

Capítulo

1

Redes Complexas na Modelagem de Redes de Computadores

Marcelo G. Almiron, Heitor S. Ramos, Eduardo M. R. Oliveira, João G. M. de Menezes, Daniel L. Guidoni, Pedro O. Stancioli, Felipe D. da Cunha, André L. L. de Aquino, Raquel A. F. Mini, Alejandro C. Frery e Antonio A. F. Loureiro

Abstract

The present short course aim at presenting the theory of complex networks applied to computer networks models. This theory has been successfully employed in a wide variety of areas in which systems can be suitably represented by structures formed by interconnected elements, namely networks. Network models are present in social, information and biological sciences and in technological areas. Thus, concepts related to characterization measurements, complex network models, and statistical methods necessary to understand and explain the computer networks behavior will be presented.

Resumo

O presente minicurso pretende abordar a teoria de redes complexas aplicada aos modelos de redes de computadores. Essa teoria tem sido aplicada em diversas áreas do conhecimento em que os sistemas podem ser representados por estruturas formadas por elementos interconectados. Essas estruturas são conhecidas como “redes” e estão presentes nas ciências sociais, ciência da informação, ciências biológicas e em áreas ligadas à tecnologia. Nesse sentido serão apresentados conceitos de medidas de caracterização, modelos de redes complexas e métodos estatísticos necessários para o entendimento do comportamento das redes de computadores.

1.1. Introdução

Em contraste com os estudos anteriores que focavam em grafos pequenos e se interessavam em propriedades individuais dos seus vértices e arestas, estudos recentes presentes na literatura especializada reportam a análise estatística de propriedades de grafos de larga

escala (Newman 2003). Nesses estudos se destacam a presença de grafos que representam redes com centenas a milhões e até mesmo bilhões de vértices. Esse aumento de escala leva à necessidade da criação de novas técnicas e teorias, em particular as redes complexas, para melhor entendimento desses novos cenários (Newman 2003)

Além da escala, a descoberta de algumas propriedades como o *small world* e *scale free* que são capazes de caracterizar diversas redes reais, têm impulsionado um crescente interesse de pesquisa em diversas áreas de aplicação (Wang & Chen 2003). No contexto de redes de comunicação, em especial as redes de computadores, este interesse pode ser observado em uma recente chamada de trabalho para uma edição especial da revista *Computer Communications* da editora Elsevier (*Special Issue of Computer Communications on Complex Networks*) em nome do editor associado Matthieu Latapy, pesquisador senior do laboratório LIP6 da Universidade Pierre and Marie Curie em Paris, França. Para esta chamada, foram enviados trabalhos até 7 de fevereiro de 2010 e a publicação está prevista para meados de setembro de 2010. Nesta chamada, os editores estimulam a submissão de trabalhos relacionados à modelagem através da teoria de redes complexas nas seguintes áreas:

- roteamento na Internet (IP, Sistemas autônomos (AS), ou outros níveis);
- grafos que representam as aplicações Web, como por exemplo as ligações entre blogs;
- redes ponto a ponto (P2P), *overlays* e *exchanges*;
- correio eletrônico e outros tipos de comunicação entre os usuários;
- redes sociais on-line como o Facebook ou Flickr;
- processo de disseminação de vírus e *worms*;
- grafos dinâmicos que representem mobilidade e redes de sensores sem fio.

Esta lista não é restritiva e dá uma boa idéia da potencial de aplicabilidade da teoria das redes complexas na área de comunicação de dados. Os editores vislumbram que o estudo de propriedades não triviais extraídas dos grafos que representam essas redes pode trazer profundas reflexões na área de comunicação de dados e que vários esforços têm se concentrado nas medições, análises modelagem e outros aspectos dessas redes. Segue a transcrição de parte do texto dessa chamada de trabalhos.

“It is now clear that the study of such objects brings deep insight on communications, and much effort is done in their measurement, analysis, modeling, and other issues. Significant results have been obtained, and this area of research is nowadays very active”.

1.1.1. Redes complexas e redes reais

Diversos fenômenos encontrados na natureza podem ser representados através da teoria de redes complexas. Esses sistemas apresentam características em comum, por exemplo, eles podem ser representados através de grafos, isto é, vértices e arestas representam a interconexão entre os diversos elementos desses sistemas. As redes sociais podem ser modeladas dessa maneira de modo que as pessoas são representadas por vértices e as relações que elas estabelecem são representadas por arestas. São exemplos desses sistemas as redes de relacionamento na Internet como o Facebook e o Flickr, entre outros, as redes de relacionamentos das organizações e as redes de citações de trabalhos científicos. Da mesma maneira, redes de comunicação como e-mail, redes telefônicas, redes sem fio, Internet e a Web também têm sido estudadas com o mesmo ferramental oferecido pelas redes complexas. Em aplicações em diversas áreas do conhecimento, como por exemplo as redes de interações entre proteínas, estruturas do cérebro e redes de transportes também são objetos de aplicação da teoria das redes complexas. A vasta amplitude de aplicações em diversas áreas do conhecimento sugere que as redes complexas apresentam um grande potencial a ser explorado e é uma área de pesquisa bastante ativa atualmente (Costa et al. 2007, 2008).

Uma importante característica comum aos sistemas estudados pela teoria das redes complexas, relatada no trabalho de Costa et al. (2007), é que tais sistemas não são completamente aleatórios mas possuem uma arquitetura um pouco mais estruturada. Por exemplo, a topologia de redes completamente diferentes como as interações entre proteínas e a Internet apresentam um padrão de comportamento semelhante: apresentam uma estrutura conhecida por ‘livre de escala’ (*scale-free networks*). Dessa maneira, redes formadas por estruturas completamente diferentes apresentam características topológicas semelhantes. Baseados neste fato, Costa et al. (2007) apresentam um desafio muito interessante para estudo das redes complexas: encontrar as leis fundamentais que geram, modelam e caracterizam tais redes para auxiliar o entendimento do comportamento de sistemas complexos.

No sentido de apresentar as bases para o estudo do comportamento das redes de computadores através das redes complexas o presente minicurso está estruturado como segue. Na seção 1.2 é apresentada uma breve revisão da teoria dos grafos, base do estudo das redes complexas, e a definição da notação empregada no resto deste texto. A seção 1.3 descreve as principais medidas da teoria das redes complexas, úteis para a caracterização de diversos comportamentos dos sistemas complexos. Na seção 1.4 os principais modelos oriundos da teoria das redes complexas estão definidos. Tais modelos como o *small world* e *scale free* são úteis para a caracterização da geração e do comportamento das redes complexas. A seção 1.5 apresenta o ferramental estatístico necessário para o estudo e análise do comportamento de redes reais através da teoria de redes complexas, bem como diversos exemplos da aplicação dessa teoria no campo das redes de computadores. Por fim, a seção 1.6 traz as considerações finais deste trabalho.

1.2. Revisão da teoria de grafos

Um grafo é um formalismo matemático que serve para representar objetos e relações entre eles. Esta simples estrutura encontra-se em uma grande diversidade de aplicações como

circuitos elétricos, estradas, vários tipos de redes, ecossistemas, relações sociais, interação molecular, bases de dados, estruturas de controle de programas e sistemas de permissões de sistemas de segurança, são alguns exemplos da grande variedade de aplicações da teoria dos grafos.

Esta área tem seu foco nas relações de propriedades topológicas de grafos com aquelas derivadas de matrizes de representação ou caracterizações dos mesmos.

Na continuação vamos definir formalmente um grafo, para logo descrever suas características, e finalmente definir grupos de grafos com propriedades que são de especial interesse para este trabalho.

Um tipo de grafo bastante utilizado para modelagem de redes de computadores é o grafo direcionado simples, definido a seguir.

Definição 1.2.1 (Grafo direcionado simples). Um grafo direcionado simples é uma dupla $\vec{G} = (V, \vec{E})$ onde V é o conjunto finito de vértices, e \vec{E} é o conjunto de arestas. Cada aresta é um par (u, v) de vertices em V , com $u \neq v$.

O termo “simples” indica que não é permitido expressar que um vértice mantenha relacionamento com ele mesmo. Em diversos tipos de redes, como as redes de sensores sem fio, isto faz muito sentido já que não estamos interessados em modelar autocomunicação pelo meio. Um caso particular deste tipo de grafos, é o grafo não direcionado simples. A única diferença com o grafo definido na definição 1.2.1, é que as arestas são representadas como um conjunto, e não como um par ordenado de vértices.

Definição 1.2.2 (Grafo não direcionado simples). Um grafo não direcionado simples é uma dupla $G = (V, E)$ onde V é o conjunto finito de vértices, e E é o conjunto de arestas. Cada aresta é um conjunto $\{u, v\}$ de vertices em V , com $u \neq v$.

Neste grafo não direcionado simples, cada aresta $\{u, v\}$ não direcionada, representa a presença das arestas direcionadas (u, v) e (v, u) no grafo direcionado simples. A figura 1.1 apresenta estes tipos de grafos. Na figura 1.1(a), o gráfico corresponde ao grafo direcionado $\vec{G}_1 = (V, \vec{E})$ com $V = \{a, b, c, d, e\}$ e $\vec{E} = \{(a, c), (a, e), (b, d), (c, d), (d, e)\}$. Já na figura 1.1(b), o gráfico representa o grafo não direcionado $G_2 = (V, E)$ com $V = \{f, g, h, i, j\}$ e $E = \{\{f, i\}, \{g, j\}, \{h, i\}\}$.

Uma das representações mais utilizadas para grafos, tanto direcionados quanto não direcionados, é a matriz de adjacência. Este conceito é fundamental para continuar construindo definições e propriedades sobre grafos.

Definição 1.2.3 (Matriz de adjacência). A matriz de adjacência é uma representação de grafos. Para um grafo $\vec{G} = (V, \vec{E})$, é uma matriz $A(\vec{G})$ de dimensão $|V| \times |V|$, definida por

$$A(\vec{G})[i, j] = \begin{cases} 1 & \text{se } (i, j) \in \vec{E}, \\ 0 & \text{caso contrario.} \end{cases}$$

Nesta definição, e no resto do texto, expressões do tipo “ $|V|$ ” representam a cardinalidade do conjunto V . É importante esclarecer que quando dizemos que a matriz de

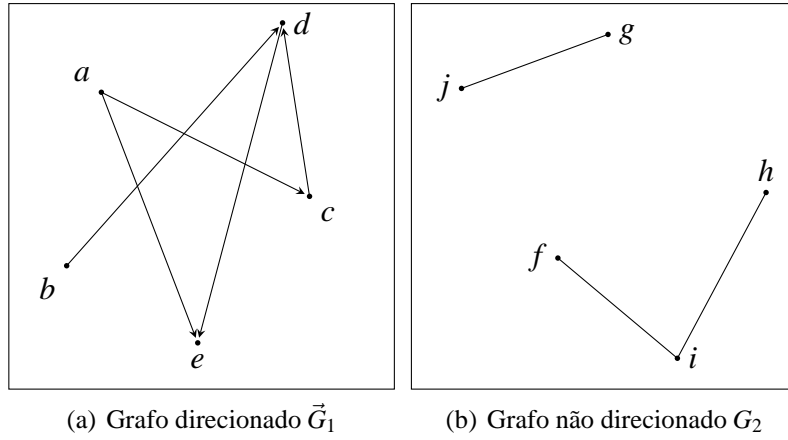


Figura 1.1. Representação gráfica de grafos simples

adjacência é uma representação, estamos afirmando que o grafo \vec{G} e sua matriz de adjacência A contêm a mesma informação, isto é, são equivalentes em conteúdo. Baseados no grafo da figura 1.1(a), construímos a matriz de adjacência correspondente e obtemos

$$A(\vec{G}_1) = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Neste caso, para indexar, utilizamos como exemplo a função de ordem lexicográfico $I_1(x): V \rightarrow \mathbb{N} \cup 0$, isto é $I_1(a) = 0$, $I_1(b) = 1$, $I_1(c) = 2$, $I_1(d) = 3$ e $I_1(e) = 4$. Agora criamos uma nova função de indexação $I_2(x)$, definida por extensão da seguinte maneira: $I_2(f) = 0$, $I_2(g) = 1$, $I_2(h) = 2$, $I_2(i) = 3$ e $I_2(j) = 4$. A matriz de adjacência associada ao grafo da figura 1.1(b) é a seguinte:

$$A(G_2) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Para denotar as relações diretas que um vértice mantém com outros, vamos introduzir os conceitos de adjacência e grau de um vértice.

Definição 1.2.4 (Grafo valorado). Grafo valorado é um grafo em que cada aresta tem um valor (peso) associado. Formalmente, um grafo valorado $G = (V, E)$ consiste de um conjunto V de vértices, um conjunto E de arestas e uma função $\Pi: E \rightarrow \mathbf{P}$.

Utilizamos \mathbf{P} (conjunto de pesos) de maneira genérica para representar o contradomínio da função Π . O conjunto \mathbf{P} pode ser convenientemente especificado para o conjunto dos números reais (\mathbb{R}) ou inteiros (\mathbb{Z}), por exemplo.

Para grafos valorados, a matriz de adjacência pode apresentar valores diferentes de 0 e 1 como apresentado na definição 1.2.3. Para esta situação os valores presentes na matriz serão atribuídos de acordo com a função Π para a respectiva aresta.

Definição 1.2.5 (Subgrafo). Um grafo $H = (V', E')$ é dito ser um subgrafo de $G = (V, E)$ se e somente se:

- cada vértice de H é também um vértice de G , ou seja, $V' \subseteq V$;
- cada aresta de H é também uma aresta de G , ou seja, $E' \subseteq E$; e
- cada aresta de H tem os mesmos nós terminais em G , ou seja se $(u, v) \in E'$ então $(u, v) \in E$.

Definição 1.2.6 (Adjacência e grau). Num grafo direcionado $\vec{G} = (V, E)$, o vértice v é adjacente a u se $A[u, v] = 1$. O número de vértices adjacentes a u chama-se grau de u , e diferenciamos entre grau de saída (*output*)

$$k_u^{\text{out}}(\vec{G}) = \sum_{v \in V: v \neq u} A[u, v]$$

e grau de entrada (*input*)

$$k_u^{\text{in}}(\vec{G}) = \sum_{v \in V: v \neq u} A[v, u].$$

Denotamos o grau de u simplesmente por $k_u(\vec{G}) = k_u^{\text{out}}(\vec{G}) + k_u^{\text{in}}(\vec{G})$.

Definição 1.2.7 (Vizinhança). A vizinhança de um vértice u é definida como $N_u = \{u \mid e_{uv} \in E\}$ e $k_u = |N_u|$.

Os grafos “conexos” apresentam propriedades interessantes para as redes de computadores e para definí-los é necessário, primeiro, introduzir o conceito de caminho simples. Notemos que até agora, definindo conceitos para grafos direcionados simples, estamos também incluindo a definição para grafos não direcionados, já que os grafos direcionados são mais gerais.

Definição 1.2.8 (Caminho simples). O caminho simples entre v_i e v_j , vértices do grafo direcionado $\vec{G} = (V, \vec{E})$, é qualquer sequência de arestas

$$(v_i, v_{ij_1}), (v_{ij_1}, v_{ij_2}), \dots, (v_{ij_{n-2}}, v_{ij_{n-1}}), (v_{ij_{n-1}}, v_j),$$

que denotaremos $v_i \rightsquigarrow v_j$. Cada v_{ij} denota um vértice intermediário do caminho desde v_i até v_j . Aqui, o comprimento do caminho é n .

Note que os nós intermediários v_{ij} são irrelevantes para a definição de caminho simples. No sentido mais específico, vamos definir um grafo conexo para grafos não direcionados como segue.

Definição 1.2.9 (Grafo conexo). Um grafo não direcionado simples $G = (V, E)$ é dito conexo se existe um caminho simples $u \rightsquigarrow v$ de qualquer vértice $u \in V$ a qualquer vértice $v \in V$, com $v \neq u$.

Esta definição pode se aplicar a grafos direcionados simples, mas temos que fazer primeiro uma transformação. Vamos dizer que um grafo direcionado simples \vec{G} é conexo se sua “bi-orientação” é um grafo não direcionado simples G conexo. A bi-orientação é uma função sobre a matriz de adjacência que para cada elemento $A[i, j] = 1$, atribui adjacência para o seu par simétrico, isto é, $A[j, i] = 1$. Outro conceito de interesse é a conectividade forte em grafos direcionados simples, que definimos a seguir:

Definição 1.2.10 (Grafo fortemente conexo). Um grafo direcionado simples $G = (V, E)$ é dito fortemente conexo se existe um caminho simples $u \rightsquigarrow v$ de qualquer vértice $u \in V$ a qualquer vértice $v \in V$, com $v \neq u$.

Dentro da teoria espectral de grafos uma definição básica é a de característica polinomial. A partir deste conceito é que vamos definir os autovalores de um grafo, definição necessária para introduzir a definição principal desta seção: a conectividade algébrica.

Definição 1.2.11 (Característica polinomial). A característica polinomial de um grafo \vec{G} é o determinante $\det(xI - A(\vec{G}))$, onde I é a matriz identidade de dimensão $|A| \times |A|$, e x uma variável real.

Quando expandimos o determinante da definição 1.2.11, vemos que a característica polinomial é um sistema de equações lineares. Resolvendo este sistema, podemos achar os autovalores do grafo, definidos a continuação.

Definição 1.2.12 (Autovalores). Os autovalores de um grafo $\vec{G} = (V, \vec{E})$ são as raízes da característica polinomial.

Notemos que para um grafo $\vec{G} = (V, \vec{E})$, existem exatamente $|V|$ autovalores, como veremos em breve. Estes autovalores determinam o espectro do grafo definido da seguinte maneira:

Definição 1.2.13 (Espectro). O espectro de um grafo $\vec{G} = (V, \vec{E})$ é o multiconjunto de autovalores associados. Seja $|V| = m$, então existem m autovalores associados a \vec{G} .

O espectro de um grafo fornece caracterizações do mesmo. Mas é por meio do espectro do laplaciano do grafo que obtemos a caracterização de maior interesse neste trabalho: a conectividade algébrica. Com o laplaciano de um grafo é possível determinar se um grafo é conexo apenas com operações matriciais. Porém não se conhece ainda muito sobre as propriedades espectrais do laplaciano de um grafo, trabalhos relativamente novos nesta área sugerem que há muito mais por descobrir (ver Gross & Yellen 2003, seção 6.5.6). A definição do laplaciano de um grafo é introduzida da seguinte maneira:

Definição 1.2.14 (Laplaciano de um grafo). O laplaciano L de um grafo \vec{G} define-se como $L = D - A(\vec{G})$, onde D é a matriz diagonal de graus dos vertices de \vec{G} .

O segundo menor autovalor do laplaciano de um grafo fornece informação útil a respeito da conectividade do grafo. Antes de conhecer estas propriedades, vamos definir formalmente a conectividade algébrica introduzida por Fiedler (1973).

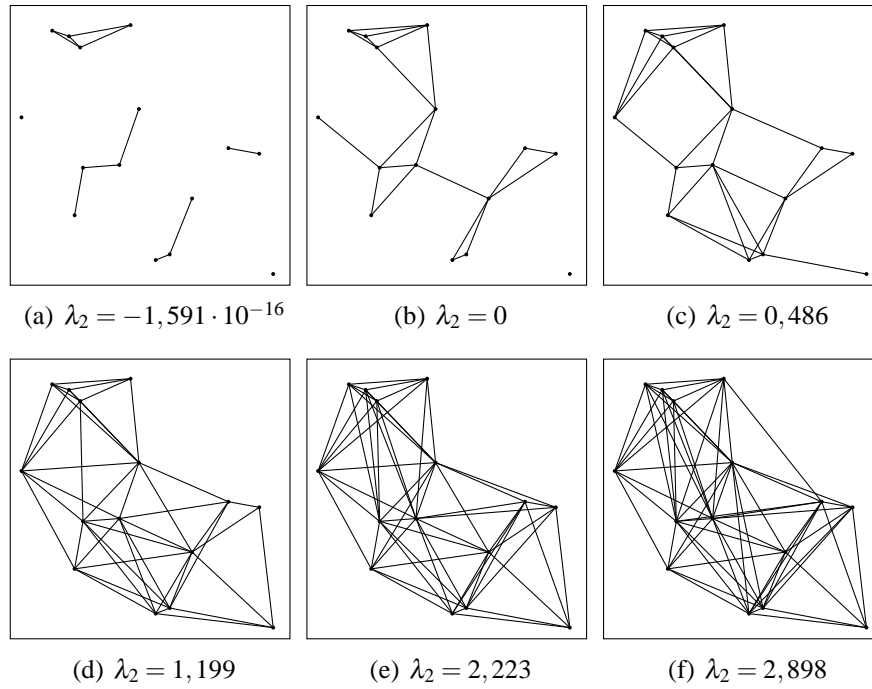


Figura 1.2. Comportamento da conectividade algébrica

Definição 1.2.15 (Conectividade algébrica). A conectividade algébrica do grafo $\vec{G} = (V, \vec{E})$ com laplaciano L de autovalores $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ é λ_2 .

Como antecipamos, a conectividade algébrica é o segundo menor autovalor do laplaciano associado ao grafo.

O principal resultado a respeito da conectividade algébrica é apresentado no seguinte teorema.

Teorema 1.2.1. A conectividade algébrica é positiva ($\lambda_2(L_{\vec{G}}) > 0$) se e somente se o grafo \vec{G} é fortemente conexo.

Uma característica interessante deste teorema (a prova encontra-se em Fiedler 1973) é que permite determinar se um grafo é conexo ou não apenas com operações matriciais. Para ferramentas como R e Ox, isto representa uma vantagem, já que as operações matriciais estão implementadas com eficiência. Por outro lado, deveríamos analisar o impacto da precisão numérica nos resultados obtidos. Por exemplo, para um grafo conexo, a conectividade algébrica poderia ser 10^{-20} . Um valor negativo muito próximo de zero, implicaria que não teremos classificado o grafo corretamente segundo a propriedade de ser conexo ou não. Porém a conectividade algébrica tem um significado mais forte. Quanto menor o seu valor, “mais longe” está o grafo de ter conectividade, e quando o valor cresce passando a ser positivo, quanto maior for, “mais forte” é sua conectividade, isto é, a eliminação de algumas arestas do grafo não tira a propriedade do mesmo ser conexo. A figura 1.2 mostra o comportamento da conectividade algébrica segundo o grafo.

1.3. Caracterização de redes complexas

Nessa seção serão apresentadas diversas medidas que tipicamente são utilizadas nos estudos que envolvem a teoria das redes complexas. Essas medidas são capazes de expressar as características topológicas mais relevantes das redes. Neste trabalho foi feita uma classificação das medidas em: medidas relacionadas à distância, medidas de centralidade, medidas espectrais, medidas egocêntricas, medidas de identificação de comunidades e medidas de hierarquia. A lista aqui apresentada não esgota todas as medidas presentes na literatura mas dá uma boa visão das medidas mais frequentemente utilizadas. O trabalho de Costa et al. (2007) é uma referência bastante abrangente sobre esse tema.

1.3.1. Medidas relacionadas à distância

De uma maneira geral, distância está associada a quantidade de arestas contidas em um caminho que conecta dois vértices u e v , denotada por $d_{u,v}$. Essa definição é válida para grafos não valorados. Observe que para grafos dirigidos em geral $d_{u,v} \neq d_{v,u}$, dado que o caminho em um sentido é, em geral, diferente do caminho no sentido oposto. Um caminho geodésico, ou caminho mínimo (*shortest path*) é qualquer caminho que interligue os vértices u e v e tenha distância mínima, denotada por d_G . Nota-se que podem existir diversos caminhos mínimos entre quaisquer dois vértices em um grafo. O conceito de distância pode ser facilmente estendido para grafos valorados, levando em consideração o peso associado às arestas. Quando não há um caminho entre dois vértices u e v atribui-se distância infinita, ou seja, $d_{u,v} = \infty$.

Uma medida relacionada à distância que aparece com muita frequência no contexto de redes complexas é o comprimento médio dos caminhos mínimos, ou, em inglês, *average path length* (L) do grafo $G = (V, E)$, definido por:

$$L = \frac{1}{n(n-1)} \sum_{i \in V} \sum_{j \in V} SP_{ij},$$

onde $n = |V|$ é o número de vértices do grafo e SP é o comprimento do caminho geodésico entre os vértices i e j .

Para evitar que esta soma divirja, ou seja, a soma acima ir para infinito, apenas vértices que estão conectados são contabilizados. Mais formalmente, $i, j \in V'$, onde $V' = \{u \in V \mid \exists u \rightsquigarrow v \in V, u \neq v\}$. Esta medida tem uma importância destacada na teoria de redes complexas pois é utilizada para caracterização das redes *small world*.

Algumas outras medidas estão relacionadas à caminhos geodésicos mas, neste trabalho, foram consideradas como medidas de centralidade que estão apresentadas na próxima seção.

1.3.2. Medidas relacionadas a agrupamentos e ciclos

Nesta seção serão apresentadas algumas medidas que caracterizam a tendência de algumas redes reais a formarem conjuntos de vértices agrupados.

Uma medida que caracteriza agrupamento e que possui um destaque especial na teoria de redes complexas é o coeficiente de agrupamento (*clustering coefficient*) para um vértice u , também conhecido por transitividade. Para grafos não dirigidos, é definido

como

$$CC_u = \frac{2|\{e_{vw}\}|}{k_u(k_u - 1)},$$

onde $v, w \in N_i$, $e_{vw} \in E$ e $k_u(k_u - 1)/2$ é o número máximo de arestas entre vértices na vizinhança de u . Essa medida também pode ser utilizada na caracterização da propriedade *small world*.

O coeficiente de agrupamento da rede é então definido como

$$CC = \frac{1}{|V|} \sum_{u \in V} CC_u.$$

Para grafos dirigidos, o coeficiente de agrupamento é um conceito mais complexo e foge do escopo desse trabalho. Algumas propostas para tal estão presentes no trabalho de Costa et al. (2007).

Uma outra medida é o coeficiente do “clube dos ricos” (*Rich-club coefficient*). Essa medida foi proposta por Zhou & RJ (2004) ao perceberem que a topologia da Internet obedece a um padrão que os *hubs* estão bem conectados entre si. Esse mesmo padrão em redes de citação científica onde pesquisadores influentes tendem a formarem grupos e publicarem artigos conjuntamente. A medida proposta por Zhou & RJ (2004) captura esse comportamento e é definida como

$$\phi(g) = \frac{1}{|\mathcal{R}(g)|(|\mathcal{R}(g)| - 1)} \sum_{i,j \in \mathcal{R}(g)} a_{i,j},$$

onde $\mathcal{R}(g)$ é o conjunto que denota o “clube dos ricos” de grau g na rede representada pelo grafo $G = (V, E)$, ou seja, $\mathcal{R}(g) = \{v \in V \mid k_v > g\}$, e a_{ij} é o valor correspondente aos vértices i e j na matriz de adjacência de G . Essa medida é semelhante ao coeficiente de agrupamento para a fração de vértices que apresentam grau maior que g .

1.3.3. Medidas de centralidade

Métricas de centralidade são utilizadas para quantificar a intuição de que, sob algum aspecto, as redes possuem elementos posicionados mais ao centro de sua estrutura do que outros. Na teoria dos grafos, a idéia de classificar um vértice por sua centralidade estrutural foi introduzida em 1948 por Alex Bavelas, resultando na proposição do primeiro índice de centralidade para redes conectadas, o índice de Bavelas.

Medidas de centralidade têm o intuito de estimar a importância de um dado vértice, ou seja, ranqueá-lo segundo sua importância topológica. Nós em posições centrais geralmente têm grande importância estrutural e, em casos onde há fluxo de dados, são agentes de escoamento vitais. Geralmente, quanto maior for a participação de um vértice ou de uma aresta em caminhos no grafo, maior será sua importância. Desta forma, pode-se saber o quão importante é, por exemplo, um elemento computacional para uma rede, ou uma pessoa em uma rede social, inferindo-se esta importância a partir da posição desta entidade frente às demais. É importante ressaltar que o termo centralidade está intimamente ligado à perspectiva das características consideradas pela métrica, pois, por

exemplo, um elemento pode ser considerado mais central por estar ligado a outros elementos importantes, ou por estar mais próximo de todos os outros nós, ou ainda por estar presente em muitos caminhos quando se percorre o grafo.

Existem diversas métricas de centralidade baseadas em diferentes características do grafo como, por exemplo, os conceitos de distância entre nós, ou do grau do nós são amplamente explorados pelas métricas *closeness*, *degree centrality*, *eccentricity* e *centroid*. Além disto, outra característica base usada pelas métricas de centralidade são os menores caminhos do grafo, utilizados, por exemplo, pelo *shortest-path betweenness centrality*, também conhecido como *betweenness*, *stress centrality* e *reach*. Existem ainda métricas que utilizam a importância dos vértices de sua vizinhança para o cálculo da importância própria como é o caso da métrica *eigenvector*, ou ainda as métricas correlatas de classificação de páginas da *World-Wide Web* *hub Score* e *authority*. Ainda no contexto da *Web* uma métrica de centralidade amplamente estudada é o *PageRank*TM, que é a base da máquina de busca do Google.

A utilização de métricas de centralidade no projeto de algoritmos no contexto de comunicação em redes de computadores vem sendo assunto de estudos recentes na literatura como, por exemplo, a possibilidade de se utilizar a centralidade como forma de se balancear carga e assim evitar a sobrecarga de nós mais centrais. A centralidade também pode ser utilizada para diminuir a latência empregando um tráfego de dados ciente da centralidade dos nós (Krause et al. 2006).

1.3.3.1. Distância e vizinhança

Nesta seção serão apresentados índices que ranqueiam os nós por sua centralidade baseado em características básicas como a noção de vizinhança e distância dentro de um grafo.

Degree centrality É o índice de centralidade mais simples e, para grafos não direcionados, é representado como $C_D(v)$, que corresponde ao número de arestas ligadas ao vértice v . O *degree centrality* é usualmente apresentado de maneira normalizada, pois, desta forma, pode-se ter uma noção do quão distante este nó está das centralidades mínimas e máximas. Para tanto, a versão normalizada é calculada por $C_{Dn}(v) = k(v)/n - 1$, em que $k(v)$ é o grau do vértice v e n é o número de vértices do grafo. Para grafos direcionados, a métrica se divide de maneira a representar o grau de entrada, *in-Degree centrality* $C_{Di}(v)$ e o grau de saída, *out-Degree centrality* $C_{Do}(v)$. O *degree centrality* é uma métrica local, ou seja, é necessário apenas o conhecimento da quantidade de nós da vizinhança de primeiro nível (nós diretamente conectados) para que esta métrica seja calculada.

Eccentricity Esta métrica é definida como a distância máxima entre um vértice $u \in V$ e qualquer outro vértice $v \in V$ da rede, i.e,

$$Ecc(u) = \max_{v \in V} d_G(u, v).$$

Esta métrica pode ser empregada em um típico problema de localização, encontrar um vértice cuja distância máxima para todos os outros nós é mínima. Tal problema pode ser solucionado determinando-se o vértice u tal que $Ecc(u)$ seja mínimo. Em outras palavras, o diâmetro do grafo é o valor máximo de Ecc enquanto o raio é o valor mínimo.

Closeness Inicialmente, é importante denotar o somatório das distâncias geodésicas de um vértice $u \in V$ para um outro vértice qualquer $v \in V$, i.e., $\sum_{v \in V \setminus u} d_G(u, v)$. Este conceito de somatório das distâncias é empregado, por exemplo, em problemas de localização. O problema de localização mini-soma consiste na idéia de se minimizar a distância total percorrida num sistema de distribuição, ou seja, uma abordagem para a solução deste problema de localização é encontrar o conjunto de vértices com o menor valor de *closeness*.

O *Closeness* de um vértice u cresce à medida que o total da soma das distâncias diminui e é definido como a recíproca da soma das distâncias geodésicas para todos os outros vértices $v \in V$, ou seja,

$$C_{CL}(u) = \frac{1}{\sum_{v \in V \setminus u} d_G(u, v)}.$$

Esta métrica apresenta um problema quando aplicada em grafos desconexos, pois a distância entre vértices de componentes distintos é, normalmente, definida como infinito. Desta forma, o valor do *closeness* $1/\infty$ para todos os vértices será o mesmo

Em Valente & Foreman (1998) são propostas duas métricas baseadas no *closeness*: *integration*, que mede o quão bem conectado um membro da rede está e *radiality*, que mensura o quanto as ligações de um nó expandem-se pela rede. Vértices com *integration* alto ficam, mais cedo, cientes de informações trafegadas pela rede pois, na média, eles estão mais próximos dos demais vértices, enquanto vértices com valor alto de *radiality* supostamente são bons emissores de informação.

1.3.3.2. Caminhos mínimos

Nesta seção serão mostrados índices de centralidade que fazem uso dos caminhos mínimos de um grafo para definir o valor de centralidade de um vértice. Caminhos mínimos são usualmente definidos para vértices, mas podem ser aplicados também para arestas, tanto que algumas métricas que existiam apenas para centralidade de vértice, mas posteriormente foram propostas para considerar centralidade de aresta.

Stress centrality Esta é a métrica mais simples que faz uso da enumeração dos caminhos mínimos de um grafo, originalmente proposta em Shimbel (1953) para mensurar a quantidade de trabalho que um elemento precisa suportar numa rede. Para um vértice u , esta métrica é denotada pelo número de caminhos mínimos que contém u , exceto aqueles que comecem, ou terminem em u . Formalmente, o *stress centrality* para u é definido como

$$C_S(u) = \sum_{s \neq u \in V} \sum_{t \neq u \in V} \sigma_{st}(u),$$

onde $\sigma_{st}(u)$ é o número de caminhos mínimos que contém o vértice u . Analogamente, esta métrica pode ser definida para uma aresta e qualquer como:

$$C_S(e) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(e),$$

onde $\sigma_{st}(e)$ é o número de caminhos mínimos que contém a aresta e .

Shortest-path Betweenness Centrality Esta métrica, usualmente chamada apenas de *betweenness* foi proposta em Anthonisse (1971) e Freeman (1977). Basicamente é uma variação do *stress centrality* com o acréscimo da visão geral de todos os caminhos mínimos presentes no grafo. Formalmente, para um vértice u , esta métrica é definida como:

$$C_B(u) = \sum_{s \neq u \in V} \sum_{t \neq u \in V} \frac{\sigma_{st}(u)}{\sigma_{st}},$$

onde σ_{st} representa a quantidade de caminhos mínimos entre dois vértices quaisquer s e t , enquanto $\sigma_{st}(u)$ representa a quantidade de caminhos mínimos em σ_{st} que passem pelo vértice u . Assim como no *stress centrality*, os caminhos que comecem ou terminem em u são excluídos da contagem.

Ao contrário do *closeness*, o *betweenness* não possui problemas ao ser aplicado em grafos desconexos, pois a fração referente a qualquer par de vértices s e t para os quais não existe caminho será considerada 0.

O algoritmo mais eficiente conhecido o cálculo do *betweenness* de um grafo possui complexidade de espaço $O(n + m)$ e de tempo $O(nm)$ para grafos não valorados e $O(nm + n^2 \log n)$ para grafos valorados, apresentado no trabalho de Brandes (2001).

Portanto, o cálculo exato desta métrica pode se tornar uma tarefa impraticável para grafos com milhões de nós e, desta forma, para estes casos pode-se trocar precisão no resultado da métrica por uma diminuição do custo computacional para o cálculo da mesma. Para certas aplicações, o valor exato da centralidade de cada vértice não é importante, contanto que os vértices mantenham-se ranqueados por centralidade da mesma forma como estariam se o cálculo exato fosse feito. Em Brandes & Pich (2007) é proposta a idéia do uso de nós pivô como forma de se atenuar o custo computacional para o cálculo do *betweenness*. Ao contrário do algoritmo proposto em Brandes (2001) em que cada um dos nós precisa fazer uma busca em profundidade, apenas um pequeno conjunto de pivôs executam a busca e o resultado é extrapolado com a ajuda de métodos estatísticos, que posteriormente são melhorados em Geisberger et al. (2008).

1.3.3.3. Medidas aplicadas à Web

Devido ao imenso tamanho da World-Wide Web (WWW), um problema considerável é entregar a um usuário aquilo que mais se adequa aos seus interesses quando o mesmo faz uma busca. A modelagem através de grafos se ajusta à estrutura da WWW, uma vez que nós são as páginas e as arestas são modeladas como os *hyperlinks*. Desta forma, a máquina de busca pode escolher o resultado que é mais apropriado à pesquisa de um

usuário a partir de um índice de centralidade que ranqueia os vértices do grafo da WWW. Nesta seção serão apresentados os índices *PageRank*TM e *hubs & authorities*.

***PageRank*TM** Intuitivamente, o *PageRank*TM se baseia num modelo do comportamento de usuário que navega pelas páginas e, de vez em quando, desiste daquela e vai buscar outra página diferente. A probabilidade de desistência é o chamado *damping factor*. Este índice, proposto em Brin (1998), leva em consideração apenas a característica topológica da rede, ou seja, uma página terá sua importância definida apenas pela posição que ela ocupa no grafo. O *PageRank*TM é uma métrica que é definida baseada na centralidade dos vizinhos e em sua quantidade.

Inicialmente todas as páginas começam com um valor de centralidade igualmente definido entre 0 e 1. Desta forma, sem considerarmos o modelo de navegação do usuário, o *PageRank*TM de uma página u é:

$$PR(u) = \frac{1-d}{N} + d \sum_{v \in B_u} \frac{PR(v)}{L(v)},$$

em que d é o *damping factor*, N é o número de páginas, B_u é o conjunto de páginas que tem *links* apontando para u e $L(v)$ é o número de *hiperlinks* que apontam para v . Vários estudos já foram feitos para avaliar o comportamento do algoritmo variando-se o *damping factor*, mas geralmente assume-se seu valor como 0.85.

Hubs & Authorities Proposto em Kleinberg (1999) e usualmente chamado de *HITS*, este índice atribui dois valores a um vértice referentes às grandezas propostas, *hub* e *authority*. Diferentemente do *PageRank*TM, ele considera o conteúdo da página aliado à forma com que a mesma está ligado às outras no cálculo destas grandezas. Segunda a definição do autor, um bom *hub* é aquele que aponta para para muitos *authorities* bons, enquanto um bom *authority* é aquele apontado por muitos *hubs* bons

Inicialmente, todos as páginas começam com os valores de *hub* e de *authority* igual a 1. Uma vez que o cálculo do *authority* depende do valor *hub* e vice-versa, o algoritmo atualiza cada um destes valores para uma página u da seguinte forma:

$$C_{Auth}(u) = \sum_{i=1}^n C_{Hub}(i),$$

em que i é uma página conectada a u e n é o número total de páginas conectadas a u . Desta forma, o valor *authority* de uma página é a soma de todos os valores *hub* das páginas que apontam para ela. Da mesma maneira,

$$C_{Hub}(u) = \sum_{i=1}^n C_{Auth}(i),$$

em que i é uma das páginas a qual u se conecta e n é o número total de páginas a que u se conecta. Desta forma, o valor *hub* é a soma dos valores *authority* das páginas a que ela se liga. O algoritmo obtém valores que convergem após infinitas execuções (Kleinberg 1999) de cada um dos somatórios mostrados acima, um após o outro, sendo necessário, ao final, normalizar a matriz de adjacências do grafo original.

1.3.4. Medidas egocêntricas

Redes egocêntricas, ou redes de vizinhança são redes constituídas apenas de um elemento conhecido como *ego* e aqueles aos quais ele está ligado, i.e, seus vizinhos chamados *alters*. Informações tais como a maneira com que os *alters* estão conectados podem ser coletadas pelo *ego*.

A teoria que dá suporte ao conceito de redes egocêntricas supõe que a estrutura composta por *ego* e *alters* pode ser amostrada de um grande conjunto de elementos e, os resultados de análises sobre a amostra podem ser generalizados para toda a rede, tornando esse estudo atraente. Porém, se por um lado a utilização de uma pequena parte da rede composta apenas de um vértice (*ego*) e de sua vizinhança (*alters*) embute simplicidade, por outro lado, a análise dos resultados fica usualmente restrita à conectividade, ou densidade (Burt 1995).

Um ramo de estudos das redes egocêntricas é a proposição e aplicação de métricas que considerem esses aspectos. Métricas como, por exemplo, *eigenvector* e o *closeness* não são aplicáveis uma vez que elas avaliam a interação entre um *alter* e os demais elementos de toda a rede, desconsiderando assim o conjunto restrito à rede egocêntrica.

Um exemplo de uma métrica que utiliza o conceito das redes egocêntricas é apresentado a seguir.

Ego-Betweenness Métrica proposta em Everett & Borgatti (2005) que, para determinados tipos de grafos, apresenta alta correlação com o *shortest-path betweenness*, mas que apresenta custo computacional significativamente menor. Embora não quantificada no artigo, o autor mostra a existência de correlação entre o valor do *betweenness* de um vértice e o valor em sua versão egocêntrica, para este mesmo vértice.

Ao contrário do *shortest-path betweenness*, a sua versão egocêntrica é computacionalmente muito mais simples de se calcular. Para cada nó i , extrai-se a matriz de adjacências A que contém i juntamente com seus vizinhos imediatos, i.e., aqueles vértices que estão a um salto de i . Portanto, os caminhos mínimos contidos no grafo induzido por esta matriz terão tamanhos um, que ligam o *ego* aos *alters* e dois, que ligam dois *alters* passando pelo *ego*. Entre cada par de nós não adjacentes nesta matriz existe um caminho mínimo de tamanho 2, que serão os únicos considerados no cálculo desta métrica. Para se extrair o número de caminhos de tamanho dois entre i e j faz-se $A^2[1 - A]_{i,j}$. O somatório das posições da matriz que não possuem zero, dividido por dois, é o valor do *ego-betweenness* de i .

1.4. Modelos de redes complexas

A observação de padrões comuns no comportamento de redes, mesmo as de natureza diferentes, estimulou diversos pesquisadores a desenvolverem modelos que descrevem e caracterizam tais comportamentos. Nesta seção são apresentados os modelos que frequentemente aparecem nos estudos das redes de computadores como os grafos aleatórios, *small world*, *scale free*, comunidades e as redes geográficas.

1.4.1. Grafos aleatórios

Um dos modelos de redes mais simples e antigo é o modelo de grafos aleatórios Bolobás (2001), Janson et al. (1999). Grafos aleatórios foram inicialmente estudados por Solomonoff e Rapoport (Solomonoff & Rapoport 1951) e extensivamente estudado por Paul Erdős e Alfréd Rényi (Erdős & Rényi 1959, 1960). Erdős e Rényi estudaram dois diferentes modelos de grafos aleatórios. Esses dois modelos são normalmente diferenciados pela sua notação: $G_{n,m}$ e $G_{n,p}$:

- $G_{n,m}$ é o conjunto de todos os grafos aleatórios que consistem de n nós e m arestas. Para gerar uma amostra de um grafo aleatório pertencente ao conjunto $G_{n,m}$ basta adicionar m arestas aleatórias, uniformemente distribuídas, entre os n vértices inicialmente desconexos.
- $G_{n,p}$ é o conjunto de todos os grafos aleatórios que consistem de n vértices, onde cada par de vértices é conectado de acordo com uma probabilidade p . Para gerar um grafo aleatório segundo esse modelo, adiciona-se arestas entre todos os pares de vértices inicialmente desconexos segundo uma probabilidade p .

É importante observar que na geração de grafos aleatórios segundo o modelo $G_{n,m}$, a quantidade de arestas no grafo é fixa. Entretanto, a quantidade de arestas na criação de um grafo aleatório segundo o modelo $G_{n,p}$ pode variar, mas em média, o valor é o mesmo para n vértices e a probabilidade p . É importante observar que em ambos os modelos, dependendo da quantidade de arestas a ser adicionada no grafo ou da probabilidade de criação de arestas entre dois vértices, o grafo resultante pode ser desconexo. Além disso, grafos aleatórios $G_{n,p}$ possuem uma distribuição de graus binomial. Ou seja, a probabilidade p_k que um vértice qualquer seja conectado por exatamente k outros vértices é

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}.$$

Entretanto, dependendo da aplicação, é necessário saber quantas arestas são criadas no grafo inicialmente desconexo. Para isso, seja n a quantidade de vértices e p a probabilidade de criação de arestas, a quantidade de arestas criadas no grafo é, em média, $pn(n-1)/2$. Uma característica importante de grafos aleatórios é o pequeno comprimento médio dos caminhos mínimos (L) entre os pares de nós. Isso acontece porque as arestas são criadas no grafo de maneira aleatória. Entretanto, o coeficiente de agrupamento do grafo é, tipicamente, pequeno.

A figura 1.3 ilustra a criação de um grafo aleatório para diferentes valores de probabilidade. Na figura 1.3(a), quando a probabilidade de criação de arestas é 0, nenhuma aresta é criada no grafo. Quando a probabilidade de criação de arestas é 0.1 ou 0.15 (Figura 1.3(b) e Figura 1.3(c), respectivamente) as arestas são criadas de maneira aleatória entre os vértices. É interessante observar que se $p < 1$, o grafo resultante pode ser desconexo.

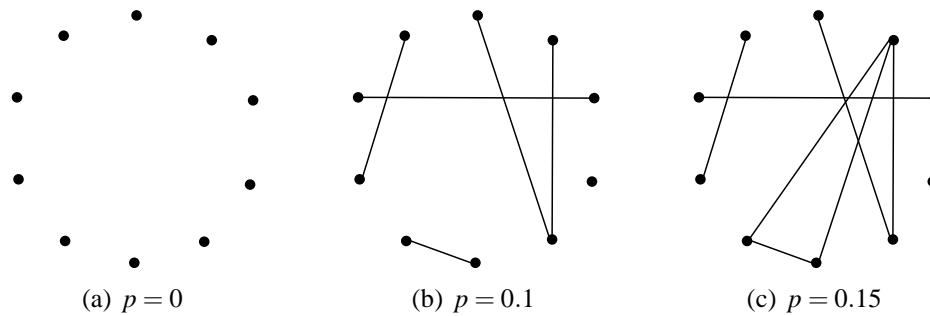


Figura 1.3. Exemplo de criação de um grafo aleatório segundo $G_{n,p}$

1.4.2. *Small world*

Redes *small world* recebem esse nome por analogia ao fenômeno *small world*, também conhecido como “seis graus de separação”. O fenômeno *small world* corresponde à hipótese de que em uma rede social a distância para conectar duas pessoas quaisquer em qualquer lugar do mundo é, em geral, pequena. Em 1967, Stanley Milgram testou experimentalmente essa hipótese e descobriu que, em média, apenas 6 pessoas separam quaisquer pessoas no mundo (Milgran 1967).

Estudos recentes definem uma forma de mensurar redes com características *small world* de forma genérica (Watts & Strogatz 1998). Para uma rede ser chamada de *small world*, ela deve possuir as seguintes características: (i) pequeno comprimento médio de caminhos mínimos entre nós e (ii) alto coeficiente de agrupamento. Entretanto, como mensurar o que é um pequeno comprimento médio de caminhos mínimos ou alto coeficiente de agrupamento? Para isso, devemos criar um grafo aleatório com as mesmas características do grafo original, em termos de número de nós e arestas. Após isso, encontre as métricas de caminho médio e mínimo e coeficiente de agrupamento de ambos os grafos. Uma rede é considerada *small world* se o comprimento médio de caminhos mínimos é próximo em relação ao grafo aleatório equivalente, entretanto, possuindo valores de coeficiente de agrupamento muito maior.

De maneira formal, seja CC e L os valores de coeficiente de agrupamento e caminho médio mínimo da rede que deseja-se saber se possui características *small world*. Além disso, seja CC_a e L_a as mesmas variáveis para o grafo aleatório equivalente ao grafo anterior. Para uma rede possuir características *small world*, as seguintes inequações devem ser verdadeiras: $CC \gg CC_a$ e $L/L_a \sim 1$. A Tabela 1.1 ilustra o caminho médio mínimo e o coeficiente de agrupamento de diferentes redes *small world* em comparação com grafos aleatórios gerados com as mesmas características das redes.

	L	L_a	CC	CC_a
Atores	3.65	2.99	0.79	0.00027
Rede de Energia	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Tabela 1.1. Exemplos empíricos de redes *small world* Watts & Strogatz (1998)

Na literatura de redes complexas, existem vários modelos de geração de redes com características *small world*. Como exemplo temos os modelos de Watts & Strogatz (Watts & Strogatz 1998), Newman & Watts (Newman & Watts 1999) e o modelo de Kleinberg (Kleinberg 2000). Para gerar uma rede com características *small world* nesses modelos, parte-se de um grafo regular (grafos em que os nós apresentam grau constante). Grafos regulares apresentam altos comprimentos médios dos caminhos mínimos e coeficiente de agrupamento. A idéia é adicionar atalhos na rede, diminuindo o comprimento médio dos caminhos mínimos entre os nós mantendo altos valores de coeficiente de agrupamento.

Watts & Strogatz Model O primeiro modelo para criação de uma rede com características *small world* foi proposto por Watts & Strogatz. Nesse modelo, parte-se de um grafo regular e, para cada aresta do grafo, a aresta em questão é reposicionada no grafo de maneira aleatória segundo uma probabilidade pré-definida, mantendo um dos seus pontos finais. Dessa forma, a quantidade de arestas do grafo resultante é a mesma. Entretanto, o grau de cada vértice se altera. Os autores mostraram que para valores pequenos de probabilidade uma rede regular pode ser transformada em uma rede com características *small world*. Quando a probabilidade de reposicionamento das arestas é grande, o grafo regular é transformado em um grafo aleatório, possuindo pequenos valores de comprimento médio dos caminhos mínimos e coeficiente de agrupamento.

A Figura 1.4 ilustra a criação de uma rede com características *small world*. Quando $p = 0$ (Figura 1.4-(a)), nenhuma aresta do grafo original é reposicionada e têm-se um grafo regular. Quando a probabilidade de reposicionamento de arestas possui valores $0 < p \ll 1$ (Figura 1.4-(b)), pode-se perceber que algumas arestas do grafo foram reposicionadas, servindo como atalhos na rede. Nesse caso, o caminho médio mínimo entre os nós diminui, mantendo altos valores de coeficiente de agrupamento. É importante observar que quando uma aresta é reposicionada, um de seus pontos finais é mantido e o outro, é reposicionado de maneira aleatória na rede. Quando a probabilidade de reposicionamento de arestas é próxima de 1, o grafo apresenta características de grafos aleatórios, devido ao grande número de arestas reposicionadas (Figura 1.4-(c)).

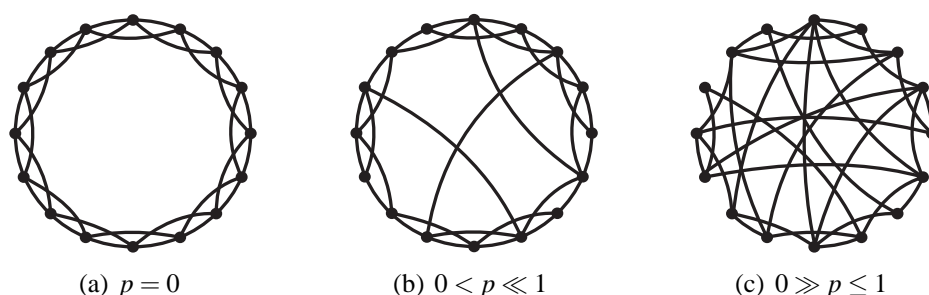


Figura 1.4. Watts & Strogatz model na criação de uma rede com características *small world*.

Newman & Watts Model No modelo anterior, para gerar um rede com características *small world*, as arestas do grafo eram reposicionadas segundo uma probabilidade p . En-

tretanto, em alguns tipos de redes, o reposicionamento de arestas não se aplica. Dessa forma, Newman & Watts propuseram um modelo que, ao invés de reposicionar as arestas do grafo segundo uma probabilidade, novas arestas são adicionadas ao grafo. Nesse modelo, para cada aresta existente no grafo regular, uma nova aresta é adicionada segundo uma probabilidade p . Além disso, um dos pontos finais da aresta adicionada é o mesmo de um dos pontos finais da aresta que foi verificada. Os autores mostraram que para valores intermediários da probabilidade de adição de arestas, a nova rede criada também apresenta características *small world*.

Nesse modelo, a quantidade de arestas no grafo aumenta de acordo com a probabilidade utilizada. Por exemplo, seja $p = 0.1$ a probabilidade de adição de arestas. Se o grafo original possuir 100 arestas, o grafo resultante possuirá, em média, 110 arestas. As 10 arestas adicionadas serão os atalhos na rede, diminuindo o caminho médio mínimo entre os nós mantendo altos valores de coeficiente de agrupamento. É importante observar que quando a probabilidade de adição de arestas é 1, a rede resultante possuirá o dobro de aresta, já que para cada aresta do grafo uma nova aresta será criada.

A Figura 1.5 ilustra a adição de arestas no modelo de Newman & Watts. Quando $p = 0$, nenhuma aresta é adicionada no grafo original (Figura 1.5(a)). Quando o valor de p é maior do que 0, o modelo adiciona algumas arestas no grafo original, transformando-o em um grafo com características *small world*. Quando o valor de p é próximo de 1, o grafo original é transformado em um grafo aleatório. É importante observar que para todos os valores de probabilidade, as arestas segundo o modelo de Newman & Watts são adicionadas no grafo original, e não reposicionadas.

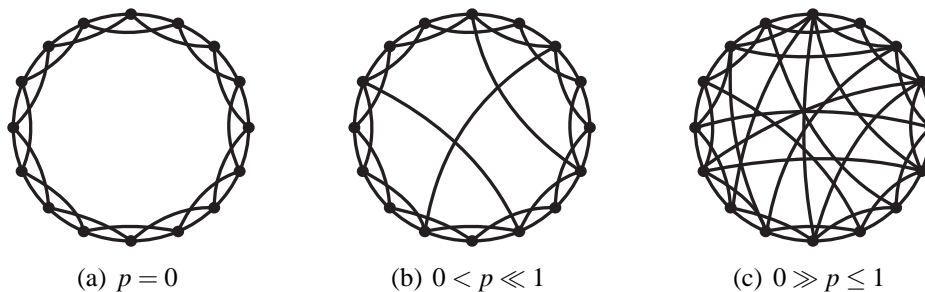


Figura 1.5. Newman & Watts model na criação de uma rede com características *small world*.

Kleinberg Model Os modelos vistos anteriores para reposicionamento ou adição de arestas seguem uma probabilidade fixa p para todas as arestas do grafo original, independente se o novo ponto final da aresta que será adicionada no grafo está perto ou longe da aresta verificada. Kleinberg propôs um modelo onde a distância entre os nós é um fator para a criação de atalhos no grafo. Essa distância pode ser entendida como distância euclidiana ou distância em saltos entre os dois nós em questão. Quando a distância em número de saltos for utilizada, será utilizado a quantidade de saltos no caminho mínimo entre os dois nós. Dessa forma, segundo o modelo de Kleinberg, uma aresta é adicionada na rede segundo uma probabilidade $p = 1/d_{i,j}^r$, onde $d_{i,j}$ é a distância entre os dois nós e r é um parâmetro estrutural na criação de novas arestas, ou seja, o valor de r definirá a

importância da distância na criação de uma aresta entre os nós i e j . Dessa forma, para cada aresta existente no grafo original, um novo vértice escolhido de maneira aleatória será avaliado por $p = 1/d_{i,j}^r$ e, em caso afirmativo, uma nova aresta será criada na rede.

É importante observar que o modelo de Kleinberg necessita de dois parâmetros na definição na adição de uma aresta na rede, $d_{i,j}$ e r . Dessa forma, se dois nós estão distantes mas o valor de r é próximo de zero, a probabilidade de criação de uma aresta entre eles é alta. Entretanto, se o valor de r for suficientemente maior do que 1, a probabilidade será baixa. É interessante observar que o valor de r pode ser entendido como o custo para a criação de um atalho na rede. Se dois nós estão geograficamente ou em número de saltos distantes, a criação de um atalho entre eles terá um custo maior se a ligação entre eles for feita, por exemplo, de maneira cabeada.

1.4.3. Scale free

As Redes Livre de Escala (RLEs) (Newman 2003, Li et al. 2005, Keller 2005), ou *Scale-free Networks*, foram introduzidas por Barabási e Albert em 1999 (Barabási & Albert 1999) para modelar topologias de redes em que a distribuição da conectividade dos nós segue uma lei de potência (Clauset et al. 2007). Desde então, podem ser verificados na literatura diversos outros trabalhos que modelam os mais variados sistemas, sejam naturais ou tecnológicos, como RLEs. Como exemplos de redes livre de escala, pode-se citar a Internet (Faloutsos et al. 1999), a *World Wide Web* (Albert et al. 1999), serviços de redes sociais online (Leskovec et al. 2008), redes de colaboração científica (Newman 2001), cadeias alimentares (Camacho et al. 2002), redes de contatos sexuais (Lilijeros et al. 2001) e rotas ferroviárias (Faloutsos et al. 1999).

A distribuição da variável aleatória X que define a conectividade dos nós de uma rede segue uma lei de potência se, dado a sua função distribuição acumulada $F(x) = P(X \leq x)$ e a sua função distribuição acumulada complementar $\bar{F}(x) = P(X > x)$, $\bar{F}(x) = 1 - F(x) \approx cx^{-\alpha}$ para alguma constante $0 < c < \infty$ e índice de cauda, ou *tail index*, $\alpha > 0$ (Li et al. 2005). Para $1 < \alpha < 2$, F tem variância infinita e média finita mas, para $0 < \alpha \leq 1$, tanto a variância quanto a média são infinitas. Além disso, uma propriedade interessante dessa distribuição é que $\log P(X > x) \approx \log(c) - \alpha \log x$, fazendo com que o gráfico de \bar{F} em escalas logarítmicas seja uma reta com inclinação $-\alpha$ para altos valores de x . Para exemplos e maiores detalhes sobre como identificar leis de potência em dados empíricos, consulte Clauset et al. (2007).

1.4.3.1. Modelos

Por serem redes particulares, com características que não podem ser reproduzidas por modelos de redes aleatórias como, por exemplo, o modelo de Erdős & Rényi (1960), as RLEs demandam que outros modelos sejam propostos. Assim, Barabási & Albert (1999) propuseram o primeiro modelo que explica tais características. Esse modelo, comumente chamado de modelo Barabási-Albert ou simplesmente modelo BA, incorpora duas novas características para gerar a rede: crescimento (*growth*) e conexão preferencial (*preferential attachment*). A primeira estipula que o número de nós da rede cresce com o tempo, enquanto a segunda estipula que quanto mais conectado for um nó, maiores são as chan-

ces dele receber novas conexões. Esse fenômeno, também conhecido como “ricos ficam mais ricos”, vai fazer com que a rede tenha poucos nós com muitas conexões, também chamados de *hubs*, e muitos nós com poucas.

A geração da rede segundo o modelo BA se dá a partir de uma rede conexa com $n_0 > 2$ nós iniciais v_1, \dots, v_{n_0} . Logo em seguida, um novo nó v_{n_0+1} é adicionado à rede e conecta-se a um nó $v_j, 0 < j \leq n_0$ de acordo com uma probabilidade que é proporcional ao grau d_j de v_j . Formalmente, a probabilidade $p_{i,j}$ de um novo nó v_i se conectar à um nó existente v_j é:

$$p_{i,j} = \frac{d_j}{\sum_{u=1}^n d_u}.$$

Em Leskovec et al. (2008), o modelo BA foi extensivamente validado a partir de quatro coleções de dados reais: FLICKR, DELICIOUS, YAHOO! ANSWERS e LINKEDIN. Foi verificado que, mesmo comparado com variações mais sofisticadas, o modelo BA é aquele que melhor explica as bases de dados analisadas. No entanto, o modelo BA leva em consideração o conhecimento global da rede por parte dos nós, o que não é realista para a maior parte das redes. Também em Leskovec et al. (2008) e em Vázquez (2003), os autores propõem uma versão local para o modelo BA, em que o processo de criação das arestas não depende do conhecimento global da rede. As arestas são criadas com o propósito de fechar triângulos, considerando unicamente da vizinhança do nó que ingressara na rede.

Além disso, há outras características inerentes à esse tipo de rede ou à variações dela que o modelo falha em capturar. Devido a isso, outros modelos para geração de RLEs foram propostos na literatura. Em Chen & Shi (2004), foram propostos duas variações do modelo BA que consideram a realocação e a deleção de arestas da rede. Em Holme & Kim (2002), os autores definiram um passo adicional ao modelo BA, em que cada aresta criada pelo modelo BA gera uma segunda aresta adicional que fecha um triângulo. Esse modelo é interessante porque cria uma rede com características de uma RLE e de uma rede *small-world*. Em Leskovec et al. (2007) é proposto o modelo *forest fire*, que além das características previamente descritas, também é capaz de reproduzir duas novas observações temporais feitas a partir de análises de redes reais: densificação e diminuição do diâmetro. Densificação estipula que o número de arestas cresce super-linearmente em relação ao crescimento do número de nós, ao mesmo tempo em que o diâmetro da rede diminui com o tempo.

Outro cenário interessante que foi investigado é aquele em que as redes possuem pesos nas arestas. Em McGlohon et al. (2008), os autores propuseram o modelo *butterfly*, em que foi verificado o efeito de fortificação da rede com o tempo, que estipula que a relação entre o número de arestas de um grafo e o peso total das mesmas é super linear, seguindo uma lei de potência. Tal efeito também é válido para explicar a relação entre o grau de um nó e o peso total das suas arestas. Em Du et al. (2009), os autores propuseram o modelo *PAC*, que reproduz os padrões dos pesos dos triângulos em grafo, mostrando que as distribuições do peso maior, do intermediário e do menor em um triângulo seguem leis de potência em que os expoentes não variam com o tempo. Por fim, o modelo RTG Akoğlu & Faloutsos (2009) é um modelo linear de quatro parâmetros que serve como uma interessante alternativa para gerar RLEs com todas as características descritas nesta se-

ção. Para mais informações sobre modelos geradores de RLEs, consulte Mitzenmacher (2004), Chakrabarti & Faloutsos (2006), Zhou & Lipowsky (2005).

1.4.3.2. Propriedades em Redes de Computadores

Uma característica imediata que pode-se imaginar de uma rede de computadores que tenha a topologia de uma RLE é a alta tolerância à falhas. Como foi mostrado em Albert et al. (2000), como a grande maioria dos nós possui baixo grau, a remoção de nós aleatórios da rede, mesmo que em grande escala, mantém a rede conectada. No entanto, esse tipo de rede é altamente vulnerável a ataques, ou seja, a remoção dos poucos *hubs* da rede a desconecta. Além disso, em Motter et al. (2002) é mostrado que redes com topologia RLE são mais robustas a ataques feitos a arestas de longo alcance, isto é, arestas que conectam nós que de outra maneira seriam separados por uma longa distância em saltos, que redes com topologias aleatórias e com conectividade segundo a distribuição de Poisson. Além disso, foi mostrado que RLEs são mais vulneráveis a ataques feitos a arestas de curto alcance que de longo alcance, pois tais arestas são responsáveis por um maior tráfego, uma vez que conectam dois nós de alto grau e, conseqüentemente, possuem um alto valor de *betweenness*.

Outra característica interessante de redes de computadores com topologia de RLEs é quanto à busca por nós ou por informações. Em Adamic et al. (2001), os autores mostraram que um simples algoritmo local de navegação, que escolhe como próximo nó na busca o nó de maior grau, tem sucesso na busca a um custo sublinear em relação ao tamanho da rede. Isso permite que nós remotos de redes informais, como sistemas P2P por exemplo, sejam alcançados através de arestas locais e, conseqüentemente, permite que pessoas e sistemas de comunicação consigam distribuir informação e acessar os recursos que desejam. Nessa direção, Sarshar & Roychowdhury (2004) propuseram um protocolo que permite a formação de topologias livre de escala mesmo em redes *ad-hoc* altamente dinâmicas, como redes P2P em que nós freqüentemente se conectam e desconectam da rede. Guclu et al. (2008) propuseram um modelo de busca para um cenário em que os nós desse tipo de rede possuem informação limitada, ou seja, em que há um limite para o número de vizinhos que um nó pode ter, o que faz com que não haja *hub* majoritário na rede.

1.4.4. Comunidades

Algumas redes, em especial as redes sociais, biológicas e tecnológicas, como a Internet, tendem a apresentar uma estrutura modular (Girvan & Newman 2002). Tal estrutura modular é caracterizada pela existência de um conjunto ou comunidade de vértices, onde a maioria das conexões se dá entre vértices pertencentes a uma mesma comunidade, enquanto que conexões entre vértices pertencentes à comunidades diferentes são menos frequentes. A formação de tais comunidades pode ser dar por algum tipo interesse, idade, trabalho, páginas mais frequentemente visitadas e assim por diante.

Girvan & Newman (2002) propuseram um modelo para a geração de redes com tal característica, onde esse modelo pode ser visto como uma generalização do modelo para a geração de grafos aleatórios. Nesse modelo, a primeira etapa consiste na classificação de

m vértices em c comunidades. Os passos seguintes consistem na seleção de dois vértices e adição de uma aresta com probabilidade p_{in} , caso ambos os vértices pertençam a mesma comunidade, ou p_{out} , caso os vértices pertençam a comunidades diferentes. É importante observar que a escolha dos valores de p_{in} e p_{out} deve ser feita de forma cuidadosa, de forma que a distinção entre comunidades seja o mais nítida possível.

Por outro lado, o método tradicional para extração de estruturas de comunidade de uma rede é conhecido como *análise de agrupamentos* ou *agrupamento hierárquico*. Nesse método, é atribuído um fator conhecido como *connection strength* a pares de vértices na rede de interesse. Em seguida, iniciando com n vértices que não possuem qualquer aresta entre si, adiciona-se arestas de acordo com a ordem decrescente do fator *connection strength* entre vértices. Pode-se interromper esse processo em qualquer momento de forma a examinar a estrutura do componente formado pela adição de arestas até então. Tais componentes são tidos como as comunidades ou agrupamentos. Quando todas as arestas forem adicionadas, todos os vértices estarão conectados a todos os outros e existirá apenas uma comunidade. A escolha para os valores do fator *connection strength* pode levar em conta questões como distância entre vértices, fluxo de dados, entre outros.

Uma questão interessante relacionada as redes que apresentam estruturas de comunidades é: dado um vértice pertencente a uma rede, é possível identificar a qual comunidade o mesmo pertence. Algoritmos que respondam a tal questão pode ser de fundamental importância para alguns tipos de redes, como por exemplo, a *World Wide Web*.

1.4.5. Geográficas

O modelo de redes geográficas (ver Gastner & Newman 2006) incorpora um elemento a mais que os outros modelos apresentados não consideram: o posicionamento dos vértices no espaço. Desta maneira, a probabilidade de uma aresta existir depende do posicionamento de cada um dos vértices, e especialmente a distância dentre eles.

O modelo mais simples de rede geográfica é o GRG (*Geometric Random Graph*). Neste modelo, os vértices são distribuídos aleatoriamente em um espaço Ω . A conectividade entre dois nós quaisquer u e v é dada como função da distância Euclidiana $\delta(u, v)$ e uma constante que determina o limiar de comunicação r ,

$$(u, v) \in E(G) \iff \delta(u, v) \leq r,$$

onde $E(G)$ é o conjunto de arestas do grafo não direccionado G . O modelo GRG segue uma regra rígida baseada na distância Euclidiana, o que é útil para modelar comunicações sem fios em ambientes ideais.

Outros modelos são usualmente empregados para modelar a conectividade como uma função decrescente em relação à distância. Por exemplo, instâncias de redes geográficas podem ser geradas seguindo a seguinte função de probabilidade:

$$\Pr((u, v) \in E(G)) \sim e^{-\varepsilon \delta(u, v)},$$

onde ε fixa o a escala de comprimento das arestas no grafo.

Outras maneiras de modelar o posicionamento dos vértices no espaço podem ser utilizadas. Em particular, uma metodologia que facilita muito a modelagem é expressar o

posicionamento dos vértices por meio de processos pontuais estocásticos (ver seção 1.5.1 para maior detalhamento sobre processos pontuais).

1.5. Redes complexas na modelagem de redes de computadores

Nesta seção trataremos, inicialmente, dos métodos estatísticos necessários para a aplicação da teoria das redes complexas e em seguida discutiremos diversas aplicações

1.5.1. Métodos estatísticos

Consideremos uma rede de sensores sem fios (ver Seção 1.5.2.4), onde os nós sensores estão posicionados de maneira aleatória numa área de interesse. Neste caso temos redes espaciais, também conhecidas por redes geográficas. Agora, do ponto de vista da modelagem, podemos nos fazer as seguintes perguntas:

- Como podemos modelar o posicionamento?
- O que significa precisamente “posicionados de maneira aleatória”?
- Como podemos determinar se há conectividade entre dois nós quaisquer da rede?
- Há algum modelo de rede complexa capaz de capturar a estrutura desta rede?

Podemos utilizar uma modelagem no espaço Euclidiano bi- ou tri-dimensional onde o posicionamento de um sensor qualquer é representado pela tupla (x, y) ou (x, y, z) , respectivamente, onde x , y e z são valores numéricos. O tipo de valor numérico das coordenadas Euclidianas depende da nossa modelagem. Para simplificar, de agora em diante, vamos considerar a modelagem bi-dimensional numa região $W = [0, \ell]^2 \subset \mathbb{R}^2$, portanto consideramos o posicionamento de cada sensor como sendo uma tupla (x, y) com $0 \leq x \leq \ell$ e $0 \leq y \leq \ell$.

Para modelar um conjunto de sensores distribuídos aleatoriamente seguindo alguma lei de probabilidade, utilizamos processos pontuais espaciais, que são distribuições de probabilidade capazes de descrever a localização espacial de pontos (ver Baddeley 2007, Berthelsen & Møller 2002, Møller & Waagepetersen 2007, e as referências aí citadas). O processo pontual mais utilizado na área de redes de sensores é o URP (*Uniform Random Placement*)

O URP é conhecido na literatura de processos pontuais como *Processo Pontual de Poisson*. Ele é tido como referência de aleatoriedade, para fins de comparação com outros processos pontuais. A forma construtiva de definir o URP com intensidade $\eta > 0$ sobre a região W consiste das seguintes etapas:

1. Observar n , ocorrência da variável aleatória N com distribuição de Poisson e parâmetro $\lambda = \eta \ell^2$, isto é, $\Pr(N = k) = e^{-\lambda} \lambda^k / k!$.
2. Observar $2n$ ocorrências de variáveis aleatórias independentes e identicamente distribuídas no intervalo $[0, \ell]$, por exemplo, $(x_1, y_1, \dots, x_n, y_n)$.
3. Posicionar n pontos na região W nas coordenadas $(x_1, y_1), \dots, (x_n, y_n)$.

Este processo é considerado como referência pois ele possui (de fato, ele é caracterizado por) as seguintes propriedades:

1. Sejam $A_1, A_2 \subset W$ disjuntos, então $\Pr(N_{A_1} = k_1, N_{A_2} = k_2) = \Pr(N_{A_1} = k_1) \Pr(N_{A_2} = k_2)$, isto é, há independência entre as variáveis aleatórias que descrevem o número de pontos em regiões disjuntas.
2. A probabilidade de observar $k \geq 0$ pontos em uma região $A \subset W$ segue uma distribuição de Poisson com média $\lambda_A = \eta \mu(A)$, onde $\mu(A)$ é a área do conjunto A , isto é, $\Pr(N_A = k) = e^{-\lambda_A} \lambda_A^k / k!$, onde N_A é o número de pontos no conjunto A .

Tal como dito anteriormente, o URP é o modelo mais frequentemente encontrado na modelagem de deposição de nós sensores. Há, contudo, outros modelos de processos pontuais aptos a descrever situações onde os pontos tendem a se aglomerar (processos atrativos) ou a se repelir (processos repulsivos). Frery et al. (2008) propoem um modelo composto que, através de um único parâmetro, exhibe os três comportamentos. No que segue, empregaremos apenas o URP. A figura 1.6 ilustra três casos de URP com intensidade variável. A figura 1.7 ilustra três eventos do mesmo modelo de URP, onde constatamos que o número de sensores varia de evento para evento, mesmo sendo a intensidade constante.

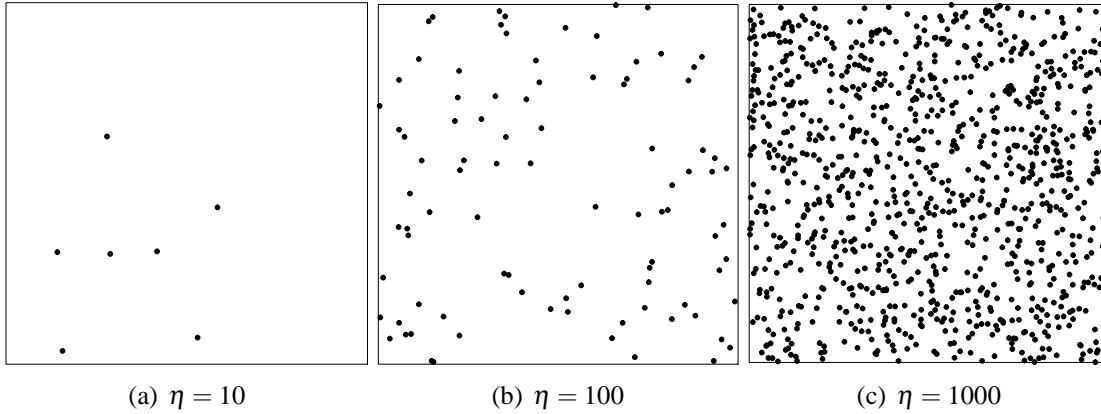


Figura 1.6. Três eventos do modelo URP definido sobre a mesma janela $W = [0, 1]^2$ com intensidade η variável

Uma vez depositos os n sensores na região W segundo o URP, precisamos estipular o modelo de conectividade induzido pela localização espacial dos nós. Por simplicidade pode-se adotar o modelo UDG (*Unit Disk Graph*). Outros modelos mais realistas podem ser encontrados em Tselishchev et al. (2010). O modelo UDG consiste em considerar que os nós u e v , localizados nos pontos (x_u, y_u) e (x_v, y_v) respectivamente, são vizinhos no grafo de conectividade, isto é, se comunicam, se a distância Euclidiana entre eles é no máximo δ_c , isto é

$$(u, v) \in E \iff (x_u - x_v)^2 + (y_u - y_v)^2 \leq \delta_c^2.$$

O modelo de comunicação induzido pelo UDG é bilateral, isto é, se u se comunica com v , então necessariamente v se comunica com u . Há outros modelos de conectividade, inclusive unilaterais.

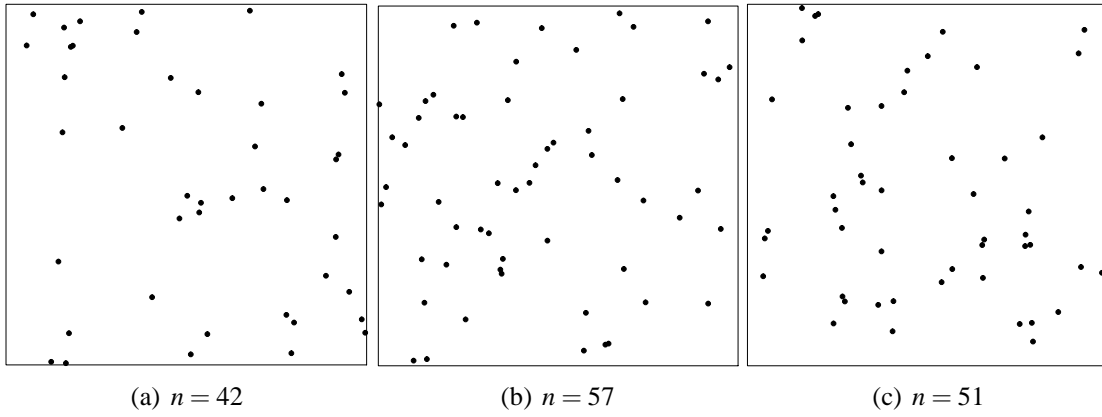


Figura 1.7. Três eventos do modelo URP com intensidade $\eta = 50$ definido sobre a janela $W = [0, 1]^2$

Com os dois ingredientes descritos acima, isto é, com o modelo de deposição URP e o modelo de comunicação UDG, já podemos simular redes de sensores sem fios. Basta, para tanto, estipular a área de interesse W através do lado ℓ , a intensidade do processo η e o raio de comunicação δ_c . A tripla (ℓ, η, δ_c) estipula ou indexa um modelo da classe dos modelos geográficos, nos quais a topologia não depende dos dados mas apenas da localização. Esta classe de modelos é discutida em mais detalhes na seção 1.4.5. A figura 1.8 ilustra três grafos de comunicação induzidos por três raios de comunicação diferentes sobre o mesmo conjunto de $n = 50$ pontos.

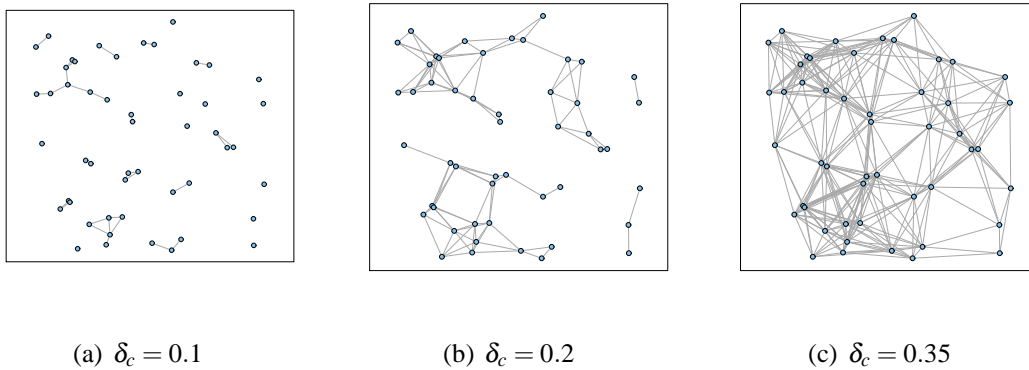


Figura 1.8. Três eventos do modelo URP com intensidade $\eta = 50$ definido sobre a janela $W = [0, 1]^2$

Tal como construído até o momento, um grafo de conectividade como os mostrados na figura 1.8 é um evento de um processo aleatório caracterizado pela tripla (ℓ, η, δ_c) . Medições efetuadas sobre um desses eventos serão eventos de variáveis aleatórias. Por exemplo, se simularmos uma sequência de grafos independentes g_1, g_2, \dots, g_N cuja distribuição é caracterizada pela tripla (ℓ, η, δ_c) , e em cada um deles medirmos o grau médio do grafo $k_m(g_i)$, teremos uma sequência de eventos $k_m(g_1), k_m(g_2), \dots, k_m(g_N)$. Essa sequência de valores poderá ser analisada como sendo eventos de variáveis aleatórias independentes, também caracterizadas pela tripla (ℓ, η, δ_c) . Em outras palavras, neste ponto

do texto dispomos de um modelo estocástico para a variável aleatória G , definida sobre o espaço amostral genérico Ω com valores no conjunto de grafos de conectividade \mathcal{C} , i.e., $G: \Omega \rightarrow \mathcal{C}$. Esse modelo é indexado pela tripla (ℓ, η, δ_c) , isto é, a probabilidade de observar um grafo particular $\Pr(G = g)$ depende apenas de (ℓ, η, δ_c) . A seguir definiremos variáveis aleatórias reais sobre esses grafos, e procederemos a analisá-las. Sem perda de generalidade, iremos considerar sempre o mesmo suporte do processo pontual, W , fixando $\ell = 1$ e, portanto, indexando o modelo apenas pela dupla (η, δ_c) .

Considere o grafo aleatório $G = (V, E)$, onde V é o conjunto de nós (produzido por um URP de intensidade η sobre a janela quadrada de lado unitário) e V é o conjunto de arcos (induzido pelo raio de conectividade δ_c).

A seguir iremos montar um experimento Monte Carlo (Cipra 2000, Dongarra & Sullivan 2000, Robert & Casella 2000) para verificar a hipótese “não sei nada” em uma boa diversidade de situações. Idealmente, a nossa hipótese de trabalho deveria ser verificada analiticamente para todos os valores de (η, δ_c) e, com isso, acabaria a dúvida. A técnica Monte Carlo é uma ferramenta de aplicabilidade geral, que só deve ser usada quando outras abordagens falham ou seriam muito demoradas para dar uma resposta aceitável.

O primeiro passo consiste em discretizar o conjunto de todas as situações de interesse formando, assim, o conjunto $\Theta = [\eta_1, \dots, \eta_{M_1}] \times [\delta_{c1}, \dots, \delta_{cM_2}]$. O tamanho desse conjunto, isto é, o valor $M_1 M_2$, e a abrangência dos intervalos deverão ser tão grandes quanto possível, mas sem comprometer a execução da experiência pelo tempo demandado.

O segundo passo consiste em montar os algoritmos de simulação, isto é, as técnicas que irão produzir os eventos do grafo aleatório G para cada $\theta \in \Theta$. Na nossa experiência, recomendamos empregar funções disponíveis através dos pacotes `spatstat` (para simulação de processos pontuais) e `igraph` (para computação com grafos) da plataforma R (R Development Core Team 2009).

A estrutura do nosso experimento Monte Carlo é mostrada no algoritmo 1.

Algoritmo 1 Estrutura do experimento Monte Carlo para análise de grau médio, máximo *betweenness*, diâmetro e vulnerabilidade de grafos aleatórios

```

1: for cada  $\eta \in [\eta_1, \dots, \eta_{M_1}]$  do
2:   for cada  $\delta_c \in [\delta_{c1}, \dots, \delta_{cM_2}]$  do
3:     for cada  $r \in \{1, \dots, N_R\}$  do
4:       Gerar o evento  $g_r$  do grafo aleatório  $G(\eta, \delta_c)$ 
5:       Calcular e armazenar  $(k_m, B_{\max}, d, V)(g_r)$ 
6:     end for
7:   end for
8: end for
9: Analisar resultados e procurar padrões

```

Para poder implementar o algoritmo 1 precisamos estipular o número de replicações N_R . Em situações relativamente simples podemos lançar mão da regra de bolso $N_R = 30$ como mínimo ou $N_R \geq 100$ para termos resultados mais confiáveis. Essa regra

de bolso se baseia em uma confiança um pouco cega na validade do teorema central do limite. Para fazermos uma análise quantitativa mais cuidadosa, podemos seguir a apresentação feita por Díaz-Emparanza (1996).

Vamos a seguir considerar a situação de querermos estimar uma proporção p , desconhecida, através da repetição de N experimentos dicotômicos independentes e identicamente distribuídos. Pelo fato de serem dicotômicos, o resultado possível é ou “Verdadeiro” ou “Falso”, e não há nenhuma outra possibilidade. Atribuíamos o valor 1 ao primeiro resultado possível, e 0 ao segundo. Podemos estimar p através de

$$\hat{p} = \frac{1}{N} \sum_{i=1}^N X_i, \quad (1)$$

onde X_i é a variável aleatória que modela o experimento número i . É importante frisar que \hat{p} é uma variável aleatória, pois é o resultado de transformar as variáveis aleatórias X_1, \dots, X_N . Dizemos que \hat{p} é um *estimador* de p ; uma vez realizada a experiência e observado um valor particular, por exemplo, se fizemos o experimento e observarmos dezoito sucessos em cinquenta tentativas, não temos mais um estimador mas sim uma *estimativa*: teremos $\hat{p}(\omega) = 19/50$, onde ω denota um ponto do espaço amostral Ω .

O que acontecerá se repetirmos o experimento? Tipicamente teremos outra estimativa. Se fosse possível repetir infinitas vezes o experimento, ou percorrer todo Ω , poderíamos ter uma idéia da variabilidade do nosso estimador \hat{p} , mas essa abordagem não é conveniente e muito raramente é viável.

Ao invés de procedermos por enumeração sobre todo o espaço amostral Ω , podemos obter um bom conhecimento da qualidade do procedimento de estimação (ou inferência) analisando as propriedades de \hat{p} , variável aleatória, que decorrem da sua definição dada na equação (1).

Uma preocupação subjacente a qualquer experiência empírica é a sua *precisão*. No nosso contexto, a precisão de um estimador está associada ao seu viés, que é definido como a diferença entre a esperança do estimador e o valor que ele estima, isto é

$$B(\hat{p}, p) = E(\hat{p}) - p, \quad (2)$$

onde B é pelo termo em Inglês *bias*, e E denota a esperança matemática. O viés de um estimador tipicamente depende do tamanho da amostra N e do verdadeiro valor do parâmetro. Um estimador que possui viés nulo para qualquer N e qualquer valor do parâmetro, é denominado *não viesado* (*unbiased* em Inglês).

Embora o viés seja uma propriedade muito importante, ao ponto de ser um ativo objeto de pesquisa na estatística (ver, por exemplo, Cox & Snell 1968, e a miríade de artigos que seguiram a esse trabalho pioneiro), ele não fornece uma idéia da variabilidade do estimador. Essa variabilidade é um elemento fundamental, junto com o viés, para poder estipular comparações entre estimadores. Mediremos a variabilidade de um estimador através da acurácia; dizemos que o estimador \hat{p} tem acurácia a ao nível $1 - \alpha$, com $\alpha \in (0, 1)$ se

$$\Pr(|\hat{p} - p| < a) \geq 1 - \alpha. \quad (3)$$

É importante interpretar os elementos desta desigualdade.

Voltemos à nossa experiência imaginária de obter infinitas observações independentes da variável aleatória \hat{p} , com p e N fixos. Algumas estimativas ficarão bem próximas de p , outras nem tanto e umas poucas irão ficar bem distantes do verdadeiro valor do parâmetro. Fixando uma tolerância a (a nossa acurácia), e seguindo uma abordagem frequentista, podemos ter uma idéia de quão frequentemente o estimador \hat{p} se afasta do verdadeiro valor p menos do que a . Conhecendo a distribuição do estimador, podemos inclusive calcular a probabilidade disso ocorrer; temos então a equação (3). Intuitivamente, se aumentarmos o tamanho da amostra N , o nosso estimador ficará cada vez mais confinado ao intervalo em volta do verdadeiro valor e, com isso, a probabilidade da equação (3) aumentará com N . Valores típicos dessa probabilidade são 90/100, 95/100, 99/100 e 999/1000, que correspondem a $\alpha = 1/10, 5/100, 1/100$ e $1/1000$, respectivamente.

Podemos pensar de forma dual, isto é, ao invés de aumentar o tamanho da amostra N até alcançarmos a probabilidade (ou α) alvo, estipular a probabilidade (ou o valor de α) e calcular o tamanho mínimo da amostra que nos dará a acurácia a desejada. Como exemplo, suponhamos que desejamos $a = 1/100$ e $\alpha = 5/100$, isto é, desejamos calcular o tamanho da amostra N tal que

$$\Pr(|\hat{p} - p| < 10^{-2}) \geq 95/100 \quad (4)$$

para todo valor de $p \in (0, 1)$. Reescrevendo a equação (4) temos

$$\Pr(-10^{-2} < \hat{p} - p < 10^{-2}) \geq 95/100. \quad (5)$$

Lembrando que a variância de \hat{p} é $p(1-p)/N$,

$$\Pr\left(-\frac{10^{-2}}{\sqrt{\frac{p(1-p)}{N}}} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} < \frac{10^{-2}}{\sqrt{\frac{p(1-p)}{N}}}\right) \geq 95/100. \quad (6)$$

O Teorema Central do Limite (recomendamos ver a formulação de James 1981) afirma que, sob condições bastante gerais, a soma de variáveis aleatórias convenientemente padronizada é uma variável aleatória cuja distribuição pode ser aproximada por uma lei gaussiana. Em particular, se X_1, \dots, X_N são variáveis aleatórias independentes e identicamente distribuídas com esperança μ e variância σ^2 , então

$$\Pr\left(\frac{\frac{1}{N}\sum_{i=1}^N X_i - \mu}{\sigma} \in I\right) \rightarrow \int_I \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\} dt, \quad (7)$$

quando $N \rightarrow \infty$, para qualquer intervalo I . A integral do lado direito encontra-se tabuada na maioria dos livros de estatística, e está implementada em virtualmente todas as plataformas de software de análise de dados.

Se for possível considerar válido o Teorema Central do Limite para a variável aleatória

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}},$$

então é imediato que $\Pr(-1,96 \leq Z \leq 1,96) \approx 95/100$ e, com isso,

$$1,96 = \frac{10^{-2}}{\sqrt{\frac{p(1-p)}{N}}},$$

e daí

$$N = (1,96)^2 10^4 p(1-p). \quad (8)$$

A equação (8) não resolve o nosso problema pois, segundo ela, para determinar o tamanho da amostra necessária para estimar p ao nível de confiança 95/100 com precisão 10^{-3} precisamos do valor de p . A saída para este aparente dilema circular advém do uso de simulação: podemos substituir p na equação (8) por uma primeira estimativa $\tilde{p} = M^{-1} \sum_{j=1}^M X_j$ baseada em M replicações, com M um número aceitável de replicações. Por “aceitável”, entenda-se o maior número possível de replicações cujo tempo de obtenção for compatível com o estudo.

Se na nossa simulação exploratória determinarmos $\tilde{p} = 1/10$, então o experimento Monte Carlo que precisamos realizar requer $N = (1,96)^2 10^4 (9/10)/10 \approx 3457$ replicações.

Voltando à equação (8), vamos escrevê-la com os valores teóricos ao invés dos numéricos empregados para criar o exemplo:

$$N = \left(\frac{t_{\alpha/2}}{a} \right)^2 p(1-p). \quad (9)$$

A equação (9) fornece uma relação útil entre o número mínimo de replicações N necessário para ter precisão a ao nível de confiança $1 - \alpha$ na estimação da proporção p , onde t_v é o quantil de ordem v da distribuição gaussiana padrão. Ela permite calcular o número mínimo de replicações em função de a e de α , ou conhecer a precisão em função da significância e do número de replicações, isto é, o intervalo de confiança.

A noção de “intervalo de confiança” é frequentemente mal empregada. É comum ler que “a estimativa estará em um intervalo de tamanho a ao redor do verdadeiro valor com probabilidade $1 - \alpha$ ”, que é uma interpretação errada. Uma vez observada a estimativa, ela não é mais uma variável aleatória e, portanto, qualquer cálculo de probabilidade ao seu respeito não faz sentido. A interpretação correta (do ponto de vista frequentista) é “se repetirmos infinitas vezes a experiência sob as mesmas condições, em uma proporção $1 - \alpha$ desses ensaios a estimativa estará dentro do intervalo de tamanho a centrado no verdadeiro valor p ”. A diferença é sutil, porém contundente.

O procedimento descrito acima pode ser facilmente adaptado para determinar o número mínimo de replicações necessário para alcançar uma certa precisão com um certo nível de confiança. Na maioria dos casos, será necessário o conhecimento da distribuição do estimador, bem como a realização de pelo menos uma experiência de simulação exploratória.

Mais detalhes podem ser vistos em Bustos & Frery (1992).

A seguir daremos algumas dicas sobre o significado do passo 9 do algoritmo 1, isto é, sobre o que é *analisar* dados de um ensaio Monte Carlo. Em primeiro lugar, os

dados de um ensaio Monte Carlo devem ser tratados como dados de qualquer experiência empírica, isto é, devem ser convenientemente catalogados e organizados.

Toda análise de dados é formada por operações que podem ser categorizadas em (i) análise qualitativa, e (ii) análise quantitativa. A análise qualitativa costuma ser exploratória, enquanto a quantitativa é mais confirmatória. A primeira procura hipóteses a serem verificadas ou descartadas com a segunda. Ambas podem ser feitas com ferramentas gráficas ou quantitativas. Não se trata de atividades lineares, no sentido de uma preceder a outra; são operações que se realizam conforme o analista vai conhecendo os dados à procura de informações relevantes. Um bom analista dispõe de um amplo ferramental, e de experiência para guiar as atividades.

Everitt (2005) apresenta uma introdução às principais técnicas de análise de dados multivariados, tanto para análise qualitativa quanto quantitativa, utilizando R. Para aplicar qualquer um destes procedimentos, como mencionado anteriormente, é importante manter os dados organizados corretamente para conseguir aplicar as diferentes operações de análise de dados sem a necessidade de trocar de formatação. A formatação de dados estatísticos usual é o *data frame*, e ele tem a forma de uma matriz, onde cada fila representa uma instância do experimento, e cada coluna pode ser um *fator*, que caracteriza a instância “simulada”, ou uma medida de interesse registrada para essa instância. Uma boa referência sobre a organização dos dados (e sobre análise de dados em geral) é Chambers (2008). Recomendamos aos leitores ler especificamente a seção 5 do capítulo 6, que trata sobre este tópico em particular.

1.5.2. Aplicações

A seguir, apresentamos alguns resultados obtidos pela comunidade de pesquisadores que vêm empregando a teoria de redes complexas em diferentes áreas de pesquisa relacionadas com redes de computadores. Cada uma das seções seguintes é dedicada a tipos de redes diferentes. A classificação escolhida é baseada na modelagem e propósitos destas redes. A seção 1.5.2.1 apresenta os resultados relacionados com a Internet. A aplicação de teoria de redes complexas em redes *peer-to-peer* são apresentadas na seção 1.5.2.2. Na seção 1.5.2.3 consideramos aplicações no contexto de redes ad-hoc. Finalmente, a seção 1.5.2.4 considera aplicações em redes de sensores sem fios.

Cada uma das seções mencionadas acima considera a modelagem dessas redes do ponto de vista de redes complexas. A figura 1.9 apresenta algumas das áreas onde a teoria de redes complexas já foi empregada com êxito.

1.5.2.1. Internet

Faloutsos et al. (1999) apresentam o primeiro trabalho de modelagem, do ponto de vista topológico, da Internet. Como os autores citam neste trabalho, a compreensão da topologia da Internet traz benefícios, dentre eles:

- conhecendo melhor a topologia, seremos capazes de projetar protocolos mais eficientes que tirem proveito das propriedades topológicas presentes na Internet;
- é possível obter modelos de redes mais precisos para efetuar simulações, e;

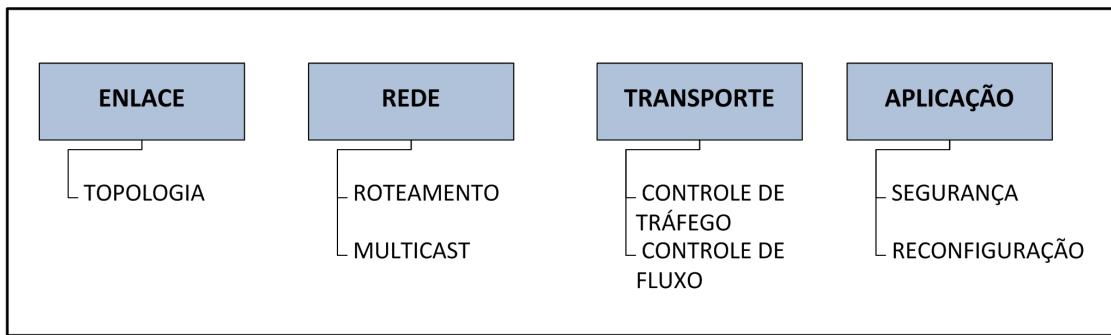


Figura 1.9. Taxonomia de áreas de aplicação de redes complexas

- podemos derivar estimadores de parâmetros topológicos para a análise de protocolos, assim como também obter predições da evolução da topologia da Internet no futuro.

Siganos et al. (2003) apresentam um estudo da topologia da Internet ao nível de sistemas autônomos, propondo modelos de leis de potências da forma $y \propto x^a$ nesta hierarquia, onde a é uma constante e x e y são medidas de interesse. Os autores coletaram informações das tabelas de roteamento BGP quase diariamente por um período de 5 anos (desde 1999 até 2002), e em cada instância do tempo foram validados os modelos de leis de potência para diferentes medidas topológicas. Em particular foram consideradas as seguintes:

- D_d : a função de distribuição acumulada complementar de um grau, que é a porcentagem de nós que possuem grau maior ou igual a d ;
- r_v : a ordem de um nó, v , é o índice na ordem decrescente de graus;
- $P(h)$: o “número de pares” de nós, é o número total de pares ordenados de nós que estão como máximo a h hops de distância;
- $NN(h)$: o número médio de nós em uma vizinhança de h hops;
- λ : o autovetor do grafo de comunicações;
- i : o autovalor de ordem i do autovetor λ ordenado.

Baseado nestas medidas, as seguintes leis de potência foram validadas:

- expoente de *rank*: dado um grafo, o grau d_v de um nó v é proporcional ao *rank* r_v à potência de uma constante \mathcal{R} , isto é, $d_v \propto r_v^{\mathcal{R}}$;
- expoente de grau: dado um grafo, a função de distribuição acumulada complementar de um grau D_d de um grau d é proporcional ao grau à potência de uma constante \mathcal{D} , isto é, $D_d \propto d^{\mathcal{D}}$;

- auto-exponente: dado um grafo, os autovalores λ_i são proporcionais ao ordem i à potência de uma constante \mathcal{C} , isto é, $\lambda_i \propto i^{\mathcal{C}}$.

O trabalho de Siganos et al. (2003) apresenta uma caracterização da topologia de sistemas autônomos da Internet através de leis de potência. Apesar de que os autores consideraram o fator temporal sobre estas leis, elas não são suficientes para representar todas as características topológicas inter-domínio. As leis apresentadas aqui (e outras de este e trabalhos posteriores) são uma condição necessária mas não suficiente para obtermos uma modelagem que permita conhecer a evolução topológica da Internet.

1.5.2.2. Redes *peer-to-peer*

No mundo real é possível observar a existência de vários tipos de redes sociais, como redes de conhecimento e colaboração, onde estudos realizados em redes sociais têm sido aplicados em outros tipos de redes, como as tecnológicas e as biológicas. Isso decorre do fato de que a grande maioria das redes sociais apresentam propriedades desejáveis para esses outros tipos de redes, como por exemplo, o efeito *small world*, *scale free* e a existência de comunidades. Muitos modelos interessantes foram propostos de forma a obter redes com determinadas propriedades em uma escala global. No entanto, em redes reais, todo o controle e o gerenciamento sobre a estrutura da rede está implícita sobre as escolhas locais realizadas pelos pares de forma distribuída e não-supervisionada. Cada par escolhe seu vizinho e liga-se ao mesmo obedecendo apenas seus interesses pessoais, sem levar em consideração a formação de uma rede com uma estrutura ou propriedade desejada. Diante disso, no trabalho proposto por Carchiolo et al. (2008) é apresentado um modelo para o crescimento e evolução de redes *peer-to-peer* inspiradas na dinamicidade de redes sociais, modelo conhecido como *PROSA*. Esse modelo é capaz de gerar redes com características desejáveis, tais como *small world* e estruturas de comunidade, além de manter os mecanismos de comunicação entre pares e gerenciamento de rede simples e intuitivos. Os mecanismos utilizados no gerenciamento de ligações e roteamento de consultas induzem a formação de uma forte estrutura de comunidades, a qual pode explicar o porque a busca e recuperação de recursos no *PROSA* é tão rápida e eficiente.

A principal consideração realizada pelo modelo *PROSA* é que as propriedades globais de muitas redes sociais devem-se ao comportamento local dos pares. A principal hipótese é que é possível obter características globais semelhantes à uma rede social real em uma rede P2P se o comportamento local dos pares for similar ao comportamento de pessoas em uma rede social real. Por essa razão, *PROSA* é fortemente inspirado pelas redes de relacionamentos entre pessoas. Dessa forma, a rede P2P formada pelo modelo *PROSA* apresenta características como caminho mínimo médio baixo entre os pares da rede, alto coeficiente de agrupamento e a presença de fortes estruturas de comunidade. Os pares e as ligações no *PROSA* evoluem como pessoas e relacionamentos entre pessoas em uma rede social, ou seja, cada par liga-se a um certo número de outros pares, repassam buscas para tais pares e ocasionalmente tomam conhecimento de novos pares e estabelece ligações com eles. Pares enviam mensagens somente para os seus parentes e vizinhos; mensagens para os pares desconhecidos são repassadas através de rotas que provavelmente irão encontrá-los, olhando apenas para as conexões e conhecimentos lo-

cais. O resultado é uma rede com gerenciamento totalmente descentralizado e distribuído, a qual apresenta muitas propriedades estruturais de redes sociais locais.

Já o protocolo *SWOP* proposto em Hui et al. (2006) utiliza-se das propriedades presentes nas redes do tipo *small world* para a criação e manutenção de uma rede estruturada *peer-to-peer*. Nesse trabalho, duas questões fundamentais aos sistemas *P2P* são trabalhadas:

- como melhorar o desempenho na busca por objetos dentro da rede;
- como uma rede *P2P* pode manipular a demanda elevada por objetos populares e altamente dinâmicos.

Para solucionar a primeira questão, os autores propuseram a criação de uma rede *P2P* com as propriedades de uma rede *small world*. Ou seja, ao se utilizar tal modelo na criação de uma rede *P2P* a rede passaria a apresentar um baixo caminho mínimo médio entre nós aleatórios, resultando na melhoria da busca por objetos dentro da rede. Já a segunda questão é de fundamental importância, pois alguns objetos podem ser extremamente populares e requisições por tais objetos podem ser recebidas dentro de um intervalo de tempo relativamente pequeno. Esse tipo de tráfego pode sobrecarregar o nó que possui tal objeto e conseqüentemente muitos usuários podem não ter acesso a esse objeto. Na tentativa de solucionar tal problema, os autores tiraram vantagem do alto grau de agrupamento existente nas redes do tipo *small world* de forma que os nós dentro de uma vizinhança podem se auto-organizar e replicar o objeto altamente popular entre eles e dessa forma tornar o sistema fortemente robusto em situações de tráfego intenso.

1.5.2.3. Redes ad-hoc

Uma rede de computadores *ad hoc* é uma rede que não precisa de infraestrutura especial para os nós estabelecerem comunicação, pois, nesse tipo de rede um nó pode atuar como roteador ou transmissor de pacotes para outros nós. Esses nós estão distribuídos em um espaço e a comunicação entre eles é feita através do meio sem-fio. As comunicações fim-a-fim são possíveis através de comunicações salto a salto entre o emissor e o receptor da informação. Assim, por ser um tipo de rede diferente das tradicionais, as redes *ad hoc* tem despertado grande interesse de pesquisa, e uma área bastante emergente é a modelagem de uma rede *ad hoc* como uma rede complexa. No trabalho de Krause et al. (2004), os autores apresentam as redes *ad hoc* como uma rede complexa. Nesse modelo, os N nós podem assumir posições aleatórias na rede $(x, y) \in [0, L] \times [0, L]$, onde L representa o valor limite para as variáveis x e y (este modelo é equivalente ao URP descrito na seção 1.5.1). A modelagem de ligação entre os nós sensores nesse tipo de rede está relacionada com alguns fatores como a potência do rádio do nó sensor e a existência de interferência no meio. Esses dois fatores determinam o raio de transmissão de um sensor, e dois nós estão conectados quando se encontram no raio de transmissão um do outro. Por isso, nesse mesmo trabalho são propostos alguns modelos de redes que modelam a ligação entre os nós de maneira diferente.

- **Rede com potência de transmissão constante:** Esse é um modelo no qual a potência de transmissão P é a mesma para todos os nós sensores da rede. Considera-se também que todas as ligações são bidirecionais. Assim, se o valor de P é pequeno, a rede não é fortemente conectada. Se o valor P é muito grande a rede é fortemente conectada. Entretanto, nessa configuração o controle de acesso ao meio é prejudicado e a probabilidade de colisão em uma transmissão tende a ser grande. A escolha de P depende diretamente do tamanho da rede.
- **Rede com valor mínimo de grau para os nós:** Nesse modelo, é definido a priori um valor para o grau mínimo k_{min} de cada nó sensor. Em seguida, todos os nós enviam pacotes do tipo “Hello” e “Hello-reply” com o intuito de descobrir diretamente quem são os vizinhos, definindo assim suas ligações. Assim, o nó ajusta a sua potência de transmissão de forma a alcançar pelo menos os k_{min} vizinhos seus.
- **Rede com o grau alvo constante:** Nesse tipo de modelo, é definido um valor para o grau alvo k_{alvo} que cada nó deve possuir. Assim, os nós ajustam a potência de seus rádios de forma que cada nó tenha exatamente k_{alvo} vizinhos. Um dos requisitos é garantir a forte conectividade entre os nós da rede. Essa garantia de conectividade exclui algumas topologias da rede, por apresentarem nós desconexos.
- **Rede com o grau alvo *scale-free*:** Esse modelo se baseia na distribuição de graus *scale-free*. Essa é uma distribuição na qual os nós são independentes e a escolha de seu grau k_i^{alvo} é feita de forma aleatória dentro de um intervalo.

Nesse contexto, alguns trabalhos são encontrados na literatura que utilizam essa abordagem para resolver problemas no cenário das redes *ad hoc*. Assim, em (Yen & Cheng 2005) é apresentado um trabalho que modela uma rede ad-hoc como uma rede complexa, com o intuito de estabelecer uma relação entre coeficiente de clusterização da rede e a quantidade de terminais escondidos na rede. A existência desses terminais implica na ocorrência de colisões nas transmissões, o que compromete o desempenho do protocolo CSMA. Assim, para o cálculo do coeficiente de clusterização foi assumido um modelo de localização uniforme dos nós e, para a cobertura foi utilizada a convenção de Torus. Após a modelagem analítica verificou-se que o número de terminais escondidos está relacionado ao número de nós na rede e a probabilidade de existir um link entre cada um deles.

Dando continuidade aos trabalhos que consideram as redes *ad hoc* como redes complexas, em Danon et al. (2005) é apresentada uma avaliação das recentes abordagens encontradas na literatura para identificação de comunidades em termos da sensibilidade na busca das comunidades e do custo computacional. Esses métodos são apresentados e classificados. Em seguida, é comparado o desempenho dos métodos aplicados a um cenário de redes *ad hoc*, no qual a estrutura de comunidades é conhecida. Assim, é apresentada uma análise da fração de comunidades corretamente identificadas para cada abordagem, como também o custo computacional de cada uma delas. Verificou-se que os algoritmos que apresentam melhor desempenho conhecem a priori a quantidade de comunidades existentes e apenas as identifica, o que sugere a necessidade de projetos de novas abordagens que primam pela redução do tempo para realizar essa identificação e que aumentem a exatidão da busca.

1.5.2.4. Redes de sensores sem fios

As redes de sensores sem fio (RSSF) são um tipo especial de redes ad-hoc composta de dispositivos sensores que atuam em conjunto no sentido de monitorar, instrumentar e, eventualmente controlar aspectos do mundo físico. As RSSFs apresentam diversas restrições no consumo de energia, dado que são compostas de dispositivos autônomos que funcionam alimentados por baterias e que a vida útil dessa bateria varia, principalmente, com a quantidade de dados transmitido pelo sensor. Dessa forma, em uma determinada rede, alguns sensores podem deixar de funcionar devido ao descarregamento da bateria enquanto outros sensores ainda estão aptos a funcionar. Alguns sensores também podem apresentar falhas e deixarem de funcionar enquanto outros sensores permanecem em pleno funcionamento. O ambiente monitorado e o fenômeno podem mudar constantemente fazendo com que a rede tenha que se adaptar a essas mudanças. Devido aos diversos desafios presentes nessas redes, há um esforço crescente de pesquisa desde o surgimento das RSSFs no final da década de 1990.

Neste contexto, Pásztor et al. (2010) apresentam um esquema de reprogramação seletiva de redes de sensores móveis que monitoram alguns aspectos da vida de animais. Esta abordagem aproveita a estrutura de comunidade entre os indivíduos que estão sendo monitorado para disseminação de código de maneira eficiente. O objetivo desta aplicação é monitorar diferentes aspectos de comportamento através da implantação de nós sensores nos próprios indivíduos. Neste cenário, os programas de aplicação utilizados em cada nó sensor dependem do comportamento do animal o qual o sensor está localizado. Suponhamos que estamos interessados em estudar o comportamento de diferentes espécies de texugos¹. Certamente vai existir um subconjunto que estará com maior frequência nas tocas, e neste caso estaremos interessados em estudar o ambiente nas tocas. Por outro lado, aqueles que não frequentam muito as tocas estarão percorrendo caminhos na floresta, e neste outro caso podemos estar interessados em conhecer os fatores (o clima, por exemplo) que determinam a escolha de um caminho ou outro.

Dado que é inviável do ponto de vista prático reprogramar os nós sensores capturando os animais, a reprogramação *online* neste caso é a única solução aceitável. Notemos que nesta situação é necessário identificar o comportamento para determinar a aplicação que será necessária num determinado nó. Para isso, os autores propõem um algoritmo de detecção de comunidades totalmente distribuído. O sistema consta de três passos:

1. o código de aplicação e as restrições que determinam a reprogramação é enviado ao nó *sink*;
2. é feito um *broadcast* das restrições para todos os nós da rede;
3. o código é disseminado, através de um algoritmos ciente da estrutura social, somente para os nós que satisfazem as restrições.

¹ Animais de pernas curtas e atarracados, carnívoros que pertencem à família dos mustelídeos (Mustelidae, a mesma família de mamíferos dos furões, doninhas, lontras, e muitos outros tipos de carnívoros). Fonte: Wikipedia.

Líder	Alvo	Próximo	Distância
A	Sim	<i>Sink</i>	2
B	Não	C	2
C	Sim	C	1
D	Sim	–	–

Tabela 1.2. Exemplo de tabela de roteamento para um nó D em Pásztor et al. (2010)

Na fase de disseminação, os indivíduos socialmente centrais são utilizados como disseminadores, já que potencialmente tendem a conhecer maior quantidade de indivíduos, satisfazendo restrições. Estes líderes formam uma hierarquia de roteamento, e as tabelas de roteamento têm a forma ilustrada na tabela 1.2. Nesta tabela, a coluna “Alvo” indica se há algum nó na comunidade, correspondente ao líder da linha da tabela, que satisfaz a restrição. Direcionamos aos leitores interessados em mais detalhes a ler o artigo de Pásztor et al. (2010) e as referências incluídas nesse texto.

Outra aplicação interessante é o atendimento de requisições em redes de sensores sem fios. Como atender requisições quando estamos frente a uma rede dinâmica sem nenhum sistema de localização? Existem duas amplas categorias de soluções na literatura:

1. *flooding* ou *flooding* controlado: cada nó (re)transmite o pacote de requisição uma única vez;
2. caminhos aleatórios: o caminho começa em um nó fixo, e em cada passo o pacote de requisição é enviado a um vizinho. No caso mais simples esse vizinho é escolhido aleatoriamente de maneira uniforme dentre a vizinhança atual. Esse método é conhecido como “caminho aleatório simples”.

As estratégias de *flooding* são ótimas em termos da latência (ou quantidade de saltos), mas têm o problema de não serem eficientes em termos de energia, aspecto fundamental para o projeto de redes de sensores sem fios. Por outro lado, as técnicas de caminhos aleatórios simples são melhores pois poupam energia permitindo que o pacote de requisição somente seja repassado para um único nó. Porém, esta técnica possui perda de desempenho em relação à latência quando o dados de interesse estão muito afastados do nó *sink*, já que esses caminhos podem conter ciclos.

Zuniga et al. (2010) apresentam uma combinação de estas duas técnicas chamada “caminho aleatório sem repetição”. Nesta forma de caminho aleatório, o caminho começa com um nó fixo e em cada passo o vizinho com menor quantidade de visitas é escolhido para repassar o pacote de requisição. Em caso de mais de um nó ter a mesma mínima quantidade de visitas, o próximo nó é selecionado dentre esse conjunto de maneira aleatória uniforme.

Dois aspectos podem ser estudados para avaliar o desempenho deste novo algoritmo:

- tempo de cobertura: o número esperado de passos para um caminho começado em u visitar a rede toda;

- tempo de acerto: o número esperado de passos para um caminho começado em u visitar por primeira vez v .

Os resultados apresentados por Zuniga et al. (2010) confirmam que o mecanismo proposto aumenta a probabilidade de achar nós não visitados, e consequentemente o tempo de cobertura é menor. Para mais detalhes sobre este trabalho, os leitores devem continuar a leitura de Zuniga et al. (2010) e das referências contidas nesse trabalho.

Helmy (2003) estudou o efeito da propriedade de *small world* no contexto das redes ad-hoc e de sensores sem fio. Em seu estudo, ele classifica as redes de sensores sem fio como grafos espaciais (grafos geográficos) que tendem a apresentar maior quantidade de agrupamentos e comprimento dos caminhos médios maiores quando comparados com os grafos aleatórios. Dessa maneira, ele observa que a adição de poucas interconexões de longa distância (chamadas de atalhos) pode reduzir drasticamente o comprimento médio dos caminhos mínimos e, assim, apresentar características de *small world*. Ele defende que esses atalhos não precisam ser totalmente aleatórios (o que seria impraticável em redes geográficas), mas podem estar confinados a um número limitado de saltos. Neste trabalho, são destacadas algumas propriedades interessantes do grafo apresentar a propriedade de *small world* como por exemplo a diminuição da latência para (os caminhos ficam menores). Os autores sugerem que alguns sensores podem aumentar suas potências de transmissão para conseguir alcançar o efeito desejado.

Nessa linha, Guidoni et al. (2008) propõem e avaliam dois modelos para geração de redes de sensores sem fio que apresentam a propriedade *small world*. Eles sugerem o uso de duas categorias de sensores, L e H, sendo o segundo possui maior capacidade de energia, processamento, memória e potência de transmissão. Dessa maneira apenas uma pequena porcentagem do total de nós da rede seria composta de sensores H (sensores mais caros). Os autores mostram que redes de sensores produzidas através desse modelo apresentam bom desempenho com relação ao consumo de energia e latência. Os modelos utilizados para geração de redes *small world* são adequações dos modelos apresentados na seção 1.4.2 para o contexto de redes de sensores. O leitor interessado deve consultar o trabalho (Guidoni et al. 2008) e as referências ali contidas para mais detalhes.

Maia et al. (2009) propuseram um protocolo para reprogramação para redes de sensores sem fio que aproveita os benefícios da propriedade *small world*. Os autores utilizam um modelo heterogêneo de redes de sensores sem fio semelhante ao do trabalho de Guidoni et al. (2008), e a tarefa de reprogramação aproveita esta infraestrutura de comunicação para reprogramar os nós sensores. Os autores demonstram que utilizando o protocolo *OAP – SW* proposto, a tarefa de reprogramação de sensores se torna mais eficiente em termos de quantidades de mensagens necessárias, o tempo e o consumo de energia necessários para reprogramar a rede.

Nem toda métrica de centralidade presente na teoria de redes complexas (ver seção 1.3.3) é apropriada para todo contexto, em Ramos et al. (2010) é proposto o *sink betweenness* (SBet), variação do *shortest-path betweenness* que considera os menores caminhos entre nós sensores e nó *sink*. Formalmente SBet, é definido como:

$$SB(v) = \sum_{\substack{v,t \in V \\ v \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

onde s é o *sink*, σ_{st} é o número de caminhos mínimos entre o *sink* e o nó t , e $\sigma_{st}(v)$ é o número de caminhos mínimos entre o *sink* e o nó t que passam pelo nó v .

Os autores desenvolvem um estudo estatístico e demonstram que para diversos cenários típicos das redes de sensores sem fio o *sink betweenness* se mostra bastante correlacionado com o consumo de energia devido à tarefa de retransmissão de pacotes (oriunda do roteamento). Dessa maneira, os autores concluem que o *sink betweenness*, uma vez estimado, pode ser utilizado no projeto de algoritmos de roteamento e de camada de enlace para, por exemplo, aliviar o efeito de sobrecarga de trabalho dos nós mais próximos ao *sink*.

Para mostrar a aplicabilidade desta métrica no contexto de redes de sensores sem fio, o trabalho de Oliveira et al. (2010) propõe dois algoritmos de roteamento, um baseado em árvore e outro *gossip*. Com o intuito de gerar árvores de roteamento que contenham menor quantidade de arestas, e assim economizar energia, cada nó escolhe como seu pai na árvore àquele com maior centralidade. Esta abordagem baseia-se na hipótese de que os nós de maior centralidade tendem a agregar o maior número de rotas e, assim, favorecer a fusão de dados (Nakamura et al. 2007, 2009).

Por outro lado, no roteamento *gossip* proposto, a probabilidade de envio é relacionada a quão central é um nó e, com base nisso, o intuito foi de se evitar os nós mais centrais. Neste caso, o principal objetivo é melhorar o balanceamento de carga fazendo com que os nós mais centrais não sejam sobrecarregados.

Os resultados de simulação mostram ganho ao se utilizar o *sink betweenness* no roteamento quando se comparado a alguns algoritmos clássicos de roteamento. No roteamento em árvore, apresentou-se diminuição do número de arestas da árvore de roteamento, e, para o algoritmo *gossip* é mostrada uma melhor distribuição de carga de roteamento.

1.6. Conclusão

Neste trabalho foi apresentada uma introdução à teoria das redes complexas, suas principais métricas de caracterização e seus modelos mais comumente empregados em aplicações de redes tecnológicas como as redes de computadores. Também foi apresentada uma introdução aos métodos estatísticos imprescindíveis para o desenvolvimento de estudos que envolvam as técnicas discutidas nesse trabalho. Por fim foram apresentadas diversas aplicações da teoria das redes complexas no contexto de redes de computadores, mais especificamente em: (i) Internet, (ii) redes P2P, (iii) redes ad-hoc e (iv) redes de sensores sem fio.

Um grande benefício que a teoria de redes complexas tem para oferecer na modelagem de redes de computadores está na caracterização da topologia através de um conjunto finito de medidas. Deste conjunto de medidas, muitas delas podem ser caras de se obter num modelo distribuído, e, podemos lançar mão de métodos estatísticos para dar suporte nos diferentes modelos de redes complexas. Estes métodos podem ser utilizados para criar estimadores que, sob uma família de cenários com algumas propriedades comuns, descrevem propriedades mais fortes possivelmente a um menor custo computacional.

Apesar de não haver intenção de esgotar o tema e nem a lista de aplicações em que já foram empregadas as técnicas de redes complexas, o presente trabalho demonstra o potencial que essa teoria tem apresentado para o contexto de modelagem das redes de computadores. Um indicativo desse potencial pode ser observado em uma recente chamada de trabalho para uma edição especial da revista *Computer Communications* da editora Elsevier (*Special Issue of Computer Communications on Complex Networks*) em nome do editor associado Matthieu Latapy, pesquisador senior do laboratório LIP6 da Universidade Pierre and Marie Curie em Paris, França.

Referências

- Adamic, L., Lukose, R., Puniyani, A. & Huberman, B. (2001), 'Search in power-law networks', *Physical Review E* **64**(46135).
- Akoglu, L. & Faloutsos, C. (2009), Rtg: A recursive realistic graph generator using random typing., in W. L. Buntine, M. Grobelnik, D. Mladenic & J. Shaw-Taylor, eds, 'ECML/PKDD (1)', Vol. 5781 of *Lecture Notes in Computer Science*, Springer, pp. 13–28.
- Albert, R., Jeong, H. & Barabási, A. (2000), 'Error and attack tolerance of complex networks', *Nature* **406**(6794), 387–482.
- Albert, R., Jeong, H. & Barabási, A.-L. (1999), 'Diameter of the World Wide Web', *Nature* **401**, 130–131.
- Anthonisse, J. M. (1971), The rush in a directed graph, Technical Report BN 9/71, Stichting Mathematisch Centrum, 2e Boerhaavestraat 49 Amsterdam.
- Baddeley, A. (2007), Spatial point processes and their application, in W. Weil, ed., 'Stochastic Geometry', Vol. 1892 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 1–75.
- Barabási, A. & Albert, R. (1999), 'Emergence of scaling in random networks', *Science* **286**, 509–512.
- Berthelsen, K. K. & Møller, J. (2002), 'A primer on perfect simulation for spatial point processes', *Bulletin of the Brazilian Mathematical Society* **33**(3), 351–367.
- Bolobás, B. (2001), *Random Graphs*, 2nd edn, Academic Press.
- Brandes, U. (2001), 'A faster algorithm for betweenness centrality', *Journal of Mathematical Sociology* **25**, 163–177.
- Brandes, U. & Pich, C. (2007), Centrality estimation in large networks, in 'International Journal of Bifurcation and Chaos, Special Issue on Complex Networks? Structure and Dynamics', Vol. 17, pp. 2303–2318.
- Brin, S. (1998), The anatomy of a large-scale hypertextual web search engine, in 'Computer Networks and ISDN Systems', Vol. 30, pp. 107–117.
URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.109.4049>

- Burt, R. (1995), *Structural Holes: The Social Structure of Competition*, Harvard University Press.
URL: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0674843711>
- Bustos, O. H. & Frery, A. C. (1992), 'Reporting Monte Carlo results in statistics: suggestions and an example', *Revista de la Sociedad Chilena de Estadística* **9**(2), 46–95.
URL: <http://sites.google.com/site/acfrery>
- Camacho, J., Guimerà, R. & Nunes Amaral, L. A. (2002), 'Robust patterns in food web structure', *Physical Review Letters* **88**(22), 228102.
- Carchiolo, V., Malgeri, M., Mangioni, G. & Nicosia, V. (2008), 'Emerging structures of p2p networks induced by social relationships', *Computer Communications* **31**(3), 620–628.
- Chakrabarti, D. & Faloutsos, C. (2006), 'Graph mining: Laws, generators, and algorithms', *ACM Computing Surveys* **38**(1), 2.
- Chambers, J. M. (2008), *Software for Data Analysis: Programming with R*, Statistics and Computing, Springer.
- Chen, Q. & Shi, D. (2004), 'The modeling of scale-free networks', *Physica A: Statistical Mechanics and its Applications* **335**(1-2), 240 – 248.
- Cipra, B. A. (2000), 'The best of the 20th century: Editors name top 10 algorithms', *SIAM News* **33**(4), 1–2.
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. (2007), Power-law distributions in empirical data. code available at <http://www.santafe.edu/aaronc/powerlaws/>.
URL: <http://arxiv.org/pdf/0706.1062>
- Costa, L., Oliveira Jr., O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., Viana, M. P. & da Rocha, L. E. C. (2008), Analyzing and modeling real-world phenomena with complex networks: A survey of applications.
URL: <http://arxiv.org/abs/0711.3199>
- Costa, L., Rodrigues, F. A., Travieso, G. & Villas Boas, P. R. (2007), 'Characterization of complex networks: a survey of measurements', *Advances in Physics* **56**, 167–242.
- Cox, D. R. & Snell, E. J. (1968), 'A general definition of residuals (with discussion)', *Journal of the Royal Statistical Society B* **30**, 248–275.
- Danon, L., Diaz-Guilera, A., Duch, J. & Arenas, A. (2005), 'Comparing community structure identification', *Journal of Statistical Mechanics-Theory and Experiment* p. 10.
- Díaz-Emparanza, I. (1996), Selecting the number of replications in a simulation study, *Econometrics* 9612006, EconWPA.
URL: <http://ideas.repec.org/p/wpa/wuwpem/9612006.html>

- Dongarra, J. & Sullivan, F. (2000), 'Guest editors' introduction: The top 10 algorithms', *Computing in Science and Engineering* **2**(1), 22–23.
- Du, N., Faloutsos, C., Wang, B. & Akoglu, L. (2009), Large human communication networks: patterns and a utility-driven generator, in 'KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, New York, NY, USA, pp. 269–278.
- Erdős, P. & Rényi, A. (1959), 'On radom graph', *Publicationes Mathematicae* **6**, 290–297.
- Erdős, P. & Rényi, A. (1960), 'On the evolution of random graphs', *Publications of The Mathematical Institute of The Hungarian Academy of Sciences* **5**, 17–61.
- Everett, M. & Borgatti, S. P. (2005), 'Ego network betweenness', *Social Networks* **27**(1), 31 – 38.
URL: <http://www.sciencedirect.com/science/article/B6VD1-4F29STG-1/2/acad221427b2d051625939c7b00b8d97>
- Everitt, B. S. (2005), *An R and S-PLUS Companion to Multivariate Analysis*, Springer-Verlag.
- Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999), On power-law relationships of the internet topology, in 'SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication', ACM, New York, NY, USA, pp. 251–262.
- Fiedler, M. (1973), 'Algebraic connectivity of graphs', *Czechoslovak Mathematical Journal* **23**(98), 298–305.
- Freeman, L. C. (1977), 'A set of measures of centrality based on betweenness', *Sociometry* **40**(1), 35–41.
- Frery, A. C., Ramos, H., Alencar-Neto, J. & Nakamura, E. F. (2008), Error estimation in wireless sensor networks, in 'ACM Symposium on Applied Computing', ACM, Fortaleza, CE, Brazil, pp. 1923–1927.
- Gastner, M. T. & Newman, M. E. (2006), 'The spatial structure of networks', *The European Physical Journal B - Condensed Matter and Complex Systems* **49**(2), 247–252.
- Geisberger, R., Sanders, P. & Schultes, D. (2008), Better approximation of betweenness centrality, in 'Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments'.
URL: <http://www.siam.org/proceedings/alnex/2008/alnex08.php>
- Girvan, M. & Newman, M. E. J. (2002), 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences of the United States of America* **99**(12), 7821–7826.
- Gross, J. L. & Yellen, J. (2003), *Handbook of Graph Theory - Discrete Mathematics and Its Applications*, 1 edn, CRC.

- Guclu, H., Kumari, D. & Yuksel, M. (2008), Ad hoc limited scale-free models for unstructured peer-to-peer networks, *in* 'P2P '08: Proceedings of the 2008 Eighth International Conference on Peer-to-Peer Computing', IEEE Computer Society, Washington, DC, USA, pp. 160–169.
- Guidoni, D. L., Mini, R. A. F. & Loureiro, A. A. F. (2008), On the design of heterogeneous sensor networks based on small world concepts, *in* 'MSWiM '08: Proceedings of the 11th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems', ACM, pp. 309–314.
- Helmy, A. (2003), 'Small worlds in wireless networks', *IEEE Communications Letters* **7**(10), 490–492.
- Holme, P. & Kim, B. (2002), 'Growing scale-free networks with tunable clustering', *Physical Review E* **65**(026107).
- Hui, K. Y. K., Lui, J. C. S. & Yau, D. K. Y. (2006), 'Small-world overlay p2p networks: construction, management and handling of dynamic flash crowds', *Comput. Netw.* **50**(15), 2727–2746.
- James, B. (1981), *Probabilidade: um Curso em Nível Intermediário*, Projeto Euclides, Instituto de Matemática Pura e Aplicada, Rio de Janeiro.
- Janson, S., Luczak, T. & Rucinski, A. (1999), *Random Graphs*, 1st edn, John Wiley.
- Keller, E. F. (2005), 'Revisiting scale-free networks', *BioEssays* **27**(10), 1060–1068.
- Kleinberg, J. (2000), The small-world phenomenon: an algorithm perspective, *in* 'STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing', ACM, New York, NY, USA, pp. 163–170.
- Kleinberg, J. M. (1999), 'Authoritative sources in a hyperlinked environment', *Journal of the ACM* **46**(5), 604–632.
URL: <http://dx.doi.org/10.1145/324133.324140>
- Krause, W., Glauche, I., Sollacher, R. & Greiner, M. (2004), 'Impact of network structure on the capacity of wireless multihop ad hoc communication', *Physica A - Statistical Mechanics and Its Applications* **338**(3-4), 633–658.
- Krause, W., Scholz, J. & Greiner, M. (2006), 'Optimized network structure and routing metric in wireless multihop ad hoc communication', *Physica A: Statistical Mechanics and Its Applications* **361**(2), 707–723.
- Leskovec, J., Backstrom, L., Kumar, R. & Tomkins, A. (2008), Microscopic evolution of social networks, *in* 'KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, New York, NY, USA, pp. 462–470.

- Leskovec, J., Kleinberg, J. & Faloutsos, C. (2007), 'Graph evolution: Densification and shrinking diameters', *ACM Transaction on Knowledge and Discovery from Data* **1**(1), 2.
- Li, L., Alderson, D., Doyle, J. & Willinger, W. (2005), 'Towards a theory of scale-free graphs: Definition, properties, and implications', *Internet Mathematics* **2**(4), 431–523.
- Liljeros, F., Edling, C., Amaral, L., Stanley, E. & Åberg, Y. (2001), 'The web of human sexual contacts', *Nature* **411**, 907–908.
- Maia, G., Guidoni, D. L., Aquino, A. L. & Loureiro, A. A. (2009), Improving an over-the-air programming protocol for wireless sensor networks based on small world concepts, in 'MSWiM '09: Proceedings of the 12th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems', ACM, New York, NY, USA, pp. 261–267.
- McGlohon, M., Akoglu, L. & Faloutsos, C. (2008), Weighted graphs and disconnected components: patterns and a generator, in 'KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, New York, NY, USA, pp. 524–532.
- Milgran, S. (1967), 'The small world problem', *Psychology Today* pp. 60–67.
- Mitzenmacher, M. (2004), 'A brief history of generative models for power law and log-normal distributions', *Internet Mathematics* **1**(2), 226–251.
- Møller, J. & Waagepetersen, R. P. (2007), 'Modern statistics for spatial point processes', *Scandinavian Journal of Statistics* **34**(4), 643–684.
- Motter, A., Nishikawa, T. & Lai, Y. (2002), 'Range-based attacks on links in scale-free networks: Are long-range links responsible for the small-world phenomenon?', *Physical Review E* **66**(065103).
- Nakamura, E. F., Loureiro, A. A. F. & Frery, A. C. (2007), 'Information fusion for wireless sensor networks: Methods, models, and classifications', *ACM Computing Surveys* **39**(3), 9.
- Nakamura, E. F., Ramos, H. S., Villas, L. A., de Oliveira, H. A. B. F., de Aquino, A. L. L. & Loureiro, A. A. F. (2009), 'A reactive role assignment for data routing in event-based wireless sensor networks', *Computer Networks* **53**(12), 1980–1996.
- Newman, M. (2003), 'The structure and function of complex networks', *SIAM Review* **45**(2), 167–256.
- Newman, M. E. (2001), 'The structure of scientific collaboration networks.', *Proc Natl Acad Sci U S A* **98**(2), 404–409.
- Newman, M. E. J. & Watts, D. J. (1999), 'Scaling and percolation in the small-world network model', *Physical Review E* **60**(6), 7332–7342.

- Oliveira, E. M. R., Ramos, H. S. & A.F. Loureiro, A. (2010), Centrality-based routing for wireless sensor networks, in 'ICCCN 2010 Track on Wireless Networks and Emerging Technologies (WNET)', Zurich, Switzerland. Artigo submetido, em processo de avaliação.
- Pásztor, B., Motolla, L., Mascolo, C., Picco, G. P., Ellwood, S. & Macdonald, D. (2010), Selective reprogramming of mobile sensor networks through social community detection, in 'Proceedings of the 7th European Conference of Wireless Sensor Networks (EWSN)', Vol. 5970 of *Lecture Notes in Computer Science*, pp. 178–193.
- R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org>
- Ramos, H. S., Frery, A. C., Almiron, M. G., Oliveira, E. M. & Loureiro, A. A. F. (2010), 'A new topological measure of centrality and its application to the energy hole problem in wireless sensors networks', *Special Issue of Computer Communications on Complex Networks*. Artigo submetido, em processo de avaliação.
- Robert, C. P. & Casella, G. (2000), *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York.
- Sarshar, N. & Roychowdhury, V. (2004), 'Scale-free and stable structures in complex ad hoc networks', *Physical Review E* **69**(026101).
- Shimbel, A. (1953), 'Structural parameters of communication networks', *Bulletin of Mathematical Biology* **15**(4), 501–507.
URL: <http://dx.doi.org/10.1007/BF02476438>
- Siganos, G., Faloutsos, M., Faloutsos, P. & Faloutsos, C. (2003), 'Power laws and the AS-level internet topology', *IEEE-ACM Transactions on Networking* **11**(4), 514–524.
- Solomonoff, R. & Rapoport, A. (1951), 'Connectivity of random nets', *Bulletin of Mathematical Biophysics* **13**, 107–117.
- Tselishchev, Y., Boulis, A. & Libman, L. (2010), Experiences and lessons from implementing a wireless sensor network mac protocol in the castalia simulator, in 'IEEE Wireless Communications & Networking Conference 2010 (WCNC 2010)', Sydney/Australia.
- Valente, T. W. & Foreman, R. K. (1998), 'Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network', *Social Networks* **20**(1), 89–105.
URL: <http://www.sciencedirect.com/science/article/B6VD1-3SX6NDT-5/2/6195078c6b2163bb7bc8a98caf174973>
- Vázquez, A. (2003), 'Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations', *Physical Review E* **67**(5), 056104+.

- Wang, X. F. & Chen, G. (2003), 'Complex networks: small-world, scale-free and beyond', *Circuits and Systems Magazine, IEEE* **3**(1), 6–20.
- Watts, D. J. & Strogatz, S. H. (1998), 'Collective dynamics of 'small-world' networks', *Nature* **393**(6684), 440–442.
- Yen, L. H. & Cheng, Y. M. (2005), 'Clustering coefficient of wireless ad hoc networks and the quantity of hidden terminals', *IEEE Communications Letters* **9**(3), 234–236.
- Zhou, H. & Lipowsky, R. (2005), 'Dynamic pattern evolution on scale-free networks', *Proceedings of the National Academy of Sciences USA* **102**(29), 10052–10057.
- Zhou, S. & RJ, R. M. (2004), 'The rich-club phenomenon in the Internet topology', *IEEE Communications Letters* **8**(3), 180 – 182.
- Zuniga, M., Avin, C. & Hauswirth, M. (2010), Querying dynamic wireless sensor networks with non-revisiting random walks, in 'Proceedings of the 7th European Conference of Wireless Sensor Networks (EWSN)', Vol. 5970 of *Lecture Notes in Computer Science*, pp. 49–64.