algoritmo heurístico A* (A-star) [23], que busca um caminho mais curto entre dois pontos específicos.

Como veremos mais adiante, o cálculo de distâncias tem papel importante no cálculo de certas medidas de centralidade.

3.4.3 Clustering e estrutura de comunidades

O *coeficiente de clustering local* de um nó v , denotado por C_v , mede o quanto os nós se agrupam na vizinhança de v . Sendo G'_v o subgrafo induzido pelos vizinhos de um vértice v de G , contendo n_v vértices e m_v links, definimos

$$C_v = \frac{m_v}{\binom{n_v}{2}}, \quad (3.1)$$

ou seja, é uma medida de o quanto os vizinhos de v estão próximos de serem uma clique. Para o grafo como um todo, define-se o *coeficiente de clustering global*, denotado por \bar{C} , como a média de C_v calculada sobre os vértices do grafo.

Redes sociais costumam exibir grupos com interesses comuns, chamados de *comunidades*. De maneira mais geral, comunidades são grupos de nós (ou seja, subgrafos) cuja densidade de links é maior dentro de cada grupo do que entre os grupos. Critérios mais precisos, entretanto, são variados. Por exemplo, podemos definir comunidades como sendo as cliques ou n -cliques maximais. Outra definição diz que G' é uma comunidade se a soma dos graus dentro de G' é maior que a soma dos graus entre G' e o restante do grafo. Dependendo do algoritmo, pode haver ou

não sobreposição de comunidades. Um exemplo de rede particionada em comunidades é mostrado na Figura 8.

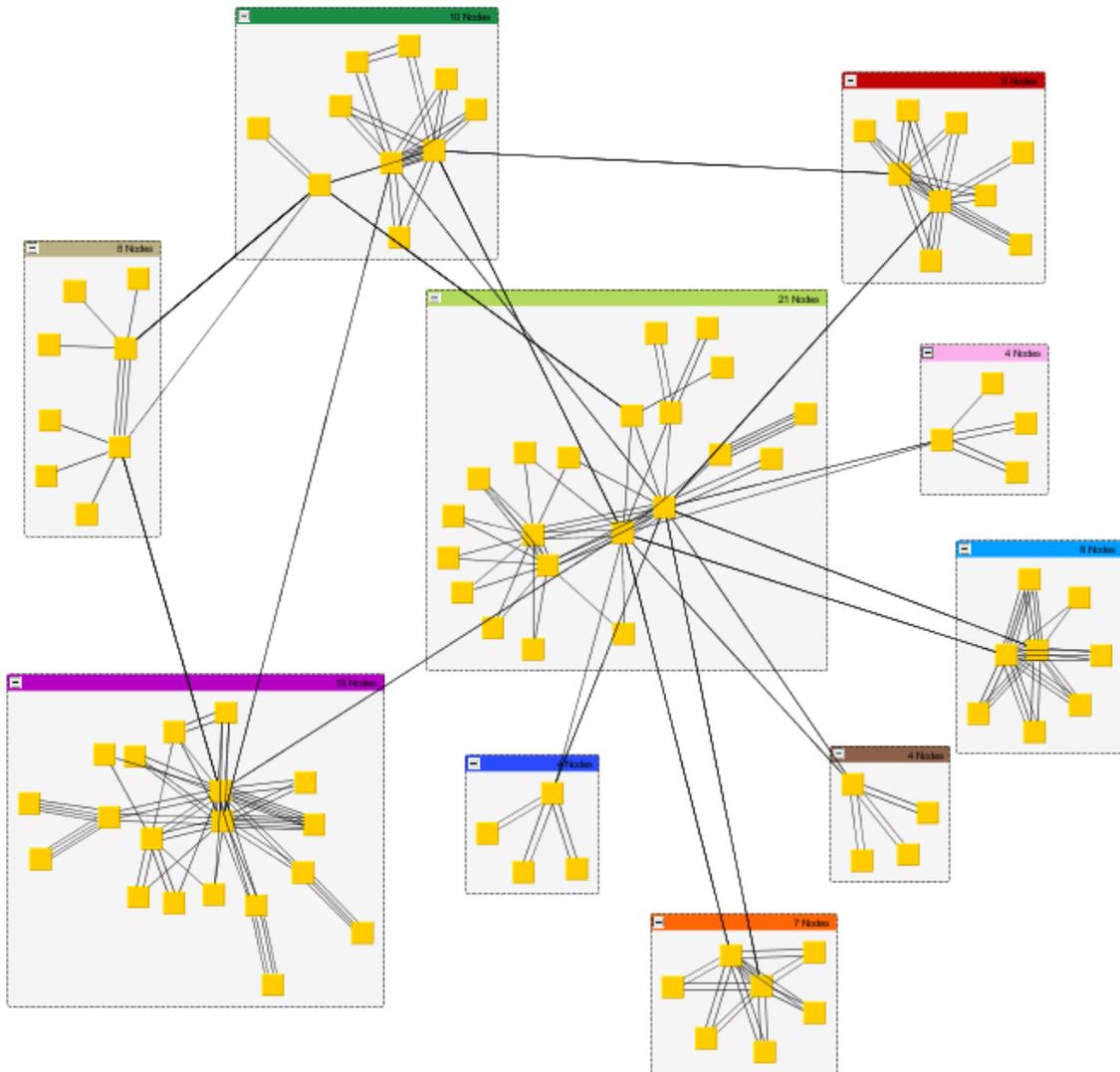


Figura 8: Comunidades em uma rede de roteadores.

Alguns métodos para a obtenção da estrutura de comunidade de uma rede serão abordados no capítulo 5, incluindo alguns que fazem o uso de medidas de centralidade.

3.4.4 Motifs

Quando um subgrafo conexo aparece em um grafo com frequência significativamente maior que em um grafo aleatório equivalente, chamamos esse

subgrafo de *motif*. Os possíveis *motifs* de n nós são todos os possíveis subgrafos conexos de K_n . Para uma rede dirigida, há 13 possíveis *motifs* de 3 nós, mostrados na Figura 9.

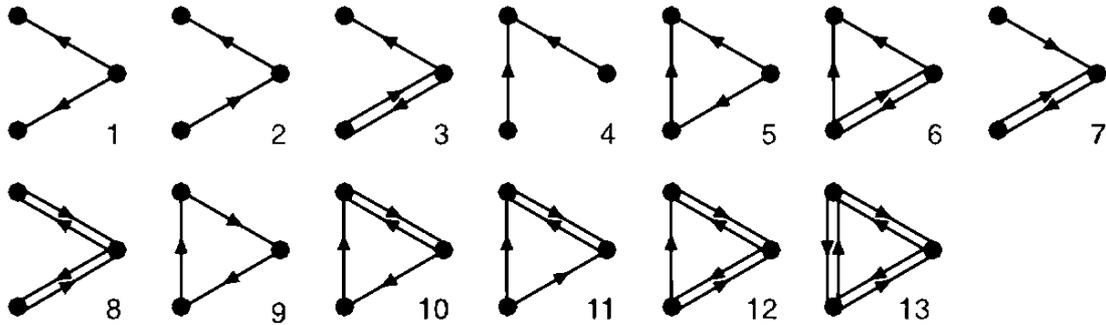


Figura 9: Os 13 motifs possíveis de 3 nós².

Denotemos por M_n o número de possíveis *motifs* de n nós. Para um grafo orientado, temos, por exemplo, que $M_2 = 2, M_3 = 13, M_4 = 199, M_5 = 9364$ e $M_6 = 1.530.843$ [24].

Uma análise possível é medir a frequência estatística de cada um desses *motifs*, quando comparada à correspondente num grafo aleatório, através de seu *escore-Z*, definido como

$$Z_M = \frac{n_M - \langle n_M^{rand} \rangle}{\sigma_M^{rand}}, \quad (3.2)$$

onde n_M é o número de ocorrências do motif M no grafo, e $\langle n_M^{rand} \rangle$ e σ_M^{rand} são respectivamente, a média e o desvio-padrão da ocorrência de M num grafo aleatório equivalente. A partir daí podemos traçar o perfil de uma rede de acordo com os escores- Z de seus *motifs*. Algumas redes apresentam perfis semelhantes, podendo ser agrupadas em “superfamílias” [1], como mostra a Figura 10. Cada gráfico nessa sequência apresenta os 13 possíveis motifs de 3 nós, no eixo horizontal, e Z_M no eixo vertical. Vemos, por exemplo, que tanto em redes de páginas da WWW quanto em redes sociais (terceiro gráfico), comparativamente ao esperado em um grafo aleatório, o número de desvios-padrão acima ou abaixo da

² http://www3.nd.edu/~cse/2013sp/60884/docs/network_motifs.pdf

média é semelhante para cada um dos motifs. Em especial, para os dois tipos de rede, há excesso dos motifs de nº 9 em diante, principalmente o nº 13, que está quase meio desvio padrão acima da média; ao mesmo tempo, os motifs de nº 4 a 6 estão subrepresentados.

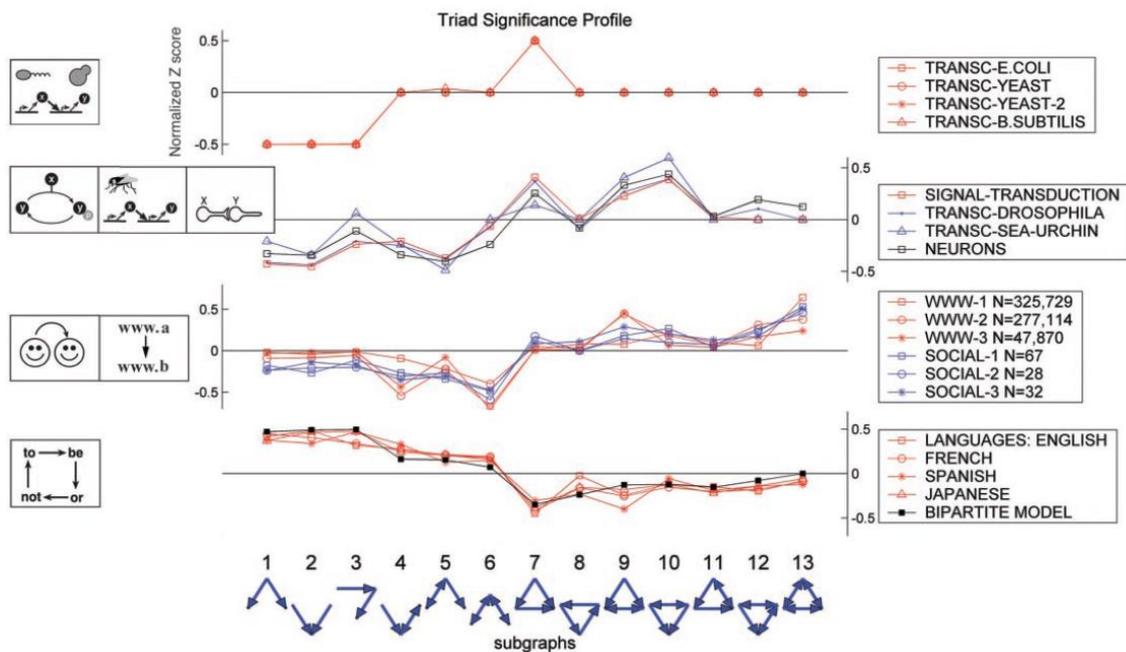


Figura 10: Redes classificadas em “superfamílias” [1] de acordo com os escores-Z de seus motifs.

3.4.5 Mediação e Centralidade

Alguns nós têm uma posição relativamente privilegiada devido à sua posição. Um exemplo óbvio são os pontos de articulação. Se modelarmos uma rede profissional como um grafo, pessoas em pontos de articulação atuam como *mediadores* de informação, podendo obter vantagens devido a essa posição [25].

A existência, em várias redes, de pontos obviamente importantes nos motiva a tentar quantificar a importância, ou *centralidade*, de um nó qualquer (ou, por extensão, de um link qualquer) dentro da estrutura da rede. Medidas de centralidade e suas aplicações serão vistas em profundidade nos capítulos 4 e 5.

3.4.6 Processos

Uma rede pode ter atributos dinâmicos que variam com o tempo seguindo certos processos. Por exemplo, pode-se estudar a evolução do contágio de uma doença através de um grafo em que os nós são pessoas. Esse tipo de processo é conhecido como *difusão*. Um exemplo é o modelo de Bass [26], em que os elementos podem ter dois estados (0 ou 1), e $F(t)$ representa a proporção de elementos no estado 1 no tempo t , definida por $F' = \frac{p+qF}{1-F}$, onde p é a taxa com que elementos passam espontaneamente para o estado 1, e q é a taxa com que elementos “imitam” o estado de outro elemento já no estado 1. O modelo pode ser enriquecido de várias formas. Um exemplo é o modelo SIS (suscetível-infectado-suscetível), usado em estudos epidemiológicos, onde cada elemento pode ser suscetível ao contágio (S) ou infectado (I), equivalentes respectivamente aos estados 0 e 1 do modelo de Bass, porém nesse modelo pode acontecer de um elemento infectado se recuperar e voltar ao estado de suscetível, com uma certa probabilidade.

Outro tipo de processo importante é a ocorrência de *falhas* em nós ou links, que correspondem matematicamente à retirada de um conjunto de nós ou links da rede. Esse é um ponto de especial interesse no âmbito das redes de dados, pois falhas podem ter um impacto que pode se traduzir em indisponibilidade de serviços e consequente perda de receita.

Falhas, a princípio, são consideradas aleatórias; quando são frutos de ação intencional, recebem o nome de *ataques*. Um atacante normalmente não escolherá seu alvo aleatoriamente, preferindo a possibilidade onde causará o máximo de impacto. Analistas de segurança de rede podem tentar prever essa preferência e preparar estratégias de defesa de acordo. Engenheiros de rede podem, por exemplo, tentar reforçar a estrutura dos links mais suscetíveis a ataques, ou prover mais redundância a esses links.

4. Medidas de Centralidade

Ao se analisar uma rede, um aspecto importante é decidir quais são os nós ou os links mais importantes (ou *centrais*). As *medidas de centralidade* são uma forma de quantificar essa importância. Neste capítulo, definiremos mais precisamente as medidas de centralidade.

4.1 Introdução

Nas definições, considera-se um grafo $G = (V = \{v_i\}, E = \{e_i\})$, com $|V| = n$, $|E| = m$, matriz de adjacência A .

Uma *medida de centralidade de nó* é uma função $c_X: V \rightarrow \mathbb{R}$ que possui certas propriedades, descritas a seguir. Primeiro, c_X deve ser um índice estrutural de G , ou seja, se G e H são grafos isomorfos por ϕ , então $c_X(\phi(v)) = c_X(v), \forall v \in V$. Outra propriedade, matematicamente menos precisa e mais dependente do contexto, é que a relação de ordem entre $c_X(v_i)$ e $c_X(v_j)$ deve refletir a percepção de que v_i é *mais central* que v_j , em algum sentido; para a maioria das medidas, quanto mais central for o vértice v , maior o valor de $c_X(v)$, porém pode valer o oposto. De maneira análoga, podem-se definir *medidas de centralidade de aresta* como $c_X: E \rightarrow \mathbb{R}$, ou ainda combinar os valores dos vértices ou arestas a fim de se obter uma *medida de centralização* do grafo como um todo. Usaremos a notação \mathbf{c}_X para referenciar o vetor coluna contendo as centralidades de todos os vértices, ou seja, $\mathbf{c}_X = (c_X(v_1), \dots, c_X(v_n))^T$.

A noção de centralidade, em várias aplicações, é associada à importância do elemento na estrutura. Esperam-se, por exemplo, altos³ índices de centralidade de nó para uma pessoa influente num certo círculo social, uma pessoa com cargo de chefia em uma organização, uma página da Internet com muitos visitantes ou um

³ Considerando, obviamente, que quanto maior o valor da centralidade, mais central o nó.

roteador que media um grande fluxo de dados. Analogamente, para as avenidas principais de uma cidade, reações químicas chave em uma célula ou fibras óticas ligando continentes, são esperadas altas centralidades de aresta.

Um grafo estrela S_n , onde um nó central v_0 é conectado aos n demais nós, sem nenhuma conexão entre estes últimos, é um exemplo onde a centralidade é óbvia. Vemos então que o *grau* é um exemplo de medida de centralidade.

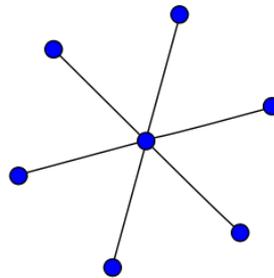


Figura 11: Grafo estrela S_6 .

Para redes em geral, entretanto, o grau nem sempre captura adequadamente a importância de um nó. Tomemos o grafo da Figura 12 como exemplo. O nó v em destaque é o de menor grau em todo o grafo, porém é visível que v tem um papel importante na estrutura. Em relação aos demais nós, v está mais próximo, em média, de um outro nó qualquer, ou seja, ele exibe maior *proximidade*; além disso, qualquer caminhos entre um dos 4 nós à esquerda e um dos 4 nós à direita (ou vice-versa) passa por v , conferindo a v um caráter de *intermediação*.

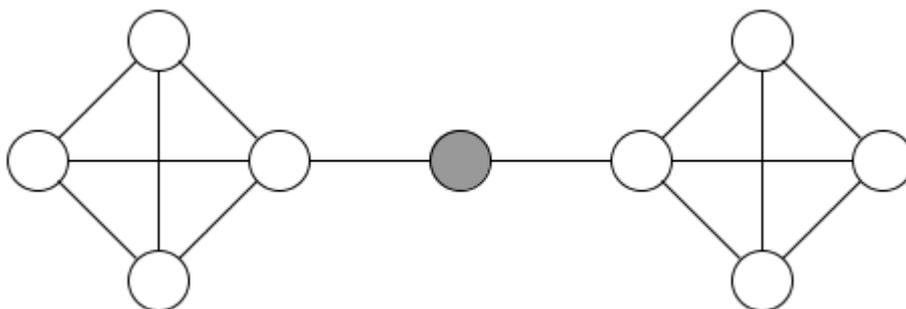


Figura 12: Neste grafo, o nó em destaque tem papel central evidente, ainda que seu grau seja mínimo.

Essas três características (grau, proximidade e intermediação) são a motivação para as medidas básicas de centralidade, que serão apresentadas na seção 4.2.

Essas mesmas três características são usadas para categorizar várias outras medidas de centralidade. Algumas dessas medidas são descritas na seção 4.3, onde também é apresentada uma classe separada: a das medidas *delta*.

4.2 As três medidas básicas: grau, proximidade e intermediação

A seguir, serão apresentadas de uma maneira informal as três medidas de centralidade fundamentais: *grau*, *proximidade* e *intermediação*⁴, a partir das ideias que as motivaram. As definições seguem as feitas por Freeman [2]. Borgatti e Everett [27] usam essas medidas para categorizar outras encontradas na literatura em centralidades similares à de grau, similares à de proximidade e similares à de intermediação⁵.

É comum que medidas de centralidade partam de uma medida *absoluta* e sejam normalizadas por uma cota superior $r \geq c_x$, de modo que $c'_x = c_x/r \leq 1$; c'_x então é dita uma medida *relativa*. A escolha natural de r é o supremo dos valores possíveis para a centralidade. Muitas vezes, entretanto, não existe o nó com a centralidade máxima. Por exemplo, enquanto em K_n todos os nós têm a mesma centralidade de grau máxima, a centralidade de proximidade máxima é apenas obtida pelo ponto central de um grafo estrela, como veremos mais adiante. Então, às vezes toma-se uma medida inicial da centralidade que é então normalizada pelo valor máximo dentre as centralidades dos elementos, de modo que $c'_x = 1$ para o conjunto de elementos com centralidade máxima dentro da rede. É o que faz, por exemplo, o software yEd.

⁴ Em inglês, *degree centrality*, *closeness centrality* e *betweenness centrality*.

⁵ Em inglês, *degree-like*, *closeness-like* e *betweenness-like*.

Vamos, então, às definições formais.

4.2.1 Centralidade de grau⁶

Numa rede social, uma medida que chama imediatamente a atenção é a quantidade de amigos que uma pessoa possui. A medida de centralidade mais básica reflete essa ideia:

“Um nó importante está conectado com muitos nós.”

Definição: Para um grafo não-orientado, a *centralidade de grau* de um vértice v é dada por seu grau, ou seja,

$$c_D(v_i) = k_i \quad (4.1)$$

A partir de A , podemos calcular $c_D(v_i) = \sum_j a_{ij}$ e o vetor de centralidades como $\mathbf{c}_D = A\vec{\mathbf{1}}$. Analogamente, para grafos orientados, definimos centralidades de grau de entrada ($c_D^{in}(v) = k_i^{in}$ e $\mathbf{c}_D^{in} = A^T\vec{\mathbf{1}}$), de saída ($c_D^{out}(v) = k_i^{out}$ e $\mathbf{c}_D^{out} = A\vec{\mathbf{1}}$) e total ($c_D^{tot}(v) = k_i^{in} + k_i^{out}$ e $\mathbf{c}_D^{tot} = (A + A^T)\vec{\mathbf{1}}$). A complexidade de tempo, nesse caso, é $O(n^2)$. Numa representação esparsa, a complexidade é $O(m)$.

Se o grafo é simples, $c_D \leq n - 1$, e portanto podemos definir a medida relativa

$$c'_D(v) = \frac{k_i}{n - 1} \quad (4.2)$$

Um exemplo de aplicação desse tipo de centralidade é mostrado em [28]: analisando as interações (arestas) entre as proteínas (nós) da bactéria *Helicobacter pylori*, é verificada uma correlação entre o grau e a letalidade da proteína se retirada do sistema.

⁶ Originalmente chamada por Freeman de *Point Centrality*.

4.2.2 Centralidade de proximidade

Em vários contextos, mais importante que ter muitas conexões é não estar longe demais dos demais nós. Nessa perspectiva, temos a seguinte motivação:

“Um nó importante está próximo dos outros nós.”

Definição: A *centralidade de proximidade* de um vértice v é dada pelo recíproco da soma das suas distâncias aos demais nós⁷, ou seja,

$$c_C(v) = \frac{1}{\sum_{a \in V} d(v, a)}. \quad (4.3)$$

A extensão dessa medida para grafos orientados é feita considerando apenas distâncias sobre caminhos orientados. Essa medida não tem significado em um grafo desconexo, pois aí existem v_i, v_j tais que $d(v_i, v_j) = \infty$, torna o somatório infinito e, por consequência, $c_C(v) = 0$ para qualquer v . Em grafos sem peso, $(c_C)^{-1} \leq n - 1$, e então podemos definir $c'_C = (n - 1)c_C$.

O cálculo dessa centralidade envolve o cálculo das distâncias entre todos os pares de nós, que pode ser feita em tempo $O(n^3)$ usando o algoritmo de Floyd-Warshall; em redes esparsas, é possível efetuar o cálculo em $O(mn + n^2 \log n)$ usando o algoritmo de Johnson.

4.2.3 Centralidade de intermediação

Retomemos a ideia de mediação, mencionada anteriormente; por exemplo, pontos de articulação farão parte de todos os caminhos entre nós de blocos diferentes; uma cidade que faz parte de várias rotas comerciais tem certamente uma vantagem estratégica. Isto motiva a seguinte ideia:

“Um nó importante faz parte de muitos caminhos.”

⁷ A definição original [2] calcula apenas a soma, o que dá uma medida de *lonjura* ao invés de proximidade, cujo valor é *menor* quanto mais central for o nó.

Definição: Dados $v, a, b \in V$, sejam g_{ab} o número de geodésicas entre a e b , e g_{avb} o número dessas que passam por v . A *centralidade de intermediação* do vértice v é

$$c_B(v) = \sum_{a,b \neq v} \frac{g_{avb}}{g_{ab}}. \quad (4.4)$$

De outra forma, é a proporção de caminhos entre a e b que passa por v , somada sobre todos os pares possíveis. Novamente, a extensão para grafos orientados é feita considerando apenas caminhos orientados.

Teorema 4.1: Seja um grafo não-orientado qualquer. O valor máximo de c_B é $\binom{n-1}{2}$, e é atingido apenas pelo ponto central de um grafo estrela.

Prova: Fixado v , temos que $\max \frac{g_{avb}}{g_{ab}} = 1, \forall a, b \neq v$ e há $\binom{n-1}{2}$ pares (a, b) possíveis, portanto $\max c_B = \sum_{i=1}^{\binom{n-1}{2}} 1 = \binom{n-1}{2}$.

Seja agora um grafo estrela S_n , sendo v_0 seu nó central. Para $a, b \neq v_0$ dados, há apenas a geodésica (a, v_0, b) , portanto $\frac{g_{av_0b}}{g_{ab}} = 1$. Os demais nós não fazem parte de nenhuma geodésica, e então $c_B(v \neq v_0) = 0$.

Seja agora um grafo G com n nós ($G \neq S_{n-1}$), sendo v o nó de maior c_b . Partindo do grafo vazio, adicionamos as arestas de que ligam v aos seus vizinhos. Até esse ponto, a maior parte conexa do grafo é uma estrela e $c_B(v_k) = \binom{n-1}{2}$. Cada nova aresta adicionada entre nós a, b criará um caminho direto entre a e b , fazendo com que $\frac{g_{avb}}{g_{ab}} = 0$. Como o somatório é sobre $\binom{n-1}{2}$ pares, $c_B(v_k) < \binom{n-1}{2}$. ■

Com base nesse resultado, podemos definir a centralidade de intermediação relativa como

$$c'_B(v) = \frac{c_B(v)}{\binom{n-1}{2}} \quad (4.5)$$

Brandes [29] descreve um algoritmo para calcular a centralidade de intermediação de todos os nós de um grafo, com uma complexidade de tempo de $O(nm + n^2 \log n)$, para grafos com pesos, e $O(nm)$, para grafos sem pesos.

4.3 Outras medidas de centralidade

Das três medidas básicas, diversas outras podem ser derivadas. Faremos aqui um apanhado de medidas encontradas na literatura, seguindo a classificação de centralidades sugerida por Borgatti e Everett [27], separando as medidas em centralidades similares à de grau, similares à de proximidade e similares à de intermediação. Separada dessa há também a classe das centralidades delta, introduzidas por Latora e Marchiori [30].

4.3.1 Medidas similares à de grau

Um aspecto que diferencia a centralidade de grau é a possibilidade de poder calculá-la a partir de A . Um resultado conhecido é que as potências da matriz de adjacência fornecem informações sobre os caminhos.

Teorema 4.2:

$$(A^p)_{ij} = \text{número de passeios de tamanho } p \text{ entre } v_i \text{ e } v_j. \quad (4.6)$$

Prova: Por indução em p . Para $p = 1$, o resultado segue da definição de A . Da multiplicação $A^{p+1} = A^p A$, temos que

$$(A^{p+1})_{ij} = \sum_{k=1}^p (A^p)_{ik} A_{kj} \quad (4.7)$$

Basta então observarmos que um passeio de tamanho $p + 1$ entre v_i e v_j é constituído por um passeio de tamanho p entre v_i e $v_k \neq v_j$ (sendo que o número

desses passeios é de $(A^p)_{ik}$, pela hipótese de indução), mais uma aresta entre v_k e v_j , que existe se e somente se $A_{kj} = 1$. ■

Observamos que, para grafos orientados, A^p contém as quantidades de passeios orientados “para frente”, enquanto $(A^T)^p$ contém as quantidades de passeios orientados “para trás”. Portanto, ao se utilizar as potências de A , como faremos nas definições a seguir, pode ser conveniente substituir A por A^T , dependendo do contexto.

Se pensarmos $c_D(v)$ como a quantidade de passeios de tamanho 1 saindo de v , podemos expandir essa noção para passeios de um tamanho k qualquer. Isso motiva a definição a seguir.

Definição: a *centralidade de k -passeio* é dada por

$$C_{k\text{-walk}} = W_k \vec{1}, \quad (4.8)$$

onde $W_k = \sum_{j=1}^k A^j$, ou seja, $W_k[i, j]$ contém o número de passeios de tamanho k ou menos entre v_i e v_j . Note-se que $c_{1\text{-walk}} = c_D$. Assumindo que a complexidade de tempo da multiplicação de duas matrizes $n \times n$ é $O(M(n))$, a complexidade de tempo do cálculo dessa centralidade é de $O(kM(n))$.

Diferentes restrições podem ser colocadas quanto à contagem dos passeios. Em [27] são citadas medidas em que são considerados apenas passeios disjuntos por aresta, ou então disjuntos por vértice.

Outra expansão possível é fazer uma combinação linear dos números de passeios de vários tamanhos, ou mesmo todos os tamanhos possíveis, por exemplo, atribuindo um peso α^k aos passeios de tamanho k . Tem-se então a seguinte centralidade:

Definição: A *centralidade de Katz* [31] é dada por

$$\mathbf{c}_{\text{Katz}} = K \vec{1}, \quad (4.9)$$

onde $K = \sum_{k=1}^{\infty} \alpha^k A^k$.

Teorema 4.3: Seja λ_1 o maior autovalor de A . Se $|\alpha| < 1/\lambda_1$, então

$$K \rightarrow (I - \alpha A)^{-1} - I. \quad (4.10)$$

Prova: Pelo Corolário 2.4, $\alpha A \rightarrow 0$ (ou seja, αA é convergente) se $\|\alpha A\| < 1$, ou seja, se $\|A\| < 1/|\alpha|$. Pelo Teorema 2.3, $\lambda_1 < \|A\|$. Portanto αA é convergente se $\lambda_1 < 1/|\alpha|$. Assim, pelo Teorema 2.5,

$$\sum_{k=0}^{\infty} (\alpha A)^k = I + \sum_{k=1}^{\infty} \alpha^k A^k = (I - \alpha A)^{-1} \quad (4.11)$$

Subtraindo I dos dois lados da igualdade, segue o resultado. ■

Portanto, se $|\alpha| < 1/\lambda_1$, podemos escrever

$$\mathbf{c}_{\text{Katz}} = ((I - \alpha A)^{-1} - I)\vec{\mathbf{1}}, \quad (4.12)$$

e portanto a complexidade de tempo do cálculo dessa centralidade é de $O(M(n))$.

Uma variação possível é substituir $\vec{\mathbf{1}}$ por um vetor β que represente um grau externo de importância atribuído a cada nó.

Outra maneira de interpretar a centralidade é pensar que a centralidade de um nó seja uma função da centralidade dos seus vizinhos, ou seja, que

“Um nó importante tem vizinhos importantes.”

Se considerarmos que a centralidade $c_x(v_i)$ é proporcional (por um fator α) à soma das centralidades de seus vizinhos, temos que

$$c_x(v_i) = \alpha \sum_{j=1}^n a_{ji} c_x(v_j) = \alpha A(:, i) \mathbf{c}_x. \quad (4.13)$$

Em notação matricial, $\mathbf{c}_X = \alpha A \mathbf{c}_X$, ou seja, α^{-1} é um autovalor de A e \mathbf{c}_X é o autovetor correspondente. Pelo Teorema de Perron-Frobenius, temos que um autovetor associado a λ_1 pode ser tomado de modo a termos todas as entradas positivas. Bonacich [32] foi quem primeiro recomendou que o autovetor principal \mathbf{v}_1 associado a λ_1 fosse usado como uma medida de centralidade.

Definição: A *centralidade de autovetor* é dada por

$$\mathbf{c}_{\text{eig}} = \mathbf{v}_1. \quad (4.14)$$

onde \mathbf{v}_1 é o autovetor unitário positivo associado a λ_1 .

Uma vantagem desse método é a possibilidade de uso de métodos numéricos (como o método das potências) para a determinação do autovetor.

Um problema da abordagem usada na centralidade de autovetor é que, se um nó importante é vizinho de um grande número de nós, todos eles ganham importância. Por exemplo, se eu tenho minha página hospedada no *Yahoo!*, eu ganho importância em função disso - assim como milhões de outros sites. Além disso, nesse esquema, uma página poderia inflar artificialmente sua importância gerando uma grande quantidade de links [3]. Seria razoável sugerir que a importância do nó fosse “diluída” entre seus links. Essa é a principal ideia do algoritmo *PageRank*. Similarmente ao que foi feito com \mathbf{c}_{eig} , montamos a equação

$$c_X(v_i) = \alpha \sum_{j=1}^n a_{ij} \frac{c_X(v_j)}{k_j^{\text{out}}} \quad (4.15)$$

Um problema óbvio nessa equação são os nós com $k_j^{\text{out}} = 0$; nesse caso, porém, $a_{ij} = 0$, então podemos ajustar $k_j^{\text{out}} = 1$ (ou outro valor não-nulo qualquer). Assim, em forma matricial, temos

$$\mathbf{c}_X = \alpha A D^{-1} \mathbf{c}_X, \quad (4.16)$$

onde D é uma matriz diagonal tal que $D_{ii} = \max\{k_i^{out}, 1\}$. Portanto, a centralidade *PageRank* é um autovetor da matriz αAD^{-1} . Para uma discussão mais aprofundada dessa centralidade, ver [33] e [3].

4.3.2 Medidas similares à de proximidade

Outra possibilidade de se atribuir centralidade é primeiro designar o nó (ou conjunto de nós) mais central, e a partir daí usar a distância de cada nó até esse nó (ou conjunto de nós) central.

Definição: A *centralidade de centroide* é dada por

$$c_{cent}(v) = d(v, c(G)), \quad (4.17)$$

onde $c(G)$ é o centroide do grafo. A escolha natural é o próprio centro do grafo (o conjunto de vértices de menor excentricidade), porém outro critério poderia ser usado, como até mesmo outra medida de centralidade. Usando o algoritmo de Bellman-Ford (de complexidade de tempo $O(mn)$) para o cálculo de distâncias entre um único nó de origem e os demais nós da rede, e sendo $\mathcal{C}(n)$ a complexidade de tempo da determinação do centroide, a cálculo dessa centralidade pode ser feita em tempo $O(mn + \mathcal{C}(n))$.

4.3.3 Medidas similares à de intermediação

Lembramos que a centralidade de intermediação original considera apenas as geodésicas. Entretanto, muitas vezes a interação entre os nós ocorrem por outros caminhos, além desses. Por exemplo, notícias, boatos ou doenças tendem a se difundir por vários caminhos.

Ford e Fulkerson [34] desenvolveram um modelo de *fluxo* entre dois nós. Inicialmente, para todo $ab \in E$, define-se uma função $\kappa : E \rightarrow \mathbb{R}_+$ chamada de *capacidade de fluxo*; assim, o valor de $\kappa(a, b)$ pode ser visto intuitivamente como a

quantidade de bens ou informação que chega ao nó j vinda do nó i por unidade de tempo.

Define-se então um *conjunto de corte* (ou simplesmente *corte*) entre a e b como um conjunto de links que desconecta a e b , e a capacidade de fluxo do corte é a soma das capacidades de fluxo dos pares que compõem esse corte. Dentre todos os cortes possíveis entre i e j , aquele com capacidade mínima é chamado de *corte mínimo*. O seguinte teorema, cuja prova pode ser encontrada em [35], é apresentado:

Teorema 4.4 (Fluxo Máximo-Corte Mínimo): Sejam $a, b \in V$. O fluxo máximo entre a e b é igual à capacidade do corte mínimo entre a e b .

Baseados nesses conceitos, Freeman, Borgatti e White [36] apresentaram a seguinte medida:

Definição: Dados $v, a, b \in V$, sejam m_{ab} o fluxo máximo entre a e b , e m_{avb} o fluxo máximo entre a e b passando por v . A *centralidade de fluxo* do vértice v é

$$c_F(v) = \sum_{a,b \neq v} m_{avb} \quad (4.18)$$

e respectiva medida relativa é

$$c'_F(v) = \frac{\sum_{a,b \neq v} m_{avb}}{\sum_{a \neq b} m_{ab}} \quad (4.19)$$

4.3.4 Medidas delta

As medidas de centralidade *delta* [30] partem do seguinte pressuposto:

“Um nó importante afeta a rede se retirado dela.”

Para a definição dessa classe de centralidades, é necessário que seja definida uma *medida de coesão* para o grafo como um todo. Assim, um nó seria

mais central quanto mais decaísse essa coesão em consequência de sua retirada da rede. Denotemos por $P(G)$ essa medida de coesão.

Definição: seja $G' = G - v$. A *centralidade delta* é dada por

$$c_{\Delta}(v) = \frac{P(G) - P(G')}{P(G)}, \quad (4.20)$$

Uma escolha simples para P é usar o número de arestas, ou seja, $P(G) = m$. Uma alternativa é usar a *eficiência* [37] da rede, definida por

$$\mathcal{E}(G) = \frac{1}{n(n-1)} \sum_{\substack{u,v \in V \\ u \neq v}} \frac{1}{d(u,v)} \quad (4.21)$$

A eficiência assume que a interação (como um fluxo de informação) entre dois nós é tão eficiente quanto menor é a distância entre eles.

Definição: Seja $G' = G - v$. A *centralidade de informação* é dada por

$$c_{\text{inf}}(v) = \frac{\mathcal{E}(G) - \mathcal{E}(G')}{\mathcal{E}(G)}, \quad (4.22)$$

Embora não tenha sido citado em [30], podemos atribuir c_{Δ} a conjuntos de nós ou vértices, incluindo conjuntos mistos.

5. Aplicações em redes de dados

Apresentaremos nesse capítulo duas aplicações das medidas de centralidade às redes de dados. A primeira remete à detecção de *estruturas de comunidades* em uma rede, e a segunda à avaliação de seus *pontos de vulnerabilidade*.

5.1 Introdução

As medidas de centralidades encontram aplicações nos mais diversos contextos. Nosso interesse específico no presente texto é quanto às redes de dados. Cabe, nesse ponto, apresentar algumas terminologias. Nesse contexto, *dados* são informações de qualquer tipo codificadas em dígitos binários (*bits*), ou seja, em seqüências de zeros e uns. Os dados costumam ser agrupados em *pacotes* com tamanho variável de bits. O formato mais usado para codificação desses pacotes é o protocolo IP (*Internet Protocol*⁸). Por esse motivo, muitas vezes essas redes são chamadas de *redes IP*. Ao trafegarem entre os pontos de origem e destino dos dados, os pacotes são direcionados por dispositivos denominados *roteadores*. Diversos outros elementos podem compor uma rede, como *estações, servidores, switches e firewalls*, porém a rede formada pelos roteadores e suas interconexões é a parte mais importante de uma rede de dados. Não por acaso, o conjunto dos principais roteadores de uma rede de dados é denominado o *backbone* (“espinha dorsal”) da rede.

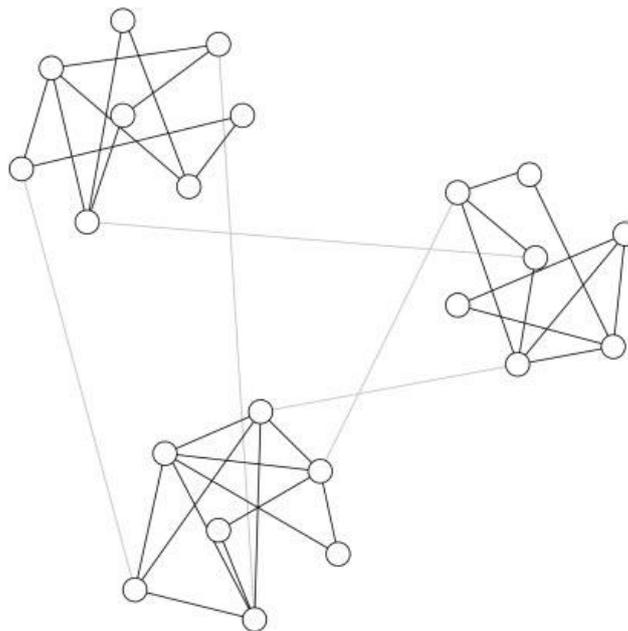
Aqui apresentaremos duas aplicações das medidas de centralidade às redes de dados. A seção 5.2 trata da detecção de *estruturas de comunidade* em uma rede. Usaremos um algoritmo conhecido para verificar a existência de comunidades em uma rede de roteadores.

⁸ <https://tools.ietf.org/html/rfc791>

Na seção 5.3, apresentamos um algoritmo para a identificação de pontos de vulnerabilidade em uma rede, a partir da simulação de um ataque aos nós da rede.

5.2 Detecção de comunidades

Em diversas redes, é típica a existência de grupos de nós que apresentam uma maior interação dentro do grupo, em comparação com as interações dentro de cada grupo, como no exemplo da Figura 13. Por exemplo, em redes sociais, certos grupos podem ter interesses em comum, como um artista ou time preferido, ou uma área do conhecimento. Denominamos esses grupos *comunidades*. Esse conceito é facilmente estendido para redes biológicas, tecnológicas e outras.



**Figura 13: uma rede com sua estrutura de comunidades evidenciada [4].
A densidade dos links dentro de cada comunidade (linhas escuras)
é maior que a densidade de links entre as comunidades (linhas claras).**

Vários métodos são propostos na literatura para a detecção de comunidades. Um método simples é, para algum $n \in \mathbb{N}$ dado, considerar como uma comunidade cada subgrafo maximal em que a maior geodésica é no máximo n^9 . Porém, é fácil haver sobreposição de comunidades grandes (caso n seja alto) ou

⁹ Em [42], usa-se para tal o termo *n-clique*; evitaremos esse termo para evitar confusão com o termo *k-clique*, que designa uma clique com k vértices.

proliferação de comunidades pequenas (caso n seja baixo). Outro método comum é usar cortes mínimos (já citados na seção 4.3.3) para separar o grafo. Nesse método, devemos saber de antemão quantas comunidades queremos, pois cada exclusão de corte mínimo separará em duas alguma comunidade existente até a iteração anterior.

No método da *Aglomerção Hierárquica*, inicialmente monta-se uma matriz W (chamada *matriz de pesos*) cujos pesos w_{ij} representam de alguma forma a força da interação entre os nós v_i e v_j . Exemplos de peso são o número de caminhos disjuntos por nó (ou por vértice) e a matriz K usada no cálculo da centralidade de Katz. A seguir, há uma abordagem *bottom-up*: inicia-se com a rede vazia e são adicionadas arestas em ordem decrescente de pesos, gerando uma árvore (chamada de *dendograma*). Um corte na árvore abaixo de uma determinada aresta dá a estrutura de comunidade existente antes da adição dessa aresta.

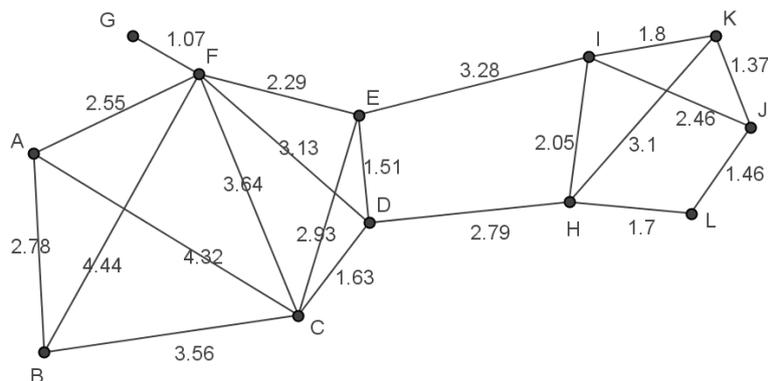


Figura 14: Rede de exemplo com distâncias.

Vejamos um exemplo de aplicação do Algoritmo na rede da Figura 14, em que são mostradas as distâncias entre nós adjacentes. Como peso, usaremos o recíproco da menor distância entre dois nós. Estes estão calculados na matriz da Figura 15.

j\i	A	B	C	D	E	F	G	H	I	J	K	L
A		0,36	0,231	0,193	0,207	0,392	0,276	0,125	0,123	0,095	0,101	0,103
B	0,36		0,281	0,193	0,154	0,225	0,181	0,125	0,102	0,09	0,09	0,103
C	0,231	0,281		0,613	0,341	0,275	0,212	0,226	0,161	0,132	0,133	0,163
D	0,193	0,193	0,613		0,662	0,319	0,238	0,358	0,209	0,168	0,17	0,223
E	0,207	0,154	0,341	0,662		0,437	0,298	0,233	0,305	0,174	0,197	0,167
F	0,392	0,225	0,275	0,319	0,437		0,935	0,169	0,18	0,125	0,136	0,131
G	0,276	0,181	0,212	0,238	0,298	0,935		0,143	0,151	0,11	0,118	0,115
H	0,125	0,125	0,226	0,358	0,233	0,169	0,143		0,488	0,316	0,323	0,588
I	0,123	0,102	0,161	0,209	0,305	0,18	0,151	0,488		0,407	0,556	0,267
J	0,095	0,09	0,132	0,168	0,174	0,125	0,11	0,316	0,407		0,73	0,685
K	0,101	0,09	0,133	0,17	0,197	0,136	0,118	0,323	0,556	0,73		0,353
L	0,103	0,103	0,163	0,223	0,167	0,131	0,115	0,588	0,267	0,685	0,353	

Figura 15: Pesos calculados na rede de exemplo.

Conforme a Figura 16, iniciamos com os nós A a L da rede. O maior peso é entre F e G, então adicionamos o nó 1 e o conectamos a F e G, “aglomerando-os”. Similarmente, criamos os nós 2 (aglomerando J e K) e 3 (aglomerando D e E). O próximo peso, na ordem, é entre J e L; então o nó 4 aglomera L e o nó 2 (abaixo do qual está J). O processo segue até o nó 11, que aglomera toda a rede. Fazendo um corte abaixo de 11, dividimos a rede em duas comunidades: uma contendo os nós de A a G, e a outra contendo os restantes.

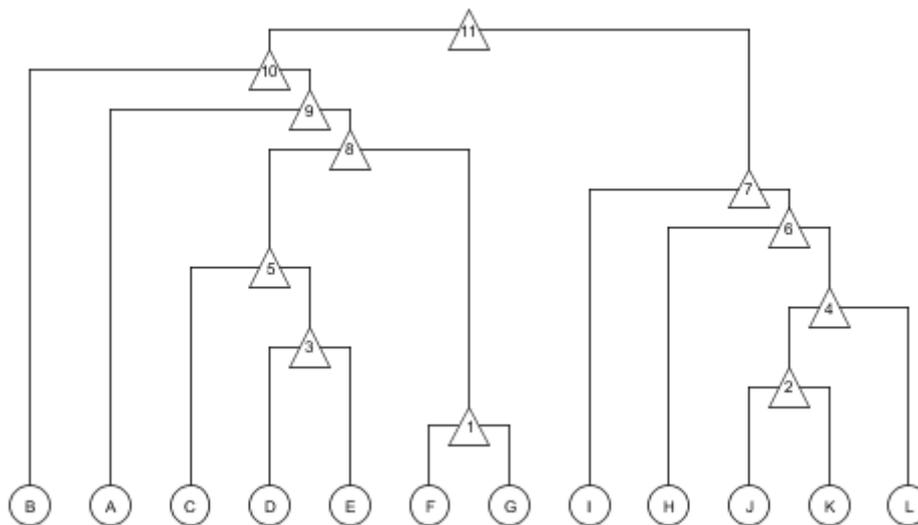


Figura 16: Dendrograma resultante.

5.2.1 O algoritmo de Girvan-Newman (GN)

Girvan e Newman [4] propuseram um metodo de deteccao de comunidades que usa uma versao da centralidade de intermediacao adaptada aos *links* da rede.

Definicao: Dados $a, b \in V$ e $l \in E$, sejam g_{ab} o numero de geodesicas entre a e b , e g_{alb} o numero dessas que contem l . A *centralidade de intermediacao* do link l e

$$c_{BL}(l) = \sum_{a,b} \frac{g_{alb}}{g_{ab}}. \quad (5.1)$$

A partir da definicao 5.1, esperamos que os links de maior c_{BL} sejam justamente aqueles interligam comunidades. O algoritmo de Girvan-Newman (o qual, deste ponto em diante, chamaremos apenas de *algoritmo GN*), e descrito a seguir.

Algoritmo 5.1 (GN):

- 1) Calcular c_{BL} para todos os links da rede;
- 2) Remover o link com maior c_{BL} ;
- 3) Recalcular c_{BL} para todos os links afetados por essa remocao;
- 4) Repetir a partir do passo 2 ate nao sobrarem mais links.

Diferentemente da Aglomeracao Hierarquica, o algoritmo GN usa uma abordagem *top-down*: cada componente conexa comeca como uma unica comunidade, que vai se subdividindo com o aumento no numero de iteracoes.

Usando o algoritmo de Newman [38], para uma rede de n nos e m links, c_{BL} e calculado em $O(mn)$ (se o grafo e esparso, ou seja, se $m \sim n$, entao esse tempo fica em $O(n^2)$). O recalculo do passo 3 faz o algoritmo completo rodar em $O(m^2n)$ (ou $O(n^3)$ para um grafo esparso), no pior caso. Entretanto, normalmente apos algumas iteracoes a rede se divide em componentes, e o recalculo de c_{BL} fica restrito as componentes afetadas por cada remocao de link. Ainda assim, essa complexidade torna a utilizacao do algoritmo pouco pratica para a utilizacao em

redes muito grandes, da ordem de alguns milhares de nós. É tentador questionar a necessidade do recálculo do passo 3; porém, pode acontecer de um link secundário aumentar sua c_{bl} após a remoção de um link importante, superando o valor de outros links.

5.2.2 Modularidade

No algoritmo GN, assim como no método dos cortes mínimos, poderíamos seguir removendo links até a rede ficar vazia; no caso da aglomeração hierárquica, poderíamos adicionar links até esgotá-los. Enfim, não se havia definido um critério de parada para os algoritmos.

Uma *partição* é uma divisão específica de uma rede em comunidades. Um dos conceitos encontrados na literatura para medir a “qualidade” de uma partição é a *modularidade* [39] de uma rede.

Definição: Seja G uma rede com m links e matriz de adjacência A , na qual foi feita uma partição em comunidades G^* . A *modularidade* de G^* é dada por

$$Q(G^*) = \frac{1}{2m} \sum_i \sum_j \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (5.2)$$

onde $k_i = \deg v_i$ e $\delta(c_i, c_j) = 1$ se v_i e v_j estão na mesma comunidade, e 0 caso contrário.

O termo $k_i k_j / 2m$ corresponde ao valor esperado para a_{ij} . Em vez do grau k_i , pode ser utilizado um peso qualquer. A modularidade representa a média, calculada sobre as comunidades existentes, da diferença entre os pesos existentes entre v_i e v_j e o valor esperado para estes pesos em um grafo aleatório equivalente. Em outras palavras, o valor de $Q(G^*)$ dá a diferença entre a quantidade de links existente e esperada dentro das comunidades. Assim, um valor alto de modularidade indica que a densidade de links dentro das comunidades é maior que o esperado ao acaso, indicando uma boa partição da rede.

5.2.3 Comunidades em uma rede de dados

Os roteadores de uma rede de dados podem ser agrupados em comunidades a fim de atender a diferentes propósitos. Vários protocolos de roteamento utilizam conceitos que equivalem ao de comunidade, ou se assemelham a ele. Como exemplos desses conceitos, citamos (com os protocolos que os utilizam): *áreas* (OSPF e IS-IS), *regiões* (MSTP) e *sistemas autônomos* (BGP). Além disso, muitas vezes é importante definir grupos de roteadores, de modo que cada grupo fique sob a administração de uma determinada equipe.

Como exemplo inicial, consideremos o caso da Figura 17, uma rede pequena hipotética cuja estrutura de comunidades é bastante aparente.

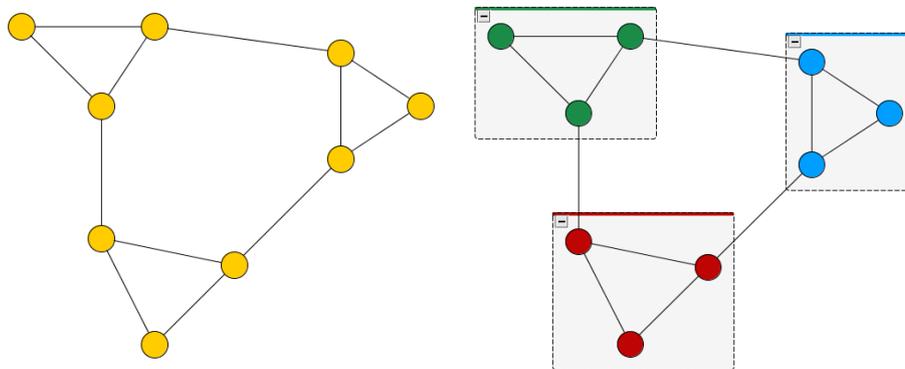


Figura 17: Uma pequena rede antes e depois do algoritmo GN executado pelo software *yEd*.

Verifiquemos, entretanto, os resultados para a modularidade, a fim de evidenciar o porquê da escolha dessa partição em particular. Fazemos, então, algumas iterações do algoritmo GN e calculemos a modularidade Q em cada ponto onde se gere uma nova partição, como mostrado na Figura 18.

Nota-se que a modularidade inicialmente cresce mas, a partir de certo ponto, passa a decrescer. Isso acontece porque estamos retirando da rede os links

intercomunidades, que estão em regiões de menor densidade de links. Posteriormente à definição da modularidade, foram desenvolvidos outros métodos matriciais para o cálculo da modularidade máxima, e a partir daí surgiram algoritmos eficientes que possibilitam a descoberta de comunidades em redes com tamanhos da ordem de milhões de nós. Detalhes desses métodos, entretanto, fogem ao escopo do presente trabalho. Um exemplo de um método desse tipo é apresentado em Newman [40].

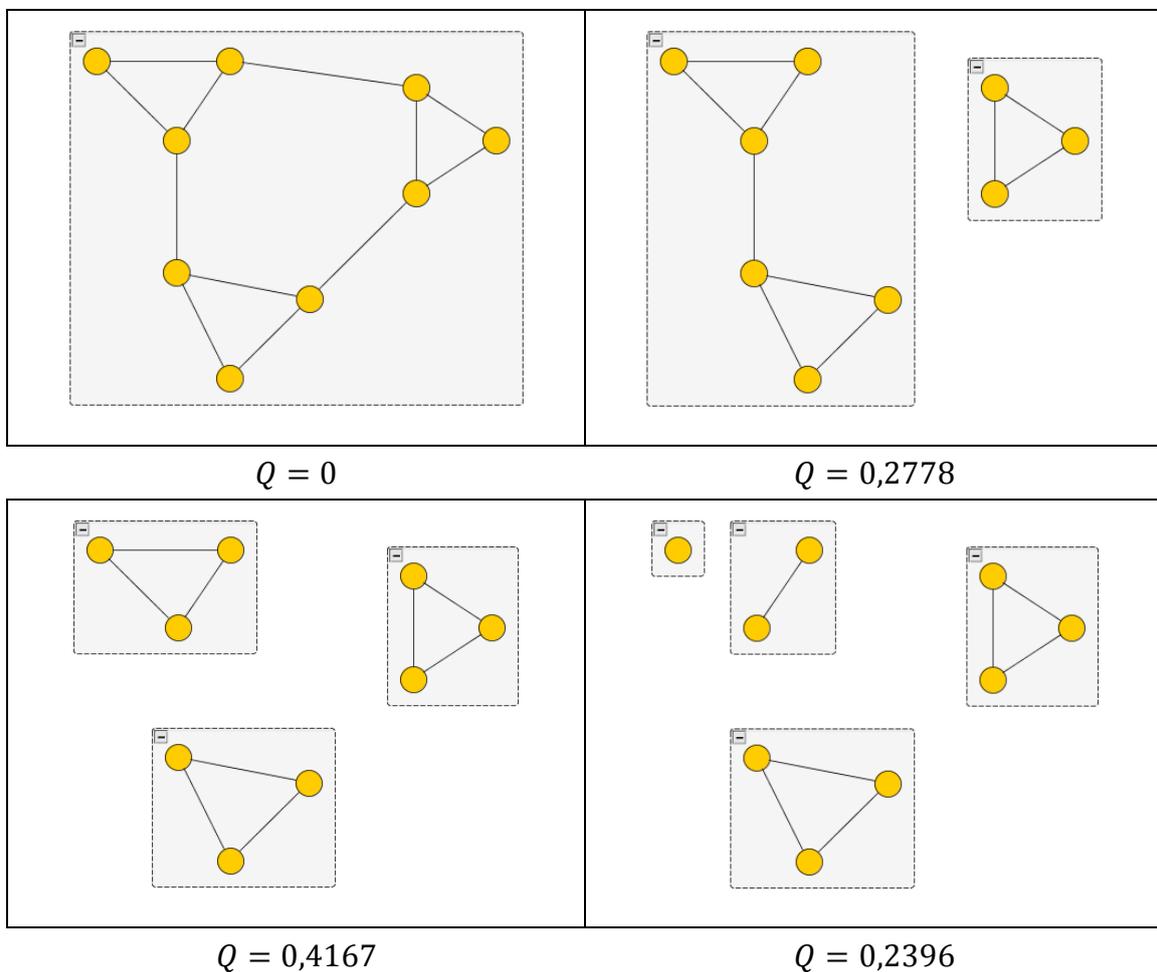


Figura 18: Modularidade das partições.

Apresentamos, por fim, uma rede adaptada de uma porção do backbone de uma empresa de tecnologia. Na Figura 19, aparece a rede não-particionada. Após

aplicado o algoritmo GN, a correspondência com a real divisão de administração dos grupos de roteadores foi quase total: a única diferença foi que os dois pequenos grupos nos cantos superiores da Figura 20 estão, na realidade, juntos com o grupo maior (na parte central, acima). Quem sabe essa não seria uma divisão mais natural da administração dos roteadores?

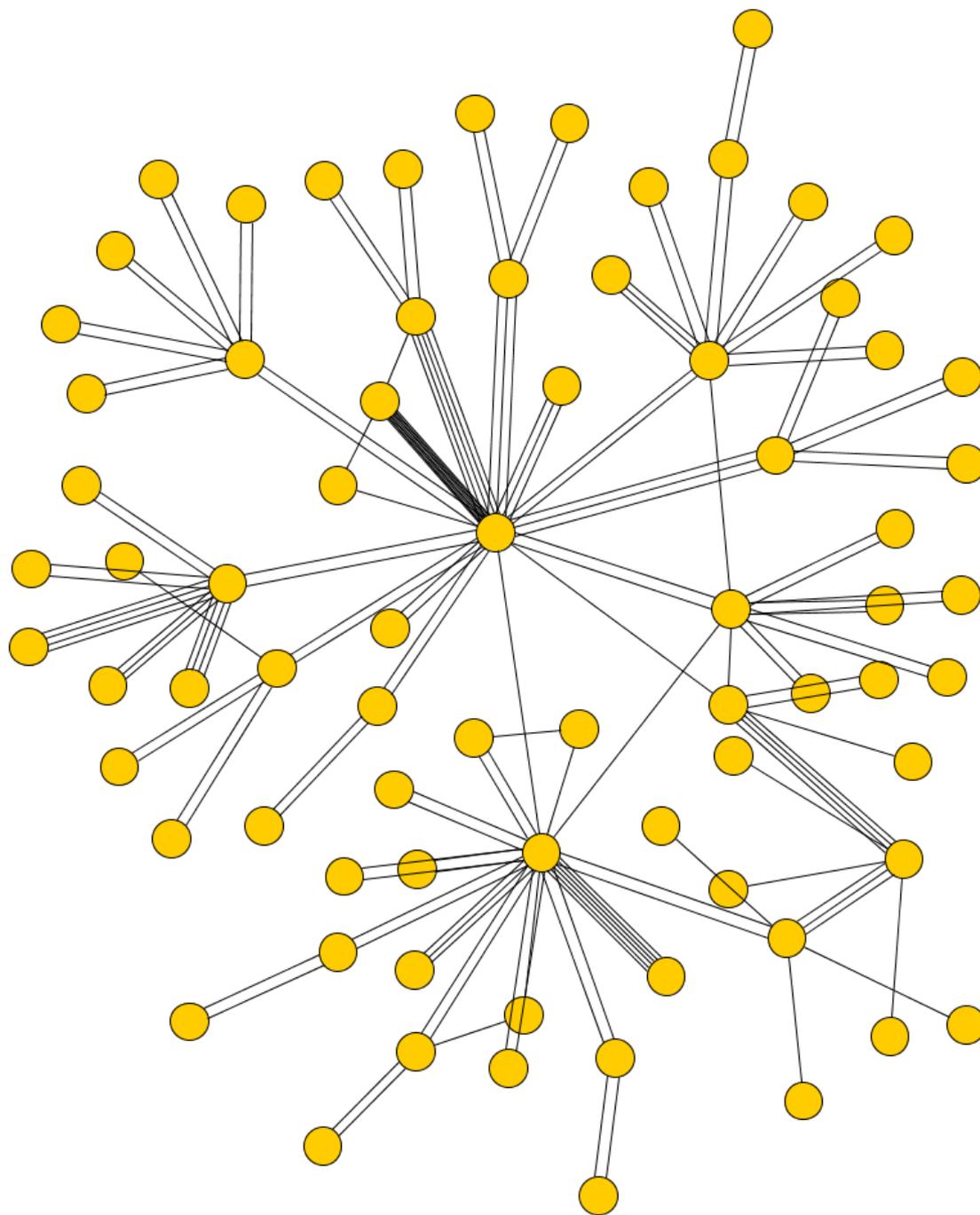
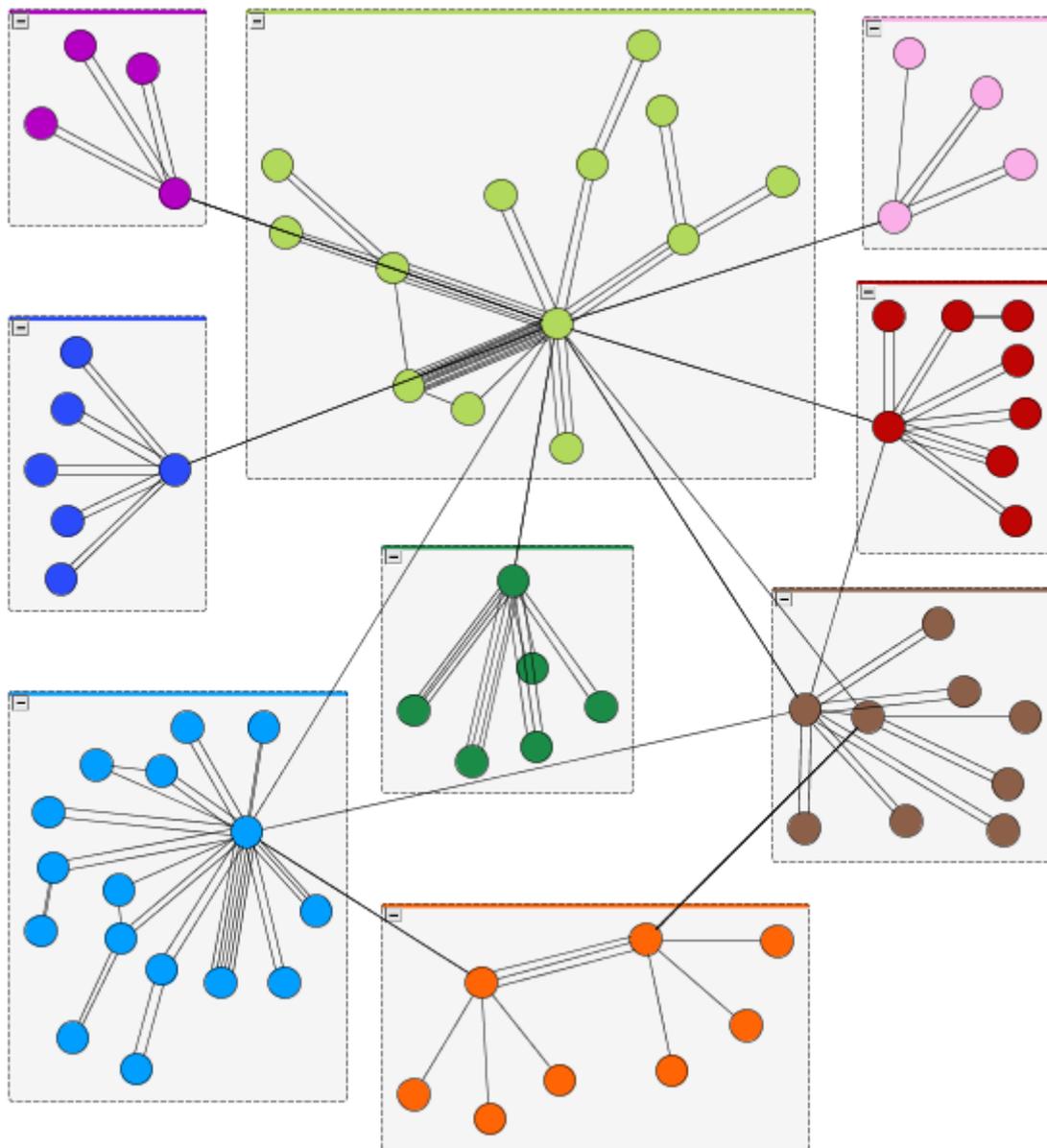


Figura 19: Rede de dados não-particionada.



**Figura 20: A mesma rede, após a aplicação do algoritmo GN.
Na rede real, as três comunidades da parte superior são uma só.**

5.3 Pontos de vulnerabilidade

5.3.1 Falhas e ataques

Já citamos, na seção 3.4.6, vários processos que podem ocorrer em uma rede. De especial interesse, no contexto das redes IP, são os processos de *falhas e ataques*, que aludem, respectivamente, à *confiabilidade* e à *vulnerabilidade* de uma rede.

Falhas são normalmente tratadas como eventos aleatórios aos quais uma rede está sujeita. Elas podem ocorrer em nós (como falhas de hardware ou problemas de energia), mas a grande maioria das ocorrências é de falhas em links. Quanto mais *confiável* é uma rede, menor é o impacto oriundo de falhas aleatórias. Em Silva [5], pode ser encontrado um estudo sobre a confiabilidade em redes, onde é evidenciada uma relação entre centralidade e confiabilidade. No caso da ampliação de uma rede, buscava-se a inserção de um link de modo a aumentar a confiabilidade da rede. Os melhores resultados foram obtidos quando inserido um link entre o nó menos central e o mais distante dele.

Ataques, por sua vez, são fruto de ação intencional e maliciosa, e podem visar tanto links quanto nós. Por exemplo, links podem ser vítimas de atos de vandalismo ou sabotagem; no caso de uma guerra, as instalações onde se localizam os nós podem ser alvo de bombardeios. Dizemos que a rede é *vulnerável* a esses ataques.

Os exemplos citados são um tanto extremos. Hoje em dia, entretanto, com a crescente ubiquidade da Internet e maior disponibilidade de conhecimento técnico, tornou-se comum a ocorrência dos chamados *cyberataques*, que são ataques que fazem uso exclusivamente dos protocolos de comunicação de dados para causar impacto. Um tipo comum de cyberataque é o *denial of service* (DoS), que ocorre quando o atacante (ou grupo de atacantes) esgota a capacidade de processamento de um nó (como um servidor ou roteador), incapacitando-o para

sua operação normal. Neste trabalho, focaremos em ataques similares a esses, ou seja, trataremos de situações onde os nós de uma rede serão alvo de ataques.

5.3.2 Análise da vulnerabilidade de uma rede de dados

Paradoxalmente, a fim de proteger uma rede, é necessário prever que alvos um atacante pode preferir, de modo a causar o máximo de impacto. É intuitivo pensar que o ataque deve ser feito nos nós mais importantes. As medidas de centralidade vistas no capítulo anterior podem ser usadas para aferir essa importância.

A pergunta que motiva a análise é: supondo que um determinado número de nós será atacado, *quais* devem ser esses nós? Ou seja, quais são seus principais pontos de vulnerabilidade?

Usaremos a rede da Figura 21, adaptada a partir de uma rede real de roteadores.

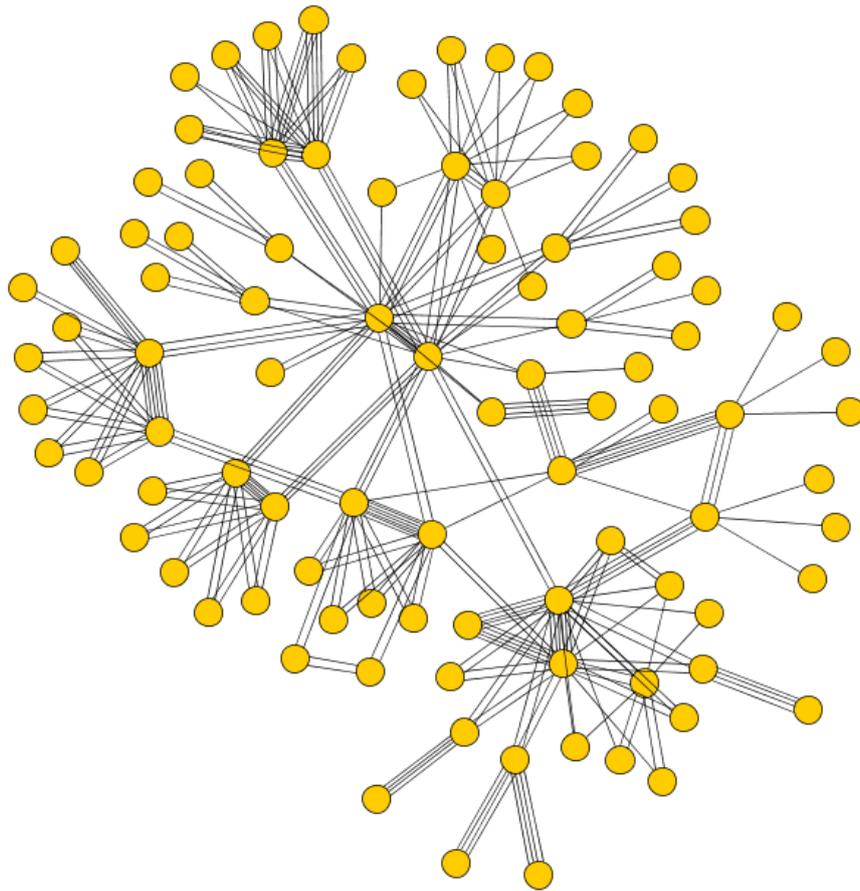


Figura 21: Rede com 94 nós, que serão "alvo" de simulações de ataque.

Nossa análise consistirá em assumir a importância de nós a partir de suas centralidades, a fim de escolher os “alvos” mais atraentes para um possível ataque. Usaremos o software *yEd*, que implementa o cálculo das centralidades de grau, proximidade, intermediação e autovetor, possibilitando boa visualização dos resultados, permitindo alterar o aspecto (tamanho ou cor) do nó conforme seu índice de centralidade. O resultado na análise das medidas de centralidade da rede de exemplo está na Figura 22.

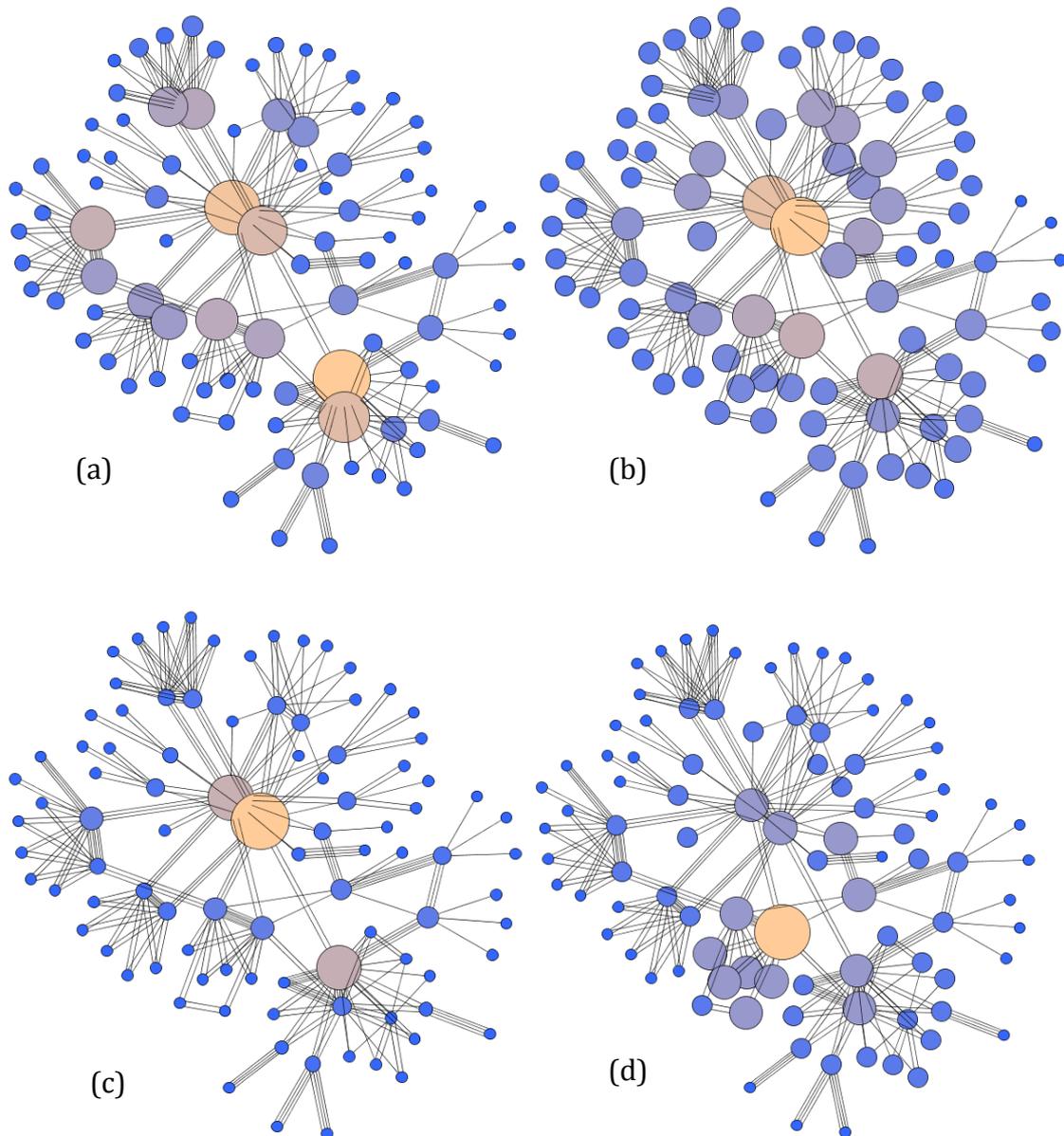


Figura 22: Centralidades calculadas na rede de exemplo: grau (a), proximidade (b), intermediação (c) e autovetor (d). Nós mais claros e maiores são mais centrais.

Resta decidir um critério que meça a eficiência de um ataque. Suponhamos que uma rede com N nós sofreu um ataque, após o qual ficou particionada em c componentes conexas C_1, \dots, C_c com $|C_i| = n_i$; então $p_i = n_i/N$ é a probabilidade de um nó estar em C_i .

Digamos que haja um nó s que seja o provedor de um serviço. Dessa forma, um cliente, ou seja, um nó, não terá acesso a esse serviço se não estiver na mesma componente conexa que s . Então o valor esperado para a proporção p_{imp} de nós impactados é

$$\overline{p_{\text{imp}}} = \sum_{i=1}^c \left[p(s \in C_i) \frac{N - n_i}{N} \right] = \sum_{i=1}^c [p_i(1 - p_i)] = 1 - \sum_{i=1}^c p_i^2 \quad (5.3)$$

O impacto é máximo quando $\overline{p_{\text{imp}}}$ é máximo, ou seja, quando $\sum_{i=1}^c p_i^2$ é mínimo, dada que todos $p_i \geq 0$ e que $\sum_{i=1}^c p_i = 1$. Isso ocorre quando todos $p_i = \max p_i = 1/c$. Assim, podemos tomar como medida de eficácia do ataque o tamanho n_i da maior componente (que é onde ocorre o valor máximo de p_i) ou o número de componentes conexas c . Entretanto, fixado c , o valor máximo de n_i pode variar grandemente (entre $\lceil N/c \rceil$ e $N - c + 1$), enquanto que, se fixarmos n_i , sabemos que $c \geq N/n_i$; portanto o tamanho da maior componente conexa acaba sendo uma métrica mais adequada.

De maneira informal, faz muita diferença ter a rede separada em duas componentes ($c = 2$) aproximadamente do mesmo tamanho, ou ter uma componente grande e outra bem menor. No primeiro caso, é quase certo que metade dos nós será impactada; no segundo, pode ser que a maior parte da rede seja impactada, mas a probabilidade de isso acontecer é pequena – é muito mais provável que o provedor esteja no componente maior. Por exemplo, assumindo $N = 100$, no primeiro caso temos $p_1 = p_2 = 0,5$ e portanto $\overline{p_{\text{imp}}} = 0,5$ (que é o valor máximo); no segundo, assumindo por exemplo $p_1 = 0,1$, temos que $\overline{p_{\text{imp}}} = 0,18$.

Com isso, descrevemos, a seguir, um algoritmo de simulação de ataque.

Algoritmo 5.2:

- 1) Calcular a centralidade dos nós da rede;
- 2) Remover o nó de maior valor de centralidade da rede¹⁰ no momento;
- 3) Obter o tamanho da maior componente conexa;
- 4) Repetir a partir do passo 2.

No yEd há recursos para executar facilmente as operações dos passos 2 e 3. Note-se que a centralidade é calculada apenas uma vez, ou seja, na segunda iteração, será retirado o nó com a segunda maior centralidade, e assim por diante. Na Figura 23 temos um gráfico com os resultados das 10 primeiras iterações dos ataques.

¹⁰ Em caso de empate, é escolhido um nó aleatório dentre aqueles com o maior valor de centralidade.

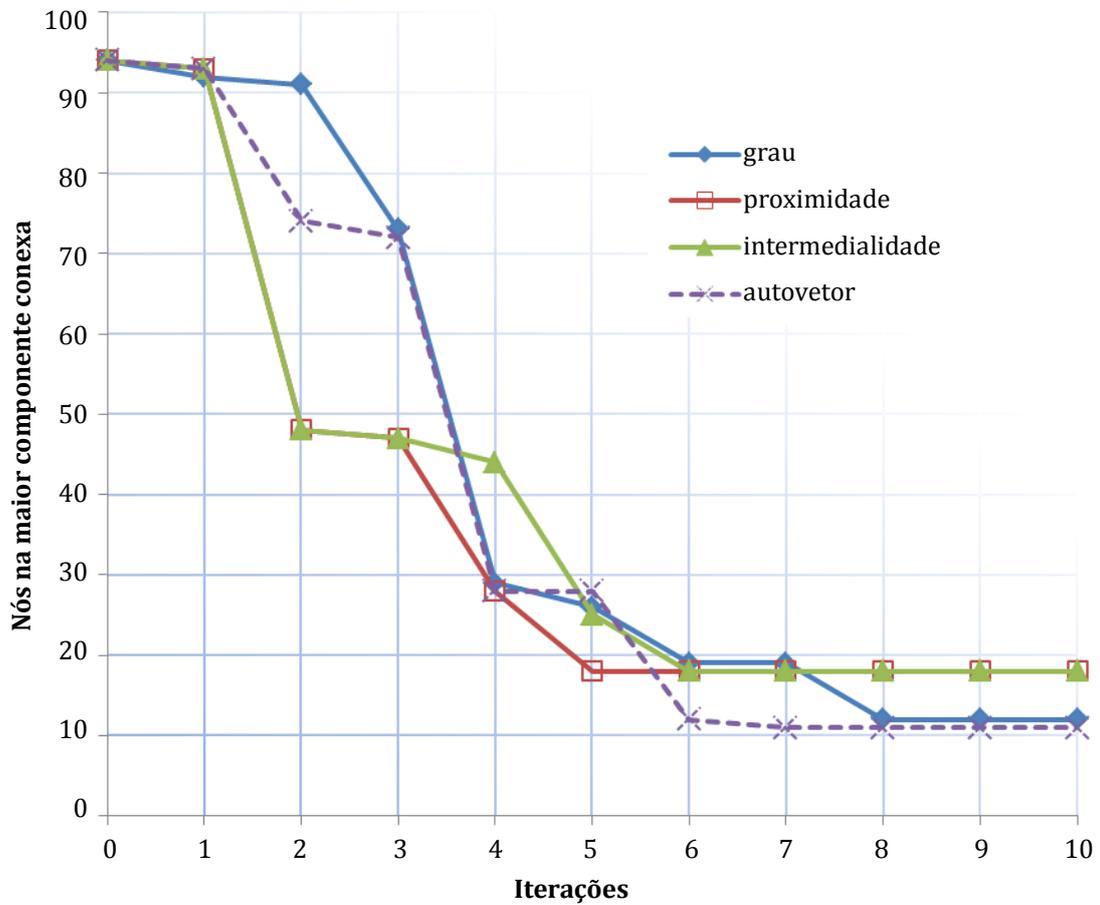


Figura 23: Tamanho da maior componente conexa da rede em função do número de iterações do Algoritmo 5.2.

A partir do 6º ataque, a rede já está bem fragmentada em qualquer um dos casos. Isso evidencia, portanto, que esse algoritmo pode servir de guia para direcionar procedimentos de segurança, indicando onde deve ser reforçada a segurança dos nós da rede.

Observamos que os ataques iniciais que têm mais impacto são aqueles em que são escolhidos os nós de maior proximidade e maior intermediação. Também deve-se levar em consideração que o cálculo da centralidade de intermediação é menos custoso ($O(n^2 \log n)$ [29]) em relação à de proximidade ($O(n^3)$ [41]).

Há de se pensar que poderia haver melhora se a centralidade fosse recalculada a cada iteração, considerando que a configuração da rede é alterada

com a retirada de um nó que ocorre no passo 2. Esse algoritmo com recálculo é descrito a seguir (as alterações relevantes estão grifadas).

Algoritmo 5.3:

- 1) Calcular a centralidade dos nós da rede;
- 2) Remover o nó de maior valor de centralidade *da maior componente conexa*¹¹;
- 3) Obter o tamanho da maior componente conexa;
- 4) Repetir a partir do *passo 1 (recálculo das centralidades)*.

No caso da centralidade de proximidade, o recálculo deve ser feito separadamente em cada componente conexa. Entretanto, a única centralidade que apresentou melhora significativa foi a de intermediação, como indica o gráfico da Figura 24.

Deve-se levar em conta, entretanto, que se o recálculo é executado k vezes, a complexidade do algoritmo é de $O(kn^2 \log n)$. Assumindo que a quantidade de nós atacados para se obter um impacto significativo seja proporcional a n , a nova complexidade é $O(n^3 \log n)$, o que prejudica a escalabilidade da sua aplicação.

¹¹ Novamente, em caso de empate, é escolhido um nó aleatório dentre aqueles com o maior valor de centralidade.

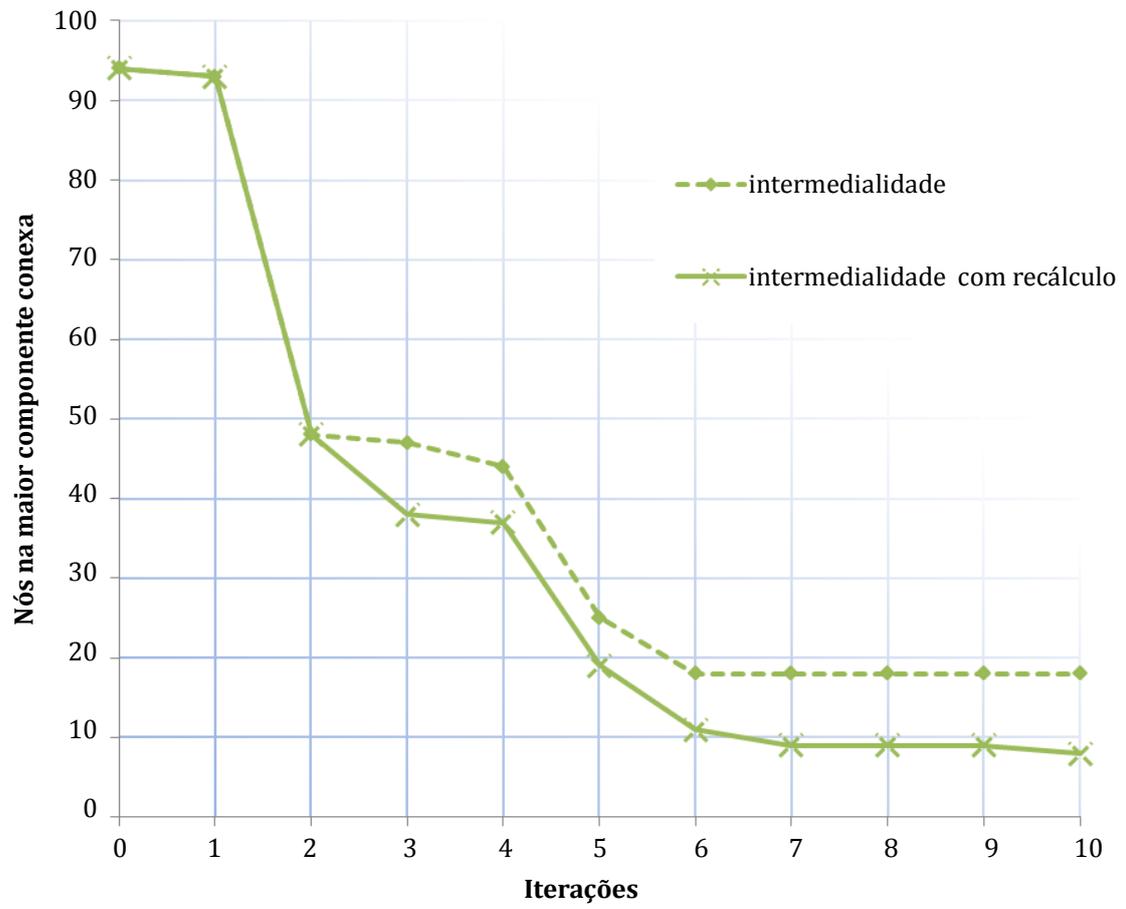


Figura 24: Comparação do resultado, para a centralidade de intermediação, do Algoritmo 5.2 (sem recálculo) e Algoritmo 5.3 (com recálculo).

6. Conclusão

A Análise de Redes é uma área bastante ampla, com pesquisa intensa, impulsionada pelas diversas aplicações e também pela evolução dos recursos tecnológicos disponíveis. Neste trabalho apresentamos uma visão geral da área, mostrando que há uma estreita ligação entre a Análise de Redes e a Teoria dos Grafos, desde os primórdios (com problemas como o das pontes de Königsberg ou das árvores geradoras de Boruvka) até conceitos mais sofisticados como o *PageRank* e a modularidade de uma partição de uma rede em comunidades.

Dentro dessa área, o conceito de centralidade tem destaque. Fizemos um apanhado das principais medidas de centralidade conhecidas, partindo das três centralidades básicas (grau, proximidade e intermediação), e mostrando como as outras medidas podem surgir a partir delas.

Vimos também que as centralidades podem ter aplicações que, à primeira vista, não parecem ter relação com seu propósito original. Um exemplo é o algoritmo de Girván-Newman, que usa a centralidade de intermediação de link para obter a estrutura de comunidades de uma rede. No caso das redes de dados, mostramos que esse algoritmo pode ser útil para dividir uma rede de roteadores em grupos, de modo que cada grupo fique sob a responsabilidade de uma equipe.

Uma rede de dados está permanentemente sujeita a eventos que incapacitam seus elementos, sejam eles nós ou links. Chamamos esses eventos de *falhas*, se são aleatórios, e de *ataques*, quando são intencionais. Assim, é importante avaliar tanto a confiabilidade (resistência a falhas) quanto a vulnerabilidade (suscetibilidade a ataques) de uma rede.

Atualmente, os *cyberataques* são uma ameaça constante às redes de dados, em especial os seus nós (servidores e roteadores). No presente trabalho, focamos na questão da identificação dos nós mais vulneráveis de uma rede, apresentando um algoritmo que faz uso das medidas de centralidade para simular um desses

ataques. Emulando o ponto de vista do atacante, partimos do pressuposto que os nós de maior centralidade são os alvos cuja incapacitação traria o maior impacto. Mostramos, também, que um bom critério para a avaliação do impacto de um ataque é o tamanho da maior componente conexa após efetivação deste ataque.

Verificamos que principalmente as medidas de centralidade de proximidade e intermediação são as que previram melhor a fragmentação da rede, sendo a de intermediação uma melhor escolha devido ao seu menor custo computacional. Portanto, esse algoritmo pode servir de guia para direcionar procedimentos de segurança.

O algoritmo também foi modificado de modo a recalcular a centralidade na maior componente conexa a cada iteração. Foi verificada uma melhora para o caso centralidade de intermediação. A complexidade desse algoritmo modificado, entretanto, prejudica sua escalabilidade.

Bibliografia

- [1] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, e U. Alon, "Superfamilies of evolved and designed networks.," *Science*, vol. 303, no. 5663, pp. 1538–42, Mar. 2004.
- [2] L. C. Freeman, "Centrality in Social Networks Conceptual Clarification," vol. 1, no. 1968, pp. 215–239, 1978.
- [3] K. Bryan e T. Leise, "The \$25,000,000,000 eigenvector: The linear algebra behind Google," *SIAM Rev.*, no. 812, pp. 1–11, 2006.
- [4] M. Girvan e M. E. J. Newman, "Community structure in social and biological networks.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 12, pp. 7821–6, Jun. 2002.
- [5] T. S. A. Silva, "Um estudo de Medidas de Centralidade e Confiabilidade em Redes," CEFET/RJ, 2010.
- [6] R. Diestel, *Graph Theory*, 4^a ed. Springer, 2010.
- [7] B. Datta, *Numerical linear algebra and applications*, 2nd ed. 1995.
- [8] L. A. Adamic, "The Political Blogosphere and the 2004 U . S . Election : Divided They Blog," pp. 36–43, 2005.
- [9] J. Nešetřil, E. Milkova, e H. Nešetřilová, "Otakar Borůvka on minimum spanning tree problem Translation of both the 1926 papers, comments, history," *Discrete Math.*, vol. 233, pp. 3–36, 2001.
- [10] J. Moreno, *Who shall survive?: A new approach to the problem of human interrelations*. Washington: Nervous and Mental Disease Publishing Co., 1934.
- [11] P. Erdős, "On random graphs I," *Publ. Math. Debrecen*, 1959.
- [12] P. Erdős e A. Rényi, "On the evolution of random graphs," *Magy. Tud. Akad. Mat. Kut. Int. Közl.*, 1960.

- [13] J. Grossman, "The Erdős Number Project." [Online]. Available: <http://www.oakland.edu/enp/>. [Accessed: 24-Jul-2013].
- [14] J. Travers e S. Milgram, "An experimental study of the small world problem," *Sociometry*, 1969.
- [15] D. J. Watts e S. H. Strogatz, "Collective dynamics of 'small-world' networks.," *Nature*, vol. 393, no. 6684, pp. 440–2, Jun. 1998.
- [16] V. E. Krebs, "Uncloaking Terrorist Networks," *First Monday*, vol. 7, no. 4, pp. 1–12, 2002.
- [17] V. Hougham, "Sociological Skills Used in the Capture of Saddam Hussein." [Online]. Available: <http://www.asanet.org/footnotes/julyaugust05/fn3.html>. [Accessed: 13-May-2013].
- [18] N. Committee on Network Science for Future Army Applications, *Network Science*. 2006.
- [19] B. Bollobás, "Random Graphs." Cambridge, 2001.
- [20] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, e D. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, no. 4–5, pp. 175–308, Feb. 2006.
- [21] E. Bender e E. R. Canfield, "The Asymptotic Number of Labeled Graphs Degree Sequences," vol. 296301, pp. 296–307, 1978.
- [22] A.-L. Barabási e R. Albert, "Emergence of Scaling in Random Networks," *Sci. Mag.*, vol. 286, no. October, pp. 509–512, 1999.
- [23] P. E. Hart, N. J. Nilsson, e B. Raphael, "A formal basis for the heuristic determination of Minimum Cost Paths," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, no. 2, pp. 100–107, 1968.
- [24] F. Harary e E. Palmer, *Graphical enumeration*. London, 1973.
- [25] R. Burt, *Neighbor networks: Competitive advantage local and personal*. 2010.

- [26] F. Bass, "A new product growth model for consumer durables," *Manage. Sci.*, vol. 15, no. 5, pp. 215–227, 1969.
- [27] S. P. Borgatti e M. G. Everett, "A Graph-theoretic perspective on centrality," *Soc. Networks*, vol. 28, no. 4, pp. 466–484, Oct. 2006.
- [28] H. Jeong, S. P. Mason, A. L. Barabási, e Z. N. Oltvai, "Lethality and centrality in protein networks.," *Nature*, vol. 411, no. 6833, pp. 41–2, May 2001.
- [29] U. Brandes, "A faster algorithm for betweenness centrality*," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, Jun. 2001.
- [30] V. Latora e M. Marchiori, "A measure of centrality based on network efficiency," *New J. Phys.*, vol. 9, no. 6, pp. 188–188, Jun. 2007.
- [31] L. Katz, "A New Status Index Derived from Sociometric Index," *Psychometrika*, pp. 39–43, 1953.
- [32] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification.," *J. Math. Sociol.*, vol. 2, pp. 113–120, 1972.
- [33] M. Newman, *Networks: an introduction*. 2009.
- [34] D. Ford and D. Fulkerson, *Flows in networks*. 1962.
- [35] C. Papadimitriou e K. Steiglitz, *Combinatorial optimization: algorithms and complexity*. New York: Dover Publications Inc., 1998.
- [36] L. C. Freeman, S. P. Borgatti, e D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," *Soc. Networks*, vol. 13, no. 2, pp. 141–154, Jun. 1991.
- [37] V. Latora e M. Marchiori, "Efficient Behavior of Small-World Networks," *Phys. Rev. Lett.*, vol. 87, no. 19, p. 198701, Oct. 2001.

- [38] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Phys. Rev. E*, vol. 64, no. 1, p. 016132, Jun. 2001.
- [39] A. Clauset, M. E. J. Newman, e C. Moore, "Finding community structure in very large networks," pp. 1–6.
- [40] M. E. J. Newman, "Modularity and community structure in networks.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 23, pp. 8577–82, Jun. 2006.
- [41] U. Kang, "Centralities in Large Networks: Algorithms and Observations."
- [42] R. Luce, "Connectivity and generalized cliques in sociometric group structure," *Psychometrika*, vol. 15, pp. 169–190, 1950.

Índice Remissivo

- aglomeração hierárquica, 47
- agrupamento, 27
- algoritmo
 - de Girvan-Newman, 47, 48, 61
 - de simulao de ataque, 57, 59, 60, 61
- aresta, 12
- ataque, 31
- bloco, 14
- caminho, 14
- centralidade, 21, 30, 32, 55, 58, 59, 61
 - absoluta, 34
 - de autovetor, 41, 57
 - de centroide, 42
 - de fluxo, 43
 - de grau, 33, 34, 35, 57
 - de informao, 44
 - de intermediao, 33, 34, 37, 47, 57, 59, 61
 - de Katz, 40, 47
 - de k-passeio, 39
 - de proximidade, 33, 34, 36, 57
 - delta, 43, 44
 - relativa, 35
- centralizao, 32
- ciclo, 14, 24
- clique, 14
- componente, 14, 58, 59
 - gigante, 24
 - comunidade, 27, 28, 46, 49, 50
- conexo preferencial, 24
- corte, 43, 46
- cyberataque, 54, 61
- difuso, 31
- distncia, 14
- eficincia, 44
- Erds, nmero de, 19
- falha, 31
- fenmeno do mundo pequeno, 21, 23
- fluxo, 43
- geodsica, 14, 37
- grafo
 - aleatrio, 19, 23, 29, 49
 - centro, 14
 - completo, 14
 - conexo, 14
 - definio, 12
 - desconexo, 14, 36
 - dimetro, 14
 - ER. *Consulte* grafo aleatrio
 - estrela, 37
 - no-orientado, 13
 - orientado, 13, 36
 - peso, 13
 - simples, 13, 35
- grau, 13, 33
 - de entrada, 13
 - de sada, 13, 42

distribuição de, 26

lista

- de adjacência, 15
- de incidência, 15

loop, 12, 24

matriz

- convergente, 16
- de adjacência, 13, 35, 38
- de incidência, 15
- de pesos, 47
- diagonal, 15
- geodésica, 15
- laplaciana, 15

mediação, 30

modelo

- da configuração, 24
- de Bass, 31

modularidade, 49, 51, 61

motif, 28, 29, 30

multigrafo, 13, 24

n-clique, 14, 27, 46

PageRank, 22, 41, 42, 61

passeio, 13, 39, 40

ponte, 14

ponto de articulação, 14, 30, 37

rede

- confiabilidade, 54
- IP, 26, 45
- vulnerabilidade, 54, 61

vértice, 12