UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

MATHEUS PALHARES VIANA

Caracterização e modelagem de redes biológicas
geográficas

São Carlos
2011

MATHEUS PALHARES VIANA

Caracterização e modelagem de redes biológicas
geográficas

*Tese de Doutorado apresentada ao Programa
de Pós-Graduação em Física do Instituto de
Física de São Carlos da Universidade de São
Paulo, para obtenção do título de Doutor em
Ciências.*

*Área de concentração: Física Aplicada
Opção: Física Computacional*

*Orientador: Prof. Dr. Luciano da Fontoura
Costa*

Versão original

São Carlos
2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTES TRABALHOS, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Viana, Matheus P.

Caracterização e modelagem de redes biológicas geográficas/
Matheus Palhares Viana; orientador Luciano da Fontoura Costa –
São Carlos, 2011.

103p.

Tese (Doutorado em Ciência - Área de concentração: Física
Aplicada - Opção: Física Computacional) - Instituto de Física de
São Carlos da Universidade de São Paulo.

1. Rede biológica. 2. Rede geográfica. 3. Teoria de grafo. 4.
Visualização 3D. 5. Processamento de imagem. I. Título.

FOLHA DE APROVAÇÃO

Matheus Palhares Viana

Tese apresentada ao Instituto de Física de
São Carlos da Universidade de São Paulo
para obtenção do título de Doutor em
Ciências.

Área de Concentração: Física Aplicada -
Opção: Física Computacional.

Aprovado(a) em: 23.03.2011

Comissão Julgadora

Prof(a). Dr(a). Luciano da Fontoura Costa

Instituição: IFSC/USP

Assinatura

Prof(a). Dr(a). Luiz Antonio Baccalá

Instituição: EP/USP

Assinatura

Prof(a). Dr(a). André Carlos Ponce de Leon Ferreira de Carvalho

Instituição: ICMC/USP

Assinatura

Prof(a). Dr(a). Roberto de Alencar Lotufo

Instituição: UNICAMP

Assinatura

Prof(a). Dr(a). Alexandre Gonçalves Evsukoff

Instituição: UFRJ

Assinatura

Agradecimentos

Gostaria de desejar meus sinceros votos de agradecimentos

- ao meu orientador Prof. Dr. Luciano da Fontoura Costa, pela dedicação e amizade ao longo destes inúmeros anos.
- à minha querida família, pelo amor, amizade, confiança e respeito.
- aos meus grandes amigos Gerson e Ivan.
- aos meus amigos de grupo, Bruno, João, Lucas, Renato², Mauro, Diego, André, Débora, Mônica e Krissia.
- à FAPESP por ter fomentado minha pesquisa.
- à Erica, por tudo.

REDUCTIONISM is the most natural thing in the world to grasp. It's simply the belief that "a whole can be understood completely if you understand its parts, and the nature of their 'sum'." No one in her left brain could reject reductionism.

-Douglas Hofstadter, Gödel, Escher, Bach: an Eternal Golden Braid

Resumo

VIANA, M P. **Caracterização e modelagem de redes biológicas geográficas**. 2011. 103p. Tese (Doutorado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

Nesta tese apresentamos uma metodologia de mapeamento capaz de gerar representações em termos de grafos para sistemas biológicos de conectividade complexa. Tais sistemas são inicialmente armazenados na forma de imagens digitais e em seguida submetidos a um pré-processamento com objetivo de padronizar as imagens. As imagens pré-processadas são então utilizadas para gerar modelos tridimensionais dos sistemas de interesse. Um algoritmo de propagação de rótulos é utilizado para extrair os esqueletos dos modelos volumétricos e estes esqueletos são por fim, representados por um grafo, composto por vértices e arestas. Os vértices e arestas desse grafo armazenam propriedades do sistema original, como posição, comprimento e diâmetro, bem como as características topológicas de tais sistemas. Finalmente, os grafos resultantes são estudados através da teoria das redes complexas, dentro de um contexto específico para cada sistema. Nossos procedimentos foram aplicados com sucesso a diferentes sistemas biológicos, como artérias carótidas, árvores arteriais, estruturas mitocondriais e poros em amostras de solo.

Palavras-chave: *Rede biológica. Rede geográfica. Teoria de grafo. Processamento de imagem. Visualização 3D.*

Abstract

VIANA, M P. **Characterization and modelling of biological networks**. 2011. 103p. Tese (Doutorado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

In the present work, we developed a mapping methodology able to build a graph representation for biological branched systems. Initially, such systems are stored as digital images and then they undergo a pre-processing in order to standardize the images. The pre-processed images are used to build tridimensional models of the interested systems. A label-propagation algorithm is used to extract the skeleton from the volumetric models and these skeletons are then represented by a graph, composed by nodes and edges. The nodes and edges of these graphs store properties of the original system, such as spatial position, lengths and diameter, as well as the topological features of such systems. Finally, the graphs are studied by using the complex networks theory within a specific context for each system. Our procedures were applied successfully to different biological systems, such as carotid artery, arterial trees, mitochondrial structure and pores in soil samples.

Keywords: *Biologic network. Geographical network. Graph theory. Image processing. 3D visualization.*

Lista de Símbolos

$\mathcal{V}$	Conjunto de vértices da rede
N	Número de vértices da rede (número de elementos do conjunto $\mathcal{V}$)
$\mathcal{E}$	Conjunto de conexões da rede
E	Número de conexões da rede (número de elementos do conjunto $\mathcal{E}$)
$\mathbf{A}$	Matriz de adjacência
k_i	Grau do vértice i
ki_i	Grau de entrada do vértice i
ko_i	Grau de saída do vértice i
P_k	Probabilidade de que um vértice tenha grau k
c_i	Coefficiente de clusterização do vértice i
b_i	Centralidade do vértice i
$a_{i,j}$	Elemento da matriz de adjacência
$w_{i,j}$	Peso da aresta (i, j)
$R_{i,j}$	Resistência da aresta (i, j)
$l_{i,j}$	Comprimento de arco da aresta (i, j)
$r_{i,j}$	Distância Euclideana entre i e j
$\ell_{i,j}$	Comprimento do menor caminho que liga i e j
N_k	Número de vértices com grau k
$\Xi(k)$	Probabilidade de que um vértice de grau k receba uma conexão
$\Xi(i, j)$	Probabilidade de que uma conexão seja criada entre i e j
$P_{k k'}$	Probabilidade conjunta de que uma aresta conectada a um vértice de grau k também esteja conectada a um vértice de grau k'
knn_i	Grau médio dos primeiros vizinhos de i
$knn(k)$	Média do grau médio dos primeiros vizinhos dos vértices com grau k
$p(i, j)$	Probabilidade de que um agente localizado em i alcance j após um passo
$p_\ell(i, j)$	Probabilidade de que um agente localizado em i alcance j após ℓ passos
$\mathcal{A}_i$	Acessibilidade do vértice i
σ_i	Entropia de Shannon do vértice i
τ	Tempo de exploração
Ω_i	Conjunto de vértices acessíveis a partir de i
β	Probabilidade de infecção no modelo SIR
μ	Probabilidade de recuperação no modelo SIR

Sumário

1	Introdução	19
1.1	Motivação e Objetivos	21
1.2	Organização da Tese	21
2	Redes complexas	23
2.1	Introdução	24
2.2	Modelos	27
2.2.1	Rede Regular	28
2.2.2	Modelo ER	28
2.2.3	Modelo WS	29
2.2.4	Modelo BA	30
2.2.5	Modelo GE	31
2.3	Correlações em Redes	32
2.4	Caminhadas Aleatórias em Redes	34
2.5	Acessibilidade	36
2.5.1	Detecção de bordas	39
2.6	Propagação de doenças em redes geográficas	42
3	Mapeando estruturas biológicas em redes	45
3.1	Pré-processamento	46
3.2	Propagação de rótulos	48
3.2.1	Bifurcações	50
3.2.2	Loops	52
3.2.3	Limitações da propagação de rótulos	53
3.3	Aperfeiçoamento do software	54
3.3.1	Regularização utilizando divergente do campo gradiente	54
3.3.2	Afinamento	56
3.3.3	Diâmetro	58
4	Aplicações	61
4.1	Complexidade geométrica de artérias carótidas	61
4.2	Diâmetro ao longo da árvore arterial de embriões de galinha	65

4.3	Estrutura mitocondrial	67
4.4	Topologia e dinâmica das redes de poros do solo	72
4.5	Outros estudos relacionados	77
4.5.1	Obtenção de diversidade topológica através da otimização da acessibilidade	77
4.5.2	A rede de teoremas da Wikipédia	80
4.5.3	Definição de borda em redes complexas	82
5	Considerações finais	85
5.1	Contribuições	86
5.1.1	Software	86
5.1.2	Complexidade geométrica de artérias carótidas	86
5.1.3	Diâmetro ao longo da árvore arterial de embriões de galinha	86
5.1.4	Estrutura mitocondrial	87
5.1.5	Topologia e dinâmica das redes de poros do solo	87
5.1.6	Outros estudos relacionados	88
5.2	Desenvolvimentos futuros	88
	Referências	90
	Apêndice 1: Decomposição <i>k-shell</i>	102

1 Introdução

Na última década, pesquisadores têm utilizado a teoria de redes complexas para desvendar a maneira como os elementos de diferentes sistemas complexos, naturais e artificiais, se relacionam (1–3, 23). Nesta abordagem, a união de todos os elementos de um sistema e as relações entre eles são descritas por meio de uma rede, na qual os vértices representam os elementos básicos e as arestas representam suas interações. O estudo da organização topológica dessas redes tem sido o principal objetivo de grande parte dos trabalhos desta área. Por exemplo, verificou-se que algumas redes de natureza completamente diferentes, como a rede WWW, redes de interação de proteínas e redes de citações científicas possuem o mesmo padrão de distribuição de grau, i.e. número de vizinhos (veja seção 2.1) (2). Estas redes são chamadas livres de escala, já que nelas a distribuição de grau segue uma lei de potência que não possui uma escala típica bem definida (3). Outras quantidades, como coeficiente de clusterização e centralidade apresentam padrões similares nestas redes (4), remetendo certa universalidade as características topológicas desses sistemas. Nos últimos anos, o foco dos trabalhos tem mudado de maneira gradual, do estudo puramente topológico das redes complexas para o estudo dos processos dinâmicos que ocorrem nessas redes. Ao estudar alguns processos dinâmicos, os pesquisadores observaram que resultados inesperados podem emergir graças ao comportamento coletivo dos elementos da rede. Foi feita então a ponte entre a teoria das redes complexas e os conceitos físicos de sistemas complexos e criticalidade (5). Como resultado, evidências do paradigma *estrutura-função* têm sido observadas. Por exemplo, puderam ser determinar as características estruturais necessárias para que em determinadas redes, (i) fenômenos de sincronização sejam observados (6) e (ii) a disseminação de uma infecção possa ser controlada (2).

É interessante observar que mesmo com o crescente aumento do número de trabalhos envolvendo processos dinâmicos em redes, ainda é pequeno o número de trabalhos que abordam redes geográficas, em especial redes geográficas tridimensionais baseadas em sistemas biológicos. Entendemos por redes geográficas aquelas em que cada vértice possui uma coordenada espacial associada (veja seção 2.1). A presente tese, tem como principal objetivo desenvolver um mecanismo de mapeamento capaz de representar sistemas biológicos ramificados através de redes, bem como estudar estas redes utilizando os conceitos da teoria de redes complexas. Observe que o surgimento de redes geográficas neste tipo de descrição é imediato, uma vez que, durante o mapeamento os canais dos sistemas reais são mapeados em arestas e as intersecções entre os canais são mapeadas em vértices, os quais recebem as mesmas coordenadas espaciais das intersecções que os originaram.

Redes geográficas compreendem uma classe importante das redes complexas, as quais estão imersas em um espaço físico 2D ou 3D. Elas são úteis para a descrição de diferentes tipos de sistemas reais, como sistemas de transporte (7–9, 104), sistemas de galerias subterrâneas (10), sistemas de canais nos ossos (11, 12), entre outros. As redes que representam esses sistemas compartilham de uma propriedade comum a grande parte das redes geográficas: elas são altamente esparsas. Isso ocorre porque as restrições impostas pelo espaço físico limitam a conectividade entre os vértices (9). De fato, o número de arestas nessas redes é da ordem do número de vértices. Outra característica desses sistemas é que em geral, o surgimento de uma aresta é um fenômeno que leva em conta o balanço entre dois fatores: (i) o benefício que a nova aresta irá trazer pra rede frente a alguma propriedade de interesse, e (ii) o custo que ela terá ao sistema como um todo. Consequentemente, a ocorrência de arestas longas em redes geográficas torna-se um evento raro. Por esse motivo, o diâmetro (D) dessas redes (veja seção 2.1) é similar ao das redes regulares, no qual $D \sim N^{1/d}$, onde d é a dimensão do sistema e N o número de vértices. Por esse motivo, dizemos que as redes regulares não apresentam efeito *mundo pequeno*, ao contrário da redes livres de escala citadas acima, onde o diâmetro escala com o número de vértices de acordo com a lei $D \sim \log \log N$.

A ausência de leis de potência e efeito *mundo pequeno* em grande parte dos sistemas geográficos podem causar a errônea impressão de que estes são menos complexos do que aqueles sistemas descritos por redes livres de escala. Entretanto, as redes geográficas apresentam diversas características não-triviais, como correlações (13) e topologias próximas ao limiar de percolação (14), que não são explicadas facilmente. Graças a essas características, cada vez mais tem crescido o número de trabalhos abordando sistemas geográficos reais por meio da teoria de redes. Entretanto, quase não há trabalhos que explorem possíveis estratégias de mapeamento dos sistemas reais em redes geográficas, nem mesmo trabalhos que descrevam de maneira satisfatória como tal mapeamento foi realizado. Muitos deles limitam-se a dizer que uma confluência entre canais no sistema real é representada por vértice na rede, enquanto os canais são representados pela próprias arestas. Esta descrição minimalista esconde todo arsenal de ferramentas necessárias para converter um determinado sistema biológico, muitas vezes armazenado na forma de imagens digitais, em uma rede. Dada a importância desta etapa, acreditamos que é necessário um estudo sobre este tema, o qual é um dos temas desta tese.

De maneira geral, as informações estruturais a respeito dos sistemas biológicos podem ser obtidas a partir de diferentes técnicas de imageamento, como por exemplo *microtomografia computadorizada*, *microscopia de convolução* e *sequenciamento histológico*. Como resultado, todas essas técnicas produzem um conjunto de imagens digitais que pode ser então armazenado em um computador. Antes que essas imagens possam ser utilizadas para a geração de redes geográficas, é necessário aplicar nelas uma diversidade de técnicas de processamento de imagens com o objetivo de padronizar, eliminar ruído e detectar os

objetos de interesse. Esta tese consiste também na apresentação das rotinas necessárias para tal processamento.

Em seguida, as imagens são utilizadas para criação de modelos tridimensionais dos dados biológicos. As técnicas computacionais necessárias nesta etapa serão abordadas com poucos detalhes nesta tese, uma vez que a biblioteca de visualização utilizada aqui possui uma documentação completa e acessível (15). Os modelos tridimensionais podem então ser mapeados em redes geográficas. Iremos descrever todo procedimento necessário para realizar tal mapeamento. Este procedimento foi aplicado com sucesso em uma grande variedade de sistemas reais (12, 16–20, 104), os quais serão apresentados na seção de resultados desta tese.

1.1 Motivação e Objetivos

A principal motivação desta tese deve-se ao fato de que grande parte dos trabalhos sobre redes geográficas já desenvolvidos não descrevem com precisão a metodologia de mapeamento utilizada na geração das redes. Tendo em vista que este é um passo fundamental para caracterização dos sistemas reais, pretendemos implementar e validar uma técnica eficiente e confiável para realizar tal tarefa. Portanto, temos como principal objetivo *apresentar um conjunto de rotinas que seja capaz de gerar representações em termos de redes para uma ampla quantidade de sistemas biológicos ramificados*. Em seguida, iremos estudar cada uma dessas representações dentro de um contexto específico, capaz de ajudar-nos a entender melhor as propriedades do sistema biológico real.

1.2 Organização da Tese

No primeiro capítulo apresentamos uma introdução aos conceitos de redes complexas utilizados ao longo da tese. O segundo capítulo é dedicado a questão da visualização e processamento dos dados biológicos, bem como ao procedimento de mapeamento por nós desenvolvido. Aplicações e resultados são apresentados no terceiro capítulo. No último capítulo são apresentadas as considerações finais deste trabalho.

2 Redes complexas

Enquanto os pixels de uma imagem e os átomos em um cristal podem ser descritos por malhas regulares, muitos sistemas reais apresentam tamanha desordem que, para suas descrições, se faz necessário o uso de uma ferramenta mais poderosa e abrangente. Tais desordem foram primeiro observadas na década de 70 na interação entre pessoas através do experimento do sociólogo americano Stanley Milgram (21) que observou que distância topológica entre quaisquer duas pessoas é de aproximadamente seis indivíduos. Este número é muito menor do que o esperado caso a rede de interação fosse descrita por uma malha regular. Já no início deste século, novas observações foram realizadas em sistemas como a *World Wide Web* (22), a *Internet* (23) e redes metabólicas (24). Foi observado nestes sistemas que a distribuição de probabilidade de que um vértice tenha k vizinhos segue uma lei de potência, outro fato que não pode ser explicado quando utilizamos uma estrutura de malha regular como modelo.

Os resultados encontrados nestes trabalhos deram claros indícios de que a teoria de redes conhecida até aquele momento era insuficiente para explicar os dados provenientes dos sistemas reais. Nem mesmo os modelos de redes aleatórias desenvolvidos pelos matemáticos Paul Erdős e Alfréd Rényi (modelo ER) (25, 26) podiam explicar os dados empíricos em sua totalidade. No fim dos anos 90, teve início então, uma revolução no estudo das *redes complexas* ou seja, no estudo das redes que apresentavam características que não podiam ser explicadas pelos modelos regulares ou ER. Neste período, muitos modelos foram desenvolvidos, sendo que o primeiro foi proposto por Watts e Strogatz (modelo WS) (27) com o objetivo de reproduzir as características qualitativas do experimento de Stanley Milgram. Neste modelo, os autores introduziram uma probabilidade de redirecionamento com a qual era possível interpolar os dois modelos clássicos: regular e ER. Quando esta probabilidade é aproximadamente zero, a ordem do modelo regular prevalece, enquanto que para valores próximos a unidade a desordem domina e o modelo ER emerge. Para valores intermediários, as características da rede *small-world*, como ficou conhecida, correspondem aquelas observadas no experimento de Milgram.

Ainda que os resultados do experimento de Milgram tenham sido explicados pelo modelo *small-world*, ele não era capaz de reproduzir as leis de potência recentemente observadas na distribuição de grau das redes investigadas pelos irmãos Faloutsos (22) e por A.-L. Barabási e seus colaboradores (23). Em 1999, Albert A.-L. Barabási e R. Albert propuseram um modelo dinâmico (modelo BA) (28) que continha um ingrediente especial, a *anexação preferencial*, que diz basicamente que a probabilidade de um vértice receber conexões é proporcional ao número de conexões que ele já tem. Este ingrediente aliado ao

fato da rede evoluir ao longo do tempo com a inserção de novos vértices, foram capazes de explicar as leis de potência observadas nos sistemas reais.

Nos anos que se sucederam, grande parte dos trabalhos na área das redes complexas foram desenvolvidos com objetivo de caracterizar a estrutura topológica de novas redes reais ou então, de propor modelos capazes de descrever melhor determinados sistemas. Neste período muitos sistemas de natureza completamente diferentes foram estudados sob o mesmo ponto de vista, o da teoria de redes complexas. Desde redes de citações científicas (29, 30), redes de contatos sexuais (31), redes de interação entre proteínas (32), redes de cadeia alimentar (33), até redes de transporte aéreo (34) e redes de galerias construídas por formigas (10). Verificou-se que muitos desses sistemas apresentam características comuns, indicando certa universalidade em suas organizações e/ou fatores de otimização durante suas construções. Já nos últimos anos, o foco dos trabalhos tem mudado de maneira gradual da pura descrição topológica para o estudo dos processos dinâmicos que ocorrem nessas redes (35). Essa mudança de paradigma tem ajudado os pesquisadores a entender melhor a questão *forma-função* acerca das redes complexas ou seja, como a topologia de um sistema afeta os processos dinâmicos que ocorrem nele.

Muitas dessas redes não possuem quaisquer relação com o espaço físico real. A visualização desses sistemas é realizada por meio de algoritmos que usam os vínculos de adjacência para gerar coordenadas para os vértices. Entretanto, existe uma classe importante de redes na qual os vértices possuem posições espaciais bem definidas. São as *redes geográficas* e o número de trabalhos que exploram as características desse tipo de sistema tem aumentado de maneira expressiva (13). Neste trabalho, iremos lidar preferencialmente com redes geográficas, já que elas são ideais para representar sistemas biológicos imersos no espaço 3D. Podemos considerar que a presente tese está localizada na fronteira entre o estudo puramente topológico e o estudo dos aspectos dinâmicos das redes geográficas. Acreditamos que nossa maior contribuição refere-se a questão de como mapear sistemas biológicos armazenados na forma de imagens em redes geográficas, e nas questões relativas a forma-função de tais redes. No restante desta seção faremos uma breve revisão dos conceitos básicos da teoria de redes, bem como dos modelos clássicos citados acima e utilizados ao longo deste trabalho. Alguns conceitos mais recentes como a medida de acessibilidade e o seu uso para detecção de bordas em redes serão abordados com mais detalhes.

2.1 Introdução

Uma rede $\mathcal{G}(\mathcal{V}, \mathcal{E})$ é basicamente um conjunto de N vértices $\mathcal{V} = \{1, 2, \dots, i, i+1, \dots, N\}$ e E arestas $\mathcal{E} = \{(i, j) \mid i \in \mathcal{V}, j \in \mathcal{V}\}$. Normalmente, esta rede é descrita por uma matriz,

chamada *matriz de adjacência* $\mathbf{A}$, na qual os elementos $a_{i,j}$ assumem valores binários que indicam a presença ($a_{i,j} = 1$) ou ausência ($a_{i,j} = 0$) de aresta entre os vértices i e j . Em geral, as matrizes de adjacência são simétricas, exceto em redes direcionadas quando devemos levar em conta a direção das arestas. Existem certas quantidades que podem ser utilizadas para caracterizar os vértices de uma rede, sendo que a mais fundamental é o seu *grau* k , ou seja, a medida do seu número de vizinhos. Na figura 2.1 mostramos uma rede sintética com $N = 20$ vértices e $E = 27$ arestas. Nela, o vértice 4 tem grau $k_4 = 2$ enquanto o vértice 8 tem $k_8 = 5$. No caso das redes direcionadas, o número de vizinhos é contabilizado levando em conta as arestas que chegam e que saem de um dado vértice. Por exemplo, as arestas que saem definem o *grau de saída*, denotado por ko_i . As arestas que chegam definem o grau de entrada, denotado por ki_i . Com exceção da seção 2.5.1, iremos trabalhar apenas com redes não direcionadas ao longo desta tese. A distribuição de probabilidade de que um dado vértice tenha grau k é chamada *distribuição de grau* e será denotada por P_k . Outra quantidade definida para cada vértice é o *coeficiente de clusterização*, que mede a fração de vizinhos que também são vizinhos entre si. Esta quantidade é expressa por:

$$c_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (2.1)$$

onde E_i é o número de conexões compartilhadas pelos vizinhos do vértice i . Como exemplo, os vértices 5 e 8 da figura 2.1 possuem coeficiente de clusterização $c_5 = 0.2$ e $c_8 = 0.4$, o que nos diz que os vizinhos do vértice 8 estabelecem mais conexões entre si do que os vizinhos do vértice 5. Os vértices também podem ser caracterizados por coordenadas espaciais nos casos em que a rede está inserida em um espaço d dimensional. Nestes casos, dizemos que a rede é do tipo geográfica. Este tipo de rede compreende uma grande parte das aplicações atuais da área das redes complexas e uma boa revisão sobre este tema pode ser encontrada em (13).

No estudo de redes, sejam elas geográficas ou não, é comum nos depararmos com o conceito de caminhos. Chamamos de *caminho de comprimento* ℓ qualquer sequência de vértices $\{v_1, v_2, \dots, v_\ell\}$ tal que $v_i \in \mathcal{V}$ e $(v_i, v_{i+1}) \in \mathcal{E}$. Utilizamos este conceito para definir a métrica de distância topológica em redes, de tal maneira que a distância entre dois vértices quaisquer i e j é dada pelo comprimento do menor caminho que os une e é denotada por $\ell_{i,j}$. A distância média entre um vértice i e todos os outros da rede é chamada *comprimento do menor caminho médio* (CMCM) e é denotada por ℓ_i . Tomando a média de ℓ_i sobre todos os vértices i , obtemos o *diâmetro da rede* $D(\mathcal{G})$, que fornece uma estimativa de quantas arestas existem em média entre dois vértices quaisquer da rede. Na rede da figura 2.1, o CMCM do vértice 2 é $\ell_2 = 3.85$, enquanto o diâmetro da rede é $D = 3.485$.

Na métrica topológica, cada aresta sempre contribui com uma unidade para o com-

Figura 2.1 – Rede sintética com $N = 20$ vértices e $E = 27$ arestas. Nesta rede, o menor caminho entre os vértices 2 e 17 possui comprimento $\ell_{2,17} = 5$ e é dado por $\{2, 5, 6, 9, 16, 17\}$. O comprimento médio do menor caminho para o vértice 2 é $\ell_2 = 3.85$ e o diâmetro da rede é $D = 3.485$.

primento de um caminho. Entretanto, em determinadas situações, podemos definir um tipo diferente de métrica para o qual o comprimento do caminho perde a relação com o número de arestas entre dois vértices. Essas situações são aquelas em que estamos trabalhando com *redes com peso*, nas quais cada aresta (i, j) possui um valor associado $w_{i,j}$. Em geral, esses valores representam uma característica do sistema real que a rede representa. Por exemplo, os pesos podem estar associados a distância euclidiana (no caso de redes geográficas), o fluxo de pessoas ou a corrente máxima entre diferentes vértices da rede. Nestes casos, dizemos que o comprimento de um caminho corresponde a soma dos pesos das arestas desse caminho. Na figura 2.2 apresentamos uma versão com pesos da rede da figura onde é exemplificado este conceito. Nesta figura, o peso de uma aresta é proporcional à sua espessura e neste caso, o menor caminho entre os vértices 2 e 17 passa por 10 vértices e seu comprimento é $\ell_{2,17} = 13$.

Determinar os menores caminhos entre diferentes pontos de uma rede é uma tarefa de grande importância, já que em geral o tráfego de informações ocorre através das menores distâncias (36, 37) com objetivo de minimizar o tempo e os custos envolvidos. Sendo assim, podemos imaginar que o número de menores caminhos que passam por um determinado vértice nos diz algo a respeito da sua importância na rede. Essa ideia foi introduzida em 1980 por L. C. Freeman que estabeleceu uma medida para capturar esta importância (38). Tal medida ficou conhecida como *centralidade* (betweenness). Matematicamente, a centralidade do vértice i , b_i , é definida pela equação 2.2, onde $\rho(m, i, q)$ é o número de menores caminhos que passam por i e que ligam o vértice m ao vértice q , enquanto $\rho(m, q)$ é o número de menores caminhos que ligam m e q . Outras versões para as medidas de centralidade podem ser encontradas em (36, 37, 39).

$$b_i = \frac{\sum_{m,q} \rho_\ell(m, i, q)}{\sum_{m,q,i} \rho_\ell(m, i, q)} \quad (2.2)$$

Figura 2.2 – Versão com pesos da rede da figura 2.1. Aqui os pesos são proporcionais à espessura das arestas. Diferente do que vemos na figura 2.1, nesta versão com peso o menor caminho entre os vértices 2 e 17, dado por $\{2, 5, 6, 9, 8, 10, 11, 12, 16, 17\}$, passa por 10 vértices. Além disso, o seu comprimento é $\ell_{2,17} = 13$, já que passa por oito arestas com peso unitário e pela aresta $(16, 17)$, cujo peso é $w_{16,17} = 5$.

Em 2004, M. Barthélemy (40) estudou a correlação existente entre o grau e a centralidade dos vértices em redes com diferentes topologias. Seus resultados indicaram a existência de uma relação do tipo $b_i \sim k_i^\nu$ entre essas grandezas, onde o expoente ν não é universal e depende fortemente da topologia da rede. Enquanto $\nu = 2$ para os grafos aleatórios do modelo ER, este expoente pode assumir valores diferentes ($\nu \leq 2$) nas redes livres de escala geradas pelo o modelo BA. Isso indica que, apesar de ser completamente local, o grau de um determinado vértice está intimamente relacionado à função que ele desempenha globalmente. É interessante então buscar por generalizações desta quantidade que sejam capazes de capturar ainda mais informações sobre as vizinhanças de cada vértice. Nas seções seguintes veremos como a medida de acessibilidade pode ser utilizada para este propósito. Porém, antes daremos mais detalhes a respeito de cada um dos modelos de rede discutidos no decorrer desta tese.

2.2 Modelos

Atualmente, existem na literatura diversos modelos para geração de redes complexas. Embora capturem características gerais de muitos sistemas físicos, como leis de potência ou efeito *small-world*, a alta complexidade destes modelos os tornam verdadeiramente úteis para descrever apenas sistemas específicos, como a rede de proteínas ou a Internet. Nesta seção iremos revisar alguns modelos clássicos utilizados para geração de redes. Maiores detalhes e outros modelos podem ser encontrados em (35, 41).

2.2.1 Rede Regular

O modelo de rede regular, também conhecido por malha, ou grid, talvez seja um dos mais antigos modelos de rede já estudados. Ele é útil para modelar a posição de átomos em um sólido cristalino ou os pixels em uma imagem digital, por exemplo. Basicamente, trata-se de uma rede geográfica onde N vértices ocupam posições discretas e igualmente espaçadas em um espaço de d dimensões. Cada um desses vértices estabelece k_0 conexões com seus vizinhos mais próximos. Sendo assim, a distribuição de probabilidade de que um vértice desta rede tenha k vizinhos, é dada por:

$$P_k = \delta_{k,k_0}, \quad (2.3)$$

onde $\delta_{k,k_0} = 1$ se $k = k_0$ e 0 nos outros casos. É importante notar que estamos considerando uma rede infinita ou na presença de condições periódicas de contorno, já que efeitos de borda não são levados em conta na distribuição de P_k . É possível mostrar que o diâmetro das redes regulares segue a relação:

$$D \sim N^{1/d}, \quad (2.4)$$

enquanto seu coeficiente de clusterização para $d = 1$ é dado por (considerando condições periódicas de contorno):

$$c = \frac{3(k_0 - 2)}{4(k_0 - 1)}. \quad (2.5)$$

2.2.2 Modelo ER

O modelo Erdős-Rényi (ER), como ficou posteriormente conhecido, foi na verdade proposto pelo matemático Edgar Gilbert em 1959 (42). Ainda no mesmo ano, os matemáticos Paul Erdős e Alfréd Rényi publicaram sua própria versão do modelo (43), que continha alguns aspectos diferentes daquele proposto inicialmente. Ao longo deste trabalho, vamos considerar basicamente o modelo proposto por Gilbert, que assume a existência de um conjunto de N vértices, inicialmente desconectados. Em seguida, cada uma das possíveis $N(N - 1)/2$ conexões, é criada com probabilidade p . Várias características desse modelo são conhecidas analiticamente, como por exemplo, é possível mostrar que a distribuição de grau é dado pela seguinte lei binomial:

$$P_k = \frac{(N - 1)!}{k!(N - 1 - k)!} p^k (1 - p)^{N-1-k}. \quad (2.6)$$

Além disso, o diâmetro das redes construídas segundo este modelo é proporcional ao logaritmo do número de vértices, ou seja:

$$D = \frac{\ln N}{\ln \langle k \rangle}. \quad (2.7)$$

É importante citar que este tipo de relação $D \sim \log N$ ficou conhecido como efeito *small-world*, já que ele representa o encurtamento das distâncias na rede, quando comparadas aos modelos regulares. O coeficiente de clusterização para redes ER é dado por:

$$c = \frac{\langle k \rangle}{N}. \quad (2.8)$$

Assim, vemos que o coeficiente de clusterização tende a zero para redes que possuem muitos vértices. Esta característica não era observada em redes reais, cujos coeficientes de clusterização são, em geral, elevados. Isso inviabilizou a criação de aplicações relevantes para este modelo muito embora, ele seja utilizado como comparação (modelo nulo) em diversos trabalhos até os dias atuais.

2.2.3 Modelo WS

Embora o diâmetro das redes ER explicasse algumas observações feitas sobre sistemas reais, a discrepância entre os valores esperados e observados para o coeficiente de clusterização permanecia inexplicada até 1998. Foi neste ano que Duncan J. Watts and Steven Strogatz propuseram um modelo que mantinha a dependência logarítmica do diâmetro com o número de vértices porém, com alto coeficiente de clusterização. O modelo inicia-se com uma rede regular com grau médio k_0 e em seguida, cada aresta é redirecionada aleatoriamente com probabilidade p . Quando $p = 0$, nada acontece e temos a própria rede regular. Quando $p = 1$ todas conexões são redirecionadas e o modelo ER é recuperado. Nestas redes, a distribuição de grau é dada por:

$$P_k = \sum_{n=0}^{f(k, k_0)} \frac{n!(1-p)^n p^{k_0/2-n} (pk_0/2)^{k-k_0/2-n}}{(k_0/2)!(n-k_0/2)!(k-k_0/2-n)!} e^{-pk_0/2}, \quad (2.9)$$

onde $f(k, k_0) = \min(k - k_0/2, k_0/2)$. Enquanto o diâmetro do modelo WS segue a lei logarítmica do modelo ER, o coeficiente de clusterização neste caso é:

$$c = \frac{3k_0(k_0 - 1)}{2k_0(2k_0 - 1) + 8pk_0^2 + 4p^2k_0^2}. \quad (2.10)$$

Foi observado que para valores moderados de p , da ordem de $p \sim N^{-d}$ as redes obtidas reproduziam bem as características dos sistemas reais estudados até aquele período.

2.2.4 Modelo BA

Foi em 1999 que novas observações colocaram em cheque a aplicabilidade do modelo WS. Graças aos irmãos Faloutsos (22) e a Albert-Lazlo Barabási e seus colaboradores (23) que extraíram medidas de diversas redes reais, mostrando que a distribuição de grau destas redes não eram previstas pela equação 2.9. Em particular os dados reais indicavam a ausência de um grau típico para as redes reais. Enquanto nos modelos RE, ER e WS, a distribuição de grau possui um pico bem definido, indicando que grande parte dos vértices possuem aproximadamente o mesmo número de vizinhos, na redes reais esta escala típica não era observada. Ao contrário, eles observaram que grande parte dos vértices possuíam poucos vizinhos e alguns poucos vértices (*hubs*) possuíam muitas conexões. No mesmo ano, A.-L. Barabási, R. Albert e H. Jeong propuseram um modelo (BA) (28) que continha um ingrediente especial, a anexação preferencial e que o tornava capaz de reproduzir as distribuições de grau das redes reais.

O modelo BA assume que a rede evolui ao longo do tempo e a cada instante um novo vértice é adicionado ao sistema. Inicialmente, um conjunto de m vértices é completamente conectado (todos com todos). No instante seguinte, um novo vértice é adicionado a rede e estabelece m conexões com os vértices já existentes. A probabilidade de que o novo vértice se conecte com um vértice existente de grau k é dada por $\Xi(k) = k/2mt$, onde t é o instante de tempo no qual o novo vértice foi inserido. Esta anexação preferencial privilegia a conexão dos vértices mais recentes com aqueles de maior grau. A distribuição de grau nas redes geradas pelo modelo BA é dada por:

$$P_k = \frac{2m(m+1)}{k(k+1)(k+2)} \sim k^{-3}. \quad (2.11)$$

A lei de potência da equação acima explica a falta de escala típica observada nas distribuições das redes reais. Graças a isso, essas redes ficaram conhecidas como *redes livres de escala*. Também foi mostrado que o diâmetro dessas redes é menor do que aqueles observados nos modelos ER e WS e por isso dizemos que essas redes apresentam efeito *ultra small-world*, sendo escrito como (44):

$$D \sim \ln \ln N. \quad (2.12)$$

O coeficiente de clusterização também tende a zero para este tipo de rede quando o número de vértices é muito grande, embora seja de maneira mais lenta do que o que é observado no modelo ER. Aqui a relação entre c e N é dada por:

$$c \sim \frac{(\ln N)^2}{N}. \quad (2.13)$$

Ao longo dos últimos anos, muitas adaptações do modelo BA original, bem como

novas propostas, foram desenvolvidas e aplicadas com sucesso para reproduzir sistemas específicos, como redes de aeroportos (45), redes de rodovias (46), redes de proteínas (47) entre outros. Para uma revisão sobre os modelos de crescimento, veja (35, 41), bem como as referências ali contidas.

2.2.5 Modelo GE

Em muitos sistemas reais modelados por redes geográficas, os vínculos impostos pelas restrições espaciais, bem como a posição relativa dos vértices, são fundamentais para determinar a topologia e guiar os processos dinâmicos que ocorrem em tais sistemas (13). Se imaginarmos que existe um custo por unidade de comprimento (9), associado a construção das conexões em uma rede geográfica, veremos que nessas redes, grande parte das conexões serão de curto alcance. Sendo assim, esperamos que o diâmetro da rede não escale com o logaritmo do número de vértices, como nos modelos ER, WS e BA, vistos anteriormente. É mais natural esperar um comportamento similar aquele observado nas redes regulares, onde $D \sim N^{1/d}$ (13). Esse é exatamente o resultado observado para diversas redes geográficas reais, como redes de rodovias, redes de distribuição de energia elétrica, redes de distribuição de água, redes de canais nos ossos, redes de poros presentes no solo, entre outras. É importante observar que existem algumas redes geográficas, como as redes de aeroportos e redes de transporte marítimo, nas quais, embora os vértices tenham posições geográficas (aeroportos e portos), as conexões (linhas aéreas e rotas marítimas) não são restritas por vínculos espaciais e portanto, o comportamento $D \sim \ln N$ é observado nesses casos (13).

Aqui iremos apresentar o modelo de rede geográfica proposto por Waxman (48) na qual a restrição espacial é levada em conta durante o processo de conexão dos vértices. Inicialmente, consideramos N vértices dispostos aleatoriamente em uma região d dimensional de volume Ω . Em seguida, os vértices i e j são conectados com probabilidade:

$$\Xi(i, j) = A e^{-r_{i,j}/r_0}, \quad (2.14)$$

onde $r_{i,j}$ corresponde a distância euclidiana entre os vértices i e j . A quantidade r_0 define um comprimento típico para as conexões e A controla a densidade total de conexões. A rede é finalizada quando todos os possíveis pares (i, j) são testados. Valores típicos para r_0 são da ordem $r_0 \sim \sqrt{\Omega/N}$, para o qual a rede se comporta como uma rede regular. Versões dinâmicas deste modelo, bem como outros modelos de redes geográficas, podem ser encontrados em (13).

2.3 Correlações em Redes

Estritamente falando, no limite termodinâmico ($N \rightarrow \infty$), nenhum dos modelos aleatórios descritos acima produz redes que apresentam correlações (41). Ao contrário do que se observa em redes reais, onde correlações estão presentes e possuem papel fundamental nos processos dinâmicos e no entendimento da organização hierárquica de determinados sistemas, como a internet (49, 50). O tipo de correlação mais comum é conhecido como *correlação de grau* e é frequentemente estudado pela distribuição de probabilidade conjunta $P_{k'|k}$. Esta é a probabilidade de que uma aresta conectada a um vértice de grau k se conecte, na outra extremidade, a um vértice de grau k' . Quando esta correlação de grau não é observada, a probabilidade condicional não depende de k e pode ser escrita como:

$$P_{k'|k}^{nc} = \frac{k' P_{k'}}{\langle k \rangle}. \quad (2.15)$$

Nos casos em que $P_{k'|k}$ apresenta dependência com k , dizemos que existe correlação. Muitas vezes, é difícil estimar o valor da distribuição conjunta devido a flutuações estatísticas (35, 49) e uma alternativa normalmente utilizada é estudar o grau médio dos vértices vizinhos. O grau médio dos vizinhos do vértice i , é dado por:

$$k_{nn,i} = \frac{1}{k_i} \sum_{j=1}^N a_{i,j} k_j. \quad (2.16)$$

Tomando a média sobre os vértices com mesmo grau, temos:

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i=1}^N k_{nn,i} \delta_{k_i,k}, \quad (2.17)$$

onde N_k é o número de vértices com grau k . A quantidade definida pela equação 2.17 está relacionada com a probabilidade conjunta de acordo com:

$$k_{nn}(k) = \sum_{k'} k' P_{k'|k}. \quad (2.18)$$

Assim, em redes sem correlação de grau, $k_{nn}(k)$ é uma constante. Nos casos em que esta quantidade cresce a medida que o valor de k aumenta, dizemos que a rede é assortativa, e os vértices tendem a estabelecer conexões com vértices com mesmo número de vizinhos. Quando o valor de $k_{nn}(k)$ diminui a medida que k aumenta, dizemos que a rede é dissortativa e os vértices com poucos vizinhos tendem a se conectar com vértices de alto grau. Neste último caso, se enquadra a rede da internet estudada em (49), onde os autores

mostraram que $k_{nn}(k) \sim k^{-0.5}$. Embora os modelos descritos nas seções anteriores não apresentem correlação de grau, existem propostas de modelos capazes de gerar este tipo de correlação, por exemplo, através do método de variáveis ocultas (51) e via otimização funcional (50).

Além daquelas envolvendo o número de vizinhos dos vértices, foram observadas outros tipos de correlações em redes reais, mais especificamente em redes geográficas com peso. Por exemplo, em redes de aeroportos, verificou-se que existe forte correlação entre o grau e o tráfego em um vértice. Esta relação é expressa por:

$$k_i^\beta \sim \sum_{j=1}^N a_{i,j} w_{i,j}^{(\text{traffic})}, \quad (2.19)$$

onde o peso da aresta (i, j) , $w_{i,j}^{(\text{traffic})}$, corresponde ao número de passageiros transportados entre os aeroportos i e j . Foi observado que $\beta \approx 1.7$ para a rede de aeroportos dos Estados Unidos, contendo $N = 675$ aeroportos (vértices) e $E = 3253$ linhas aéreas (conexões) (35). Ainda para esta rede, foi observado correlações existentes entre o grau e o comprimento total das conexões de um vértice, neste caso, temos:

$$k_i^\alpha \sim \sum_{j=1}^N a_{i,j} w_{i,j}^{(\text{distance})}. \quad (2.20)$$

Aqui, o peso da aresta (i, j) , $w_{i,j}^{(\text{distance})}$, corresponde a distância euclidiana que separa os aeroportos i e j , ou seja: $w_{i,j}^{(\text{distance})} = r_{i,j}$. O valor do expoente foi estimado em $\alpha = 1.4$ (35).

É esperado que o tráfego em aeroportos mais conectados seja maior do que em aeroportos menos conectados. Entretanto, a equação 2.19 revela uma relação super-linear para estas duas quantidades, indicando uma correlação não trivial entre a dinâmica e topologia da rede de aeroportos. O mesmo tipo de argumento pode ser utilizado para explicar a importância da equação 2.20, que nos indica a existência de uma correlação entre a geografia e a topologia da rede, uma clara evidência de como o vínculo espacial é capaz de alterar a organização de determinados sistemas.

Como foi dito anteriormente (seção 2.1), também foi observada a existência de correlação entre o grau e a centralidade dos vértices em determinadas redes. Esta correlação é expressa por:

$$k_i^\nu \sim b_i, \quad (2.21)$$

com $\nu = 1$ para o modelo BA, por exemplo. Embora esta relação seja verdadeira, em algumas redes, como a de aeroportos, foram observadas grandes flutuações em torno deste comportamento médio, de tal maneira que existem vértices anômalos, os quais podem ter baixo/alto grau e ao mesmo tempo apresentarem alta/baixa centralidade. É necessário

então, maneiras alternativas de estudar a topologia dos sistemas, em especial nestes casos onde as flutuações são importantes. Na próxima seção, iremos descrever como a dinâmica de caminhadas aleatórias pode ser utilizada para complementar as informações acerca da topologia das redes complexas.

2.4 Caminhadas Aleatórias em Redes

Caminhadas aleatórias são utilizadas em diferentes contextos na área das redes complexas, que vão desde a modelagem de *busca* até transporte e difusão (52–54). A grosso modo, podemos imaginar que, a cada instante de tempo, um agente localizado sobre um vértice qualquer salta para um de seus vizinhos escolhido de acordo com alguma regra. Após ℓ saltos, dizemos que o conjunto de vértices visitados pelo agente constitui um caminho de tamanho ℓ e que, este caminho por sua vez, corresponde a caminhada aleatória do agente.

Diferentes regras podem ser definidas para o agente que caminha sobre a rede. Como exemplo, podemos não permitir a transição (salto) para vértices que já foram visitados, neste caso dizemos se tratar de uma *caminhada auto-excludente* (*self-avoid random walk*). Outra estratégia é realizar as transições com certa preferência, por exemplo, privilegiando vértices com alto grau. Neste caso, temos uma *caminhada preferencial* (*preferential random walk*). Cada uma delas é utilizada para modelar um certo conjunto de problemas, sendo que o caso mais geral é aquele em que nenhuma regra é imposta sobre os agentes e portanto, eles se movem livremente sobre a rede. Neste trabalho iremos denotar este tipo de caminhada por *caminhada tradicional* (*random walk*).

A probabilidade de transição de que, no instante t , um agente localizado no vértice i mova-se para o vértice j no instante $t + 1$, é definida por $p(i, j)$. Para caminhadas tradicionais, esta quantidade pode ser escrita como:

$$p(i, j) = \frac{a_{i,j} w_{i,j}}{\sum_{j=1}^N a_{i,j} w_{i,j}}. \quad (2.22)$$

Observe que, neste trabalho, não há dependência entre a probabilidade de transição e o instante em que ela ocorre. Podemos ainda, definir a probabilidade de transição para uma caminhada de comprimento ℓ , ou seja, a probabilidade de que um agente, inicialmente no vértice i , alcance o vértice j após ℓ passos. Esta quantidade é definida por $p_\ell(i, j)$. Um resultado clássico das caminhadas tradicionais (55), diz que no limite de caminhadas longas, esta probabilidade é dada por:

$$\lim_{\ell \rightarrow \infty} p_\ell(i, j) = p_\infty(j) = \frac{k_j}{2E}. \quad (2.23)$$

Com este resultado, vemos que, quanto maior o número de conexões de um vértice, maior serão as chances de encontrar um agente neste vértice após muitos passos de caminhada. Além disso, a medida que a dinâmica evolui, a informação a respeito da origem torna-se desnecessária e o resultado final independe do vértice inicial. Entretanto, em muitos problemas da vida real, a informação a respeito da origem é importante e para leva-lá em conta na modelagem via caminhadas aleatórias, é necessário estudar o comportamento do sistema para um número finito de passos ℓ .

Figura 2.3 – Exemplo de rede onde destacamos os únicos três caminhos de comprimento $\ell = 8$ que ligam o vértice 1 ao vértice 20 (verde, rosa e azul). A probabilidade de que um agente, inicialmente em 1, realize uma caminhada aleatória correspondente a cada um desses caminhos é dado por $1/1620$, $1/2160$ e $1/4050$ para os caminhos verde, rosa e azul, respectivamente.

Para ilustrar o cálculo da probabilidade de transição, vejamos o exemplo da figura 2.3. Nesta figura destacamos um vértice de origem e um vértice de destino, 1 e 20, respectivamente. Se estivermos interessados em estimar a probabilidade de transição $p_8(1, 20)$, precisamos determinar todos os caminhos de comprimento $\ell = 8$ que ligam os vértices 1 e 20. Na rede da figura 2.3 existem três caminhos que satisfazem esse critério, o caminho verde, o rosa e o azul. Cada um deles ocorre com probabilidades diferentes, as quais dependem do tipo de caminhada escolhida. No caso das caminhadas tradicionais, podemos aplicar a equação 2.22 a cada transição. O produto das probabilidades de transição darão a probabilidade de ocorrência de um determinado caminho. Por exemplo, para o caminho azul, temos:

$$\begin{aligned} p_{\text{azul}} &= p(1, 4)p(4, 5)p(5, 6)p(6, 10)p(10, 13)p(13, 15)p(15, 16)p(16, 20) \\ &= \frac{1}{5} \frac{1}{3} \frac{1}{5} \frac{1}{3} \frac{1}{2} \frac{1}{3} \frac{1}{3} = \frac{1}{4050}. \end{aligned}$$

Fazendo o mesmo para os outros dois caminhos encontramos: $p_{\text{verde}} = 1/1620$ e $p_{\text{rosa}} = 1/2160$. O valor final de $p_8(1, 20)$ é dado pela soma das probabilidades de todos os caminhos de comprimento $\ell = 8$, ou seja, $p_8(1, 20) = p_{\text{azul}} + p_{\text{verde}} + p_{\text{rosa}} \approx 0.0013272$. Isso significa que, em média, 0.13% dos caminhos de comprimento $\ell = 8$ realizados por um agente movendo-se de maneira completamente aleatória, irão terminar no vértice 20, se o agente iniciar suas caminhadas no vértice 1. Para o caso de uma caminhada auto-

excludente, estes valores são alterados para $p_{\text{azul}}^{\text{ae}} = 1/256$, $p_{\text{verde}}^{\text{ae}} = 1/64$ e $p_{\text{rosa}}^{\text{ae}} = 1/96$, sendo que a probabilidade final de transição será $p_8^{\text{ae}}(1, 20) = 0.0299479$. Aqui o índice “ae” indica que o resultado considera caminhada auto-excludente. É importante notar que a escolha da dinâmica altera significativamente os valores das probabilidades com que cada caminho ocorre. Além disso, a mudança relativa dessas probabilidades não é idêntica para todos os caminhos. Por exemplo, na caminhada tradicional, o caminho verde é cerca de 2.5 vezes mais provável que o caminho azul. Por outro lado, quando consideramos caminhadas auto-excludentes, o verde torna-se 4 vezes mais provável.

Estimar os valores de $p_\ell(i, j)$ para um tipo qualquer de caminhada, mesmo em redes onde o número de vértices é da ordem de algumas centenas, não é fácil. Existem propostas de algoritmos exatos para executar esta tarefa porém, para $\ell > 6$ eles tornam-se inviáveis devido ao custo computacional necessário. No caso da caminhada tradicional ou preferencial, é possível estimar os valores de $p_\ell(i, j)$ através da matriz de transição $\mathbf{T}$, cujos elementos são dados por $t_{i,j} = p(i, j)$. Neste caso, os valores de $p_\ell(i, j)$ são os elementos da matriz $\mathbf{P}_\ell$, dada por:

$$\mathbf{P}_\ell = \mathbf{T}^\ell. \quad (2.24)$$

Uma alternativa mais geral é usar simulações baseadas em agentes (35), no qual realizamos um determinado número de caminhadas sobre a rede e contamos efetivamente o número delas que partem do vértice i e atingem o vértice j após ℓ passos. Observe que a medida que o número de passos exigidos aumenta, o número de caminhadas necessárias para garantir boa estimativa deve aumentar. A menos que seja especificado o contrário, a abordagem baseada em agentes será utilizada ao longo deste trabalho para estimar os valores de $p_\ell(i, j)$.

Na próxima seção iremos introduzir o conceito de acessibilidade e mostrar como ele, em conjunto com as probabilidades de transição $p_\ell(i, j)$, nos permite generalizar o conceito de grau em redes complexas.

2.5 Acessibilidade

Para introduzir o conceito de acessibilidade, vamos utilizar um problema clássico da teoria das probabilidades: suponha um dado com N faces, onde cada face possui probabilidade $p(i) > 0$ ($i \in [1, N]$) de ser sorteada. Quantas vezes em média, é preciso lançar este dado para que todas as faces sejam sorteadas ao menos uma vez? Esta quantidade será denotada ao longo deste trabalho por *tempo de exploração* e este problema é conhecido como *O Problema do Colecionador de Cupons*, o qual possui aplicações diretas em diversos ramos da ciência (56, 57). Utilizando os princípios de inclusão exclusão, é possível

mostrar que o tempo de exploração, τ , pode ser escrito pela equação 2.25 (58). É possível ainda, estudar este problema como um processo de Poisson de variáveis independentes, o que nos fornece a equação 2.26.

$$\tau = \sum_{j=1}^N \frac{1}{p(v_j)} - \sum_{j=1}^N \sum_{k=j+1}^N \frac{1}{p(v_j) + p(v_k)} + \sum_{j=1}^N \sum_{k=j+1}^N \sum_{m=k+1}^N \frac{1}{p(v_j) + p(v_k) + p(v_m)} - \dots$$

$$- \frac{1}{p(v_1) + p(v_2) + \dots + p(v_N)} \quad (2.25)$$

$$\tau = \int_0^\infty \left(1 - \prod_{j=1}^N (1 - e^{-xp(v_j)}) \right) dx \quad (2.26)$$

Agora, imagine que uma das N faces, por exemplo, a face número 1, tenha uma probabilidade de ser sorteada muito maior do que as demais. Assim, podemos dizer que $p(1) \gg p(i)$ para todo $i \in [2, N]$. Neste caso, após vários lançamentos do dado, a face 1 será sorteada na maioria das vezes. Concluimos que, embora todas as faces sejam possíveis, aproximadamente apenas uma delas é efetivamente acessível. A acessibilidade é a medida capaz de informar quão efetivo é o acesso do jogador as faces do dado e como veremos, esta característica está intimamente relacionada a homogeneidade das probabilidades $p(i)$.

O problema do colecionador de cupons pode ser facilmente descrito no contexto de redes. Imagine que, dado um vértice de referência, digamos o vértice i , é possível estabelecer univocamente, segundo um determinado critério, um subconjunto de nós $\Omega_i = \{v_1, v_2, \dots, v_{N_i}\}$ o qual pode ou não incluir o próprio vértice i . Como pode ser observado no exemplo abaixo, o critério que determina o conjunto Ω_i é, em geral, resultado de alguma dinâmica imposta sobre a estrutura da rede.

Exemplo: Ω_i é o conjunto de todos os vértices que podem ser alcançados através de uma caminhada aleatória de comprimento ℓ partindo do vértice i .

Os vértices contidos em Ω_i possuem probabilidades associadas, descritas pelo conjunto $P(\Omega_i) = \{p(v_1), p(v_2), \dots, p(v_{N_i})\}$. Veja que, para o exemplo citado acima, as probabilidades do conjunto $P(\Omega_i)$ são as próprias probabilidades de transição descritas na seção anterior, ou seja, $p(v_j) = p_\ell(i, v_j)$. A heterogeneidade do conjunto de probabilidades $P(\Omega_i)$ pode ser medido pela sua entropia de Shannon (59), dada pela equação 2.27.

$$\sigma_i = - \sum_{j=1}^{N_i} p(j) \log p(j), \quad j \in \Omega_i \quad (2.27)$$

Se estivermos interessados em estudar o acesso efetivo do vértice i , aos demais vértices

contidos em Ω_i , podemos olhar para a acessibilidade deste vértice, que é definida pela equação 2.28.

$$\mathcal{A}_i = e^{\sigma_i} \quad (2.28)$$

$\mathcal{A}_i$ assume valores no intervalo $[1, N_i]$ e é máxima no caso homogêneo, onde todas as probabilidades são iguais a $1/N_i$. Existe uma conjectura de que este caso é também o único mínimo global do tempo de exploração (equações 2.25 e 2.26) (58, 60, 61). Embora grande parte dos especialistas estejam convencidos da sua validade, a prova desta conjectura é ainda desconhecida. Ao longo deste trabalho iremos assumir que ela é válida e assim, podemos dizer que nos casos em que a acessibilidade do vértice i é máxima, o tempo de exploração deste vértice é mínimo. Este fato é confirmado quando olhamos para o gráfico da figura 2.4, onde mostramos os valores de acessibilidade em função do tempo de exploração para diferentes conjuntos Ω_i com seis elementos. Um dos elementos desse conjunto tem probabilidade ϵ , enquanto os demais tem probabilidades iguais a $(1 - \epsilon)/5$. Fica evidente que, quando $\epsilon = 1/6$, a acessibilidade atinge seu valor máximo e o tempo de exploração seu valor mínimo. Os pontos em cinza correspondem a diferentes configurações para as probabilidades do elementos de Ω_i (62).

A acessibilidade pode ainda ser interpretada como o grau efetivo do vértice de referência, uma vez que ela indica o número efetivo de vértices que se comunicam com i segundo a dinâmica utilizada. Observe que, muito embora tenhamos ilustrado o uso da acessibilidade para a dinâmica das caminhadas aleatórias, ela pode ser aplicada a qualquer situação onde, para cada nó de referência pode-se definir um conjunto de vértices e atribuir probabilidades a eles. Situações como essas são encontrados em outras dinâmicas, como propagação de doenças, sincronização neuronal, entre outras. Vejamos o exemplo mostrado na figura 2.5a, onde apresentamos a rede de comércio de uma cidade. Cada estabelecimento comercial é representado por um vértice e as arestas representam o fornecimento de mercadorias entre eles, ou seja, o vértice A estabelece uma conexão direcionada ao vértice B se A fornece algum produto ao vértice B . Na figura 2.5b temos a representação em termos de uma rede com pesos, sendo que estes correspondem a taxa de entrega dos produtos. Observe que esta rede segue uma lei de conservação na qual a soma dos pesos das arestas que saem de um vértice devem ter soma unitária. Sendo assim, podemos interpretar essas taxas de entrega como a probabilidade de um vértice receber o produto produzido por um determinado vizinho. Podemos então, investigar a acessibilidade dos vértices desta rede. Por exemplo, o conjunto dos vértices que recebe algum produto do vértice 2 é dado por $\Omega_2 = \{1, 3, 4, 5, 6, 7\}$, com taxas/probabilidades $P(\Omega_2) = \{0.02, 0.02, 0.02, 0.03, 0.03, 0.88\}$. Usando as equações 2.27 e 2.28, obtemos $\mathcal{A}_2 \approx 1.74$. Vemos que, embora o vértice 2 tenha grau $k_2 = 6$, seu grau efetivo é bem

Figura 2.4 – Valores de acessibilidade em função do tempo de exploração para 10^5 vetores aleatórios com tamanho $N = 6$. A linha preta corresponde a função limite onde uma das probabilidades do vetor é dada por ϵ e as demais são idênticas e iguais a $(1 - \epsilon)/5$.

menor, uma vez que a probabilidade de uma de suas conexões é bem mais elevada que as demais, polarizando o acesso aos seus vizinhos. Se fizermos a mesma análise para o vértice 8, veremos que seu grau e grau efetivo são idênticos a 3, uma vez que as probabilidades de suas conexões são iguais. Através desse exemplo, fica claro como a medida de acessibilidade generaliza o conceito de grau e pode ser usada para quantificar o acesso de um vértice i aos demais contidos no conjunto Ω_i que por sua vez, é definido por um processo dinâmico de interesse.

Na próxima seção daremos particular importância a dinâmica das caminhadas aleatórias e iremos demonstrar como ela, juntamente com a medida de acessibilidade, podem ser utilizadas para definir o conceito de bordas em redes complexas.

2.5.1 Detecção de bordas

O conceito de borda está diretamente relacionado ao nosso conceito de limite. Quando dizemos que determinada região pertence a borda de um objeto, sabemos que naquela região existem descontinuidades que fazem com que determinadas grandezas assumam

Figura 2.5 – (a) Relações comerciais de uma cidade e sua (b) representação em termos de uma rede com pesos. Nesta rede, cada estabelecimento comercial é representado por um vértice e dois vértices são ligados por uma conexão dirigida se um deles fornece algum tipo de produto para o outro. A espessura de uma conexão representa a taxa com que a entrega é realizada, por exemplo, 65% da produção do vértice 7 é destinada ao vértice 6.

valores diferentes daqueles observados ao longo do objeto. Por exemplo, se estamos analisando uma rede regular, como aquela mostrada na figura 2.6, percebemos que todos vértices possuem quatro vizinhos, exceto nas regiões de borda, onde eles podem ter três ou até dois vizinhos. Isso é o que chamamos de *efeitos de borda* (20). Esses efeitos são inerentes a todos os sistemas finitos, como as redes geográficas que iremos estudar neste trabalho. Com objetivo de evitar esse tipo de efeito e realizar medidas mais precisas acerca dos sistemas de interesse, em primeiro lugar é importante desenvolver um procedimento para realizar a detecção de bordas.

Figura 2.6 – Rede regular bidimensional. Enquanto os vértices verdes, que não sofrem efeito de borda, possuem quatro vizinhos, os vértices brancos na região de borda, possuem apenas dois ou três vizinhos.

A definição de borda em redes complexas está intimamente relacionada à escolha de um processo dinâmico em particular, sendo que diferentes processos podem produzir diferentes bordas. Assim, é comum dizer que um determinado conjunto de vértices representa a borda de uma rede frente a determinada dinâmica. Aqui, iremos nos limitar ao estudo de bordas geradas pela dinâmica das caminhadas aleatórias de comprimento ℓ . Essa dinâmica gera valores de acessibilidade $\mathcal{A}_i(\ell)$, onde deixamos claro a dependência com a escala de comprimento das caminhadas.

Sabendo que a acessibilidade mede o acesso efetivo de um vértice aos demais da rede, esperamos que ela seja menor para aqueles vértices localizados nas regiões de fronteira e aumenta de maneira gradual até os vértices mais centrais (18, 20). Podemos então definir um limiar para a acessibilidade, denotado por $\mathcal{A}^*$, para o qual todo vértice com valor de acessibilidade menor que $\mathcal{A}^*$ é considerado vértice de borda. Em geral, quando estamos interessados em detectar as bordas de uma rede, olhamos para a acessibilidade na escala típica do sistema, correspondente ao seu diâmetro, ou seja: $\mathcal{A}_i(\ell = D)$. Portanto, segundo o nosso critério, se $\mathcal{A}_i(\ell = D) \leq \mathcal{A}^*$ o vértice i é um vértice de borda. Na figura 2.7a apresentamos uma rede gerada através da triangulação de Delaunay de um conjunto de 1289 pontos dispostos aleatoriamente sobre o mapa do Brasil. Nesta rede, os vértices estão coloridos de acordo com sua acessibilidade média. Conforme era esperado, os menores valores de acessibilidade são encontrados nas regiões de fronteira. Nas figuras 2.7b-e apresentamos os vértices de borda detectados para diferentes valores de $\mathcal{A}^*$.

Figura 2.7 – (a) Rede gerada através da triangulação de Delaunay de um conjunto de 1289 pontos dispostos aleatoriamente sobre o mapa do Brasil. Os vértices foram coloridos de acordo com sua acessibilidade média no intervalo $\ell \in [1, D]$, onde D representa o diâmetro da rede, neste caso 17. Os vértices de borda foram identificados para diferentes limiares: (b) $\mathcal{A}^* = 58$, (c) $\mathcal{A}^* = 74$, (d) $\mathcal{A}^* = 90$ e (e) $\mathcal{A}^* = 106$.

É interessante observar que outras medidas não são capazes de capturar as informações

de borda tão bem quanto a acessibilidade (18, 20). Como exemplo, na figura 2.8a, 2.8c e 2.8e apresentamos a distribuição dos valores de grau, comprimento médio do menor caminho (CMMC) e centralidade para a mesma rede da figura anterior. Em 2.8b, 2.8d e 2.8f mostramos a borda detectada utilizando essas medidas. Observe que, diferente da acessibilidade, o grau e a centralidade são ruidosos, enquanto a medida CMMC apresenta simetria circular que inviabiliza a detecção.

Figura 2.8 – Distribuição de (a) grau, (c) centralidade e (e) comprimento médio do menor caminho. Em (b), (d) e (f) mostramos as bordas detectadas utilizando essas medidas. Observe que nenhuma delas é capaz de igualar a qualidade das bordas apresentadas na figura anterior, quando utilizamos a acessibilidade.

Vale ressaltar que um procedimento para determinar o valor ideal de $\mathcal{A}^*$ é ainda desconhecido e acreditamos que este valor depende tanto da rede quanto da aplicação desejada (18, 20).

2.6 Propagação de doenças em redes geográficas

Juntamente com as caminhadas aleatórias, a propagação de doença é um dos processos dinâmicos mais estudados na área das redes complexas (35, 63–67). Ao longo dos últimos anos, muitos trabalhos têm explorado os efeitos da topologia desordenada das redes complexas no comportamento dos modelos epidemiológicos mais bem conhecidos, como SI

(Suscetível-Infectado), SIS (Suscetível-Infectado-Suscetível) e SIR (Suscetível-Infectado-Removido) (35). Esses modelos assumem que um conjunto inicial de vértices inicialmente infectados propaga uma infecção com uma probabilidade β para seus vizinhos suscetíveis através de um processo binário de contágio. Usando o mesmo processo estes vizinhos passam a doença adiante. No modelo SIS, existe uma probabilidade μ de que um vértice infectado torne-se novamente suscetível, enquanto no modelo SIR esta probabilidade representa a chance de um vértice ser excluído do sistema devido a auto-imunização ou morte.

Uma rede é considerada homogênea quando a probabilidade de infecção $\beta_{i,j}$ é a mesma para todas as arestas (i, j) e quando todos os seus vértices possuem a mesma probabilidade de remoção μ . Veja que $\beta_{i,j}$ corresponde a probabilidade de que o vértice i infecte o vértice j . Para redes homogêneas, pode-se mostrar que a infecção sempre se propaga quando o limiar de epidemia, definido como $R_0 = \beta/\mu$ satisfaz a relação $R_0 > 1$ (64). Obviamente, as redes reais não são homogêneas, já que suas arestas e vértices possuem características próprias que irão definir os valores de $\beta_{i,j}$ e μ_i .

Neste trabalho iremos assumir que a probabilidade de remoção é idêntica para todos os vértices e assume valor μ , enquanto que os valores de $\beta_{i,j}$ irão levar em conta as propriedades geométricas da aresta (i, j) . Pensando nas redes geográficas, é razoável assumir que arestas mais longas dificultem a propagação da infecção, enquanto arestas mais espessas facilitem essa propagação. Assim, vamos utilizar a lei empírica da resistência elétrica em condutores $R_{i,j} = \rho_{i,j} l_{i,j} / d_{i,j}$, onde $\rho_{i,j}$ é a resistividade da aresta, $l_{i,j}$ é o comprimento da aresta e $d_{i,j}$ o seu diâmetro. Considerando que toda a rede é composta pelo mesmo material e que não há variações significantes de resistividade, iremos assumir que $\rho_{i,j} = \rho = 1$ para todos os canais (i, j) . Portanto, a probabilidade de que a infecção se propague ao longo da aresta (i, j) assume a forma $\beta_{i,j} = \exp(-R_{i,j})$, ou seja:

$$\beta_{i,j} = \exp\left(-\frac{l_{i,j}}{d_{i,j}^2}\right). \quad (2.29)$$

A equação 2.29 implica que a probabilidade de infecção tende a zero para arestas com alta resistência, enquanto se aproxima da unidade para canais de baixa resistência.

As simulações do modelo SIR são realizadas da seguinte maneira: i) Escolhemos um valor de μ . ii) Infectamos um vértice aleatório i . Em seguida, este vértice infecta seus vizinhos com probabilidade de infecção dada pela equação 2.29. Por fim, o vértice i é removido do sistema com probabilidade μ . (iii) O processo de contágio evolui até que não haja mais vértices infectados. (iv) Contabilizamos o número de vértices que foram infectados ao longo da simulação. Este valor é denotado por Γ_i , indicando que a propagação foi iniciada no vértice i . (v) Repetimos os passos (ii)-(iv) muitas vezes e tomamos a média sobre todas as diferentes realizações estocásticas. Em seguida, todo procedimento é realizado novamente para outro valor de μ .

Esta implementação do modelo SIR pode ser mapeada em uma percolação isotrópica (68), na qual sabemos que existe uma transição de segunda ordem de um regime não-invadido (abaixo do limiar de epidemia) para um regime invadido (acima do limiar de epidemia) a medida que a probabilidade de remoção μ diminui. Na seção de resultados iremos utilizar os conceitos descritos aqui para estudar a propagação de fungos em redes geográficas obtidas a partir de amostras de solo.

3 Mapeando estruturas biológicas em redes

Ao longo da última década, a teoria de redes complexas apresentada na seção anterior tem sido aplicada com sucesso em diversos sistemas biológicos diferentes. Entretanto, o número de trabalhos devotados aos sistemas biológicos imersos no espaço Euclideano, seja ele 2D ou 3D, ainda é bastante escasso. Ainda mais escassos são os trabalhos que descrevem detalhes do procedimento utilizado para mapear o sistema de interesse em uma rede, impossibilitando muitas vezes, a reprodução dos resultados apresentados.

Figura 3.1 – (a) imagens de uma amostra de solo obtidas por micro-tomografia computadorizada. (b) representação computacional da amostra de solo.

Os sistemas biológicos de nosso interesse, são em geral armazenados em computador através de imagens digitais, as quais são provenientes de microscopia, tomografia ou fotografia. Nos casos tridimensionais, um conjunto de imagens é utilizado para representar a estrutura do sistema em diferentes profundidades, como pode ser visto na ilustração da figura 3.1a. Em seguida, essas imagens podem ser agrupadas, gerando um volume que por sua vez, irá passar por um processo de reconstrução 3D, gerando a representação computacional do sistema de interesse (figura 3.1b). Levando em conta essas observações, podemos dizer que um software capaz de mapear as informações de um sistema biológico em uma rede, deve ser capaz de trabalhar no nível dos pixels e voxels das imagens que representam o dado sistema. Assim, se faz necessário a aplicação de diversas técnicas de processamento e análise de imagens neste tipo de software. Tais técnicas são aplicadas desde a fase de pré-processamento até a fase final de visualização da rede resultante. Nas próximas seções iremos descrever o desenvolvimento do software utilizado no decorrer

deste trabalho para realizar o mapeamento e gerar redes geográficas a partir das imagens de sistemas biológicos.

3.1 Pré-processamento

Todas as rotinas descritas ao longo deste trabalho foram desenvolvidas em linguagem C++ através da interface *Borland C++ Builder*. Em conjunto, utilizamos a biblioteca de processamento e visualização de imagens, VTK (*Visualization Toolkit*) (15), que disponibiliza uma grande quantidade de rotinas prontas.

Antes de aplicar nossa metodologia de mapeamento, é necessário que as imagens digitais passem por uma fase de pré-processamento, a fim de detectar as regiões de nosso interesse e eliminar ruídos provenientes das técnicas de captura. Nesta fase de pré-processamento, utilizamos principalmente algoritmos para (i) reamostragem, (ii) suavização, (iii) segmentação e (iv) rotulação de componentes conectadas. A seguir daremos mais detalhes sobre eles.

- *Reamostragem*: devido às limitações técnicas, os equipamentos que capturam as imagens não são, em sua maioria, isotrópicos. Sendo assim, as dimensões reais de cada voxel são diferentes nas direções x , y e z . A fim de tornar o conjunto de imagens isotrópico, é necessário fazer a reamostragem dos voxels dessas imagens. O método de reamostragem utilizado aqui foi a interpolação linear, na qual é possível aumentar ou diminuir a resolução dos eixos, criando ou removendo voxels a partir da interpolação linear dos voxels já existentes.

- *Suavização Gaussiana*: utilizada para reduzir a diferença entre as intensidades dos pixels em uma vizinhança e conseqüentemente, reduzir os ruídos presentes nas imagens. Neste trabalho utilizamos uma implementação via convolução discreta, na qual a intensidade de um voxel é atualizada para a média ponderada do seu valor e dos seus vizinhos. Uma máscara gaussiana é responsável por ponderar as intensidades, sendo que vizinhos mais próximos contribuem mais do que aqueles mais distantes. O tamanho da máscara r_m , bem como a o desvio padrão da função gaussiana, σ_m , são os dois únicos parâmetros deste algoritmo. Na figura 3.2b vemos um exemplo de aplicação, na qual foram usados os seguintes parâmetros $r_m = 3$ e $\sigma_m = 3$.

- *Suavização Mediana*: possui os mesmos propósitos da suavização anterior, com a vantagem que este tipo de suavização não causa um borramento excessivo da imagem. Neste caso o filtro utilizado é a mediana, que consiste em substituir a intensidade de um voxel pelo valor mediano das intensidades dentro de uma caixa de lado B centrada no voxel de interesse. Na figura 3.2c mostramos o resultado da aplicação do filtro da mediana sobre a figura mostrada em 3.2a com $B = 3$.

Figura 3.2 – (a) imagem de uma secção transversal do osso cortical de suíno. (b) suavização gaussiana da imagem original com $r_m = 3$ e $\sigma_m = 3$. (c) suavização mediana da imagem original com $B = 3$. (d) limiarização da imagem (c) com valor $I^* = 98$. (e) rotulação dos componentes conectados. (f) maior componente conectado.

- *Segmentação*: consiste na divisão de uma imagem em partes, sendo que as partes de interesse são determinadas pelo problema em questão. Neste trabalho estaremos particularmente interessados nos processos de segmentação capazes de separar o objeto biológico de interesse do fundo da imagem. O algoritmo mais simples para este propósito é a *limiarização*, que consiste basicamente em assumir que todos os voxels com intensidade abaixo de um valor I^* pertencem ao fundo da imagem, enquanto os voxels cuja intensidade é maior que I^* , pertencem ao objeto de interesse. Na figura 3.2d apresentamos o resultado da limiarização aplicada sobre a figura mostrada em 3.2c, com $I^* = 98$.

- *Deteção de componentes*: Muitas vezes, o processo de segmentação gera mais de uma componente conectada (figura 3.2d), ou seja, grupos de voxels (ou pixels) que possuem as seguintes propriedades: (i) possuem a mesma intensidade e (ii) são transitivos entre si. A deteção de componentes pode ser utilizada para contar o número de componentes em uma imagem, bem como para determinar o tamanho destas componentes. As vezes, é interessante eliminar componentes com tamanho pequeno, que podem ser provenientes de ruídos que não foram eliminados pelos estágios de suavização. Maiores detalhes sobre a deteção de componentes e da implementação utilizada neste trabalho, podem ser

encontrados na referência (69). Na figura 3.2e aplicamos a detecção de componentes na imagem resultante do processo de segmentação. Agora, podemos por exemplo, identificar o maior componente conectado, como mostra a figura 3.2f, e restringir nossas análises a ele.

Como resultado da fase de pré-processamento, temos um volume binário no caso de dados tridimensionais, ou uma imagem binária no caso 2D - figura 3.2f. Em ambos, os voxels (ou pixels) do fundo possuem intensidade 255 (branco) e os objetos possuem intensidade 0 (preto). Lembrando que, nosso principal objetivo é mapear esses objetos através de uma rede, a seguir iremos descrever o desenvolvimento do algoritmo que realiza esse mapeamento.

3.2 Propagação de rótulos

Figura 3.3 – (a)-(e) Evolução da propagação de rótulos sobre uma estrutura sintética. Em (f) destacamos algumas frentes de onda encontradas ao longo da estrutura.

Inicialmente, substituímos os valores dos voxels do objeto para -1 e em seguida executamos uma propagação de rótulos sobre os voxels do objeto de interesse. Aqui vamos ilustrar a ideia em três dimensões, mas o processo é geral, podendo ser aplicado sem alterações nos casos 2D.

No instante $t = 0$, escolhemos um ponto extremo do objeto e atribuímos a ele o valor 1. Em seguida, no instante $t = 1$, os vizinhos deste voxels recebem valor 2. Em $t = 2$, os vizinhos dos vizinhos recebem valor 3 e esta propagação continua até que não haja mais voxels com valor -1. Dizemos que o conjunto de voxels com valor $t + 1$ no instante t definem uma frente de onda. Na figura 3.3a-e podemos ver a propagação em diferentes

instantes de tempo. Em 3.3f mostramos as frentes de onda no instante $t = 10$, $t = 30$, $t = 40$, $t = 60$, $t = 80$, $t = 100$ e $t = 120$. Veja que os voxels que receberam valor são representados por esferas vermelhas.

Nossa hipótese é de que, em objetos com comprimento muito maior do que o diâmetro (como na figura 3.3), o eixo central passe pelo centro de massa das frentes de ondas geradas durante a propagação de rótulos. Este fato é confirmado pela figura 3.4a, onde mostramos o esqueleto gerado após a união do centro de massa das frentes de onda consecutivas. Este esqueleto pode ser denotado pela curva paramétrica $\vec{u}(t) = [x(t), y(t), z(t)]$ e cada uma de suas componentes pode ser estudada em função da variável de parametrização t , como pode ser visto na figuras 3.4b-d.

Figura 3.4 – (a) Esqueleto obtido através da união do centro de massa de consecutivas frentes de onda. O eixo medial então é representado por uma curva paramétrica $\vec{u}(t) = [x(t), y(t), z(t)]$. As componentes desta curva são mostradas em (b), (c) e (d). Em (e) mostramos a rede que representa a estrutura original.

Este primeiro passo é fundamental para o desenvolvimento de uma rotina capaz de mapear estruturas complexas em redes, já que muitas delas são basicamente constituídas por arestas com formas similares aquela da figura 3.4a. Se imaginarmos que os pontos iniciais e finais definem vértices, podemos então criar a rede mostrada na figura 3.4e,

onde o peso $w_{1,2}$ pode representar o comprimento de arco do esqueleto ($l_{1,2}$), a distância euclidiana entre os vértices 1 e 2 ($r_{1,2}$), o diâmetro médio da estrutura ao longo do esqueleto ($d_{1,2}$), entre diversas outras quantidades.

3.2.1 Bifurcações

Figura 3.5 – (a)-(h) Evolução da propagação de rótulos sobre uma estrutura sintética contendo uma bifurcação. Em (f) apresentamos o eixo medial obtido através da união do centro de massa de consecutivas frentes de onda.

O segundo passo natural é introduzir a possibilidade de bifurcações ao longo de uma estrutura. Uma maneira simples de levar em conta essa possibilidade, é monitorar a conectividade das frentes de onda ao longo da propagação. Se, em algum instante, a frente de onda se partir em uma ou mais partes, saberemos que naquela região existe uma

bifurcação. Vejamos o exemplo mostrado na figura 3.5. A propagação ocorre normalmente (3.5a-c) até o instante mostrado na figura 3.5d, onde a frente de onda, representada pelas esferas brancas, passa a ter duas componentes desconexas. Neste momento, as duas frentes de onda são armazenadas em uma lista de prioridades. Em seguida, uma das frentes de onda é retirada da pilha e sua propagação continua até que não haja mais voxels adjacentes com valor -1 (3.5e-f). Em seguida, a outra frente de onda é retirada da lista de prioridade e propagada 3.5g. O resultado é mostrado na figura 3.5h. Por fim, podemos novamente unir o centro de massa das frentes de onda consecutivas e obter o esqueleto mostrado na figura 3.5i. Veja que ao longo desse processo, é necessário armazenar a hierarquia das frentes de onda, que devem ser utilizadas no instante da geração do esqueleto.

Observe na figura 3.5i que o esqueleto não se mantém centralizado na região próxima a bifurcação. A fim de resolver este problema, introduzimos um parâmetro (R_{ec}) que define uma distância de recuo do esqueleto em regiões de bifurcação. Na figura 3.6 vemos o esqueleto resultante para três valores de R_{ec} , $R_{ec} = 0$, $R_{ec} = R$ e $R_{ec} = 2R$, onde R é o raio médio da estrutura de interesse. Podemos observar que, para $R_{ec} = R$, o esqueleto representa bem a posição da bifurcação.

Considerando os extremos do esqueleto e seu ponto de bifurcação como vértices, podemos representar a estrutura de interesse pela rede mostrada na figura 3.6c. Aqui novamente, os pesos $w_{1,2}$, $w_{2,3}$ e $w_{2,4}$ podem representar aspectos geométricos da estrutura.

Figura 3.6 – (a) Esqueletos obtidos quando consideramos diferentes valores para o recuo do ponto de bifurcação (R_{ec}). Em (b) mostramos em detalhes a região de bifurcação, bem como os valores de R_{ec} utilizados. Em (c) mostramos a rede resultante, quando consideramos as extremidades e a confluência como vértices.

Figura 3.7 – Evolução da propagação de rótulos sobre uma estrutura sintética contendo um loop. A seta na figura (g) mostra a região onde há o choque entre duas frentes de onda.

3.2.2 Loops

Obviamente, nem todas estruturas biológicas ramificadas são topologicamente equivalentes às árvores ou seja, muitas delas apresentam circuitos fechados, ou loops. Os loops surgem muitas vezes como consequência das variações da carga que flui através de um sistema, ou então como resposta a possíveis falhas desses sistemas (70, 71). Sendo assim, iremos agora levar em conta a possibilidade da existência de loops em nossas estruturas.

Neste caso, duas ou mais frentes de onda irão se chocar, indicando a presença de um circuito fechado (loop). Como exemplo, olhemos para a figura 3.7. Em (a), a frente de onda está se propagando ao longo de um ramo e em (b) ela se divide em duas, indicando

uma bifurcação. Outra bifurcação ocorre em (c) e em (e), referente aos ramos filhos da bifurcação que ocorreu em (b). Veja que em (g), a frente de onda que está se propagando irá colidir com uma frente de onda bifurcada (branca). Este choque é denotado pelos voxels marcados em verde em (h). Veja que são marcados tanto os vértices que possuem contato direto, quanto aqueles que não possuem contato, mas pertencem a mesma hierarquia de bifurcação. Em (i), terminamos a propagação da última frente de onda que estava na lista de espera.

Figura 3.8 – Em (a), apresentamos o esqueleto obtido através da união do centro de massa de consecutivas frentes de onda, bem como algumas frentes de onda encontradas ao longo da estrutura. O detalhe do esqueleto pode ser visto em (b). Em(c) apresentamos a rede resultante, quando consideramos as extremidades e as confluências como vértices.

Em 3.8a-b, podemos ver o esqueleto resultante após a conexão do centro de massa da sucessivas frentes de onda. Em 3.8c mostramos o grafo resultante, quando consideramos as extremidades e os pontos de confluência como vértices de uma rede. Mais uma vez, cada aresta desta rede pode ter associado um valor escalar referente à propriedades geométricas da estrutura original.

3.2.3 Limitações da propagação de rótulos

Embora o algoritmo descrito até o momento seja bastante eficiente e rápido, o esqueleto resultante é satisfatório apenas nos casos em que o comprimento dos ramos da estrutura é muito maior do que o seu diâmetro. Esse critério é satisfeito em muitas estruturas biológicas, como artérias, pulmões e vasos de transporte em folhas - veja figuras 3.9a-c. Entretanto, existem muitas outras estruturas onde esta condição não é satisfeita,

Figura 3.9 – Exemplos de sistemas onde o comprimento dos canais é muito maior do que seu diâmetro: (a) pulmão ([http : //www.bioeng.auckland.ac.nz](http://www.bioeng.auckland.ac.nz)), (b) artérias e (c) vascularização de uma folha ([http : //www.backyardnature.net/yucatan/casearia.htm](http://www.backyardnature.net/yucatan/casearia.htm)). Exemplos de sistemas onde o comprimento e o diâmetro dos canais são da mesma ordem: (d) mitocôndrias, (e) sistema de poros no solo e (f) sistema de canais nos ossos corticais.

como aquelas mostradas nas figuras 3.9e-f. A fim de tornar nosso algoritmo ainda mais geral e ampliar sua fronteira de aplicabilidade, é necessário utilizar novas técnicas em conjunto com a propagação de rótulos. Neste trabalho utilizamos duas técnicas auxiliares, a regularização utilizando o gradiente do campo divergente e um procedimento de afinamento. Na próxima seção daremos mais detalhes sobre cada uma delas.

3.3 Aperfeiçoamento do software

3.3.1 Regularização utilizando divergente do campo gradiente

Dada uma estrutura 3D (ou 2D), imagine que sejamos capazes de associar a cada voxel (ou pixel) desta estrutura, um valor escalar que possui a seguinte propriedade: próximo a superfície da estrutura, os valores são baixos e aumentam gradualmente a medida que nos afastamos das bordas da estrutura, como mostra a figura 3.10a. Agora, podemos calcular um campo vetorial sobre o campo escalar definido pela intensidade de

Figura 3.10 – (a) campo escalar gerado pela intensidade dos voxels de um objeto. (b) campo vetorial gerado pelo gradiente do campo escalar (a). (c) campo escalar gerado pelo divergente do campo vetorial (b). (d) campo escalar original e (e) sua respectiva segmentação. (f) campo escalar gerado pelo divergente do gradiente e (g) sua respectiva segmentação. Observe que a segmentação em (g) é mais uniforme e regular do que aquela obtida aplicando a limiarização diretamente nos dados reais (e).

cada voxel. Assumindo que estamos interessados nos vetores gradientes, podemos avaliar a direção de maior variação do campo escalar sobre cada voxel e atribuir a eles o vetor gradiente resultante. Este procedimento define um campo vetorial, como aquele mostrado na figura 3.10b. Em seguida, podemos utilizar o operador divergente para estudar como a direção dos vetores mudam sobre cada voxel, assim geramos um novo campo escalar. Este novo campo possui a importante propriedade de que o centro dos objetos terão valores bem próximos de zero, já que nestas regiões, os vetores gradientes irão apontar para diversas direções. Por outro lado, próximo as regiões de superfície, teremos altos valores de divergente, já que ali todos os vetores gradiente apontam aproximadamente, para a mesma direção. Um exemplo deste resultado por ser visto na figura 3.10c. Este procedimento de regularização é importante pois, resolve melhor a separação entre ramos, o que contribui para uma melhor segmentação do objeto de interesse, como é mostrado nas figuras 3.10d-g. Em 3.10f vemos que o objeto segmentado (mostrado em azul) não é equivalente ao objeto original, uma vez que a uniformidade dos dados reais é bem baixa. Por outro lado, olhando para 3.10g, vemos que o objeto segmentado é bastante similar

aquele regularizado (figura 3.10e).

Nossa hipótese inicial assume que sejamos capazes de criar um campo escalar com baixos valores próximo a superfície e altos valores no centro dos objetos. Uma possibilidade é utilizar o campo escalar gerado pela transformada distância (72). Basicamente, esta transformação atribui a cada voxel, o valor da distância entre ele e o voxel da borda mais próximo. Embora várias métricas possam ser empregadas para estimar a distância entre dois voxels, neste trabalho iremos considerar apenas a métrica euclidiana.

Figura 3.11 – (a) corte transversal de uma mitocôndria. (b) intensidade ao longo da linha marcada em azul em (a). Observe que a intensidade é maior na região central, possibilitando a aplicação da regularização da estrutura pelo divergente do campo gradiente, sem a necessidade do uso da transformada distância.

Uma outra possibilidade para geração do campo escalar inicial surge naturalmente quando trabalhamos com imagens obtidas através de microscopia de convolução. Neste tipo de microscopia, a intensidade dos voxels é proporcional a densidade do sistema biológico estudado. Felizmente, alguns sistemas biológicos possuem a propriedade de serem mais densos no centro do que nas regiões superficiais. Consequentemente, os tons de cinza das imagens obtidas pela microscopia já definem um campo escalar com a propriedade desejada. Como pode ser visto na figura 3.11, este é o caso, das mitocôndrias que serão estudadas neste trabalho.

3.3.2 Afinamento

A segunda técnica auxiliar, utilizada em conjunto com a propagação de rótulos, foi a *transformação do eixo central*, ou *afinamento* (thinning). Os algoritmos de afinamento possuem longa tradição na área de processamento de imagens, sendo aplicados aos mais diversos tipos de problemas, como reconhecimento de formas (73), reconhecimento de movimento (74), caracterização e reconhecimento de órgãos (75), entre outras.

Os algoritmos de afinamento possuem o propósito de converter um dado objeto em um outro, mais simples, mas que mantém a topologia do objeto original. Um bom al-

goritmo de esqueletização deve possuir as seguintes propriedades (76): i) homotopia, ii) finura (thickness) e iii) centralidade. A primeira diz que o objeto resultante deve ser topologicamente equivalente ao inicial. A segunda impõe ao objeto resultante a espessura de apenas um voxel e por fim, a centralidade garante que ele esteja centralizado no objeto original. Atualmente, existem diversos algoritmos de afinamento, entretanto poucos conseguem satisfazer as três condições citadas, sendo que a aplicação interfere diretamente na escolha do algoritmo a ser usado. O conceito de redução de um objeto a uma forma mais simples foi primeiramente proposto por H. Blum em 1967 (77) e o primeiro algoritmo de afinamento foi proposto duas décadas depois, por Kong e Rosenfeld (78). Em 1996, Ma e Sonka (79) propuseram um algoritmo que foi amplamente difundido e é utilizado até os dias de hoje, embora diversas propostas têm sido feitas para aumentar o desempenho e a qualidade dos resultados obtidos. Revisões sobre algoritmos de afinamento podem ser encontradas em (80–82).

Basicamente, o que fazem os algoritmos de afinamento é classificar os voxels de um objeto como triviais e não triviais, de acordo com critérios de adjacência. Quando um voxel é considerado trivial, ele é então removido do volume. O processo iterativo termina quando nenhum voxel restante no volume pode ser considerado trivial. Ao longo deste trabalho, iremos utilizar o algoritmo proposto por Kálman Palágyi e Attila Kuba (83), que utiliza um conjunto de máscaras para efetuar a classificação dos voxels. As máscaras primárias são mostradas na figura 3.12a. Com elas, nós geramos um conjunto de 144 máscaras secundárias, através de rotações (0° , 90° , 180° e 270°) ao longo das direções dos seis vizinhos de cada voxel (3.12b). Um exemplo do algoritmo de afinamento agindo sobre o voxel v do volume vol é mostrado abaixo:

```

for each direction
  for each angle
    for each primary-mask
      if TRIVIAL(vol,v,primary-mask,angle,direction) == True
        delete v from vol
      end
    end
  end
end

```

Algoritmo 1: Algoritmo de afinamento utilizado para classificar o voxel v contido no volume vol .

Na figura 3.13 mostramos alguns resultados obtidos utilizando este algoritmo. Fica claro pelas imagens que após o afinamento, a tarefa de detecção das arestas e pontos de bifurcações no novo objeto requer menos esforço, uma vez que o objeto resultante é bem

Figura 3.12 – (a) seis máscaras primárias da direção U e rotação de 0° . (b) direções que devem ser consideradas. (c) exemplo da máscara M4 na direção U e rotação de 90° . (d) exemplo da máscara M3 na direção E e rotação de 0° .

mais simples. Agora podemos utilizar a propagação de rótulos descrita anteriormente para descrever estes novos objetos através de redes.

3.3.3 Diâmetro

Informações adicionais, como o diâmetro da estrutura original, podem ainda ser agregadas aos esqueletos. Tomemos como exemplo, a estrutura apresentada na figura 3.14a e sua versão processada pelo algoritmo de thinning descrito acima (figura 3.14b). Utili-

Figura 3.13 – Exemplos de aplicação do algoritmo de afinamento. Em cinza temos a representação do objeto original e em vermelho o resultado do algoritmo. Observe que o objeto final possui as mesmas características topológicas do objeto original.

zando agora a metodologia da propagação de rótulos descrita anteriormente, obtemos o esqueleto mostrado em 3.14c. Gostaríamos de associar a cada ponto deste esqueleto o diâmetro da estrutura original e para isso, usaremos a aproximação de que o diâmetro em um ponto é bem aproximado pelo diâmetro do maior círculo inscrito na seção transversal naquele ponto, como mostra a figura 3.14d. O valor do diâmetro do maior círculo inscrito pode ser facilmente estimado através da transformada distância do objeto original. O resultado desta abordagem produz resultados similares aqueles da figura 3.14e.

Figura 3.14 – (a) Estrutura original, (b) resultado do processo de afinamento, (c) esqueleto obtido após aplicar a propagação de rótulos na estrutura (b). Em (d) mostramos a seção transversal em um ponto do esqueleto, bem como sua transformada distância. (e) Esqueleto no qual o diâmetro em cada ponto é proporcional ao valor da transformada distância de (a).

4 Aplicações

Ao longo deste projeto de doutorado, realizamos diversas aplicações dos métodos descritos nas seções anteriores. Tais métodos foram unificados em uma mesma interface gráfica desenvolvida em linguagem C++, sendo que o desenvolvimento desta interface segue basicamente as etapas descritas na seção anterior. Durante a conclusão de cada etapa, uma ou mais aplicações em sistemas reais foram realizadas. Nesta seção, iremos apresentar os resultados de cada uma delas.

4.1 Complexidade geométrica de artérias carótidas

Um sistema de grande interesse na medicina são as artérias carótidas. Em especial o trecho paraselar destas artérias, pois estão relacionados a ocorrência de diversas patologias vasculares. Entre elas, citamos a arteriosclerose e o aneurisma (84–87), e da síndrome da morte súbita em crianças (16, 88). Em colaboração com o Centro para Anatomia e Biologia Celular da Universidade Médica de Viena, utilizamos o algoritmo de propagação de rótulos para gerar o esqueleto de um conjunto de artérias e carótidas e estudar suas características geométricas. Nosso objetivo foi relacionar as propriedades geométricas, como curvatura e torção aos diagnósticos clínicos conhecidos. No total, foram fornecidos 31 trechos paraselar de adultos e 27 de crianças com idade média de 4,2 meses. Na figura 4.1a, mostramos a região da artéria carótida de nosso interesse, bem como alguns exemplos de reconstrução dos dados de adultos (figura 4.1b) e crianças (figura 4.1c). Detalhes sobre a técnica utilizada para realizar a aquisição das imagens podem ser encontradas na referência (89).

As imagens fornecidas não passaram por nenhum pré-processamento pois já encontravam-se binarizadas. Sendo assim, procedemos com a reconstrução tridimensional e consequentemente com a propagação de rótulos para identificação do esqueleto. Exemplos de esqueletos obtidos para diferentes artérias podem ser vistos na figura 4.2. Como foi mencionado anteriormente (veja figura 3.4, por exemplo), os esqueletos podem ser representados por curvas paramétricas $\vec{u}(t) = [x(t), y(t), z(t)]$. Sendo assim, podemos utilizar os conceitos de geometria diferencial para estudar a geometria destas curvas e indiretamente, estudar a geometria do trecho paraselar das artérias carótidas.

Em cada caso, cada componente do esqueleto foi ajustada por um polinômio de nona ordem. Em seguida, calculamos a curvatura e a torção (90) da curva paramétrica de

acordo com as equações:

$$cur(\vec{u}) = \frac{|\dot{\vec{u}} \times \ddot{\vec{u}}|}{|\dot{\vec{u}}|^3} \quad (4.1)$$

$$tor(\vec{u}) = \frac{|\dot{\vec{u}} \cdot (\ddot{\vec{u}} \times \ddot{\vec{u}})|}{|\dot{\vec{u}} \times \ddot{\vec{u}}|^2} \quad (4.2)$$

A fim de tornar as medidas invariantes à escala, utilizamos a energia de curvatura e a energia de torção (91, 92) de cada curva em nossas análises. Estas energias são definidas como segue:

$$E_{cur} = \frac{C^2}{N_c} \sum_{t=1}^{N_c} cur(\vec{u}(t))^2 \quad (4.3)$$

$$E_{tor} = \frac{C^2}{N_c} \sum_{t=1}^{N_c} tor(\vec{u}(t))^2 \quad (4.4)$$

Figura 4.1 – (a) Nomenclatura dos diferentes trechos da artéria carótida. Exemplos do trecho paraselar de (b) adultos e (c) crianças. Imagem adaptada de <http://health.allrefer.com>.

Figura 4.2 – Exemplo do esqueleto obtido para o trecho paraselar de uma artéria de (a) adulto e (b) criança.

Figura 4.3 – (a) Localização do trecho paraselar in situ. (b) estimacão do ângulo da curva posterior do sifão carótido de acordo com uma projeção (b) em perspectiva (PCp) e (c) ortogonal (PCo). Correlações entre os valores da energia de torção e (d) PCo (e) PCp. Correlações entre os valores da energia de torção e (f) PCo (g) PCp. Imagem adaptada de (16).

onde C é o perímetro da curva $\vec{u}$, N_c o número de pontos que ela possui, $cur(\vec{u}(t))$ é a curvatura no ponto t e $tor(\vec{u}(t))$ é a torção no ponto t .

Nosso primeiro resultado corresponde a comparação do nosso método com uma técnica comumente utilizadas para estudar a geometria de artérias. Tal técnica consiste em medir o ângulo formado pelo cotovelo posterior do sifão carótido, como indicado na figura 4.3a. A dificuldade exigida para estimar este ângulo na estrutura tridimensional, força os especialistas a utilizarem duas projeções bidimensionais, chamadas *perspectiva* (figura 4.3b) e *ortogonal* (figura 4.3c). Os ângulos estimados em cada uma delas são denotados por PC_p e PC_o , respectivamente. Tendo em vista que as projeções são feitas manualmente, demandam tempo e estão sujeitas as percepções do usuário, é de fato interessante estabelecer um método automático e eficaz para realizar a análise geométrica dessas artérias.

Iniciamos nosso estudo escolhendo ao acaso 30 pacientes adultos sem lesões na região paraselar. Em seguida, extraímos o esqueleto de suas artérias e calculamos os valores de E_{cur} e E_{tor} das respectivas curvas paramétricas. Os resultados foram comparados com os ângulos PC_p e PC_o , e os resultados são apresentados nas figuras 4.3d-g. Observamos uma fraca correlação entre a torção e os ângulos PC_p e PC_o , 0.39 e 0.29, respectivamente. Por outro lado, uma correlação significativa foi obtida entre a curvatura e os ângulos PC_p e PC_o , 0.54 e 0.45, respectivamente. Estes resultados indicam que a técnica atualmente utilizada para estudar a geometria de artérias, não é útil para investigar características como a torsão. Por outro lado, ela pode fornecer uma estimativa grosseira para o comportamento da curvatura (16).

O foco seguinte do nosso estudo foi comparar as energias de curvatura e torção em artérias de adultos e crianças, a fim de identificar as principais diferenças entre elas. Os

resultados estão sumarizados na figura 4.4, onde mostramos os valores de E_{cur} e E_{tor} em função da idade das artérias. Podemos perceber pela figura que a energia de curvatura é cerca de quatro vezes menor em artérias infantis do que artérias de adultos. Esta relação é ainda maior para energia de torção, que chega a um fator de 16 vezes menor para artérias infantis. Estes resultados indicam que, no período da infância e adolescência, os processos de remodelagem das artérias as conduzem a estruturas mais complexas, principalmente pelo aumento da torção. Ou seja, as artérias que estavam inicialmente acomodadas em um estado quase planar começam a se projetar no espaço tridimensional. Detalhes sobre este processo de remodelagem podem ser encontrados na referência (17). Em outra análise, investigamos a relação entre o número de dobramentos nas artérias e os valores das energias de curvatura e torção, sendo que nenhuma correlação significativa foi encontrada. Assim, nossos resultados sugerem que indivíduos com diferentes valores de E_{cur} e E_{tor} possuem os mesmos riscos de apresentarem arteriosclerose, já que esta desordem está intimamente relacionada a presença dos dobramentos. O foco seguinte do nosso estudo foi comparar as energias de curvatura e torção em artérias de adultos e crianças, a fim de identificar as principais diferenças entre elas. Os resultados estão sumarizados na figura 4.4, onde mostramos os valores de E_{cur} e E_{tor} em função da idade das artérias. Podemos perceber pela figura que a energia de curvatura é cerca de quatro vezes menor em artérias infantis do que artérias de adultos. Esta relação é ainda maior para energia de torção, que chega a um fator de 16 vezes menor para artérias infantis. Estes resultados indicam que, no período da infância e adolescência, os processos de remodelagem das artérias as conduzem a estruturas mais complexas, principalmente pelo aumento da torção. Ou seja, as artérias que estavam inicialmente acomodadas em um estado quase planar* começam a se projetar no espaço tridimensional. Detalhes sobre este processo de remodelagem podem ser encontrados na referência (17). Em outra análise, investigamos a relação entre o número de dobramentos nas artérias e os valores das energias de curvatura e torção, sendo que nenhuma correlação significativa foi encontrada. Assim, nossos resultados sugerem que indivíduos com diferentes valores de E_{cur} e E_{tor} possuem os mesmos riscos de apresentarem arteriosclerose, já que esta desordem está intimamente relacionada a presença dos dobramentos.

*qualquer curva que pode ser acomodada em um plano possui torção nula e consequentemente $E_{tor} = 0$ (90).

Figura 4.4 – Energia de (a) curvatura e (b) torção em função da idade das artérias, em meses. Imagem adaptada de (17).

4.2 Diâmetro ao longo da árvore arterial de embriões de galinha

Diversas desordens arteriais estão não só relacionadas a complexidade geométrica das artérias, mas principalmente ao estreitamento do canal arterial. Para que possam ser feitos diagnósticos cada vez mais confiáveis, é então de grande importância estabelecer um método preciso para estimar corretamente o diâmetro de sistemas arteriais. Ao longo dos últimos anos, diversas abordagens têm sido propostas, mas em geral elas não são completamente automáticas (93) ou limitam-se a análise bidimensional de projeções arteriais (94).

Aqui, vamos apresentar os resultados obtidos durante o estudo de embriões de galinha, na qual utilizamos os esqueletos do sistema arterial para estimar o diâmetro ao longo de cada ponto da árvore arterial. Um conjunto de 36 árvores arteriais de embriões de galinha no estágio de desenvolvimento HH34 nos foram enviadas pelo professor Wolfgang Weninger do Centro para Anatomia e Biologia Celular da Universidade Médica de Viena. Os dados enviados estavam armazenados em imagens já binarizadas e consequentemente, não sofreram nenhum pré-processamento. As árvores arteriais foram dividida em três regiões, sendo elas cabeça, tórax e pélvis. O posicionamento das árvores no embrião está esquematizado na figura 4.5.

Utilizando a versão do nosso software que inclui bifurcações, os esqueletos das árvores arteriais foram extraídos. Em seguida, para cada ponto do esqueleto, identificamos a seção de corte dada pela intersecção da artéria com o plano perpendicular a tangente do esqueleto, como mostra a figura 4.6b. As imagens correspondentes a estas seções de corte foram então armazenadas para análise posterior. Finalmente, calculamos o perímetro P e a área A de cada imagem. Com estes valores, estimamos o diâmetro (D_s) e o fator de

Figura 4.6 – (a) Destaque para o trecho R4, que foi utilizado durante a fase de validação do algoritmo. (b) Exemplos de três seções de corte do trecho R4. (c) Comparação dos diâmetros obtidos através do nosso algoritmo e dos diâmetros obtidos manualmente, através da contagem de pixels. (d) Comportamento do fator de forma ao longo do trecho R4.

uma diminuição do diâmetro. Nesta região, o fator de forma é alto, indicando que o canal é quase circular, fato este que é corroborado pela inspeção visual da seção de corte nesta região.

Nossos resultados são promissores no que diz respeito ao desenvolvimento de um ambiente integrado, capaz de visualizar, seccionar regiões de interesse e extrair medidas com alta precisão. Trabalhos futuros deverão se estender ao estudo comparativo da morfometria entre embriões de diferentes espécies.

4.3 Estrutura mitocondrial

A inclusão de loops em nosso algoritmo nos permitiu ir além das estruturas arteriais e explorar um número maior de sistemas biológicos ramificados. O primeiro exemplo consiste no estudo da estrutura mitocondrial, cuja quantificação é uma tarefa altamente complexa porém, de grande importância. Essa importância vem do fato de que, as variações estruturais da mitocôndrias são respostas as necessidades metabólicas dos organismos. Consequentemente, estudos puramente topológicos podem nos ajudar compreender

melhor a dinâmica dos processos celulares.

Nosso objetivo foi estabelecer um método automático, baseado em grafos, que fosse capaz de extrair informações acerca da arquitetura mitocondrial. Alguns trabalhos foram desenvolvidos nesse sentido, porém eles utilizam métodos semi-automáticos que, muitas vezes, necessitam de interferência humana (95).

Ao longo dessa seção iremos apresentar os resultados obtidos em colaboração com o Departamento de Bioquímica e Biofísica da Universidade da Califórnia São Francisco, que foi responsável por fornecer as imagens das mitocondrias de *Saccharomyces cerevisiae*, obtidas via microscopia confocal. Este organismo foi escolhido como objeto de estudo graças a sua tratabilidade genética, que permitirá no futuro, identificar os genes responsáveis por regular a organização da topologia mitocondrial.

Marcadores foram utilizados para destacar a estrutura mitocondrial de diversas células que foram em seguida digitalizadas via microscopia confocal. As imagens digitais foram então salvas em níveis de cinza, como mostra a figura 4.7a. Estas imagens foram submetidas a um pré-processamento que consistiu nas etapas de reamostragem (figura 4.7b) e suavização da mediana (figura 4.7c). Em seguida, as imagens foram submetidas ao processo de regularização do divergente do campo gradiente (figuras 4.7d-e), descrito na seção 3.3.1. Por fim, uma nova suavização da mediana foi realizada e as imagens foram binarizadas através de uma segmentação (figura 4.7f). Ao longo do desenvolvimento deste doutorado, os parâmetros de cada uma dessas etapas foi otimizado para produzir os melhores resultados possíveis. A seguir, iremos descrever brevemente estes parâmetros.

As imagens originais eram anisotrópicas, sendo que as dimensões em x e y correspondiam a 61nm, enquanto em z correspondia a 200nm. Após a reamostragem, todas as direções passaram a ter dimensão 66.66nm. Para o filtro da mediana, utilizamos o valor $B = 5$ pixels, que corresponde ao diâmetro típico dos canais mitocondriais. Com respeito a última etapa de pré-processamento, realizamos alguns estudos relacionados aos efeitos do limiar de segmentação na estrutura final das redes. Tais estudos serão apresentados a frente. A fase de regularização mostrou-se extremamente importante para a análise dos dados mitocondriais, uma vez que ela foi capaz de resolver ambiguidades provenientes dos ruídos intrínsecos da captura das imagens. Como mostramos na figura 4.8a-c, regiões aparentemente conectadas foram separadas após a regularização, enquanto algumas regiões desconectadas deram origem a esqueleto contínuos, como vemos em 4.8d-f. Veja que estas observações foram validadas utilizando uma mitocôndria construída manualmente, a qual foram incorporados os ruídos típicos do processo de digitalização. Neste caso, sabíamos previamente qual o esqueleto que deveria ser obtido pelo método. A acurácia do método nos forneceu um esqueleto com 88% de concordância com o esperado, quando comparamos o número de arestas de ambos esqueletos, e 90% de concordância, quando comparamos o comprimento total dos esqueletos.

Após o pré-processamento, as imagens foram utilizadas para criar modelos volumétricos

Figura 4.7 – Visualização 2D (primeira linha), 3D volumétrica (segunda linha) e 3D superficial (terceira linha) das etapas de pré-processamento e regularização de uma mitocôndria. (a) Dado original, (b) dado reamostrado, (c) dado suavizado (filtragem mediana), (d) cálculo do gradiente, (e) cálculo do divergente, (f) dado segmentado.

Figura 4.8 – (a) Representação tridimensional e seu esqueleto. (b) Detalhe de uma região aparentemente conectada que originou uma descontinuidade no esqueleto. (c) Detalhe da descontinuidade nos níveis de cinza do dado após a regularização. (d) Representação tridimensional e seu esqueleto. (e) Detalhe de uma região descontinua que originou um esqueleto contínuo. (f) Detalhe do mesmo trecho em níveis de cinza no dado regularizado, onde vemos que não há descontinuidade.

das mitocôndrias. Tais modelos foram submetidos ao processo de afinamento, gerando objetos simplificados para as estruturas mitocondriais. Por fim, estes objetos passaram pela fase de propagação de rótulos e seus esqueletos foram então extraídos, como mostra a figuras 4.8a e 4.8d.

Nossos resultados indicam que a abordagem adotada aqui foi altamente eficiente e robusta, fornecendo redes que representam com fidelidade a estrutura das mitocôndrias. Entretanto, o limiar de segmentação mostrou-se crucial, sendo que pequenas variações em seu valor podem causar enormes mudanças topológicas na rede final. Valores baixos tendem a gerar uma estrutura binária compacta, na qual a separação entre os canais não

fica bem estabelecida. Por outro lado, limiares altos tendem a particionar a estrutura em pequenos fragmentos, como podemos ver na figura 4.9a. As respectivas redes são mostradas na figura 4.9b. Na figura 4.9d-e vemos o comportamento do comprimento total do esqueleto mitocondrial, bem como algumas quantidades topológicas da rede em função do limiar de segmentação. Vemos que o comprimento permanece praticamente constante para baixos valores do limiar e que diminui drasticamente após um valor crítico de aproximadamente $I_c \approx 106$. Acima deste valor, o número de vértice na rede começa a aumentar, embora o número de vértices da maior componente diminua, indicando a fragmentação da rede. Para limiar maiores que I_c , também observamos uma diminuição no grau médio da rede, que converge para 1, indicando a existência apenas de pares isolados. Então, é necessário estabelecer um critério para a escolha do limiar ótimo, I^* . Como foi dito na seção 3.3.1, as imagens mitocondriais possuem a propriedade de terem intensidade alta no centro dos canais e intensidades baixas próximo as superfícies. Assumindo que tal variação de intensidade na direção superfície-centro possa ser descrita por uma função gaussiana (veja figura 3.11), é possível mostrar que o limiar ótimo seria aproximadamente dado por $I^* = I_f + 28$, onde I_f corresponde a intensidade de fundo após a etapa de regularização. Porém, graças ao desvio da função gaussiana, verificamos que o limiar ótimo corresponde a $I^* = I_f + 20$, e então utilizamos este valor para todas

Figura 4.9 – (a) Estrutura de uma mitocôndria em função do limiar de segmentação e (b) suas respectivas redes. (c) Comprimento total do esqueleto mitocondrial (número de pixels) como função de I^* . As seguintes quantidades topológicas da rede final também foram obtidas: (d) número de vértices e (e) grau médio.

as mitocôndrias estudadas neste projeto, obtendo com ele ótimos resultados.

Mais um ponto que merece atenção está relacionado a resolução do nosso método, ou seja a separação mínima ($d_{||}^{(min)}$) entre dois canais paralelos, para que eles não sejam considerados apenas um. Também existe a distância mínima ($d_{..}^{(min)}$) entre duas confluências, para que elas não sejam mapeadas no mesmo vértice. Através de amostras geométricas simples, pudemos identificar cada uma dessas distâncias, sendo elas $d_{||}^{(min)} = 5$ pixels e $d_{..}^{(min)} = 8$ pixels.

Estes resultados demonstram que a abordagem por nós desenvolvida aplica-se com sucesso ao estudo da topologia mitocondrial. Tal abordagem será de grande utilidade em problemas nos quais são estudados por exemplo, a relação entre o volume mitocondrial (V_m) e o volume celular (V_c). Problemas deste tipo são de grande importância, já que as relações de escala entre estas duas quantidades ainda não são bem compreendidas. O volume mitocondrial pode ser diretamente derivado de nossa representação em termos de rede. Considerando que o peso de cada aresta $w_{i,j}$ é dado pelo comprimento de arco desta aresta, definimos:

$$V_m = \pi R_m \sum_{i=1}^N \sum_{j=1}^N a_{i,j} w_{i,j}, \quad (4.7)$$

onde R_m é o raio médio dos canais mitocondriais. A relação entre V_m e V_c pode ser estudada por exemplo, em células que estão em fase de divisão celular. Na figura 4.10a apresentamos o exemplo de uma célula que encontra-se nesta fase, para a qual podemos facilmente separar a célula mãe da célula filha. Nas figuras 4.10b e 4.10c, vemos que o volume mitocondrial é altamente correlacionado com o volume celular nas células filhas, e que esta forte correlação não é observada nas células mães. Este resultado está relacionado ao fato de que as células mães perdem sua capacidade de manter a relação de escala ao longo da vida. Isso ocorre pois elas cedem grande quantidade de suas mitocôndrias às células filhas durante os primeiros estágios da vida, momento na qual estas células não podem ainda produzir toda mitocôndria necessária para seu desenvolvimento. Portanto, com ajuda das mães, as células filhas conseguem manter a relação de escala até se dividirem pela primeira vez, momento na qual elas irão auxiliar suas filhas a manter a relação entre V_m e V_c . Estes e outros resultados estão descritos no artigo *Mitochondrial Network Size Scaling Budding Yeast is Achieved in the Bud at the Expense of the Mother* que está em fase de finalização.

Figura 4.10 – (a) Exemplo de mitocôndria em fase de divisão. Correlação entre V_m e V_c para (b) células filhas e para (c) células mães. Os valores de V_c foram estimado ajustando um elipsoide as imagens microscópicas do envoltório celular.

4.4 Topologia e dinâmica das redes de poros do solo

Outro sistema geográfico analisado neste projeto consiste em uma amostra de solo obtida via micro-tomografia computadorizada. Estudos desse tipo de sistema são de grande relevância graças a sua importância em vários setores, já que as características estruturais do solo, como densidade e tamanho dos agregados, têm impacto direto em suas propriedades de transporte. Embora muitos trabalhos sobre este tema têm sido desenvolvidos nos últimos anos (96–98), um pequeno número deles utiliza conceitos de redes complexas (19). O principal objetivo desta seção é mostrar como é possível representar amostras de solo através de redes e obter informação relevantes sobre esses sistemas a partir do estudo das respectivas redes. Iremos apresentar uma análise das propriedades topológicas da rede em questão que revela a presença de correlações não triviais neste tipo de sistema, bem como a existência de leis de potência na distribuição de certas quantidades. No que diz respeito aos aspectos dinâmicos, iremos explorar modelos epidemiológicos para propagação de fungos na amostra de solo. Nossos resultados indicam que a heterogeneidades das conexões são fundamentais para tornar a rede resistente a propagação de doenças. Por fim, mostramos como o conceito de acessibilidade pode ser utilizado para detectar os vértices mais influentes na dinâmica de propagação de doenças nesta rede.

Nosso objeto de estudo neste caso consiste nos poros de uma amostra de solo cúbica, de lado 2cm, que foi escaneada pelo sistema de micro-tomografia Metris X-TEK Benchtop. As imagens em níveis de cinza e tamanho 1000×1000 pixels foram binarizadas de acordo com um análise estatística do histograma na região correspondente a interface solo-poro (figura 4.11a). Detalhes completos sobre o procedimento de obtenção das imagens, podem ser encontrados na referência (19). As imagens foram então utilizadas para construir um modelo tridimensional para os poros contidos na amostra, como mostra a figura 4.11b. O algoritmo de afinamento foi aplicado a imagem binária e em seguida foi realizada a

Figura 4.11 – (a) Conjunto de imagens binárias obtidos através de micro-tomografia de uma amostra de solo. Nestas imagens, os pixels pretos correspondem ao solo, enquanto pixel brancos correspondem aos poros (espaço vazios). (b) Modelo tridimensional da amostra de solo, onde os poros são representados em cinza. (c) Rede obtida após o procedimento de afinamento da estrutura mostrada em (b). Aqui, as cores das arestas são proporcionais aos seus diâmetros.

propagação de rótulos, que produziu a rede mostrada na figura 4.11c. Tal rede possui $N = 56816$ vértices e $E = 79659$ arestas.

Ao investigar as principais características topológicas desta rede, verificamos que a distribuição de grau é bem descrita por uma lei de potência, como pode ser visto na figura 4.12a. O alto valor do expoente desta lei é comparável ao de outro sistema geográfico também estudado por nós, o sistema de canais em ossos corticais (11). Observamos que outras quantidades também são descritas por leis de potência, como a distribuição de comprimento das arestas (figura 4.12b) e a distribuição de diâmetro das arestas (figura 4.12c). No último caso, verificamos a existência de dois regimes, os quais são separados por um valor crítico d^* . Abaixo desse valor, a lei de potência tem expoente $\gamma_d = -5.0$, enquanto ela assume expoente $\gamma_d = -2.5$ acima dessa valor. Nós também investigamos a presença de correlações na rede do solo. Para tanto, estudamos o comportamento do *strength de diâmetro* (s_d) e do *strength de comprimento* (s_l) como função do grau de cada vértice. Essas quantidades são definidas como:

$$s_d(i) = \sum_{j=1}^N a_{i,j} d_{i,j} \quad e \quad (4.8)$$

$$s_l(i) = \sum_{j=1}^N a_{i,j} l_{i,j}, \quad (4.9)$$

onde $d_{i,j}$ é o diâmetro da aresta (i, j) , $l_{i,j}$ é o comprimento de arco desta aresta e $a_{i,j}$ é o elemento da matriz de adjacência que descreve a rede. Na figura 4.12d-e mostramos

Figura 4.12 – Distribuição de (a) grau, (d) comprimento das arestas e (c) diâmetro das arestas. Os expoentes correspondem a inclinação das linhas pontilhadas que ajustam as distribuições. Em (d) mostramos a comportamento do strength de diâmetro como função do grau. O comportamento linear indica a ausência de correlação. Em (e) mostramos a comportamento do strength de comprimento como função do grau. O comportamento super-linear indica a presença de correlação.

Figura 4.13 – (a) Distribuição da probabilidade de infecção dada pela equação 2.29 na rede do solo. (b) Fração do número de vértices infectados em função da probabilidade de remoção μ para rede homogênea e heterogênea. Cada ponto corresponde a média sobre 100 realizações estocásticas.

os gráficos de $s_d(i) \times k_i$ e $s_l(i) \times k_i$. Veja que no primeiro caso, a relação perfeitamente linear indica ausência de correlação entre as quantidades. Por outro lado, no segundo caso observamos uma relação super-linear, indicando a presença de correlações não triviais no sistema. Correlações desse tipo haviam sido observadas em outros sistemas, como na rede de aeroportos estudada em (34, 35). Tais correlações indicam que arestas longas tendem a se conectar com vértices altamente conectados.

Para investigar como a heterogeneidade das conexões dada pelas leis de potência, e a correlação super-linear afetam os aspectos dinâmicos da rede de poros, estudamos um mo-

delo epidemiológico para propagação de doenças nesse sistema. Embora as propriedades microscópicas desses processos permanecerem pouco conhecidas, muito tem se avançado no sentido de compreender os efeitos das características físicas do solo na propagação da doença (96, 98). Estudos desse tipo são importantes pois auxiliam no desenvolvimento de estratégias eficientes de imunização dos sistemas de interesse.

Aqui iremos utilizar os conceitos descritos na seção 2.6, onde descrevemos os aspectos teóricos do modelo SIR (Suscetível-Infectado-Removido), utilizado para simular propagações de infecções em redes complexas. É importante observar que no caso da rede de poros, onde temos fungos propagando-se ao longo dos poros (19, 98), a probabilidade de remoção μ , deve ser interpretada como a quantidade de nutrientes disponíveis nos vértices. Quando $\mu \rightarrow 1$, a quantidade nutrientes é muito baixa e portanto os fungos não conseguem sobreviver e tão logo o vértice j é removido e torna-se imune a infecção. Outra possível interpretação relaciona o valor de μ com a eficiência de um produto antifúngico inserido na amostra para combater a propagação da infecção.

Nossa primeira análise consistiu em verificar como estão distribuídos os valores de $\beta_{i,j}$ ao longo da rede de poros. O gráfico da figura 4.13a mostra a probabilidade de que uma aresta escolhida ao acaso tenha probabilidade de infecção β . Veja que o valor médio desta distribuição é dado por $\langle \beta \rangle = 0.389$. Observamos que os valores de β estão bem distribuídos ao longo do intervalo $[0, 1]$. Isso sugere que o comportamento do processo de propagação de infecção na rede do solo deve ser fortemente afetado pela heterogeneidade da estrutura. Para confirmar essa hipótese, simulamos o modelo SIR para duas versões da rede do solo. Na primeira (heterogênea), a probabilidade de infecção das arestas é dada pela equação 2.29 e sua distribuição é mostrada na figura 4.13a. Na segunda (homogênea), todas as arestas possuem a mesma probabilidade de infecção, dada por $\langle \beta \rangle = 0.389$. Os resultados das simulações são mostrados na figura 4.13b.

Nossos resultados indicam que na rede homogênea o número de vértices infectado torna-se diferente de zero quando $\mu^{(homo)} \approx 0.389$ ($R_0 = 1$), enquanto na rede heterogênea nós temos $\mu^{(hete)} \approx 0.29$. Isso sugere que a rede heterogênea é cerca de 26% mais resistente a disseminação da infecção do que a rede homogênea. Este fato demonstra a importância das correlações na rede do solo, bem como a presença das leis de potência que governam a distribuição das características topológicas desta rede.

Detectando os vértices mais influentes

Como vimos, a dinâmica que rege a evolução do modelo SIR possui duas fases distintas (19). A primeira ocorre para valores elevados de μ , na qual a epidemia não consegue se propagar e o sistema não é tomado pela infecção. A outra, ocorre para valores baixos de μ , quando a rede é completamente invadida e todos os vértices são infectados antes de serem removidos. Nestes dois regimes, as propriedades locais de cada vértice não afetam de maneira significativa o panorama final da epidemia. Entretanto, na região crítica,

próxima a transição entre essas duas fases, o número de vértices infectados é altamente dependente das características locais do primeiro vértice infectado. Nosso senso comum nos força a crer que vértices mais conectados tendem a propagar mais rapidamente a doença. Entretanto, foi mostrado recentemente que este não é o que ocorre em várias redes reais. Em (99) os autores mostraram que os vértices mais influentes na propagação de doenças e informações são aqueles com os maiores *k-index*. Estes vértices definem um núcleo para a rede, sendo que os valores de *k-index* são determinados por um processo chamado *decomposição k-shell*. Mais detalhes sobre este processo pode ser encontrado no Apêndice 5.2. Observe que a influência de um vértice i , denotada por Γ_i , é medida pelo número de vértices que a doença iniciada em i , consegue infectar antes de ser extinta. Entretanto, ao contrário do que foi reportado em (99), não observamos qualquer correlação entre Γ_i e o valor de *k-index* do vértice i na rede de poros. Este resultado pode ser visto na figura 4.14a, onde mostramos a curva $\Gamma \times k\text{-index}$ em cinza. Veja que valores médios foram tomados para vértices com mesmos *k-index*. A principal razão para não termos observado a correlação entre as duas quantidades, consiste no fato de que o procedimento proposto em (99) para detectar os vértices mais influentes não leva em conta o peso das arestas, atributo crucial na propagação de doenças e informações.

Afim de abordar o problema da detecção dos vértices mais influentes levando em conta o peso $w_{i,j} = \beta_{i,j}$ das conexões, calculamos a acessibilidade de todos os vértices desta rede para o parâmetro $\ell = D$, onde o diâmetro D desta rede é aproximadamente 30. Em seguida, fizemos a curva de $\Gamma_i \times \mathcal{A}_i$, como mostra a figura 4.14a - curva preta. O resultado encontrado apresenta um comportamento quase monotônico entre as duas quantidades, indicando que vértices com alta acessibilidade são mais influentes, no sentido de que, doenças iniciadas neles possuem potencial de infectar mais indivíduos. Por outro lado, vértices com baixa acessibilidade são menos influentes e conseqüentemente, conduzem a baixos valores de Γ . Resultados semelhantes foram obtidos para o tempo de persistência da doença na rede (figura 4.14b). Enquanto os vértices mais influentes geram epidemias com maior tempo de persistência, epidemias iniciadas em vértices pouco influentes são rapidamente extintas. Como mostra a figura 4.14c, este resultado fica claro quando olhamos o perfil de infecção típico de dois vértices com valores extremos de acessibilidade.

A explicação para esta relação entre o potencial de infecção de um vértice i e a sua acessibilidade, está no fato desta última representar o número de vértices que efetivamente comunicam-se com i (conforme descrito na seção 2.5). Baixos valores de acessibilidade indicam a existência de uma bias que conduz grande parte das caminhadas para um pequeno subconjunto de Ω_i . Por outro lado, valores elevados de acessibilidade indicam que todos os vértices do conjunto Ω_i são alcançados com probabilidades semelhantes. Uma vez que, o cálculo da acessibilidade leva em conta as quantidades $\beta_{i,j}$, que definem a propagação da infecção, podemos inferir que quando o vértice inicial possui baixa acessibilidade, a superfície de contato da infecção é pequena. Conseqüentemente, a epidemia infecta pou-

Figura 4.14 – (a) Tempo de persistência da infecção como função de k -index (cinza) e da acessibilidade (preto) do vértice inicial. (b) Número de vértices infectados como função de k -index (cinza) e da acessibilidade (preto) do vértice inicial. (c) Perfil de infecção quando o vértice inicial é aquele com maior valor de acessibilidade (tracejado) e menor valor de acessibilidade (contínuo).

cos indivíduos e é rapidamente extinta. Quando a acessibilidade do vértice inicial é alta, a superfície de contato da epidemia é maior, ocasionando longo tempo de persistência e um número maior de vértices infectados. Assim, nossos resultados indicam que qualquer estratégia de imunização deve atuar principalmente sobre os vértices que possuem altos valores de acessibilidade.

4.5 Outros estudos relacionados

Durante o desenvolvimento deste projeto, também estivemos envolvidos com trabalhos que utilizam os conceitos de redes complexas para estudar sistemas que não são necessariamente geográficos, ou que não necessitam do software de mapeamento por nós desenvolvidos. Nesta seção iremos apresentar os principais resultados relacionados a tais sistemas.

4.5.1 Obtenção de diversidade topológica através da otimização da acessibilidade

Muitas vezes, as redes complexas evoluem seguindo mecanismos não-lineares, capazes de tornar os sistemas eficientes frente a determinadas quantidades de interesse. Por exemplo, em (100), Ginestra Bianconi mostrou que as redes livres de escala podem ser obtidas

através da otimização da entropia estrutural, definida como

$$\Phi_S = -\frac{1}{N} \sum_{i=1}^N k_i \log(k_i - 1). \quad (4.10)$$

Nesta seção, iremos mostrar que topologias livre de escala também podem ser obtidas através de um processo de reorganização estrutural, na qual a quantidade a ser otimizada é a acessibilidade. Além disso, mostraremos que a escala ℓ desempenha um papel fundamental na determinação da topologia resultante que emerge do processo de otimização. Quando este processo é realizado sobre a acessibilidade à escalas curtas (pequenos valores de ℓ), obtemos redes com topologia livres de escala. Por outro lado, quando escalas longas são consideradas, as redes resultante possuem a mesma topologia das redes aleatórias descritas pelos modelos de Edgar Gilbert (42), e Paul Erdős e Alfréd Rényi (25, 26).

Em primeiro lugar, consideramos uma rede com $N = 100$ vértices dispostos em um grid periódico de lado 10×10 . Assumimos que todas as arestas desta rede possuem peso idêntico e unitário ($w_{i,j} = 1, \forall(i,j)$). Em seguida, utilizamos um algoritmo de *Simulated Annealing* (SA) para buscar pela topologia que maximiza o valor da acessibilidade, dada pela equação 2.28. O procedimento adotado para a implementação deste algoritmo foi baseado na referência (101). A cada interação completa do algoritmo SA, uma aresta (i,j) é sorteada ao acaso e substituída por (i,x) , onde x é um vértice escolhido aleatoriamente. Esse processo simula o redirecionamento aleatório de arestas na rede. Em seguida, a acessibilidade média da rede é calculada para um dado valor de ℓ . Então, a nova configuração da rede é aceita com probabilidade $p_{SA} = e^{-\Delta\mathcal{A}/T}$, onde $\Delta\mathcal{A}$ é dado por $\Delta\mathcal{A} = \langle\mathcal{A}\rangle_{\text{old}} - \langle\mathcal{A}\rangle_{\text{new}}$ e T é a temperatura. Em seguida um novo redirecionamento é realizado e a acessibilidade é recalculada. Um ciclo completo é finalizado quando 10N novas configurações são aceitas ou quando 100N novas configurações são recusadas. Por fim, a temperatura é reduzida por um fator fixo α , de acordo com $T \rightarrow \alpha T$ ($\alpha < 1$) e um novo ciclo completo recomeça. Quando cinco ciclos completos consecutivos são realizados e nenhuma nova configuração foi aceita, o algoritmo SA é finalizado.

Na figura 4.15 mostramos a topologia das redes obtidas ao final do processo de otimização para diferentes valores de ℓ . Para escalas curtas, $\ell = 3$ (figura 4.15a), podemos observar a presença de hubs, que concentram grande parte das conexões, característica típica das redes livres de escala. Por outros lado, para $\ell = 7$ (figura 4.15c) todos os vértices apresentam aproximadamente o mesmo número de conexões, similar ao que ocorre nas redes aleatórias. Um caso intermediário foi obtido para $\ell = 5$, como mostra a figura 4.15b. Estes resultados qualitativos são confirmados quando olhamos para a distribuição de grau e as comparamos com a de uma rede aleatória, conforme mostram as figuras 4.15d-f. Veja que a medida que aumentamos o valor de ℓ , a distribuição fica cada vez mais próxima da curva preta, correspondente ao resultado esperado para uma rede aleatória com mesmo número de vértices e arestas. Além disso, para $\ell = 3$ vemos que a distribuição de grau

Figura 4.15 – Topologias obtidas a partir de um grid periódico (10×10) segundo a otimização da acessibilidade para (a) $\ell = 3$, (b) $\ell = 5$ e (c) $\ell = 7$. As distribuições de grau destas redes são mostradas em (d), (e) e (f), respectivamente. Em (g), (h) e (i), apresentamos as distribuições de caminhos. As curvas em preto correspondem a distribuição esperada para uma rede aleatória com o mesmo número de vértices e arestas. Os insets nas figuras (g)-(i) correspondem ao valor da acessibilidade como função da temperatura, durante o processo de otimização.

possui o comportamento $P_k \sim k^{-\gamma}$, típico das redes livres de escala. Como mostram as figuras 4.15g-i, resultados similares são obtidos quando investigamos a distribuição de caminhos nas redes otimizadas. Além disso, nota-se que o comprimento típico para $\ell = 3$ é menor do que aquele obtido para $\ell = 7$. Este fato é também um reflexo da emergência da topologia livre de escala para baixos valores de ℓ , um vez que nessas redes, o diâmetro escala com o número de vértices de acordo com $D \sim \log \log N$. Devido a esta relação, dizemos que as redes livres de escala possuem efeito *ultra small-world*, enquanto nas redes aleatórias, a relação entre D e N , dada por $D \sim \log N$, é conhecida como efeito *small-world* (44) (veja seção 2.2.4).

Os resultados descritos acima sugerem que, através da maximização da acessibilidade,

é possível obter redes com diferentes topologias, quando consideramos diferentes valores de ℓ . Vimos que redes livres de escala emergem do processo de otimização quando baixos valores de ℓ são considerados, indicando que nessas redes, a otimização ocorre a nível local. Por outro lado, para longas escalas, a topologia que maximiza a acessibilidade é aleatória. Observe que estes resultados foram obtidos para redes relativamente pequenas e que eles deverão ser confirmados em redes com um número maior de vértices. Entretanto, as simulações necessárias para isso demandam muito tempo de processamento, exigindo um estudo mais aprofundado sobre os métodos de simulação, o que foge ao tema principal desta tese.

4.5.2 A rede de teoremas da Wikipédia

A acessibilidade de um vértice está intimamente relacionada a posição que ele assume na rede. É natural imaginar que vértices mais importantes assumem papel de destaque e possivelmente estejam localizados em regiões centrais, longe das periferias. Para mostrar como a acessibilidade é capaz de capturar tal propriedade, utilizamos a rede de teoremas da Wikipédia e calculamos a acessibilidade para todos os seus vértices. Esta rede foi construída através de um sistema automático que associa um vértice a cada página da Wikipédia relacionada a um teorema matemático. Uma conexão entre dois vértices foi estabelecida se a página de um deles cita, através de hiperlinks, a página referente ao outro.

A rede resultante possui 371 vértices e 502 arestas com pesos idênticos e unitários. A acessibilidade de cada vértice foi estimada para a escala $\ell = 6$, que corresponde ao diâmetro desta rede. A figura 4.16 mostra a rede de teoremas da Wikipédia, onde cada vértice está colorido de acordo com seu valor de acessibilidade. Nesta figura, é possível observar que os teoremas localizados nas regiões periféricas apresentam baixos valores de acessibilidade. Por outro lado, aqueles posicionados no centro da rede, tendem a ter altos valores de acessibilidade. Na tabela 4.1 mostramos os vértices que apresentaram os dez maiores valores de acessibilidade, os quais estão compreendidos pelas áreas de Álgebra, Cálculo, e Topologia.

Intuitivamente, imaginamos que os teoremas mais centrais sejam aqueles demonstrados a mais tempo e cujas influências no desenvolvimento da matemática foram maiores. Para tentar abordar esta questão, levantamos as informações sobre a data de prova de cerca de 60% dos teoremas desta rede. Na figura 4.17, nós apresentamos a correlação entre a acessibilidade dos teoremas e sua data de prova. Note que os teoremas foram agrupados de acordo com o século em que a prova foi realizada. Podemos observar que existe uma correlação expressiva entre essas duas quantidades, indicando uma tendência de que os

Figura 4.16 – Rede de teoremas da Wikipédia obtida em Junho de 2009. Esta rede é formada por $E = 502$ arestas e $N = 371$ vértices, os quais estão coloridos de acordo com o valor de sua acessibilidade para $\ell = 6$. Diamantes e quadrados destacam os 15 vértices com maiores e menores valores de acessibilidade, respectivamente. Imagem adaptada de (102).

Figura 4.17 – Correlação entre a acessibilidade e o século de prova dos teoremas. Imagem adaptada de (102).

teoremas mais antigos tenham altos valores de acessibilidade.

Com este exemplo, foi possível relacionar a medida proposta neste trabalho com um dado funcional da rede de teoremas. A explicação para este fato é de que teoremas

Tabela 4.1 – Dez teoremas da Wikipédia com maiores valores de acessibilidade.

Teorema	A
Intermediate Value theorem	0.558
Banach Fixed Point theorem	0.547
Complex conjugate root theorem	0.546
Fundamental theorem of algebra	0.533
Inverse function theorem	0.531
Polynomial function theorems for zeros	0.528
Squeeze theorem	0.520
Kakutani fixed point theorem	0.519
Extreme value theorem	0.516
Rational root theorem	0.514

provados a mais tempo possuem maiores chances de influenciarem novas descobertas. Estas por sua vez, irão ser adicionadas a rede nas vizinhanças do teorema original. Dessa forma, teoremas mais importantes assumem posições mais centrais, enquanto as novas descobertas acabam sendo naturalmente alocadas nas regiões periféricas, de acordo com um crescimento em camadas.

4.5.3 Definição de borda em redes complexas

Como foi discutido na seção 2.5.1 e como pôde ser visto na aplicação anterior, a acessibilidade tem a interessante tendência de assumir valores mais baixos para os nós localizados em regiões periféricas. A importância do conceito de borda em redes complexas também pode ser observado quando analisamos a diferença entre o comportamento de outras dinâmicas na região de borda e na região central. Mais especificamente, foi investigado (103) a probabilidade de ocupação no estado estacionário ($p_\infty(i)$) da dinâmica de caminhadas aleatórias (veja seção 2.4). A correlação entre esta quantidade e o grau de entrada mostrou-se completamente diferente entre os nós localizados na borda e no centro das redes. Por simplicidade, iremos chamar a quantidade $p_\infty(i)$ de atividade.

Enquanto que, em redes não direcionadas existe uma relação direta entre a atividade e o grau de entrada[†], dada pela equação 4.11, tal relação não é observada em redes direcionadas. Para ilustrar tal fato, usaremos a rede cortical de um macaco do gênero *Macaque*, mostrada na figura 4.18a, na qual cada vértices representa uma região do córtex, e duas regiões são conectadas se a ativação de uma delas induz uma ativação na outra. Na figura 4.18b, mostramos α_i como uma função de ki_i para a versão simetrizada (sem direções) da rede cortical, onde o comportamento linear, dado pela equação 4.11, é observado.

[†] em redes não direcionadas, o grau de entrada é igual ao próprio grau ou seja, $ki_i = k_i$.

Figura 4.18 – (a) Rede cortical de um macaco do gênero *Macaque*. Esta rede direcionada possui $N = 95$ vértices e $E = 2356$ arestas. (b) Relação entre a atividade e o grau de cada vértice na rede cortical simetrizada. Observe que o resultado corresponde exatamente a previsão dada pela equação 4.11. (c) Atividade como função do grau de entrada na rede cortical direcionada. As relações lineares são parcialmente recuperadas quando separamos os vértices de borda dos demais

$$p_{\infty}(i) = \frac{k_i}{2E} \quad (4.11)$$

Um cenário completamente diferente é observado quando as direções das conexões são consideradas. Neste caso, nenhuma relação é observada entre as quantidades de interesse. Como veremos, esta relação pode ser parcialmente recuperada quando separamos as redes em vértices de borda e vértices centrais, e olhamos para os dois conjuntos separadamente. Inicialmente, a borda da rede cortical foi detectada utilizando diferentes valores para o limiar $\mathcal{A}^*$. Para cada um deles, investigamos o valor da correlação $p_{\infty}(i) \times k_i$ para os vértices de borda e para os vértices centrais. Visualmente, escolhemos o valor de $\mathcal{A}^* = 0.91$ e para este limiar, plotamos na figura 4.18c os valores de $p_{\infty}(i)$ como função de k_i para cada conjunto (bordas e centrais). O resultado surpreendente é que a relação linear observada nos sistemas não direcionados é parcialmente recuperada. Além disso, fica claro através desta figura que os vértices localizados na região de borda, ao contrário dos vértices mais centrais, possuem altos valores de ativação. Uma possível explicação para esse fato é de que as caminhadas que definem a atividade dos vértices ficam aprisionadas nos vértices de borda, elevando a atividade dos mesmos.

Esse resultado indica que a região de borda afeta drasticamente as dinâmicas que

ocorrem em redes complexas. Consequentemente, conhecer os vértices que pertencem a esta região é importante em sistemas onde existe transporte de informações, o qual pode ser modelado pela dinâmica de caminhadas aleatórias. Em particular, os sistemas direcionados, onde a atividade dos nós não pode ser prevista pela simples informação estrutural fornecida pelo grau de entrada.

5 Considerações finais

Nesta tese descrevemos a criação e aplicação de procedimentos computacionais no estudo de sistemas biológicos ramificados por meio dos conceitos de redes complexas. Lidar com tais sistemas biológicos em geral requer técnicas de processamento de imagens, uma vez que estes sistemas são armazenados na forma de imagens digitais obtidas por meio de tomografia, micro-tomografia ou microscopia confocal. O procedimento proposto neste trabalho parte do tratamento básico dessas imagens, através de uma fase de pré-processamento. Nesta fase, as imagens passam por um processo de reamostragem, eliminação de ruídos, segmentação e detecção de componentes. Em seguida, as imagens são utilizadas na reconstrução de modelos tridimensionais dos sistemas. Tais modelos são então submetidos a um processo de propagação de rótulos que tem por objetivo detectar o esqueleto das estruturas de interesse. Por fim, estes esqueletos são usados na construção de uma representação em termos de rede para o dado sistema biológico.

Nós também apresentamos o limite de validade da abordagem via propagação de rótulos e propusemos duas técnicas adicionais que melhoram o desempenho de tal algoritmo. A primeira delas consiste na regularização dos dados através do divergente do campo gradiente dos dados. A segunda está relacionada ao procedimento de afinamento que tem por objetivo produzir estruturas com apenas um voxel de espessura, mantendo a topologia do objeto original. Essas duas técnicas aliadas ao processo de propagação de rótulos nos permitiu estudar diversos sistemas biológicos ramificados, cujas as representações através de redes mantêm todas as características topológicas dos sistemas originais. A investigação de tais redes, por sua vez, nos ajudou a quantificar a organização estrutural dos sistemas biológicos.

Durante o desenvolvimento deste projeto, a teoria de redes complexas também foi utilizada para investigar outros sistemas além dos sistemas biológicos ramificados. A seguir daremos maiores detalhes sobre as principais contribuições desta tese.

5.1 Contribuições

5.1.1 Software

As rotinas descritas neste trabalho foram desenvolvidas em linguagem C++ através da interface *Borland C++ Builder*. Em conjunto, utilizamos a biblioteca de processamento e visualização de imagens, VTK (*Visualization Toolkit*) (15), que disponibiliza uma grande quantidade de rotinas prontas. A manipulação destas rotinas é realizada através de uma interface gráfica que será disponibilizada na Internet em breve.

5.1.2 Complexidade geométrica de artérias carótidas

Utilizamos a propagação de rótulos para representar trechos paraselar de artérias carótidas por meio de curvas paramétricas, que foram estudadas por meio de geometria diferencial. Mostramos a existência de correlações significativas entre a energia de curvatura (E_{cur}) dos esqueletos e medidas tradicionais realizadas manualmente. Por outro lado, não verificamos correlações entre essas medidas e a energia de torção (E_{tor}), evidenciando que técnicas comumente empregadas em diagnósticos clínicos não são eficazes para capturar lesões relacionadas a torção das artérias. Em outro estudo, observamos a ausência de correlação entre o número de dobramentos nas artérias e os valores das energias de curvatura e torção. Concluimos então que indivíduos com diferentes valores de E_{cur} e E_{tor} possuem os mesmos riscos de apresentarem arteriosclerose, já que esta desordem está intimamente relacionada a presença dos dobramentos. Esses resultados deram origem aos trabalhos *Objective Characterization of the Course of the Parasellar Internal Carotid Artery Using Mathematical Tools* (16) e *Three-dimensional Description and Mathematical Characterization of the Parasellar Internal Carotid Artery in Human Infants* (17).

5.1.3 Diâmetro ao longo da árvore arterial de embriões de galinha

Esqueleto de sistemas arteriais de embriões de galinha também foram obtidos por meio da propagação de rótulos. Mostramos que nossa abordagem permite a estimação do diâmetro arterial, além da caracterização geométrica das secções transversais em cada ponto da artéria. Nossos resultados são promissores no que diz respeito ao desenvolvi-

mento de um ambiente integrado, capaz de visualizar, seccionar regiões de interesse e extrair medidas com alta precisão.

5.1.4 Estrutura mitocondrial

A inclusão de duas técnicas adicionais (divergente do campo gradiente e afinamento), nos permitiu utilizar o algoritmo de propagação de rótulos em sistemas mais complexos, como as mitocôndrias digitalizadas via microscopia confocal. Mostramos que nossa abordagem é capaz de representar com precisão a topologia mitocondrial e extrair importantes medidas acerca de sua organização. Também estabelecemos o conjunto de parâmetros ideal para análise deste tipo de sistema e os limites de resolução que conseguimos lidar. Tal abordagem será de grande utilidade em problemas nos quais são estudados, por exemplo, a relação entre o volume mitocondrial (V_m) e o volume celular (V_c). Estamos preparando um trabalho intitulado *Mithochondrial Network Size Scaling Budding Yeast is Achieved in the Bud at the Expense of the Mother* que trata de tais problemas.

5.1.5 Topologia e dinâmica das redes de poros do solo

O último sistema estudado através da propagação de rótulos consiste em uma amostra de solo escaneada em um micro-tomografo. Geramos uma rede para representar o sistema de poros contidos nessa amostra e estudamos suas características topológicas e dinâmicas. Observamos a presença de leis-de potência na distribuição de grau, comprimentos e diâmetros das arestas. Outro resultado interessante, foi a presença de correlação super-linear entre o strenght de comprimento e o grau dos vértices desta rede. Os aspectos dinâmicos deste sistema foram investigados sob o ponto de vista do modelo SIR, utilizado para estudar a propagação de doenças em redes. Vimos que a heterogeneidade existente neste sistema induz uma resistência a propagação de uma infecção cerca de 26% maior do que na versão homogênea do mesmo sistema. Por fim, mostramos como podemos utilizar o conceito de acessibilidade para detectar os vértices mais influentes na propagação de tal infecção. Este e outros resultados foram publicados no trabalho *Biological Invasion in Soil: Complex Networks Analysis* (19).

5.1.6 Outros estudos relacionados

Os demais trabalhos que utilizam os conceitos de redes complexas possuem em comum diferentes aplicações da medida de acessibilidade. Ao contrário de medidas previamente conhecidas, mostramos que esta nova medida fornece informações tanto sobre a organização topológica, quanto sobre os aspectos dinâmicos da redes. A primeira vem do fato de que a acessibilidade pode ser entendida como o grau efetivo de cada vértices, e a segunda do fato de que ela é maximizada quando o tempo para acessar todos os vértices do conjunto Ω_i é mínimo. Através de um algoritmo de “simulated annealing”, encontramos as topologias que otimizam a medida de acessibilidade para diferentes escalas (parâmetro ℓ) e vimos que, a topologia livre de escala emerge quando consideramos baixos valores de ℓ . Por outro lado, a topologia que otimiza a acessibilidade para escalas longas, é a aleatória dos modelos de Edgar Gilbert, e Paul Erdős e Alfréd Rényi. Vimos ainda que a acessibilidade pode estar relacionada com dados funcionais das redes, como por exemplo a data de prova dos teoremas na rede de teoremas da Wikipédia. Também definimos o conceito de bordas em redes complexas, o qual é obtido por um processo de limiarização da acessibilidade. Por fim, apresentamos evidências de que, em redes direcionadas, os processos dinâmicos baseados em caminhadas aleatórias comportam-se de maneiras bastante diferentes nas regiões de borda. Abaixo segue uma lista de trabalhos relacionados a esses temas:

- (102) - *Identifying the Borders of Mathematical Knowledge*;
- (Submetido) - *Investigating multidisciplinary in Wikipedia*;
- (20) - *Characterizing topological and dynamical properties of complex networks without border effects*;
- (18) - *Border Detection in Complex Networks*.
- (104) - *On the efficiency of transportation systems in large cities*.

5.2 Desenvolvimentos futuros

Este projeto de doutorado será continuado em um projeto de Pós-doutorado (já aprovado) intitulado *Processos dinâmicos em redes geográficas*. O principal objetivo deste novo projeto é o estudo de processos dinâmicos em redes geográficas extraídas de sistemas reais. Para tanto, iremos utilizar a metodologia de mapeamento apresentada nesta tese. Enquanto a obtenção das redes segue, em linhas gerais, os mesmos passos para diferentes sistemas, a escolha do processo dinâmico que será estudado está intimamente relacionada a natureza do sistema em questão e as perguntas que queremos responder

sobre cada um deles.

Daremos particular atenção as investigações que envolvem questões sobre a relação *estrutura-função* das redes geográficas obtidas, como por exemplo, propagação de microrganismos, robustez frente a ataques e eficiência no transporte. A princípio, estes tópicos serão explorados em dois sistemas diferentes, sendo eles: (i) amostras de solo e (ii) aglomerados neuronais. Apesar da natureza distinta de cada sistema, todos eles são inicialmente representados por imagens digitais e irão passar pela mesma fase de mapeamento para serem descritos através de redes. Em seguida iniciaremos o estudo dos processos dinâmicos em si, com os quais tentaremos responder por exemplo, as seguintes questões: *como a invasão de fungos no solo depende da profundidade e como ela é afetada por fatores externos como a pressão. Qual o efeito da forma dos neurônios na formação de redes de contato e nas dinâmicas que atuam nessas redes.* É interessante observar que estas indagações aparentemente desconexas, estão intimamente ligadas aos conceitos citados anteriormente, como propagação de doenças, robustez e eficiência no transporte. Além disso, o mesmo conjunto de ferramentas pode ser utilizado para abordá-las, como por exemplo *a teoria das caminhadas aleatórias, teoria de percolação e sincronização.*

Referências

- 1 WATTS, D. J. *Small worlds: the dynamics of networks between order and randomness*. Princeton: Princeton University Press, 2003.
- 2 NEWMAN, M. E. J. *Networks: an introduction*. Oxford: Oxford University Press, 2010.
- 3 DOROGOVTSSEV, S. N.; MENDES, J. F. F. Evolution of networks. *Advances in Physics*, v. 51, n. 4, p. 1079, 2002.
- 4 SONG¹, C.; HAVLIN, S.; MAKSE, H. A. Self-similarity of complex networks. *Nature*, v. 433, n. 7024, p. 392, 2004.
- 5 DOROGOVTSSEV, S. N.; GOLTSEV, A. V.; MENDES, J. F. F. Critical phenomena in complex networks. *Reviews of Modern Physics*, v. 80, n. 4, p. 1275, 2008.
- 6 ARENAS, A.; DÍAZ-GUILERA, A.; KURTHS, J.; MORENO, Y.; ZHOU, C. Synchronization in complex networks. *Physics Reports*, v. 469, p. 93, 2008.
- 7 LAEMMER, S.; GEHLSSEN, B.; HELBING, D. Scaling laws in the spatial structure of urban road networks. *Physica A*, v. 363, p. 89, 2006.
- 8 KALAPALA, V.; SANWALANI, V.; CLAUSET, A.; MOORE, C. Scale invariance in road networks. *Physical Review E*, v. 73, n. 2, p. 026130, 2006.
- 9 GASTNER, M. T.; NEWMAN, M. E. J. Shape and efficiency in spatial distribution networks. *Journal of Statistical Mechanics*, v. 2006, p. P01015+, 2006.
- 10 BUHL, J.; GAUTRAIS, J.; SOLÉ, R. V.; KUNTZ, P.; VALVERDE, S.; DENEUBOURG, J. L.; THERAULAZ, G. Efficiency and robustness in ant networks of galleries.

European Physics Journal B, v. 42, n. 1, p. 123, 2004.

11 COSTA, L. F.; VIANA, M. P.; BELETTI, M. E. The complex channel networks of bone structure. *Applied Physics Letters*, v. 88, n. 3, p. 033903, 2006.

12 VIANA, M. P.; TANCK, E.; BELETTI, M. E.; COSTA, L. F. Modularity and robustness of bone networks. *Molecular Biosystems*, v. 5, n. 3, p. 255, 2009.

13 BARTHÉLEMY, M. Spatial networks. *Physics Reports*, v. 499, n. 1, p. 1, 2011.

14 VALVERDE, S.; MURTRA, B. C.; PERNA, A.; KUNTZ, P.; THERAULAZ, G.; SOLÉ, R. V. Percolation in insect nest networks: evidence for optimal wiring. *Physical Review E*, v. 79, n. 6, p. 066106+, 2009.

15 SCHROEDER, W. J.; AVILA, L. S.; HOFFMAN, W. Visualization with vtk: a tutorial. *IEEE Computers Graphics and Applications*, v. 20, n. 5, p. 20, 2000.

16 MENG, S.; GEYER, S. H.; COSTA, L. F.; VIANA, M. P.; WENINGER, W. Objective characterization of the course of the parasellar internal carotid artery using mathematical tools. *Surgical and Radiologic Anatomy*, v. 30, n. 6, p. 519, 2008.

17 MENG, S.; COSTA, L. F.; GEYER, S. H.; VIANA, M. P.; REITER, C.; MULLER, G. B.; WENINGER, W. Three-dimensional description and mathematical characterization of the parasellar internal carotid artery in human infants. *Journal of Anatomy*, v. 212, n. 5, p. 636, 2008.

18 TRAVEÇOLO, B. A. N.; VIANA, M. P.; COSTA, L. da F. Border detection in complex networks. *New Journal of Physics*, v. 11, p. 063019, 2009.

19 PEREZ-RECHE, F.; TARASKIN, S.; NERI, F. M.; GILLIGAN, C. A.; COSTA, L. da F.; VIANA, M. P.; OTTEN, W.; GRINEV, D. Biological invasion in soil: complex network analysis. 2009. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5201098&tag=1>. Acesso em: jan. 2011.

-
- 20 VIANA, M. P.; TRAVENÇOLO, B. A. N.; TANCK, E.; COSTA, L. da F. Characterizing topological and dynamical properties of complex networks without border effects. *Physica A*, v. 389, n. 8, p. 1771, 2010.
- 21 MILGRAM, S. The small world problem. *Psychology Today*, v. 1, n. 1, p. 60, 1967.
- 22 FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the Internet topology. Disponível em: <<http://portal.acm.org/citation.cfm?id=316229>>. Acesso em: fev. 2011.
- 23 ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Diameter of the world wide web. *Nature*, v. 401, n. 6749, p. 130, 1999.
- 24 JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z. N.; BARABÁSI, A.-L. The large-scale organization of metabolic networks. *Nature*, v. 407, n. 6804, p. 651, 2000.
- 25 ERDÖS, P.; RÉNYI, A. On the evolution of random graphs. *A Matematikai Kutató Intézet Közleményei*, v. 5, p. 17, 1960.
- 26 ERDÖS, P.; RÉNYI, A. *Random graphs*. London: Academic Press, 1985.
- 27 WATTS, D. J.; STROGATZ, S. Collective dynamics of "small-world" networks. *Nature*, v. 393, n. 6684, p. 440, 1998.
- 28 BARABÁSI, A.-L.; ALBERT, R.; JEONG, H. Mean-field theory for scale-free random networks. *Physica A*, v. 272, n. 1, p. 173, 1999.
- 29 REDNER, S. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B*, v. 4, n. 2, p. 131, 1998.
- 30 BILKE, S.; PETERSON, C. Topological properties of citation and metabolic networks. *Physical Review E*, v. 64, n. 3, p. 036106, 2001.

-
- 31 LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E.; ÅBERG, Y. The web of human sexual contacts. *Nature*, v. 411, n. 6840, p. 907, 2001.
- 32 JEONG, H.; MASON, S. P.; BARABÁSI, A.-L.; OLTAVAI, Z. N. Lethality and centrality in protein networks. *Nature*, v. 411, n. 6833, p. 41, 2001.
- 33 CAMACHO, J.; GUIMERÀ, R.; AMARAL, L. A. N. Robust patterns in food web structure. *Physical Review Letters*, v. 88, n. 22, p. 228102, 2002.
- 34 BARRAT, A.; BARTHÉLEMY, M.; PASTOR-SATORRAS, R.; VESPIGNANI, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, v. 101, n. 11, p. 3747, 2004.
- 35 BARRAT, A.; BARTHÉLEMY, M.; VESPIGNANI, A. *Dynamical processes on complex networks*. New York: Cambridge University Press, 2008.
- 36 NEWMAN, M. E. J. A measure of betweenness centrality based on random walks. *Social Networks*, v. 27, n. 1, p. 39, 2005.
- 37 ESTRADA, E.; RODRÍGUEZ-VELÁZQUEZ, J. A. Subgraph centrality in complex networks. *Physical Review E*, v. 71, n. 5, p. 056103, 2005.
- 38 FREEMAN, L. C. The gatekeeper, pair-dependency, and structural centrality. *Quality and Quantity*, v. 14, n. 4, p. 585, 1980.
- 39 COSTA; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: a survey of measurements. *Advances in Physics*, v. 56, p. 167, 2007.
- 40 BARTHÉLEMY, M. Betweenness centrality in large complex networks. *The European Physical Journal B*, v. 38, n. 2, p. 163, 2004.
- 41 COHEN, R.; HAVLIN, S. *Complex networks: structure, robustness and function*. New York: Cambridge University Press, 2010.

-
- 42 GILBERT, E. N. Random graphs. *Annals of Mathematical Statistics*, v. 30, n. 4, p. 1141, 1959.
- 43 ERDÖS, P.; RÉNYI, A. On random graphs. I. *Publicationes Mathematicae*, v. 6, p. 290, 1959.
- 44 COHEN, R.; HAVLIN, S. Scale-free networks are ultrasmall. *Physical Review Letters*, v. 90, n. 5, p. 058701, 2003.
- 45 GUIMERÁ, R.; AMARAL, L. A. N. Modeling the world-wide airport network. *The European Physical Journal B*, v. 38, n. 2, p. 381, 2004.
- 46 VILLAS-BOAS, P. R.; RODRIGUES, F. A.; COSTA, L. da F. Modeling worldwide highway networks. *Physics Letters A*, v. 374, n. 1, p. 22, 2009.
- 47 PASTOR-SATORRAS, R.; SMITH, E. D.; SOLÉ, R. V. Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, v. 222, n. 2, p. 199, 2003.
- 48 WAXMAN, B. M. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, v. 6, n. 9, p. 1617, 2002.
- 49 PASTOR-SATORRAS, R.; VÁZQUEZ, V.; VESPIGNANI, A. Dynamical and correlation properties of the internet. *Physical Review Letters*, v. 87, n. 25, p. 258701, 2001.
- 50 BERG, J.; LÄSSIG, M. Correlated random networks. *Physical Review Letters*, v. 89, n. 22, p. 228701, 2002.
- 51 BOGUÑÁ, M.; PASTOR-SATORRAS, R. Class of correlated random networks with hidden variables. *Physical Review E*, v. 68, n. 3, p. 036112, 2003.
- 52 NOH, J. D.; RIEGER, H. Random walks on complex networks. *Physical Review Letters*, v. 92, n. 11, p. 118701, 2004.

-
- 53 YANG, S.-J. Exploring complex networks by walking on them. *Physical Review E*, v. 71, n. 1, p. 016107, 2005.
- 54 COSTA, L. da F.; SPORNS, O.; ANTIQUEIRA, L.; NUNES, M. G. V.; JR, O. N. O. Correlations between structure and random walk dynamics in directed complex networks. *Applied Physics Letters*, v. 91, p. 054107, 2007.
- 55 BALLOBÁS, B. *Random graphs*. London: Academic Press, 1985.
- 56 GADRICH, T.; RAVID, R. The coupon collector problem in statistical quality control. In: INTERNATIONAL CONFERENCE OF APPLIED AND ENGINEERING MATHEMATICS, (ICAEM), 2008, London, U.K. *Proceedings...* London: IAENG, 2008. p. 905.
- 57 BERENBRINK, P.; SAUERWALD, T. The weighted coupon collector's problem and applications. In: INTERNATIONAL CONFERENCE ON COMPUTING AND COMBINATORICS., (COCOON), 15, 2009, Niagara Falls, N.Y. *Proceedings...* Berlin: Springer-Verlag, 2009. p. 449. (Lectures Notes in Computer Science, v.5609)
- 58 MYERS, A. N.; WILF, H. S. Some new aspects of the coupon collector's problem. *SIAM Review*, v. 48, n. 3, p. 549, 2006.
- 59 FELLER, W. *An introduction to probability theory and its applications*. New York: John Wiley & Sons Inc., 1971. v.2
- 60 FLAJOLET, P.; GARDY, D.; THIMONIER, L. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, v. 39, n. 3, p. 207, 1992.
- 61 NEAL, P. The generalized coupon collector problem. *Journal of Applied Probability*, v. 45, n. 3, p. 621, 2008.
- 62 VIANA, M. P.; BATISTA, J. L. B.; COSTA, L. da F. How many nodes are effectively accessed in complex networks? Disponível em: <<http://arxiv.org/abs/1101.5379>>.

Acesso em: fev. 2011.

- 63 PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic spreading in scale-free networks. *Physical Review Letters*, v. 86, n. 14, p. 3200, 2001.
- 64 NEWMAN, M. E. J. The spread of epidemic disease on networks. *Physical Review E*, v. 66, n. 1, p. 016128, 2002.
- 65 PASTOR-SATORRAS, R.; VESPIGNANI, A. Immunization of complex networks. *Physical Review E*, v. 65, n. 3, p. 036104, 2002.
- 66 PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, v. 65, n. 3, p. 035108, 2002.
- 67 NÁ, M. B.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Absence of epidemic threshold in scale-free networks with degree correlations. *Physical Review Letters*, v. 90, n. 2, p. 028701, 2003.
- 68 GRASSBERGER, P. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, v. 63, n. 2, p. 157, 1983.
- 69 HYPERMEDIA image processing reference (hipr2). Disponível em: <<http://homepages.inf.ed.ac.uk/rbf/HIPR2>>, Acesso em: dez. 2006.
- 70 KATIFORI, E.; OSI, G. J. S.; MAGNASCO, M. O. Damage and fluctuations induce loops in optimal transport networks. *Physical Review Letters*, v. 104, n. 4, p. 048704, 2010.
- 71 CORSON, F. Fluctuations and redundancy in optimal transport networks. *Physical Review Letters*, v. 104, n. 4, p. 048703, 2010.
- 72 FABBRI, R.; COSTA, L. D. F.; TORELLI, J. C.; BRUNO, O. M. 2d euclidean distance transform algorithms: a comparative survey. *ACM Computing Surveys*, v. 40, n. 1, p. 1, 2008.

- 73 FU, H.; CHI, Z. Combined thresholding and neural network approach for vein pattern extraction from leaf images. *IEE Proceedings on Vision, Image and Signal Processing*, v. 153, n. 6, p. 881, 2006.
- 74 FUJIYOSHI, H.; LIPTON, A. J. Real-time human motion analysis by image skeletonization. In: IEEE WORKSHOP ON APPLICATIONS OF COMPUTER VISION, 4, WACV, 1998, Princeton, N.J. *Proceedings...* Los Alamitos: IEEE Computer Society, 1998. p. 15.
- 75 NAF, M.; KUBLER, O.; KIKINIS, R.; SHENTON, M. E.; SZEKELY, G. G. Characterization and recognition of 3D organ shape in medical image analysis using skeletonization. In: IEEE WORKSHOP ON MATHEMATICAL METHODS IN BIO-MEDICAL IMAGE ANALYSIS,(MMBIA), 1996, San Francisco, CA. *Proceedings...* Los Alamitos: IEEE Computer Society, 1996. p. 139.
- 76 FOUARD, C.; MALANDAIN, G.; PROHASKA, S.; WESTERHOFF, M. Blockwise processing applied to brain microvascular network study. *IEEE Transactions on Medical Imaging*, v. 25, n. 10, p. 1319, 2006.
- 77 BLUM, H. A transformation for extracting new descriptors of shape. In: DUNN, W. W. (Ed.). *Models for the perception of speech and visual form*. Cambridge: MIT Press, 1967. p. 362.
- 78 KONG, T. Y.; ROSENFELD, A. Digital topology: introduction and survey. *Computer Vision, Graphics, and Image Processing*, v. 48, n. 3, p. 357, 1989.
- 79 MA, C. M.; SONKA, M. A fully parallel 3d thinning algorithm and its applications. *Computer Vision and Image Understanding*, v. 64, n. 3, p. 420, 1996.
- 80 LAM, L.; LEE, S. W.; SUEN, C. Y. Thinning methodologies - a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Learning*, v. 14, n. 9, p. 869–885, 1992.

-
- 81 ZHOU, Y.; TOGA, A. W. Efficient skeletonization of volumetric objects. *IEEE Transactions on Visualization and Computer Graphics*, v. 5, n. 3, p. 196, 1999.
- 82 KRUSZYNSKI, K. J.; LIERE, R. V. Quantifying differences in skeletonization algorithms: a case study. In: INTERNATIONAL CONFERENCE ON VISUALIZATION, IMAGING, AND IMAGE PROCESSING,(IASTED), 2005, Benidorm, Spain. *Proceedings...* Calgary: ACTA PRESS, 2005. p. 666.
- 83 PALÁGYI, K.; KUBA, A. A 3d 6-subiteration thinning algorithm for extracting medial lines. *Pattern Recognition Letters*, v. 19, n. 7, p. 613, 1998.
- 84 SAKATA, N.; JOSHITA, T.; OONEDA, G. Topographical study on arteriosclerotic lesions at the bifurcations of human cerebral arteries. *Heart Vessels*, v. 1, n. 3, p. 70, 1985.
- 85 CAPLAN, L. R.; GORELICK, P. B.; HIER, D. B. Race, sex and occlusive cerebrovascular disease: a review. *Stroke*, v. 17, p. 648, 1986.
- 86 WECHSLER, L. R.; JP, J. P. K.; DAVIS, K. R.; KAMINSKI, M. J. The prognosis of carotid siphon stenosis. *Stroke*, v. 17, n. 4, p. 714–718, 1986.
- 87 RAJU, S.; FREDERICKS, R. K. Carotid siphon stenosis. *The Journal of Cardiovascular Surgery*, v. 28, n. 6, p. 671, 1987.
- 88 WENINGER, W. J.; MÜLLER, G. B.; REITER, C.; S, S. M.; RABL, S. U. Intimal hyperplasia of the infant parasellar carotid artery: a potential developmental factor in atherosclerosis and sids. *Circulation Research*, v. 85, n. 10, p. 970, 1999.
- 89 WENINGER, W. J.; MENG, S.; STREICHER, J.; MULLER, G. B. A new episcopic method for rapid 3-d reconstruction: applications in anatomy and embryology. *Anatomy and Embryology*, v. 197, n. 5, p. 341, 1998.
- 90 CARMO, M. do. *Differential geometry of curves and surfaces*. Englewood Cliffs, NJ: Prentice-Hall, 1976.

- 91 CESAR, R. M.; COSTA, L. da F. Application and assessment of multiscale bending energy for morphometric characterization of neural cells. *Review of Scientific Instruments*, v. 68, n. 5, p. 2177, 1997.
- 92 COSTA, L. da F.; CESAR, R. M. *Shape analysis and classification: theory and practice*. Boca Raton, FL: CRC Press, 2001.
- 93 FISCHER, M. J. M.; UCHIDA, S.; MESSLINGER, K. Measurement of meningeal blood vessel diameter in vivo with a plug-in for ImageJ. *Microvascular Research*, v. 80, n. 2, p. 258, 2010.
- 94 LEE, J.; JIRAPATNAKUL, A. C.; REEVES, A. P.; CROWE, W. E.; SARELIUS, I. H. Vessel diameter measurement from intravital microscopy. *Annals of Biomedical Engineering*, v. 37, n. 5, p. 913, 2009.
- 95 EGNER, A.; JAKOBS, S.; HELL, S. W. Fast 100-nm resolution three-dimensional microscope reveals structural plasticity of mitochondria in live yeast. *Proceedings of the National Academy of Sciences of the United States of America*, v. 99, n. 6, p. 3370, 2002.
- 96 OTTEN, W.; HALL, D.; HARRIS, K.; RITZ, K.; YOUNG, I. M.; GILLIGAN, C. A. Soil physics, fungal epidemiology and the spread of rhizoctonia solani. *New Phytologist*, v. 151, n. 2, p. 459, 2001.
- 97 HARRIS, K.; YOUNG, I. M.; GILLIGAN, C. A.; OTTEN, W.; RITZ, K. Effect of bulk density on the spatial organisation of the fungus rhizoctonia solani in soil. *FEMS Microbiology Ecology*, v. 44, n. 1, p. 45, 2003.
- 98 OTTEN, W.; GILLIGAN, C. A. Soil structure and soil-borne diseases: using epidemiological concepts to scale from fungal spread to plant epidemics. *European Journal of Soil Science*, v. 42, n. 1, p. 131, 2006.
- 99 KITSACK, M.; GALLOS, L. K.; LILJEROS, S. H. F.; MUCHNIK, L.; STANLEY, H.; MAKSE, H. A. Identifying influential spreaders in complex networks. *Nature Physics*, v. 6, n. 11, p. 888, 2010.

-
- 100 BIANCONI, G. Entropy of network ensembles. *Physical Review E*, v. 79, n. 3, p. 036114, 2009.
- 101 PENNA, T. J. P. Traveling salesman problem and tsallis statistics. *Physical Review E*, v. 51, n. 1, p. R1, 1995.
- 102 SILVA, F. N.; TRAVENÇOLO, B. A. N.; VIANA, M. P.; COSTA, L. da F. Identifying the borders of mathematical knowledge. *Journal of Physics A: Mathematical and Theoretical*, v. 43, n. 32, p. 325202, 2010.
- 103 ANTIQUEIRA, L.; COSTA, L. da F. Structure-dynamics interplay in directed complex networks with border effects. In: INTERNATIONAL WORKSHOP ON COMPLEX NETWORKS, CompleNet, 2010, Rio de Janeiro, BR. *Proceedings...* Rio de Janeiro: COPPE/UFRJ, 2010.
- 104 COSTA, L. da F.; TRAVENÇOLO, B. A. N.; VIANA, M. P.; STRANO, E. On the efficiency of transportation systems in large cities. *Europhysics Letters*, v. 91, n. 1, p. 18003, 2010.

Apêndice 1: Decomposição k -shell

Dentre as diversas medidas utilizadas para caracterizar os vértices de uma rede, o k -index tem recebido particular atenção. Isso porque um trabalho recente (99) relaciona esta quantidade a influência de um dado vértice nos processos de propagação em redes. O valores de k -index são obtidos por um procedimento de decomposição em camadas, na qual os vértices de uma rede são removidos de acordo com o seu número de vizinhos. Para exemplificar este procedimento, vamos olhar para a figura 5.1. Inicialmente, removemos todos os vértices de grau 1. A estes vértices são atribuídos valores de k -index 1. Em seguida, verificamos se surgiram novos vértices de grau um. Em caso afirmativo, estes também são removidos e recebem k -index unitário. O procedimento é repetido até que nenhum vértice com apenas um vizinho seja encontrado. Passamos então para os vértices de grau 2, os quais são removidos e recebem k -index 2. Agora verificamos se surgiram vértices com grau menor ou igual a 2. Em caso afirmativo, todos eles são removidos e recebem k -index 2. Quando não houver mais vértices com grau 2, passamos para os de grau 3 e assim por diante. No fim, todos os vértices receberam um valor de k -index. Aqueles com valores mais alto definem um núcleo, chamado de k -core. Em (99) os autores sugerem que os vértices presentes no k -core de uma rede são os mais influentes nos processos de propagação estudados em redes complexas.

Figura 5.1 – (a) Exemplo de rede na qual foi aplicado o procedimento de decomposição para o cálculo dos valores de k -index dos vértices. As etapas do procedimento são ilustradas em (b), (c), (d) e (e), nas quais são removidos os vértices com grau menores ou igual a 1, 1, 2, 2 e 3. Em (f) mostramos os vértices coloridos de acordo com seus valores de k -index.