UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

MAURO MIAZAKI

Estudo da forma, função e
expressão gênica em neurociência

São Carlos
2012

MAURO MIAZAKI

Estudo da forma, função e
expressão gênica em neurociência

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

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

*Orientador: Prof. Dr. Luciano da Fontoura Costa
Co-Orientador: Prof. Dr. Sergei N. Taraskin*

Versão corrigida
(Versão original disponível na unidade que aloja o programa)

São Carlos
2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTA
TRABALHO, 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 do IFSC,
com os dados fornecidos pelo autor

Miazaki, Mauro

Estudo da forma, função e expressão gênica em neurociência / Mauro Miazaki; orientador Luciano da Fontoura Costa; co-orientador Sergei N. Taraskin - versão corrigida - São Carlos, 2012.

129 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Aplicada Computacional) - Instituto de Física de São Carlos, Universidade de São Paulo, 2012.

1. Modelagem computacional de sistemas neuronais.
 2. Processamento de imagens.
 3. Neuromorfometria.
 4. Expressão gênica.
 5. Bioinformática.
- I. Costa, Luciano da Fontoura, orient. II. Taraskin, Sergei N., co-orient. III. Título.

AGRADECIMENTOS

Ao Prof. Dr. Luciano da Fontoura Costa, pela orientação, ajuda, dedicação e incentivo, fundamentais para o desenvolvimento deste trabalho.

Ao meu co-orientador Prof. Dr. Sergei N. Taraskin, pela oportunidade em conhecer, trabalhar e aprender com seu grupo de pesquisa na Universidade de Cambridge.

Aos amigos do IFSC que me ajudaram neste trabalho: Matheus P. Viana, Bruno A. N. Travençolo, Alexandre S. Cristino e Krissia de Zawadzki.

Ao Gerson Ferreira Jr., pelo modelo LaTeX utilizado para escrever esta tese.

Ao auxílio na revisão do texto: Neusa (funcionária da biblioteca), Mônica e Matheus.

Ao auxílio dos funcionários do IFSC: Marquinhos, Sonia, Silvio, Victor, Ricardo, Giuliana, Ítalo e Thais Fernanda.

Ao Bruno, pelo cafezinho diário nos meus primeiros anos de doutorado.

Aos amigos do Grupo de Pesquisa em Visão Cibernética: André, Bruno, Carlos, César, Débora, Diego, Filipi, Francisco, Gustavo, João, Krissia, Lilian, Lucas, Luis, Matheus, Mônica, Osvaldo, Paulino, Renato F., Renato P. e Vilson.

Aos amigos da república: Daniel, Fernando, Kleber, Luis, Rafael, Ronaldo, Sidão, Takashi e Yuji.

Aos meus pais, Hissae e Yuzo.

Ao CNPq, pelo apoio financeiro no primeiro mês do doutorado.

À Fapesp, pelo suporte financeiro (processo n° 2007/50988-1).

RESUMO

Miazaki, M. **Estudo da forma, função e expressão gênica em neurociência**. 2012. 129p. Tese (Doutorado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

Durante o desenvolvimento de um neurônio, genes são ativados e desativados, a anatomia se forma e as funcionalidades emergem. Estes três componentes influenciam continuamente uns aos outros. O estudo da forma, função e expressão gênica nos neurônios e no cérebro permanece um tema desafiador e com potencial a ser explorado. Neste contexto, uma importante questão ainda a ser respondida é como quantificar o inter-relacionamento entre forma, função e genes. Para isso, foram realizadas atividades envolvendo caracterização e comparação da forma neuronal, o estudo de processos dinâmicos ocorrendo em redes de estruturas ramificadas, e a comparação entre expressões gênicas. Os dados da base pública NeuroMorpho, que possui quase 6.000 neurônios segmentados, foram caracterizados utilizando-se métodos estatísticos e foram analisados pelo conceito de morfoespaço proposto por McGhee. Outra base pública explorada foi o Mouse Allen Brain Atlas, com imagens de expressão gênica de cérebros de camundongo. Foi proposta a utilização de um método baseado em diagramas de Voronoi para a comparação da distribuição espacial de densidades de expressão gênica entre genes, com o propósito de encontrar correlações entre distribuições. Também foram gerados dados sobre raízes de feijão para o estudo da influência de sua estrutura ramificada na dinâmica de propagação de doenças, seguindo o modelo SIR (Suscetível-Infectado-Recuperado). Integrando os desenvolvimentos anteriores, foi proposto um arcabouço para mensurar a influência da expressão gênica ao longo da escala biológica. Este arcabouço permite mensurar a influência da expressão gênica (escala molecular) na morfologia dos neurônios (escala celular), avançando à escala topológica formada pelas conexões sinápticas, e alcançando o nível funcional das dinâmicas sobre essa rede. Nesse contexto, deve-se ressaltar que a influência da expressão gênica é direta sobre a morfologia e indireta sobre a topologia e a dinâmica. As informações obtidas a partir do arcabouço são relevantes na investigação de como a expressão gênica influencia todo o processo, desde o neurônio individual até o funcionamento cerebral. O arcabouço proposto fornece uma metodologia sistemática, com um conjunto de ferramentas para essas análises.

Palavras-chave: *Modelagem computacional de sistemas neuronais. Processamento de imagens. Neuromorfometria. Expressão gênica. Bioinformática.*

ABSTRACT

Miazaki, M. **Study of form, function and gene expression in neuroscience.** 2012. 129p. Tese (Doutorado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

During the development of a neuron, genes are turned on and off, the anatomy is shaped and the functionality emerges. These three components influence each other continuously. The study of form, function and gene expression in neurons and brain is still challenging and has many issues yet to be explored. In this context, an important question yet to be answered is how to quantify the inter-relationship between form, function and gene expression. In this way, we developed activities involving characterization and comparison of the neuronal form, the study of dynamical processes occurring in networks of branching structures, and the comparison between gene expressions. The data of the public database NeuroMorpho, which comprise almost 6,000 segmented neurons, were characterized using statistical methods and were analyzed by the concept of McGhee's morphospace. Another public database that was explored was the Mouse Allen Brain Atlas, with images of gene expression of mouse brains. We proposed to use a method based on Voronoi diagrams to compare the spatial distribution of the gene expression densities between genes, in order to find correlations in the distribution. We also generated data on bean roots to study the influence of their branched structures in the dynamics of disease spread, following the SIR model (Susceptible-Infected-Recovered). Integrating the previous developments, we proposed a framework to measure the gene expression influence through the biological scale. This framework allows the measurement of the gene expression (molecular scale) influence in the morphology of the neurons (cellular scale), advancing towards the topological scale formed by the synaptic connections, and reaching the functional level of the dynamics over this network. In this context, it is worth to note that the gene expression influence is direct on the morphology and indirect on the topology and dynamics. The obtained information through the framework is important on the investigation of how the gene expression influences the whole process, since the individual neuron to the cerebral functioning. The proposed framework yields a systematic methodology with a toolbox to carry out these analyses.

Keywords: *Computational modeling of neuronal systems. Image processing. Neuro-morphometry. Gene expression. Bioinformatics.*

LISTA DE FIGURAS

Figura 1.1 - Desenho de células encontradas no cerebelo de galinha	24
Figura 1.2 - Padrões de expressão gênica dos genes <i>bicoid</i> , <i>caudal</i> e <i>even-skipped</i> no embrião da <i>Drosophila melanogaster</i>	25
Figura 1.3 - Ilustração dos objetivos e indicação dos capítulos relacionados a cada um.	28
Figura 2.1 - Evolução do número de neurônios ao longo das versões na base de dados NeuroMorpho	32
Figura 2.2 - Medidas disponíveis na base de dados NeuroMorpho. São mostrados os nomes originais das medidas em inglês. A descrição dessas medidas pode ser encontrada na Tabela 2.1.	33
Figura 2.3 - Atlas de referência das fatias cerebrais selecionadas	36
Figura 2.4 - Registro (alinhamento) da imagem de expressão gênica com o atlas anatômico	36
Figura 3.1 - Representação genérica das várias regiões possíveis em um hiperespaço morfológico.	41
Figura 3.2 - Função de densidade radial	41
Figura 3.3 - Exemplo de um diagrama de Voronoi	42
Figura 3.4 - Passos da metodologia VLDA	45
Figura 3.5 - Entropia condicional	49
Figura 4.1 - Distribuição das medidas morfológicas dos neurônios reais e artificiais	53

Figura 4.2 - PCA mostrando a distribuição dos neurônios no morfoespaço, com a indicação de alguns exemplos de neurônios reais	55
Figura 4.3 - PCA com os neurônios reais e artificiais como na Figura 4.2, mas agora indicando exemplos de neurônios artificiais	55
Figura 4.4 - Coeficiente de correlação de Pearson entre pares de medidas. . . .	56
Figura 4.5 - Visualizações das projeções PCA, com os neurônios destacados por tipo de célula, regiões cerebrais e espécies	57
Figura 4.6 - Percentual cumulativo de variabilidade explicada dos dados a cada componente principal	58
Figura 4.7 - Visualização das categorias utilizando a análise de variáveis canônicas, com os neurônios destacados por tipo de célula regiões cerebrais e espécies	60
Figura 4.8 - Função de densidade radial para as células Purkinje, Stellate, Martinotti e Lateral Horn	62
Figura 5.1 - Redes individualmente normalizadas para melhor visualizar cada região	66
Figura 5.2 - PCA dos dados normalizados por região	67
Figura 5.3 - Redes normalizadas considerando todos os dados, permitindo comparar o padrão das conexões entre as regiões cerebrais	68
Figura 5.4 - PCA das redes normalizadas considerando todos os dados	69
Figura 6.1 - Influência da expressão gênica na morfologia, topologia e dinâmica	74
Figura 6.2 - Influência da concentração de expressão gênica no crescimento neuronal em uma grade de padrão listrado	76
Figura 6.3 - Neurônios artificiais que cresceram numa grade de padrão circular de distribuição de valores de intensidade de expressão gênica . . .	77
Figura 6.4 - Redes formadas considerando o padrão de grade listrado para $\Delta = 5$ e $\Delta = 10$	78
Figura 6.5 - Gráficos de dispersão das medidas morfológicas	80

Figura 6.6 - Entropia condicional das medidas morfológicas	80
Figura 6.7 - Medida de inclinação dos limites inferior e superior dos gráficos de dispersão das medidas morfológicas	81
Figura 6.8 - Correlação de Pearson das medidas topológicas	84
Figura 6.9 - Gráficos de dispersão das medidas topológicas com os maiores valores de correlação de Pearson	84
Figura 6.10 - Correlação de Pearson das três medidas de dinâmica em rede . . .	86
Figura 6.11 - Gráficos de dispersão das medidas de dinâmica com os maiores valores de correlação de Pearson	87
Figura 6.12 - Imagens de expressão gênica utilizadas	87
Figura 6.13 - Entropia condicional das medidas morfológicas considerando as duas expressões gênicas	88
Figura 6.14 - Correlação de Pearson das medidas morfológicas considerando as expressões gênicas G1 e G2	88
Figura 6.15 - Análise de inclinação das medidas morfológicas para G1 e G2, nos limites superior e inferior das distribuições	89
Figura 6.16 - Gráficos de dispersão das medidas topológicas, com destaque aos pontos que compõem os limites superior e inferior das distribuições	90
Figura A.1 - Micrótoomo e bloco de parafina contendo uma raiz	111
Figura A.2 - Exemplos de imagens das fatias de uma raiz de feijão, suas respectivas segmentações e reconstrução tridimensional	112
Figura A.3 - Projeções horizontais de raízes de nove feijões, dispostas sobre uma rede regular quadrada	115
Figura A.4 - Dependência da média da probabilidade de invasão na eficiência de transmissão e no espaçamento entre os nós da rede	118
Figura A.5 - Probabilidade de invasão versus transmissibilidade média para sistemas realísticos de raízes complexas	120
Figura A.6 - Modelos baseados em discos	121

Figura A.7 - Curvas de invasão para o sistema homogêneo, sistema de raízes reais, modelo de disco sólido, modelo de disco disperso e discos com bordas irregulares	123
Figura A.8 - Dependência da probabilidade de invasão na transmissibilidade para duas raízes, dispostas numa rede regular quadrada	126
Figura A.9 - Dependência entre o limiar de invasão e o ângulo de abertura da distribuição de rotações aleatórias das raízes	126
Figura A.10 - As funções de correlação para as mesmas raízes analisadas na Figura A.8	128

LISTA DE TABELAS

Tabela 2.1 - Descrição das medidas na base de dados NeuroMorpho	34
Tabela 4.1 - Pesos do PCA das projeções nas Figuras 4.2 e 4.3	54
Tabela 4.2 - Pesos do PCA relacionados aos neurônios reais na base de dados NeuroMorpho	59
Tabela 5.1 - Funções conhecidas dos genes relacionados às expressões gênicas analisadas	70
Tabela 6.1 - Medidas morfológicas e as respectivas funções de influência da ex- pressão gênica	79

LISTA DE ABREVIATURAS

2D	Bidimensional
20D	20-Dimensional
3D	Tridimensional
AMBA	Allen Mouse Brain Atlas
C1	Primeiro Componente
C2	Segundo Componente
CP	<i>Caudoputamen</i>
CP1	Primeiro Componente Principal
CP2	Segundo Componente Principal
CTXsp	<i>Cortical Subplate</i>
DC#	Medidas de Dinâmica - Padrão Circular - Distância entre nós #
DL#	Medidas de Dinâmica - Padrão Listrado - Distância entre nós #
DNA	<i>Desoxyribonucleic Acid</i> (Ácido Desoxirribonucléico)
G#	Expressão Gênica #
GM#	Grupo de Medidas #
ISH	<i>In Situ</i> Hybridization
M#	Medida #
MC	Medidas Morfológicas - Padrão Circular
ML	Medidas Morfológicas - Padrão Listrado

PAL	<i>Pallidum</i>
PCA	<i>Principal Component Analysis</i> (Análise de Componentes Principais)
RHP	<i>Retrohippocampal Region</i>
RNA	<i>Ribonucleic Acid</i> (Ácido Ribonucléico)
sAMY	<i>striatum-like Amygdalar Nuclei</i>
SIR	Suscetível-Infectado-Recuperado
TC#	Medidas Topológicas - Padrão Circular - Distância entre nós #
TL#	Medidas Topológicas - Padrão Listrado - Distância entre nós #
URL	<i>Uniform Resource Locator</i> (Localizador-Padrão de Recursos)
VLDA	<i>Voronoi Local Density Analysis</i> (Análise de Densidade Local por Voronoi)
VTK	Visualization Toolkit

SUMÁRIO

1	Introdução	23
1.1	Objetivos	27
1.2	Conteúdo e organização da tese	27
2	Materiais	31
2.1	NeuroMorpho	31
2.2	Neurônios artificiais	32
2.3	Allen Mouse Brain Atlas	35
3	Métodos	37
3.1	Análise de Componentes Principais	37
3.2	Análise Canônica	38
3.3	Modelagem de hiperespaço morfológico	39
3.4	Função de densidade radial	40
3.5	Diagrama de Voronoi	42
3.6	Análise de Densidade Local por Voronoi	43
3.7	Comparação entre redes utilizando Análise de Componentes Principais	46
3.8	Dinâmica integra-e-dispara	47
3.9	Entropia condicional	48

3.10	Análise de inclinação das fronteiras nos gráficos de dispersão	49
4	Hiperespaço neuromorfológico	51
4.1	Configuração do modelo de geração de neurônios artificiais	51
4.2	Resultados e discussão	52
4.2.1	Modelagem do neuromorfoespaço	52
4.2.2	Inter-relação das medidas e distribuição espacial dos neurônios . . .	54
4.2.3	Distribuição de categorias	59
4.2.4	Análise de densidade no hiperespaço	61
4.3	Considerações finais	62
5	Distribuição espacial das densidades de expressão gênica no cérebro	65
5.1	Resultados e discussão	65
5.2	Considerações finais	70
6	Influência da expressão gênica na morfologia, topologia e dinâmica	73
6.1	Crescimento neuronal sob influência da expressão gênica	73
6.2	Construção da rede neuronal	75
6.3	Resultados e discussão	78
6.3.1	Influência da expressão gênica na morfologia	78
6.3.2	Influência da expressão gênica na topologia	82
6.3.3	Influência da expressão gênica na dinâmica	84
6.3.4	Quantificação da influência da expressão de dois genes	85
6.4	Considerações finais	90
7	Conclusões	91

7.1	Hiperespaço neuromorfológico	91
7.2	Distribuição espacial das densidades de expressão gênica no cérebro	92
7.3	Propagação de doenças numa população de estruturas ramificadas	93
7.4	Influência da expressão gênica na morfologia, topologia e dinâmica	94
7.5	Principais contribuições	95
7.6	Trabalhos futuros	97

REFERÊNCIAS 99

Apêndice A Propagação de doenças numa população de estruturas ramificadas 109

A.1	Raízes	111
A.2	Modelo SIR de propagação de doenças	112
A.3	Rede de raízes	114
A.4	Resultados e discussão	116
A.4.1	Diagrama de fases	117
A.4.2	Correlações	119
A.4.3	Modelos baseados em discos	120
A.4.4	Anisotropia global	124
A.5	Considerações finais	127

1 Introdução

A neurociência moderna teve suas origens no final do século XIX, principalmente com os trabalhos de Santiago Ramón y Cajal (1, 2). Nessa época, Camillo Golgi desenvolveu um revolucionário método de coloração que pela primeira vez permitiu visualizar em detalhes a morfologia intrincada dos neurônios. Cajal utilizou e aprimorou essa metodologia para realizar seus estudos, nos quais demonstrou que o sistema nervoso central é composto por bilhões de células polarizadas (neurônios). Também mostrou que os neurônios não formam uma rede contínua. Ao contrário do que se acreditava, as células neuronais se comunicam umas com as outras através de junções especializadas (sinapses). Essas descobertas fundamentaram a elaboração da doutrina neuronal, que considera o neurônio como a unidade básica estrutural e funcional do sistema nervoso (3, 4). Atualmente, o escopo da neurociência é a investigação, experimental e teórica, do sistema nervoso central e periférico de organismos biológicos, com o intuito de desvendar os mecanismos responsáveis pela atividade nervosa (5–9). Procura-se, assim, entender como a atividade neuronal funciona de modo a permitir que organismos percebam e interajam com o mundo (10).

Apesar dos avanços tecnológicos e contínuos esforços em desvendar a complexidade do cérebro, a compreensão do sistema nervoso ainda é muito incipiente e um tema desafiador de pesquisa. Desde os trabalhos pioneiros de Cajal, a grande diversidade de morfologias neuronais (Figura 1.1) vem atraindo a atenção de pesquisadores (11–15). A multiplicidade de morfologias varia desde formas relativamente simples, tais como células bipolares, para estruturas complexas, como as células de Purkinje (16, 17). A dinâmica emergente em sistemas neuronais é, em última análise, consequência do estabelecimento de conexões sinápticas, que são em grande parte definidas pelo padrão de ramificação neuronal (18, 19), a posição relativa das células neuronais e do histórico de respostas a estímulos dinâmicos (20–22). Por exemplo, células que são muito simples e separadas umas das outras tendem a fazer um menor número de sinapses. Portanto, uma adequada compreensão dos padrões de conectividade do sistema nervoso exige a análise da morfologia neuronal (19, 23). Além disso, a dinâmica de funcionamento dos neurônios também é intrinsecamente limitada e até mesmo definida por suas respectivas formas (24–30). A razão desta diversidade é que, embora todas as células em um organismo compartilhem o mesmo DNA, cada tipo de

célula expressa apenas um subconjunto específico de genes (31). Portanto, o padrão de expressão desempenha um papel fundamental na diferenciação celular. O conjunto de genes expressos em uma célula (o padrão de expressão) é influenciado pelo ambiente, por sinais internos e sinais de outras células (21).

Figura 1.1 – Desenho de células encontradas no cerebelo de galinha. O cerebelo é uma região cerebral responsável por coordenar os músculos e manter o equilíbrio do corpo. Este desenho de Santiago Ramón y Cajal ilustra a diversidade das morfologias neuronais. Fonte: Wikipedia.

Os genes são os responsáveis por carregar a informação necessária para a produção das diversas seqüências de RNA e tipos de proteínas, fundamentais para o funcionamento celular e manutenção da vida. A produção de proteínas ocorre de forma indireta (32). Inicialmente, um trecho apropriado do DNA em um cromossomo é copiado sob a forma de RNA (processo de transcrição). Em seguida, o RNA é utilizado para a síntese da proteína (processo de tradução). Este processo de DNA $\rightarrow$ RNA $\rightarrow$ proteína ocorre em todos os seres vivos e, devido sua importância e universalidade, é denominado como o dogma central da biologia molecular (21). A expressão gênica é controlada por uma região reguladora, localizada no DNA, relativamente próxima de onde se inicia a transcrição (31). As proteínas de regulação gênica, se presentes, identificam e conectam-se a essas regiões, que podem ligar ou desligar certos genes. Além disso, as células também podem modificar a expressão de seus genes em resposta às mudanças em seu ambiente. Nos organismos multicelulares, as células desenvolveram mecanismos para utilizar essa capacidade de controle da expressão gênica de forma especializada, possibilitando a formação de tipos celulares diferenciados (21). Essas células especializadas, por sua vez, formam os diferentes órgãos e tecidos. Essa intrincada interação e regulação entre os genes, envolvendo RNAs e proteínas, é conhecida como Rede de Regulação Gênica (*Gene Regulatory*

Network) (33).

A importância dos padrões de distribuição da expressão gênica na formação dos organismos pode ser observada num caso já bastante estudado e conhecido: o desenvolvimento no eixo ântero-posterior (da cabeça à cauda) do embrião da *Drosophila melanogaster* (popularmente conhecida como mosca-da-fruta). Algumas horas após a fertilização, é possível observar o padrão de distribuição das proteínas Bicoid, Even-skipped e Caudal, expressões gênicas sintetizadas respectivamente pelos genes *bicoid*, *caudal* e *even-skipped*. A presença dessas proteínas indica que seus respectivos genes estão ativos. A proteína Bicoid concentra-se mais na região anterior, enquanto que a proteína Caudal apresenta-se mais ativa no lado oposto (região posterior). Já a proteína Even-skipped forma um padrão listrado de sete faixas perpendiculares ao eixo ântero-posterior. Essas listras são as responsáveis por regular diversos outros genes que controlam a formação de divisões em segmentos do embrião (34). Os padrões de distribuição da expressão gênica desses três genes são ilustrados na Figura 1.2.

Em relação à expressão gênica no cérebro, a comunidade científica tem gradualmente identificado as famílias de genes envolvidas nos processos complexos e dinâmicos de desenvolvimento e regionalização celular, ou seja, processos nos quais células precursoras dividem-se e originam neurônios que migram, se diferenciam e criam conexões sinápticas (35–37). Por exemplo, os estudos pioneiros na *Drosophila melanogaster* permitiram identificar que as famílias de genes homeobox *ems* e *otd* possuem papéis importantes no desenvolvimento cerebral. A partir desses estudos, os genes homólogos no camundongo também foram identificados: *Otx1*, *Otx2*, *Emx1* e *Emx2* (38, 39).

Figura 1.2 – Padrões de expressão gênica dos genes *bicoid*, *caudal* e *even-skipped* no embrião da *Drosophila melanogaster* (a). As expressões desses genes são apresentadas isoladamente em (b-d). As letras A e P indicam as extremidades anterior e posterior do embrião. Fonte: tese de doutorado de Bruno A. N. Travençolo (40) (reprodução autorizada pelo autor).

Entre as células designadas para a formação do sistema nervoso central, também há diferenciação entre suas características, de acordo com o momento e o local onde nasceram. Inicialmente, as células nervosas são geradas a partir de divisões celulares. Após essa fase, ocorre o crescimento dos seus axônios e dendritos para formar as sinapses. Em seguida, a rede formada é refinada e remodelada de acordo com o padrão de atividade elétrica na rede neuronal (21). Todo esse processo é ainda guiado pela expressão gênica, que é modulada por influências ambientais ao longo do tempo e pela distribuição espacial dos genes.

Com os recentes e rápidos progressos em tecnologias para armazenamento, gestão e transmissão de dados, atualmente tornou-se viável e cada vez mais comum o surgimento de enormes bancos de dados na web. Além disso, os crescentes movimentos populares no sentido de geração e disponibilização de conteúdo livre, como música, software e documentos, incentivaram o surgimento de enormes bancos de dados públicos on-line. Estas bem-vindas iniciativas permitem que pesquisadores de qualquer lugar do mundo e de qualquer área de pesquisa tenham fácil acesso a estes materiais e possam realizar seus estudos. Usar tais dados também permite facilmente a replicação de experiências e resultados, já que qualquer um pode acessar livremente os mesmos dados. Dois exemplos deste tipo de banco de dados são o NeuroMorpho (41, 42) e o Allen Mouse Brain Atlas (AMBA) (43). O primeiro possui reconstruções tridimensionais de cerca de 6.000 neurônios de variados tipos, regiões cerebrais e espécies. O segundo disponibiliza diversas bases de dados de imagens de expressão gênica, como o de cérebro de rato, que possui cerca de 20.000 genes mapeados.

Portanto, visando contribuir com os avanços das pesquisas em neurociência e reconhecendo o papel de fundamental importância da morfologia, função e genética neuronal nos processos dinâmicos emergentes que ocorrem no sistema nervoso, o objetivo deste trabalho é investigar questões envolvendo forma, função e expressão gênica no cérebro e nos neurônios. Para isso, foram utilizados os dados dos bancos de dados públicos NeuroMorpho e AMBA, além de neurônios artificiais gerados utilizando um modelo de estruturas ramificadas, e imagens de reconstrução tridimensional de raízes de feijão. Na Seção 1.1, são apresentados em mais detalhes os objetivos deste trabalho. Em seguida, na Seção 1.2, são descritos o conteúdo e a organização da tese.

1.1 Objetivos

Reconhecido o papel de fundamental importância da morfologia, função e genética neuronal nos processos dinâmicos emergentes que ocorrem no sistema nervoso, e visando contribuir com os avanços das pesquisas em neurociência, o objetivo deste trabalho é investigar a quantificação e a modelagem do inter-relacionamento entre forma, função e expressão gênica nos neurônios e no cérebro. Considerando o papel de fundamental importância das correlações entre expressão gênica, morfologia e função nos processos dinâmicos emergentes que ocorrem no sistema nervoso, uma intrigante pergunta é como quantificar a influência da expressão gênica no desenvolvimento da forma e da função nos neurônios e na rede neuronal? Visando contribuir para responder a essa questão, propomos como objetivo geral desta tese o início do desenvolvimento de um arcabouço que permita a quantificação dos relacionamentos entre expressão gênica, forma e função. Mais especificamente, o trabalho em questão envolve os seguintes objetivos:

- A) Caracterização e comparação da forma neuronal;
- B) Estudo de processos dinâmicos que ocorrem em redes de estruturas ramificadas (função emergente em redes a partir de indivíduos de forma irregular);
- C) Análise e comparação entre expressões gênicas;
- D) Quantificação da influência da expressão gênica na forma e na função.

Desta forma, os três primeiros são preparatórios para o objetivo D. Os objetivos são realizados por quatro atividades, apresentadas nos Capítulos 4, 5, 6 e no Apêndice A. Na Figura 1.3, os objetivos são ilustrados e são indicados os capítulos relacionados a cada um.

1.2 Conteúdo e organização da tese

Para a realização deste trabalho, foram utilizados os dados das bases públicas NeuroMorpho e AMBA, além de neurônios artificiais gerados por um modelo de estruturas

Figura 1.3 – Ilustração dos objetivos e indicação dos capítulos relacionados a cada um.

ramificadas, apresentados no Capítulo 2. Os métodos utilizados para a caracterização e análise desses dados são descritos no Capítulo 3.

A atividade apresentada no Capítulo 4 é o estudo do hiperespaço neuromorfológico, no qual é desenvolvido o objetivo A. Esta atividade utiliza o modelo proposto por McGhee para a análise da distribuição no morfoespaço de neurônios artificiais gerados por um modelo simples de estruturas ramificadas, que leva em consideração valores de parâmetros obtidos da análise de neurônios reais da base de dados NeuroMorpho. Esta atividade foi realizada em colaboração com o Prof. Sergei N. Taraskin, da Universidade de Cambridge, Inglaterra.

O objetivo B é satisfeito em duas atividades: na análise da influência da expressão gênica na forma e na função (Capítulo 6), e no estudo da dinâmica de propagação de doenças (modelo SIR) em raízes como indivíduos de estrutura ramificada (Apêndice A). Na primeira, é analisada a dinâmica de rede integra-e-dispara. Na segunda, são caracterizadas as formas morfológicas de estruturas ramificadas relevantes ao desempenho de propagação do agente patogênico na rede. Esta atividade foi realizada com a colaboração do grupo de pesquisa do Prof. Taraskin.

A análise e comparação entre expressões gênicas (objetivo C) é abordada na atividade descrita no Capítulo 5, que descreve uma nova metodologia para caracterizar a relação de densidade espacial entre expressões gênicas utilizando a Análise de Densidade Local por Voronoi.

Com o prévio preparo adquirido no desenvolvimento dessas atividades, que abordam os três primeiros objetivos, finalmente o objetivo D pode ser desenvolvido na atividade de análise da influência da expressão gênica na forma e na função (Capítulo 6). Esta última

atividade engloba todas as outras e, portanto, abrange todos os objetivos.

2 Materiais

2.1 NeuroMorpho

NeuroMorpho (41, 42) é um repositório público on-line de reconstruções de neurônios, obtidos de bancos de dados disponíveis na web e de contribuições de colaboradores. O seu propósito é facilitar o acesso e encorajar o compartilhamento de dados neuronais na comunidade científica. Novos dados são adicionados somente pelos administradores, que asseguram a padronização do formato dos dados.

O Computational Neuroanatomy Group (Krasnow Institute for Advanced Study, George Mason University), sob direção do Prof. Giorgio Ascoli, é o grupo responsável pelo desenvolvimento e manutenção do NeuroMorpho. Este repositório é integrante do consórcio Neuroscience Information Framework (NIF) (44), que envolve diversas instituições acadêmicas, como as universidades de Cornell, Yale, Stanford e California.

A primeira versão do NeuroMorpho (Alfa) foi disponibilizada em 2006, com 932 neurônios. Desde então, tem sido continuamente atualizado para incluir mais neurônios e para aperfeiçoar a funcionalidade do *website* (44) (Figura 2.1). Quando este trabalho foi desenvolvido, a versão disponível era a 4.0, com 5673 neurônios. Os dados incluem reconstruções 3D, medidas (volume, diâmetro, etc.) e informações gerais: provedor dos dados (pesquisador e laboratório), artigos de referência, endereços URL relacionados aos dados, experimento (protocolo, método de coloração, etc.), animal (espécie, idade, etc.), região e sub-região cerebral, classe e sub-classe de neurônio, e métodos e softwares utilizados na reconstrução.

Para poder estudar a morfologia dos neurônios, é necessário representá-los e caracterizá-los de forma apropriada para processamento e análise. Para isso, no próprio *website* do NeuroMorpho encontram-se disponíveis os valores de várias medidas, calculados com o software L-Measure (45), também disponível para download. Estas medidas estão ilustradas na Figura 2.2, nomeadas como aparecem na documentação do software. Os conceitos

Figura 2.1 – Evolução do número de neurônios ao longo das versões na base de dados NeuroMorpho. Desde seu lançamento em 2006, novos dados são continuamente adicionados e atualmente é a maior base de dados de morfologia neuronal disponível, contendo 5673 células na versão 4.0.

de compartimento, ramo e bifurcação estão ilustrados na Figura 2.2. Os compartimentos são segmentos representados como cilindros com diâmetro e coordenadas das extremidades. Os ramos são formados por um ou mais compartimentos sequenciais entre dois pontos, que podem ser o soma, bifurcações ou extremidades. As bifurcações são pontos nos quais um ramo se divide em dois outros ramos. A descrição das medidas é apresentada na Tabela 2.1. Na coluna “valor utilizado”, o símbolo “-” indica que a medida gera um único valor para o neurônio. Nos casos em que mais de um valor medido é gerado por neurônio, foram convenientemente utilizados dependendo da medida: a média, a soma total ou o maior valor.

2.2 Neurônios artificiais

Foi proposto um algoritmo gerador de neurônios artificiais como um modelo de referência simples, que gera estruturas artificiais em forma de árvores. Este algoritmo inicia com um simples ramo (uma reta), representado pelo vetor $\vec{\ell}_0$. A extremidade deste vetor é uma bifurcação onde dois outros vetores (ramos), $\vec{\ell}_1$ e $\vec{\ell}_2$, são adicionados à estrutura. Todos estes três vetores são coplanares e a bifurcação é simétrica de tal forma que os vetores $\vec{\ell}_1$ e $\vec{\ell}_2$ formam ângulos iguais com o vetor $\vec{\ell}_0$. O ângulo de bifurcação, θ (ângulo entre os vetores $\vec{\ell}_1$ e $\vec{\ell}_2$), é uma variável aleatória distribuída de acordo com a distribuição

Figura 2.2 – Medidas disponíveis na base de dados NeuroMorpho. São mostrados os nomes originais das medidas em inglês. A descrição dessas medidas pode ser encontrada na Tabela 2.1.

normal:

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp \left[-\frac{(\theta - \bar{\theta})^2}{2\sigma_\theta^2} \right], \quad (2.1)$$

na qual $\bar{\theta}$ é o ângulo médio e σ_θ^2 é a variância da distribuição. Valores grandes de σ_θ^2 permitem uma maior variedade de ângulos em torno do ângulo médio $\bar{\theta}$. Uma vez criados, os vetores $\vec{\ell}_1$ e $\vec{\ell}_2$ são simultaneamente rotacionados ao redor do vetor $\vec{\ell}_0$ num ângulo aleatório ϕ , que também segue a distribuição dada pela Equação 2.1, substituindo θ por ϕ . Esta rotação é redundante para a primeira bifurcação, mas se torna significativa para as seguintes, pois permite o aparecimento de estruturas 3D, ao invés de permanecer 2D. As extremidades dos vetores $\vec{\ell}_1$ e $\vec{\ell}_2$ tornam-se novos pontos de bifurcação. Por exemplo, os vetores $\vec{\ell}_{11}$ e $\vec{\ell}_{12}$ são adicionados na extremidade do vetor $\vec{\ell}_1$. E assim, novos ramos são sucessivamente adicionados.

Tabela 2.1 – Descrição das medidas na base de dados NeuroMorpho

ID	Medida	Descrição	Valor utilizado
1	Height	Altura do neurônio	-
2	Width	Largura do neurônio	-
3	Depth	Profundidade do neurônio	-
4	N_stems	Número de troncos	-
5	N_bifs	Número de bifurcações	-
6	N_branch	Número de ramos	-
7	Diameter	Diâmetro do compartimento	Média
8	Length	Comprimento do compartimento	Soma total
9	Surface	Área da superfície do compartimento	Soma total
10	Volume	Volume do compartimento	Soma total
11	EucDistance	Distância Euclidiana entre um compartimento e o soma	Maior valor
12	PathDistance	Distância do caminho: soma do comprimento dos compartimentos entre dois pontos	Maior valor
13	Contraction	Razão entre EucDistance e PathDistance	Média
14	Branch_Order	Ordem do ramo: distância topológica do ramo ao soma	Maior valor
15	Fragmentation	Número de compartimentos em um ramo	Soma total
16	Suma_surface	Área da superfície do soma	-
17	Pk_classic	$(d_1^r + d_2^r)/(b^r)$, em que $r = 1,5$ (lei de potência de Rall), e b , d_1 e d_2 são os diâmetros dos compartimentos na bifurcação	Média
18	Partition_Asymmetry	Assimetria entre duas sub-árvores: $ n1 - n2 /(n1 + n2 - 2)$	Média
19	Bif_ampl_local	Ângulo entre dois compartimentos numa bifurcação	Média
20	Bif_ampl_remote	Ângulo entre dois ramos numa bifurcação	Média

Para também incluir no modelo ramos que não são exatamente linhas retas, em cada ponto de bifurcação um dos novos ramos pode ser removido com probabilidade ρ_r . O processo de crescimento termina quando o número predefinido de ramos, N_b , é alcançado. Os comprimentos dos segmentos, $|\vec{\ell}_i| = \ell_i$, são variáveis aleatórias discretas distribuídas como:

$$g(\ell) = [1 - \rho_g (1 - \delta_{\ell, \ell_{max}})] (\rho_g)^{\ell-1}, \quad (2.2)$$

em que $\ell = 1, 2, \dots, \ell_{max}$ é o comprimento do segmento e ρ_g é a probabilidade do segmento crescer uma unidade a mais. Desta forma, segmentos mais longos são menos prováveis de serem gerados. $\delta_{\ell, \ell_{max}}$ é o Delta de Kronecker:

$$\delta_{\ell, \ell_{max}} = \begin{cases} 1, & \text{se } \ell = \ell_{max} \\ 0, & \text{se } \ell \neq \ell_{max} \end{cases}, \quad (2.3)$$

2.3 Allen Mouse Brain Atlas

A base de dados pública Allen Mouse Brain Atlas (AMBA) (46) foi criada e é mantida pelo Allen Institute for Brain Science, que também mantém outros repositórios de dados, como de cérebro humano, espinha dorsal de camundongo, cérebro em desenvolvimento de camundongo, entre outros. Os dados disponíveis são imagens de expressão gênica de cerca de 20.000 genes no cérebro de camundongos adultos da linhagem C57BL6, de 56 dias de idade. A expressão gênica foi marcada nos cérebros através do processo ISH (*In Situ* Hybridization), que foram em seguida congelados, fatiados em finas seções e fotografados. O método ISH permite localizar sequências específicas de DNA ou RNA em um tecido (47). Todos os dados produzidos, além dos relatórios técnicos e artigos estão publicamente disponíveis online.

Para este trabalho, foram escolhidos arbitrariamente os genes: Aars, Dach1, Gria4, Hpca, Man1a, Nptx1, Osbpl8 and Wbscr17. E também as regiões anatômicas: CP, CTXsp, PAL, RHP and sAMY. Também foi considerado para análise o cérebro inteiro sem o cerebellum, pois esta região apresenta uma grande variação de ramificações e tamanhos ao compará-la entre diferentes genes, o que dificultou o processo de registro. Foram utilizadas as fatias da direção sagital de F3 a F8. Na Figura 2.3, pode-se visualizar o atlas de referência destas fatias. As regiões anatômicas investigadas estão destacadas logo abaixo do atlas de cada fatia, numeradas de 1 a 6.

Para possibilitar a adequada comparação das seções correspondentes em diferentes genes, foi efetuado o registro das imagens. Primeiro, as imagens foram alinhadas ao atlas de referência (Figura 2.4) usando o software Sqirlz Morph. O alinhamento foi efetuado através da inserção de pontos de referência no atlas (tela à esquerda) que geraram um ponto correspondente na imagem de expressão gênica (tela à direita). Em seguida, estes pontos à direita foram manualmente movidos para os locais que mais se assemelhavam à anatomia exibida no atlas de referência (tela à esquerda). Desta forma, o software pode aplicar efeito de deformação (*morphing*) na imagem de expressão gênica, fazendo com que seus pontos se alinhem com os do atlas.

Figura 2.3 – Atlas de referência das fatias cerebrais selecionadas de F3 a F8 (Fatia 3 a Fatia 8) e as regiões escolhidas: (1) CP - *Caudoputamen*, (2) CTXsp - *Cortical Subplate*, (3) PAL - *Pallidum*, (4) RHP - *Retrohippocampal Region*, (5) sAMY - *striatum-like Amygdalar Nuclei*, e (6) Cérebro (todas as regiões anteriores, mais a área em cinza)

Figura 2.4 – Registro (alinhamento) da imagem de expressão gênica (direita) com o atlas anatômico (esquerda).

3 Métodos

3.1 Análise de Componentes Principais

A Análise de Componentes Principais (PCA - Principal Component Analysis) (48, 49) é um método estatístico que visa reduzir a dimensão de problemas com muitas medidas. O método PCA elimina redundâncias e transforma um sistema descrito por um conjunto de variáveis possivelmente correlacionadas em um novo sistema descorrelacionado. A orientação dos eixos no espaço original é alterada de forma que os eixos de maior dispersão dos dados se tornem os novos eixos de referência espacial.

Os dados podem ser organizados como uma matriz W de dimensões $N \times M$, onde as N linhas são as observações e as M colunas representam as medidas. Como PCA é sensível às variações de escala dos dados, é importante primeiro realizar a normalização dos dados. Cada elemento da linha i na coluna j é subtraído pela média e dividido pelo desvio padrão desta coluna:

$$X_{ij} = \frac{W_{ij} - \bar{w}_j}{s_j}, \quad (3.1)$$

em que

$$\bar{w}_j = \frac{\sum_{i=1}^N W_{ij}}{N}, e \quad (3.2)$$

$$s_j = \sqrt{\frac{\sum_{i=1}^N (W_{ij} - \bar{w}_j)^2}{N - 1}} \quad (3.3)$$

são respectivamente a média e o desvio padrão da medida j . O próximo passo é calcular a matriz de covariância V :

$$V_{ij} = \frac{\sum_{k=1}^N (X_{ki} - \bar{x}_i)(X_{kj} - \bar{x}_j)}{N - 1}, \quad (3.4)$$

em que $\bar{x}_i$ é a média da medida i . Deve-se ressaltar que como os dados foram previamente normalizados, a média é sempre igual a zero ($\bar{x}_i = 0$). Em seguida, são calculados os au-

tovalores λ e autovetores $\vec{v}_\lambda$ de V . Cada autovalor está associado a um autovetor. Assim, ao organizar os autovalores em ordem decrescente, os autovetores também são respectivamente ordenados. Esta ordenação é realizada porque quanto maior o autovalor, maior a quantidade de variação dos dados é explicada pela componente associada. Desta forma, para realizar redução de dimensionalidade minimizando as perdas de informação, basta selecionar os primeiros P valores, de acordo com quantas dimensões se pretende gerar. Portanto, os autovetores $\vec{v}_\lambda$ são ordenados de acordo com os autovalores λ , resultando na nova matriz de autovetores $\vec{p}_\lambda$. O próximo passo é realizar a transformação linear:

$$W' = (\vec{p}_\lambda X)^T . \quad (3.5)$$

A quantidade de variação dos dados explicada pelos P autovetores escolhidos pode ser quantificada pela seguinte expressão:

$$r = \frac{\sum_{j=1}^P \lambda_j}{\sum_{j=1}^M \lambda_j} . \quad (3.6)$$

3.2 Análise Canônica

A Análise Canônica (50–52), também conhecida como análise de discriminantes lineares (*Linear Discriminant Analysis* - LDA), é um método que busca encontrar a projeção que melhor separa classes de dados pré-definidas. Isto é obtido a partir da maximização da dispersão inter-classe, ou seja dispersão entre classes, enquanto minimiza a dispersão intra-classe dentro de cada classe. Considerando que cada elemento da matriz normalizada X , definida na Equação 3.1, pode ser classificado em uma classe C_i contendo n_i elementos, em que $i = 1, 2, \dots, N_c$ e N_c é o número máximo de classes. Assim, as matrizes de dispersão inter-classe (Equação 3.7) e intra-classe (Equação 3.8) podem ser definidas como:

$$S_{\text{inter}} = \sum_{i=1}^{N_c} n_i (\langle \vec{x} \rangle_i - \langle \vec{x} \rangle) (\langle \vec{x} \rangle_i - \langle \vec{x} \rangle)^T , \quad (3.7)$$

$$S_{\text{intra}} = \sum_{i=1}^{N_c} S_i , \quad (3.8)$$

em que $\langle \vec{x} \rangle_i$ é o vetor de características médio dos elementos na classe C_i , $\langle \vec{x} \rangle$ é o vetor de características médio de todos os elementos, e S_i é a dispersão das medidas dentro de

cada classe (matriz de dispersão para cada classe C_i):

$$S_i = \sum_{k \in C_i} (\vec{x}_k - \langle \vec{x} \rangle_i)(\vec{x}_k - \langle \vec{x} \rangle_i)^T. \quad (3.9)$$

Desta forma, pode-se finalmente calcular os autovalores e autovetores da matriz $S_{\text{intra}}^{-1} S_{\text{inter}}$, em que S_{intra}^{-1} é a inversa de S_{intra} . Em seguida, os autovalores devem ser organizados em ordem decrescente. Assim, podem-se selecionar os autovetores correspondentes aos maiores autovalores para a nova projeção, com dimensionalidade reduzida.

3.3 Modelagem de hiperespaço morfológico

Um hiperespaço morfológico teórico, em analogia com conceitos geométricos, pode ser entendido como um espaço n-dimensional, no qual os eixos estão associados com medidas. Em biologia, particularmente para análise morfológica, estas medidas referem-se às propriedades da forma, como comprimento, altura, largura ou volume de um organismo vivo ou estrutura. Em termos ideais, o morfoespaço pode ser construído pela modelagem das entidades biológicas a partir das variações desses parâmetros e considerando todos os possíveis indivíduos cuja existência é possível. Então, apesar de contínuo, o morfoespaço é reduzido como consequência de diversas restrições impostas por propriedades específicas dos organismos e seu habitat.

Utilizando o conceito de morfoespaço, torna-se possível definir regiões e bordas correspondentes às propriedades permitidas de geometria, função, filogenia e de desenvolvimento das entidades biológicas investigadas (53) (veja a Figura 3.1). Essas restrições podem ser divididas em extrínsecas e intrínsecas. A primeira refere-se às restrições impostas pelas leis da geometria e da física, restringindo as possíveis formas e funções. São elas:

- FIG: Formas Impossíveis Geometricamente;
- FPG: Formas Possíveis Geometricamente;
- FIFu: Formas Impossíveis Funcionalmente;
- FPFu: Formas Possíveis Funcionalmente;

Já as restrições intrínsecas referem-se à biologia de um organismo específico (leis genéticas):

- FPFi: Formas Possíveis Filogeneticamente
- FID: Formas Impossíveis de se Desenvolver
- FPD: Formas Possíveis de se Desenvolver

Como as restrições extrínsecas possuem origem distinta das restrições intrínsecas, uma não necessariamente engloba a outra, o que gera as intersecções de FPFi tanto com FPG quanto com FIG. Deve-se notar que FPD possui intersecção com FPFu e FIFu, mas não FIG, pois se a forma é geometricamente impossível, naturalmente também é inviável se desenvolver.

Além dessas regiões, pode-se definir também o morfoespaço teórico, que é o espaço das medidas extraídas dos indivíduos reais. A investigação do espaço teórico pode ajudar a desenvolver hipóteses como quais fatores ao longo dos estágios evolutivos e de desenvolvimento afetam as trajetórias subsequentes dentro do morfoespaço. Para simular uma possível representação do morfoespaço teórico, podem-se implementar algoritmos que gerem indivíduos artificiais baseados em modelos estatísticos que selecionam algumas características morfológicas e que possuam ampla variedade de valores de medidas. Evidentemente, este método é incapaz de reproduzir exatamente os processos naturais da criação da vida e desenvolvimento. Ao mesmo tempo, deve-se levar em conta que o conjunto de indivíduos empíricos adotado contém apenas uma fração dos indivíduos naturais. No entanto, ambos subconjuntos podem prover informações e estimativas sobre a densidade e localização dos dados empíricos dentro do hiperespaço teórico simulado.

Esta abordagem de morfoespaço teórico pode ser aplicada à neurociência para modelar o hiperespaço das formas neuronais utilizando-se um conjunto de medidas extraídas de um conjunto de células neuronais reais. Utilizando as medidas disponíveis na base de dados NeuroMorpho, pode-se modelar o espaço empírico e verificar o comportamento (bordas e sobreposições) de cada uma das regiões anteriormente definidas.

3.4 Função de densidade radial

A relação entre os neurônios no espaço de medidas 20D pode ser explorada utilizando a abordagem de função de densidade radial. A função radial $f(R)$ calcula o número de neurônios localizados entre as distâncias R e $R + \Delta R$ a partir de um neurônio em particular. Cada neurônio, representado por um vetor com componentes dados pelas

Figura 3.1 – Representação genérica das várias regiões possíveis em um hiperespaço morfológico.

respectivas medidas morfológicas, é utilizado como centro de uma esfera n -dimensional, cujo raio é progressivamente incrementado. Para cada passo, o número de neurônios na região considerada é computado em função de R .

Como esta função reflete o padrão de distribuição dos vizinhos ao redor do neurônio no morfoespaço 20D, é esperado que dois neurônios com características geométricas similares, portanto mapeadas próximas no espaço de características, produzirão curvas similares de $f(R)$. Além disso, devido ao tamanho finito do espaço ocupado pelos neurônios, é esperado que a função $f(R)$ possua um pico em algum valor R' . Mais especificamente, neurônios mais centrais tendem a produzir picos nos valores menores de R (regiões mais próximas do neurônio), enquanto que neurônios próximos à borda do espaço ocupado tendem a ter um pico nos maiores valores de R (regiões mais distantes), podendo corresponder a *outliers*, ou seja, valores atípicos que se encontram bastante afastados dos demais. Na Figura 3.2, é ilustrado o funcionamento da função de densidade radial.

Figura 3.2 – Considerando a distribuição de pontos no espaço 2D (a), aplicando a função de densidade radial $f(R)$ no ponto destacado em vermelho, obtém-se a curva de densidade mostrada em (b).

3.5 Diagrama de Voronoi

O diagrama de Voronoi (54) determina a região de influência (vizinhança) de um conjunto de pontos em termos de distâncias Euclidianas. Considerando P como um conjunto de pontos em um espaço Euclidiano, cada ponto $p_i \in P$, $i = 1, 2, \dots, n$, possui uma região ao seu redor cujos pontos estão mais próximos de p_i do que de qualquer outro ponto em P . As regiões determinadas por cada ponto p_i formam o mosaico de células que define um diagrama de Voronoi. Um exemplo é mostrado na Figura 3.3. Há diversos métodos disponíveis para se gerar um diagrama de Voronoi, como em (55, 56). Neste trabalho, foi utilizada a transformada de distância para determinar as células de Voronoi (57) por ser um método rápido e eficaz para imagens (espaço discreto). Cada ponto p_i localiza-se em um pixel da imagem e recebe um rótulo de identificação único l_i que é propagado em ondas circulares ao longo de distâncias sucessivas. As fronteiras entre as células são determinadas pela colisão entre ondas de rótulos. Para cada ponto p_i , é calculado $D_i = \{d_1, d_2, \dots, d_m\}$, em que $d_j < d_{j+1}$, que é o conjunto de todos os possíveis valores de distância Euclidiana entre um pixel inicial p_i e qualquer outro pixel na imagem. A propagação de rótulos inicia pela atribuição de um rótulo l_i para todos os pixels de distância d_1 em relação a cada ponto p_i . Em seguida, são considerados os pixels de distância d_2 para a segunda onda, e assim segue-se sucessivamente. É importante ressaltar que a atribuição de rótulo somente ocorre se o pixel ainda não estiver rotulado. Assim que não haja mais pixels sem rótulo, o algoritmo encerra a execução e o resultado final é um diagrama de Voronoi V com cada grupo de pixels com o mesmo rótulo l_i correspondendo a uma célula $t_i \in V$.

Figura 3.3 – Exemplo de um diagrama de Voronoi. Cada célula determina uma região de pontos (pixels, no caso de imagens) que se encontram mais próximos de um determinado ponto (destacado dentro de cada célula).

3.6 Análise de Densidade Local por Voronoi

A Análise de Densidade Local por Voronoi (VLDA - Voronoi Local Density Analysis) (52) é uma metodologia para analisar a distribuição de densidade e para descobrir possíveis relações espaciais entre duas imagens. Para analisar dois volumes de expressão gênica, as imagens correspondentes a cada fatia nos dois volumes são comparadas utilizando VLDA. Em seguida, é calculada a média dos valores resultantes para cada fatia para gerar um único valor que expressa o nível de similaridade de dois volumes de expressão gênica. Realizando este processo para todos os possíveis pares de expressão gênica, pode-se gerar uma rede para visualizar o nível de similaridade entre os padrões de expressão gênica a partir da espessura das conexões e também comparar entre redes formadas considerando regiões cerebrais.

Há três passos no método VLDA:

1. Gerar o diagrama de Voronoi V usando a imagem I_1 ;
2. Sobrepor o diagrama V sobre a imagem I_2 e calcular a taxa de densidade local $R(t_i)$ para cada célula $t_i \in V$;
3. Calcular o desvio padrão σ de todos os $R(t_i)$.

O diagrama de Voronoi V é gerado como descrito na Seção 3.5. A taxa de densidade local $R(t_i)$ pode ser expressa (52) como:

$$R(t_i) = \frac{d_2(t_i)}{d_1(t_i)} = \frac{\frac{Q(t_i)}{A(t_i)}}{\frac{1}{A(t_i)}} = Q(t_i) , \quad (3.10)$$

em que $d_1(t_i)$ e $d_2(t_i)$ são as densidades da célula t_i considerando o diagrama de Voronoi V sobre as imagens I_1 e I_2 , respectivamente. Como V_1 é gerado a partir de I_1 , há somente um ponto por célula:

$$d_1(t_i) = \frac{1}{A(t_i)} , \quad (3.11)$$

em que $A(t_i)$ é a área da célula t_i . Similarmente, a densidade da célula t_i considerando a imagem I_2 pode ser definida como:

$$d_2(t_i) = \frac{Q(t_i)}{A(t_i)} . \quad (3.12)$$

Neste caso, há uma quantidade $Q(t_i)$ dentro da célula t_i que pode assumir um valor diferente de 1. Portanto, a taxa local de densidade $R(t_i)$, definida na Equação 3.10,

é diretamente dada pela quantidade $Q(t_i)$. Assim, é possível determinar σ (passo 3 do método VLDA) calculando-se o desvio padrão da taxa de densidade local $R(t_i)$ para todas as células $t_i \in V$.

Ao invés de imagens binárias com vários pontos como as analisadas em (52), as imagens utilizadas neste trabalho possuem níveis (*grayscale* - escala/níveis de cinza) indicando a intensidade de expressão gênica. Assim, para comparar duas imagens, uma delas é binarizada para identificar os pontos de maior valor de intensidade e gerar o diagrama de Voronoi, enquanto que a outra imagem é utilizada diretamente no formato de níveis de cinza. A binarização converte os níveis de cinza em 0 (pixels de fundo) ou 1 (pixels relevantes), de acordo com o parâmetro de limiar τ . Desta maneira, $Q(t_i)$ é a soma dos valores de níveis de cinza (intensidade de expressão gênica) de todos os pixels dentro de uma célula t_i .

Portanto, σ representa o nível de similaridade entre os padrões de expressão entre dois genes, considerando densidades locais. Os valores mais baixos de σ indicam maior similaridade. Pode-se obter σ para cada fatia e gerar uma média μ , que pode ser utilizada como uma medida de similaridade entre dois volumes de expressão gênica. Assim, calculando-se μ para todos os possíveis pares de expressão gênica, pode-se criar uma rede, na qual é possível visualizar a similaridade entre padrões de expressão gênica.

Esta metodologia é ilustrada na Figura 3.4. Em (a), é mostrado o uso de VLDA. As imagens correspondentes à fatia cerebral k para as expressões gênicas G_i e G_j , $I_1 = G_{ik}$ e $I_2 = G_{jk}$, são selecionadas e uma região anatômica é identificada nessas imagens utilizando a máscara de região. Esta máscara é criada utilizando o mapa de referência de regiões cerebrais, disponibilizado junto com as imagens de expressão no *website* da base de dados Allen Mouse Brain Atlas. A imagem G_{ik} é binarizada para obter os pontos com os maiores valores de intensidade. Utilizando estes pontos, é gerado o diagrama de Voronoi $V = V_{ik}$. Em seguida, este diagrama é sobreposto sobre a imagem de intensidade de expressão gênica G_{jk} . Assim, é obtido $R(t_i)$ para cada célula t_i como a soma dos valores de níveis de cinza dentro de t_i . O próximo passo é o cálculo do desvio padrão $\sigma = \sigma_{ijk}$ da densidade das células t_i . Isto é repetido para todas as fatias (Figura 3.4(b)), gerando $\sigma_{ijk}, \sigma_{ij\{k+1\}}, \sigma_{ij\{k+2\}}, \dots, \sigma_{ijn_s}$. Desta maneira, pode-se determinar a média do desvio padrão das fatias, μ_{ij} , para o par de volumes de expressão gênica G_i e G_j . Estes passos são repetidos para todos os possíveis pares de volumes e uma rede é gerada. Os volumes de expressão gênica são representados nos nós. A espessura de cada conexão (aresta) é determinada pela média do desvio padrão entre os volumes representados pelos nós conectados: os mais similares possuem menores valores e são representados como conexões mais grossas (para ressaltar a similaridade). Esta representação permite a visualização das medidas de similaridade entre distribuições espaciais de expressão gênica.

Figura 3.4 – Passos da metodologia VLDA. Em (a), as imagens correspondentes às fatias cerebrais k de duas expressões gênicas, G_{ik} e G_{jk} , são analisadas. Primeiramente, uma região é selecionada utilizando uma máscara, para ambas as imagens. Em seguida, G_{ik} é binarizada usando um valor de limiar τ para selecionar os valores mais intensos como sinais de expressão gênica significativos. Com estes pontos, é possível gerar o diagrama de Voronoi V_{ik} , que é sobreposto sobre a imagem de níveis de cinza G_{jk} . Assim, pode-se calcular o desvio padrão da taxa de densidade local de todas as células, σ_{ijk} . A análise em (a) é repetida para todas as fatias. Em (b), é calculada a média μ_{ij} de todos os σ_{ijk} obtidos. Esta média μ_{ij} é uma medida que denota a similaridade entre os dois volumes de expressão gênica, G_i e G_j . Após executar estes procedimentos para todos os pares de expressão gênica, é possível gerar uma rede, na qual a espessura das conexões representa a similaridade entre volumes (quanto mais espessa, maior a similaridade).

3.7 Comparação entre redes utilizando Análise de Componentes Principais

A rede abordada na Seção 3.6 pode ser alternativamente visualizada utilizando Análise de Componentes Principais (PCA - *Principal Component Analysis*). Desta forma, é possível verificar quais nós (expressão gênica) estão mais próximos (são mais similares) em um gráfico de dispersão. Também é possível comparar entre redes, o que é útil para comparar entre redes de regiões.

PCA é uma transformação linear dos dados para um novo sistema de coordenadas que maximiza a variância (variabilidade nos dados) ao longo dos novos eixos (componentes principais) (48). O primeiro componente concentra a maior variância, e cada componente subsequente acumula a maior quantidade possível da variância restante. Esta característica permite a redução da dimensionalidade dos dados, enquanto minimiza a perda de informação por permitir descartar os componentes que descrevem uma menor quantidade de variância. A redução de dimensionalidade é útil na visualização de dados de alta dimensionalidade em gráficos 2D ou 3D.

PCA pode ser diretamente aplicada na matriz de pesos de cada rede. A matriz de pesos é uma matriz quadrada, na qual cada posição possui uma medida de similaridade μ_{ij} de duas expressões gênicas G_i e G_j . Desta maneira, cada linha pode ser considerada como um vetor de características de uma expressão gênica e as colunas são medidas que quantificam sua similaridade em relação a cada expressão gênica. Assim, é possível reduzir a dimensionalidade com PCA e visualizar num gráfico a inter-relação entre expressões gênicas na rede.

Para comparar entre redes, a matriz de pesos de cada rede é convertida numa linha (vetor de características). Em seguida, essas linhas formam uma nova matriz que pode ser analisada e visualizada como no caso anterior.

3.8 Dinâmica integra-e-dispara

As redes de neurônios foram utilizadas como substratos para a dinâmica integra-e-dispara (58, 59), que é um modelo bem conhecido para tempo discreto, utilizado para simular o comportamento coletivo de neurônios. É assumido na dinâmica que cada neurônio i começa com uma carga inicial $Q_i(0)$, escolhida aleatoriamente na faixa $Q_i(0) \in [0, Q_{max}]$. A cada passo no tempo, um dado neurônio dispara uma carga unitária para cada um de seus vizinhos se essa carga for maior que o limiar Q_t , que foi considerada como sendo igual para todos os neurônios. Em geral, a dinâmica global passa por um estado transiente até atingir um regime estacionário, no qual o número de neurônios disparando no instante t , $S(t)$, converge para um valor constante e as flutuações ao redor desta média são pequenas. Para detectar esse regime estacionário, é analisada a variância relativa de $N_S(t)$, dada por

$$v(T) = \frac{\sum_{t=0}^T N_S(t)^2}{\left[\sum_{t=0}^T N_S(t)\right]^2} - 1. \quad (3.13)$$

O comportamento de $v(t)$ é em geral o mesmo para conjuntos diferentes de parâmetros. Há uma pequena fase inicial na qual a variância relativa aumenta e atinge um valor máximo. Após isso, a curva da função declina monotonicamente. Portanto, pode-se assumir que o regime estacionário é alcançado no instante T_0 , no qual a variância relativa atinge valores menores que 10^{-3} . A dinâmica continua até um tempo máximo pré-determinado T_{max} . Enquanto a dinâmica evoluiu, o estado binário de cada neurônio foi gravado numa matriz. Um elemento um na posição (t, i) dessa matriz indica que o neurônio i disparou no instante t .

Desta forma, a dinâmica produz uma matriz binária de tamanho $T_{max} \times N$, na qual N é o número total de neurônios na rede. Desta rede, os dados utilizados são os gerados no regime estacionário, ou seja, uma matriz S de tamanho $T_{max} - T_0 \times N$, cujos elementos são denominados como $s_{t,i}$. Cada coluna na matriz corresponde à informação de um neurônio individual no estado estacionário para a dinâmica integra-e-dispara. Para caracterizar essa informação para um dado neurônio i , são extraídas três diferentes medidas da coluna i de S : (i) a taxa média de disparo (r_i), (ii) o intervalo máximo entre disparos (m_i), e (iii) a entropia entre disparos (H_i). A primeira medida corresponde ao número médio de disparos no intervalo de T_0 a T_{max} , que pode ser escrito como $r_i = 1 / (T_{max} - T_0) \sum_{t=T_0}^{T_{max}} s_{t,i}$. A segunda medida corresponde ao intervalo máximo entre dois disparos consecutivos, enquanto que a última corresponde à entropia dos intervalos entre disparos. Para calcular esta última medida, é primeiramente avaliada a probabilidade $p_i(d)$ de um dado neurônio i não disparar d consecutivas vezes antes de um novo disparo. Assim, foi aplicado o

conceito tradicional de entropia sobre a distribuição de $p_i(d)$ para obter

$$H_i = \sum_{d=0}^{\infty} p_i(d) \log(p_i(d)) . \quad (3.14)$$

3.9 Entropia condicional

Esta metodologia é ilustrada na Figura 3.5. Primeiro, um gráfico de dispersão é construído para cada medida, representado a posição de cada neurônio no gráfico em relação à intensidade de expressão gênica (eixo x) versus a sua medida morfológica (eixo y). Então, o eixo da intensidade de expressão gênica é dividido em M intervalos (caixas). Em cada intervalo, os pontos dentro dele são utilizados para criar um histograma considerando a divisão dos valores da medida em N partes. Em seguida, cada histograma é normalizado e sua entropia é calculada como sendo

$$e = - \sum_{i=1}^N h_i \log(h_i) , \quad (3.15)$$

em que e é a entropia do histograma h sendo avaliado, e h_i é a frequência h_i (número de elementos) na caixa i . Assim, de cada medida são obtidos M histogramas, que por sua vez são usados para calcular os valores de entropia e . Estes valores de entropia são utilizados para calcular a entropia condicional da medida considerada, sendo

$$e_c = \frac{\sum_{k=1}^M e_k N_k}{\sum_{k=1}^M N_k} , \quad (3.16)$$

no qual e_c é a entropia condicional que caracteriza a dispersão dos dados de uma medida específica, e_k é a entropia do histograma k , e N_k é o número de elementos no histograma k .

Figura 3.5 – Entropia condicional. Para cada medida i , é gerado o seu gráfico de dispersão. Em seguida, o eixo x (intensidade de expressão gênica) é dividido em partes iguais. Para cada parte, é gerado um histograma k a partir da divisão do eixo y (valores de medidas) em partes iguais. O histograma k é usado para calcular a entropia k na Equação 3.15. Os valores de entropia são utilizados para calcular a entropia condicional i (Equação 3.16).

3.10 Análise de inclinação das fronteiras nos gráficos de dispersão

A análise de inclinação foi proposta para analisar a inclinação dos limites superior e inferior dos pontos num gráfico de dispersão, pois estas informações podem ser úteis na identificação do efeito da influência da expressão gênica.

A análise de inclinação é calculada a partir do fator de inclinação:

$$s = \sqrt{m \times r} \ , \quad (3.17)$$

que é a média geométrica da inclinação m e da correlação de Pearson r . Para calcular esses valores, são determinados os pontos nos limites da distribuição a partir da divisão do eixo x em vários intervalos e da obtenção do ponto mais alto em cada intervalo para determinar o limite superior (ou o ponto mais baixo no caso do limite inferior). Em

seguida, esses pontos são utilizados para calcular a correlação:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right) , \quad (3.18)$$

em que $\bar{X}$ é a média e s_X é o desvio padrão dos valores X_i . Para calcular m , primeiro é aplicado o PCA nos pontos para determinar uma reta representativa da direção da distribuição. Utilizando o autovetor v da primeira componente, pode-se calcular:

$$m = \text{atan} \left(\frac{v_y}{v_x} \right) , \quad (3.19)$$

em que atan é a função arco tangente, e v_y e v_x são os dois valores que compõem o autovetor da primeira componente.

4 Hiperespaço neuromorfológico

Foi utilizado neste estudo o conceito de espaço neuromorfológico como espaço multidimensional definido por um conjunto de medidas da morfologia de um conjunto representativo de quase 6.000 neurônios biológicos, disponíveis na base de dados NeuroMorpho. Pela primeira vez, foi analisada uma grande base de dados com a finalidade de encontrar a distribuição geral das características geométricas. Foi utilizado o conceito de espaço de formas biológicas de McGhee (Seção 3.3) para formalizar a análise, permitindo a comparação entre estruturas em forma de árvore geometricamente possíveis, obtidos a partir de um modelo de referência simples e formas neuronais reais. Dois tipos de projeções, PCA e Análise Canônica, foram usadas para visualizar a distribuição original dos neurônios 20D (20 medidas) em espaços morfológicos 2D. Estas projeções permitiram que as características mais importantes fossem identificadas. Também foi realizada uma análise de densidade de dados no espaço de características 20D original para confirmar a estrutura aglomerada (*clustering*). Vários resultados são reportados, incluindo o fato de que neurônios reais ocupam somente uma pequena porção dentro do espaço geometricamente possível e que duas variáveis principais são suficientes para explicar cerca de metade variabilidade dos dados. Foi também descoberto que a maioria das medidas são importantes na representação da variabilidade morfológica dos neurônios reais.

4.1 Configuração do modelo de geração de neurônios artificiais

Utilizando o modelo da Seção 2.2, foram gerados $N = 6000$ neurônios artificiais considerando quase todos os possíveis valores para os parâmetros livres de acordo com os dados reais, ou seja, $1 \leq N_b \leq 8000$ (número de ramos), $0 \leq \rho_g \leq 1$ (probabilidade

do ramo crescer uma unidade a mais), $0 \leq \rho_r \leq 1$ (probabilidade de remover o ramo), $0 \leq \bar{\theta} \leq \pi$ (ângulo médio de bifurcação), $-\pi \leq \phi \leq \pi$ (ângulo de rotação da bifurcação), $\bar{\phi} = 0$ (ângulo médio de rotação da bifurcação), e $\ell_{max} = 100$ (comprimento máximo do ramo). Para as variáveis σ_θ e σ_ϕ (variância das distribuições dos ângulos de bifurcação e rotação), foram considerados os intervalos $[0, \pi/6]$ e $[0, \pi/9]$, respectivamente. Todas as variáveis foram escolhidas de forma aleatória, com exceção de N_b e $\bar{\theta}$, que foram escolhidos de acordo com a distribuição dos dados reais. É importante ressaltar que, devido à sua generalidade, o modelo possivelmente cobre FPG (Formas Possíveis Geometricamente) de uma forma quase ideal. O modelo possui este grau de generalidade porque cada um dos parâmetros morfológicos é coberto independentemente dos outros e de forma uniforme. Portanto, com um número grande o suficiente de amostras, as formas produzidas por este modelo podem incluir todos os casos, inclusive aqueles caracterizados por independência de características morfológicas. Por exemplo, mesmo que neurônios reais sejam caracterizados por segmentos dendríticos cujos comprimentos diminuam ao longo da hierarquia de bifurcação, este tipo de neurônio também seria gerado pelo modelo como consequência da escolha independente de comprimentos.

4.2 Resultados e discussão

Nesta seção, são apresentados os resultados relacionados ao espaço morfológico neuronal e sua organização. O modelo de referência simples é usado para gerar células artificiais utilizadas para obter as fronteiras do espaço teórico. Assim, os neurônios reais são distribuídos neste espaço e comparados com os artificiais. Em seguida, é analisada a distribuição espacial das categorias utilizando as projeções de PCA e Análise Canônica, além da análise em alta dimensão usando a função de densidade radial.

4.2.1 Modelagem do neuromorfoespaço

Para demonstrar a viabilidade de descobrir as fronteiras do morfoespaço teórico de formas neuronais, foi utilizado o modelo de referência da Seção 2.2 para gerar 6000 neurônios

artificiais. Sete medidas foram extraídas deles: largura, altura, profundidade, número de bifurcações, número de ramos, ordem do ramo (máxima distância topológica entre o soma e os ramos) e ângulo entre ramos. Para padronizar as escalas das medidas de comprimento entre os neurônios artificiais e reais, as três primeiras foram convertidas em medidas adimensionais: $L_1 = \text{altura}/\text{largura}$, $L_2 = \text{profundidade}/\text{largura}$ e $L_3 = \text{profundidade}/\text{altura}$. As distribuições dessas variáveis para os neurônios artificiais gerados são representadas pelas curvas vermelhas na Figura 4.1(a)-(g). Para comparação, as distribuições correspondentes aos neurônios reais são apresentadas em preto. É possível observar que são bastante similares em termos de forma e escala. Em parte, isto foi obtido pelo uso de valores experimentais disponíveis para alguns dos parâmetros livres do modelo, como o número médio de ramos (ver Figura 4.1(e)) e ângulo médio de bifurcação (g).

Figura 4.1 – Distribuição de (a) L_1 , (b) L_2 , (c) L_3 , (d) número de bifurcações, (e) número de ramos, (f) ordem do ramo e (g) ângulo entre ramos. As curvas vermelhas correspondem às distribuições dos neurônios artificiais gerados pelo modelo (Seção 2.2), enquanto que os histogramas em preto representam as distribuições dos neurônios reais. Em (d), (e) e (f), são exibidos em destaque ampliações das regiões dos picos.

Utilizando PCA, foi feita a projeção do espaço de medidas 7D para 2D, como pode ser visto na Figura 4.2. O modelo (círculos cinzas) foi capaz de cobrir adequadamente todo o morfoespaço real (circunferências pretas). A análise da distribuição dos neurônios reais

revelou que os neurônios tendem a se tornar mais tridimensionais ao se deslocar de baixo para cima na borda mais a direita da região dos neurônios reais (ou seja, o neurônio (b) é mais tridimensional que o (a), e assim por diante). Efeito similar pode ser observado nos neurônios artificiais mostrados na Figura 4.3, na qual também podem-se observar estruturas globulares densas típicas nas regiões do morfoespaço que não contêm nenhum neurônio real.

Em relação à cobertura da variabilidade dos dados, pode-se calcular que 63,7% da variação total dos dados é explicada pelos dois primeiros componentes principais (38,3% e 25,4%, respectivamente). Além de se analisar o peso dos componentes principais pelos autovalores, também é possível analisar o peso de influência das medidas dentro de cada componente principal usando os autovetores. Na Tabela 4.1, a contribuição das medidas está bastante distribuída em CP1 (primeiro componente principal), demonstrando que todas as medidas são importantes para a variabilidade captada por CP1. Já em CP2, a variabilidade está mais concentrada nas medidas L_3 e L_2 .

Tabela 4.1 – Pesos do PCA das projeções nas Figuras 4.2 e 4.3). São apresentados os pesos das sete medidas nos dois componentes principais (CP1 e CP2). Quanto maior o percentual, maior a contribuição da medida na variação dos dados do componente principal.

Measurements	CP1	CP2
L_1	11%	4%
L_2	15%	23%
L_3	11%	25%
N. bifurcações	19%	16%
N. ramos	19%	16%
Ordem do ramo	14%	8%
Ângulo entre ramos	11%	8%
Total	100%	100%

4.2.2 Inter-relação das medidas e distribuição espacial dos neurônios

O foco desta Seção é o espaço FPD (Formas Possíveis de se Desenvolver), que contém os neurônios reais. Para isso, foram utilizadas todas as 20 medidas disponíveis na base de dados NeuroMorpho.

Primeiramente, foi analisada a inter-relação entre essas medidas utilizando o coefici-

Figura 4.2 – PCA mostrando a distribuição dos neurônios reais (circunferências pretas) e artificiais (círculos cinzas) no morfoespaço, com a indicação de alguns exemplos de neurônios reais.

Figura 4.3 – PCA com os neurônios reais (circunferências pretas) e artificiais (discos cinzas) como na Figura 4.2, mas agora indicando exemplos de neurônios artificiais.

ente de correlação de Pearson (49). Os resultados são apresentados na Figura 4.4. Em particular, altos valores de correlação podem ser observados entre as medidas número de bifurcações (medida 3), número de ramos (medida 4) e ordem do ramo (medida 14). A princípio, quanto maior o valor de ordem do ramo, espera-se que o número de ramos e bifurcações também seja maior. No entanto, isto somente é verdade no caso em que a maioria das ordens sejam bem povoadas por ramos, ao contrário do que seria observado num encadeamento linear de ramos. Portanto, estas correlações aparentemente indicam que a maioria das ordens dos ramos são bem povoadas por ramos. Outras correlações particularmente altas podem ser verificadas entre largura (medida 5), altura (medida 6), profundidade (medida 7) e distância Euclidiana (medida 12), que são inerentemente esperadas.

Figura 4.4 – Coeficiente de correlação de Pearson entre pares de medidas.

Na Figura 4.5, são apresentados os gráficos de PCA dos neurônios, agrupados por (a) tipo de célula, (b) regiões cerebrais e (c) espécies. Para o tipo de célula, foram selecionados os 15 maiores grupos entre os 39 existentes para exibição no gráfico em (a). Os neurônios nestes 15 grupos representam 95% do número total de células.

Como pode ser observado na Figura 4.5(a), somente o grupo Uniglomerular Projected Neurons (círculos sólidos cianos) constitui um agrupamento compacto. Neurogliaform (quadrados amarelos), Calretinin (estrelas cianas) e Bitufted (círculos sólidos verdes) exibem a maior parte de suas células agrupadas à esquerda, enquanto que as outras categorias não se encontram em agrupamentos bem definidos. O grupo mais numeroso, células Piramidais (*Pyramidal* - círculos abertos azuis), pode ser encontrado em diversas áreas da projeção. Na parte de baixo, à direita, é possível distinguir algumas células do grupo Motoneuron (estrelas púrpuras) e algumas células Piramidais (círculos abertos azuis) formando dois subgrupos separados e espalhados.

Na Figura 4.5(b), um grande número de categorias agrupadas podem ser observadas,

Figura 4.5 – Visualizações das projeções PCA, com os neurônios destacados por: (a) tipo de célula, (b) regiões cerebrais e (c) espécies. Há 39 tipos de células (somente 15 mostradas aqui), 15 regiões e 11 espécies.

como Protocerebrum (cruzes azuis), Cercal Sensory System (quadrados cianos), Retina (triângulos vermelhos apontando para cima), Brainstem (quadrados azuis), Basal Forebrain (triângulos verdes apontando para baixo) e Olfactory Bulb (círculos sólidos verdes). Este último permaneceu bem separado dos outros e pode ser verificado que corresponde ao tipo de célula Uniglomerular. As células do Cortex Cerebral (*Cerebral Cortex* - sinais de mais pretos) correspondem principalmente às células Piramidais e incluem algumas células não reportadas. As regiões de Spinal Cord (estrelas vermelhas) e Brainstem (quadrados azuis) são compostos em sua maioria do tipo Motoneuron.

Na Figura 4.5(c), são mostradas as distinções entre os grupos de células de acordo com a espécie animal na qual foram encontradas. Podem-se distinguir três agrupamentos bem separados: drosophila (triângulos azuis apontando para a direita), humano (diamantes vermelhos) e gato (quadrados azuis). Grilo (triângulos púrpuras apontando para a esquerda), salamandras (círculos sólidos amarelos) e macacos (sinais de mais pretos) também apresentam regiões bem definidas, mas se sobrepõem com camundongo (quadrados verdes) e gato (cruzes cianos), que são as duas maiores categorias.

Na Figura 4.6, é exibida a variabilidade explicada por cada componente principal. Estes valores foram calculados utilizando os autovalores: valores mais altos contribuem mais. Neste gráfico, os autovalores foram convertidos em porcentagens e organizados como uma sequência cumulativa de barras, ressaltando a contribuição cumulativa de cada componente principal para explicar a variabilidade dos dados. Os dois primeiros autovalores utilizados nos gráficos de PCA explicam 46% da variância.

Figura 4.6 – Percentual cumulativo de variabilidade explicada dos dados a cada componente principal, organizados em ordem decrescente de contribuição.

Ao analisar a Tabela 4.2, pode-se verificar que a variabilidade está bastante distribuída entre as várias medidas. Assim, não há medidas que se destacam com altas porcentagens que dominariam a variabilidade. Portanto, são necessárias várias medidas para explicar

a variabilidade dos dados.

Tabela 4.2 – Pesos do PCA relacionados aos neurônios reais na base de dados NeuroMorpho (ver projeções na Figura 4.5): o percentual dos pesos das 20 medidas nos componentes principais (CP1 e CP2) são mostrados. Quanto maior o valor percentual, maior é a contribuição da medida na variabilidade do componente principal no qual se encontra.

Medidas	CP1	CP2
Soma Surface	0.026	0.077
Number of Stems	0.023	0.037
Number of Bifurcations	0.066	0.047
Number of Branches	0.067	0.046
Width	0.073	0.043
Height	0.074	0.029
Depth	0.066	0.051
Diameter	0.022	0.067
Length	0.089	0.026
Surface	0.060	0.062
Volume	0.024	0.043
Euclidean Distance	0.083	0.035
Path Distance	0.069	0.016
Branch Order	0.070	0.061
Contraction	0.028	0.035
Fragmentation	0.052	0.034
Partition Asymmetry	0.035	0.058
Rall's Ratio	0.019	0.050
Bifurcation Angle Local	0.028	0.090
Bifurcation Angle Remote	0.027	0.093

4.2.3 Distribuição de categorias

A análise de variável canônica é um método apropriado para visualizar e investigar a distribuição de categorias na base de dados NeuroMorpho. Na Figura 4.7, pode-se visualizar as categorias agrupadas por: (a) tipo de célula, (b) regiões cerebrais e (c) espécies. Foram utilizados os mesmos 15 tipos de células descritos na Seção 4.2.2. Como esperado, a Análise Canônica revelou uma melhor separação entre os grupos considerados.

Na Figura 4.7(a), o tipo Uniglomerular Projection Neuron (círculos cianos) permaneceu compacto numa região específica e alguns Motoneuron (estrelas púrpuras) podem ser encontrados na parte esquerda do gráfico, se espalhando do meio para baixo. Tanto

Figura 4.7 – Visualização das categorias utilizando a análise de variáveis canônicas, com os neurônios destacados por: (a) tipo de célula, (b) regiões cerebrais e (c) espécies. Há 39 tipos de células (somente 15 mostradas aqui), 15 regiões e 11 espécies.

no PCA quanto na Análise Canônica, as células não reportadas (NotReported, triângulos vermelhos apontando para baixo) se sobrepuseram a várias categorias, mas nesta última, pode-se observar um núcleo denso bem definido. Também similar ao PCA, os tipos Granule (cruzes pretas), Basket (círculos amarelos), Bitufted (círculos verdes sólidos), Somatostatin (quadrados cianos) e Stellate (triângulos pretos apontando para baixo) estão agrupados na mesma região.

Na Figura 4.7(b), pode-se observar uma boa separação de células neuronais de acordo com sua respectiva região cerebral. Cercal Sensory System (quadrados cianos), Olfactory Bulb (círculos verdes) e Brainstem (quadrados azuis) formaram grupos bem separados. Algumas regiões se separaram em dois subgrupos, particularmente as células Olfactory Bulb (círculos verdes sólidos), Protocerebrum (cruzes azuis) and Hippocampus (estrelas amarelas). Basal Forebrain (triângulos verdes apontando para baixo), Retina (triângulos vermelhos apontando para cima) e Hippocampus (círculos amarelos) se sobrepõem e formam um grande agrupamento.

A projeção que melhor permitiu distinguir os grupos foi a de espécies (Figura 4.7(c)), na qual pode-se observar que células da mesma espécie animal tendem a estarem agrupadas. Podem-se observar grupos divididos em subgrupos para drosophila (triângulos azuis apontando para a direita) e rato (cruzes cianos). Os camundongos (quadrados verdes) estão espalhados entre o agrupamento principal e outras regiões.

4.2.4 Análise de densidade no hiperespaço

Para analisar os dados diretamente no espaço de características 20D, foi utilizada a função de densidade radial, definida na Seção 3.4. Na Figura 4.8, são mostradas as curvas $f(R)$ para quatro tipos de células (Purkinje, Stellate, Martinotti e Lateral Horn) e suas respectivas localizações no PCA. Estes tipos de células foram selecionados para verificar a coerência entre as densidades no espaço 20D e as respectivas projeções 2D. As curvas de densidade radial dos neurônios dentro de seus respectivos grupos tendem a ser similares, definindo agrupamentos no espaço 20D.

É possível observar a presença de curvas *outliers* (observações numericamente distantes do resto dos dados) na Figura 4.8, em (b) e (c). No primeiro caso, pode-se facilmente identificar o ponto *outlier* correspondente na projeção do espaço 2D. Este não é o caso para as curvas *outliers* observadas na Figura 4.8(c), na qual não é possível identificar

diretamente os pontos correspondentes. Uma outra observação é que os neurônios Stellate são uma exceção entre os quatro grupos, no sentido de que suas células estão agrupadas no espaço 20D, mas exibem clusters separados no 2D.

Figura 4.8 – Função de densidade radial para as células: (a) Purkinje, (b) Stellate, (c) Martinotti e (d) Lateral Horn. Pode-se notar que células do mesmo tipo tendem a possuir curvas radiais similares, o que significa que estão localizadas próximas no espaço de características multi-dimensional 20D. Este comportamento também pode ser observado quando a dimensão é reduzida utilizando-se PCA (e), no qual células do mesmo tipo tendem a estar próximas.

4.3 Considerações finais

Diversas propriedades funcionais e de conexões do sistema nervoso são fortemente determinadas ou afetadas pela morfologia das células individuais envolvidas. Considerando que recentemente milhares de neurônios segmentados passaram a estar disponíveis no banco de dados público NeuroMorpho, é agora possível investigar propriedades morfológicas gerais de células neuronais. Mais especificamente, foram analisados todos os neurônios disponíveis no NeuroMorpho (5673 células catalogadas no momento em que este trabalho foi realizado). Foi utilizada a abordagem teórica de McGhee (morfoespaço) (53) para formalizar a análise realizada. Vinte medidas prontamente disponíveis no site do NeuroMorpho foram utilizadas para descrever o espaço morfológico dos neurônios. Para a visualização do morfoespaço, foram aplicados PCA e Análise Canônica nas 20 medidas, resultando nas projeções 2D. Sete das medidas originais foram usadas para comparar as células reais com as artificiais, geradas pelo modelo de referência proposto. Isto permitiu a comparação da região de neurônios geometricamente possíveis com os neurônios que realmente existem na natureza. Os resultados mostram os neurônios reais confinados numa única região do morfoespaço e uma grande área com células neuronais geometrica-

mente possíveis, mas que não são encontradas na natureza. Os neurônios desta região se caracterizam por um número significativamente maior de ramos.

Em relação às medidas obtidas do NeuroMorpho, foi descoberto que algumas delas estão fortemente correlacionadas. Em particular, medidas que envolvem comprimentos, como profundidade $\times$ largura e distância euclidiana $\times$ distância do caminho (*path distance*) possuem correlação de Pearson acima de 0,75. Todas estas correlações foram eliminadas utilizando PCA, que também permitiu reduzir a dimensionalidade dos dados. Descobriu-se também que as duas componentes principais dependem fortemente de quase todas as 20 medidas consideradas. Mesmo assim, os dois eixos principais explicam quase 50% da variação total no espaço de medidas original. Um resultado particularmente interessante é que, com poucas exceções, os neurônios tendem a se agrupar numa projeção PCA quando considerados por tipo, região cerebral e espécie. Estes agrupamentos foram melhorados com a utilização da Análise Canônica. Também foi verificado usando funções radiais que grupos no espaço original 20D tendem a se manter separados nas respectivas projeções 2D. Este trabalho foi publicado na revista *Frontiers in Computational Neuroscience* (60).

5 Distribuição espacial das densidades de expressão gênica no cérebro

Como similaridades entre as distribuições espaciais de densidade de expressão gênica podem indicar que são funcionalmente correlacionadas, a identificação destas similaridades é útil na sugestão de direções de investigação para descobrir interações entre expressões gênicas e funções correlacionadas. Neste Capítulo, é descrita uma metodologia baseada em diagramas de Voronoi para automaticamente analisar e procurar em imagens por possíveis relações de padrões de distribuição locais de densidades espaciais entre expressões gênicas. Foram utilizadas para testes algumas imagens de expressão gênica de cérebros de camundongo da base de dados pública Allen Mouse Brain Atlas (AMBA). Esta metodologia forneceu medidas capazes de caracterizar a similaridade de distribuição de densidades entre expressões gênicas e permitiu a visualização dos resultados usando redes e PCA.

5.1 Resultados e discussão

O parâmetro de limiar foi configurado como $\tau = 0,5$, de um intervalo entre 0 e 1. A partir da variação abrangente e sistemática de τ , foi constatado que o algoritmo é robusto para pequenas variações neste parâmetro. Os valores extremos não são úteis, pois valores muito baixos geram muitos pontos (a maior parte da imagem), enquanto que valores muito altos geram pouquíssimos pontos. Além disso, as imagens binarizadas de expressão gênica obtidas com $\tau = 0,5$ são similares às imagens segmentadas disponíveis na base de dados AMBA.

Os resultados da análise dos dados utilizando a metodologia VLDA foram normalizados de duas formas. Uma das normalizações foi realizada considerando cada rede

separadamente e a outra considerou todas as redes juntas. Essas normalizações consistiram em subtrair os dados considerados pelo seu menor valor e dividi-los pela diferença entre o maior e o menor valor. Assim, os dados assumiram valores entre zero e um (0 e 1). No primeiro caso, foi possível ressaltar a semelhança de variação dentro de cada rede e visualizar melhor as diferenças entre expressões gênicas de uma mesma região. No segundo caso, foi possível comparar redes (regiões cerebrais).

Considerando cada região individualmente, é possível ver na Figura 5.1, que as expressões gênicas em RHP possuem padrões de distribuição bastante semelhantes (ligações espessas), com exceção de *Dach1*, que apresenta ligações finas com todas as outras expressões gênicas. As outras regiões não possuem esses padrões de conexão tão uniformes, apresentando mais variações na espessura. Mas geralmente, as redes apresentaram uma ou duas expressões gênicas que as distinguiram das demais. Pode-se notar que *Wbscr17* possui o padrão de expressão gênica mais diferente nas redes do Cérebro e sAMY. Em CTXsp, *Wbscr17* e *Dach1* possuem as expressões gênicas mais distintas. *Dach1* também se destaca em PAL e RHP. Em CP, *Nptx1* possui a expressão gênica com as mais finas conexões. Os PCAs destes dados podem ser vistos na Figura 5.2, nos quais pode-se observar uma correspondência entre os pontos mais afastados dos demais e os nós que possuem as conexões mais finas nas redes.

Figura 5.1 – Cada rede foi individualmente normalizada para melhor visualizar peculiaridades dentro de cada região. Conexões mais espessas representam maior similaridade entre os padrões de expressão gênica. Por exemplo, é possível visualizar na rede da região RHP que *Dach1* possui um padrão de expressão gênica diferente dos demais (conexões finas). Por outro lado, as outras expressões gênicas em RHP são bem similares (conexões espessas).

Figura 5.2 – PCA dos dados normalizados por região, correspondentes às redes na Figura 5.1. O PCA visualmente enfatiza, pelas distâncias e proximidades entre os pontos, a similaridade entre padrões de expressão gênica em termos das medidas obtidas pelo VLDA. Assim, é possível identificar agrupamentos de padrões de expressão mais similares em cada região.

Comparando entre regiões, pode-se visualizar na Figura 5.3 que RHP e CP possuem o padrão de conexões mais distinto. Enquanto as outras regiões apresentam apenas conexões espessas entre as expressões gênicas, RHP apresenta *Dach1* com conexões fracas com todas as outras expressões gênicas. E CP exibe outro padrão diferente, no qual *Nptx1* apresenta algumas conexões finas, enquanto que *Aars* e *Wbscr17* têm conexões de espessura intermediária.

Quando se compara a Figura 5.1 com a Figura 5.3, pode-se notar que apenas RHP e CP mantêm distribuições de espessuras de arestas semelhantes. Isto se deve ao fato de que estas duas regiões possuem a maior amplitude de valores de pesos das conexões. As demais regiões também apresentam variações em seus valores, mas a diferença é pequena e perto do valor mais alto, resultando em fortes ligações quando são comparadas as regiões.

Estas pequenas diferenças somente podem ser observadas quando consideramos cada rede individualmente. Na Figura 5.4, pode-se ver o PCA das regiões exibidas na Figura 5.3, normalizados considerando todos os dados. Os dois pontos longe dos demais correspondem às redes de CP e RHP, que já tinham sido identificadas como distintas das outras ao visualizar as redes. Estes pontos também estão distantes entre si, o que reflete a diferença entre estas redes.

Figura 5.3 – As redes foram normalizadas considerando todos os dados, permitindo comparar o padrão das conexões entre as regiões cerebrais. É possível observar que CP e RHP possuem redes distintas das outras, além de também serem diferentes entre si.

Assim, os resultados indicaram altas similaridades nos padrões de distribuições espaciais de densidades entre as seguintes expressões gênicas em cada região:

- Cérebro: *Gria4*, *Hpca*, *Man1a*, *Nptx1* e *Osbpl8*;
- CP: *Aars* e *Wbscr17*, assim como *Dach1*, *Hpca* e *Osbpl8*;
- CTXsp: *Aars*, *Gria4*, *Hpca*, *Man1a*, *Nptx1* e *Osbpl8*;
- PAL: *Aars*, *Gria4*, *Hpca*, *Man1a*, *Osbpl8* e *Wbscr17*;
- RHP: *Aars*, *Gria4*, *Hpca*, *Man1a*, *Nptx1* e *Osbpl8*;
- sAMY: *Gria4*, *Hpca*, *Man1a*, *Nptx1* e *Osbpl8*;

Figura 5.4 – PCA das redes normalizadas considerando todos os dados. O PCA provê uma visão alternativa para a comparação dos padrões de conexão das redes. É possível visualizar que CP e RHP estão distantes, como esperado pela análise visual das redes na Figura 5.3. Além disso, o PCA trouxe uma nova informação ao mostrar dois grupos: PAL e CTXsp; e sAMY e Cérebro. Estes grupos não são diferentes o suficiente para serem detectados nas redes, mas o PCA captou as diferenças sutis. Aparentemente, são subgrupos, já que estes pares estão próximos uns dos outros.

Entre as regiões, as redes do Cérebro, CTXsp, PAL e sAMY são as mais semelhantes. No PCA, é possível observar que estas regiões localizam-se próximas, mas podemos distinguir dois pares: PAL e CTXsp; e sAMY e Cérebro. Aparentemente, são subgrupos, já que estes pares estão próximos uns dos outros. Como o Cérebro inclui todas as regiões, vale ressaltar que algumas regiões apresentam proporções de densidade de genes semelhantes ao cérebro inteiro.

Utilizando o software Osprey (61), foram pesquisadas as funções dos genes relacionados às expressões gênicas usadas neste trabalho. Como os genes foram escolhidos aleatoriamente, a função da maior parte deles ainda é desconhecida. O software Osprey não encontrou nenhuma interação documentada entre esses genes, mas encontrou as funções de três genes (ver Tabela 5.1): *Gria4*, *Hpca* (encontrado com o nome alternativo de *CACNA1A*) e *Man1a*. É possível visualizar que a função “crescimento e/ou manutenção celular” é realizada pelos três genes. *Hpca* e *Gria4* também possuem a função de “transporte” em comum, enquanto que *Hpca* e *Man1a* estão ambos envolvidos na função de “metabolismo”. Portanto, estes três genes possuem algumas funções similares e, além disso, encontram-se bastante próximos nos PCAs na Figura 5.2 (somente na região CP, *Gria4* encontra-se um pouco mais afastado dos demais), sugerindo que há entre estes três genes um alto nível de correlação.

Tabela 5.1 – Funções conhecidas dos genes relacionados às expressões gênicas analisadas, de acordo com o software Osprey (61). Pode-se verificar que há funções similares entre os genes.

Função	<i>Gria4</i>	<i>Hpca</i>	<i>Man1a</i>
Transporte	X	X	
Crescimento e/ou manutenção celular	X	X	X
Metabolismo		X	X
Transcrição		X	
Metabolismo de carboidrato			X
Biosíntese de proteína			X

5.2 Considerações finais

Como o número de bancos de dados de imagens de grande porte está continuamente aumentando na web, é cada vez mais importante o desenvolvimento de metodologias automatizadas eficientes e robustas para lidar com o desafio de analisar dados em tempo viável e com precisão. Neste contexto, a análise de densidade local por Voronoi é proposta como uma metodologia computacional para realizar a tarefa de comparar a distribuição espacial da densidade de padrões de expressão entre os genes. Neste trabalho, esta metodologia foi usada para estudar alguns genes em regiões do cérebro do rato para demonstrar as possibilidades de encontrar relações espaciais entre os genes, que podem ser úteis para orientar investigações genéticas. Outra vantagem desta abordagem é a possibilidade de comparar objetivamente estruturas complexas nas imagens que visualmente seriam difíceis de se analisar.

Os resultados fornecem uma maneira quantitativa de caracterizar a semelhança entre a distribuição espacial da densidade de genes. Estes resultados podem ser avaliados em uma variedade de métodos de classificação e também podem ser analisados visualmente por redes e PCAs. Desta forma, é possível distinguir os genes com distribuição espacial mais diferentes e os mais similares, assim como comparar regiões. Estas informações são úteis para orientar estudos de interação genética, uma vez que padrões espaciais de expressão gênica semelhantes possivelmente indicam funções correlacionadas. Por exemplo, no Cérebro, os genes *Gria4*, *Hpca*, *Man1a*, *Nptx1* e *Osbp18* estão bastante próximos uns dos outros (Figura 5.2), indicando uma possível relação. Comparando as regiões, pode-se ver que PAL, CTXsp, sAMY e Cérebro compartilham similaridades entre suas redes (Figura 5.4), o que também aponta uma possível direção de investigação biológica. Com a ajuda do software Osprey, foi possível encontrar funções conhecidas para *Gria4*, *Hpca* e *Man1a*. Constatou-se que há algumas funções similares entre esses genes. Além disso,

seus padrões de expressão gênica foram apontados como possivelmente similares na maioria das regiões anatômicas cerebrais, como pode ser visto nos PCAs da Figura 5.2). Este trabalho foi publicado na revista *Journal of Neuroscience Methods* (62).

6 Influência da expressão gênica na morfologia, topologia e dinâmica

As características morfológicas dos neurônios (forma) definem a base estrutural que possibilita o estabelecimento de conexões e a formação da topologia neuronal. Por sua vez, esta topologia limita e determina como os processos dinâmicos se desenvolvem. Um importante componente na determinação das características morfológicas é a expressão gênica, que influencia indiretamente a topologia e a dinâmica. Assim, uma importante questão a ser investigada é como a expressão gênica determina as características morfológicas, topológicas e funcionais (dinâmica) durante o desenvolvimento de neurônios e da própria rede neuronal. Para isso, foi proposto um arcabouço para identificar quais características (medidas) melhor refletem a influência da expressão gênica no desenvolvimento neuronal. Essa influência da expressão gênica é ilustrada na Figura 6.1. Foram utilizados neurônios gerados pelo modelo descrito na Seção 2.2, modificado para incluir a influência da intensidade de expressão gênica na morfologia. A vantagem na utilização de um modelo é o controle obtido em todo o processo, o que facilita a análise das características envolvidas desde o neurônio individual à dinâmica ocorrendo sobre a rede formada. Assim, é possível atingir um nível de compreensão que seria mais difícil de obter em dados reais, devido à dificuldade de aquisição desses dados.

6.1 Crescimento neuronal sob influência da expressão gênica

O ambiente neuronal foi representado como uma rede, com os nós dispostos em forma de grade e espaçamento Δ entre vizinhos na horizontal e vertical. Cada nó, que corres-

Figura 6.1 – A expressão gênica influencia diretamente a morfologia dos neurônios (a), que determinam a conectividade da rede neuronal (b), o que afeta o desenvolvimento de dinâmicas na rede (c). Desta forma, é possível verificar que a topologia (d) e a dinâmica (e) são indiretamente influenciados pela expressão gênica. A cada passo, a influência da expressão gênica torna-se mais distante e tende naturalmente a ficar cada vez mais difícil de se identificar.

ponde a um soma neuronal, possui um valor de intensidade η , representando a concentração local de expressão gênica. Esses nós também são considerados como regiões onde os neurônios se estabeleceram e cresceram numa orientação planar aleatória, de acordo com a influência da expressão gênica naquela localização. Foram utilizadas duas configurações de grade, uma listrada (Figura 6.2) e outra com um padrão circular (Figura 6.3), ambas de tamanho 100×100 . Na listrada, os valores de intensidade foram uniformemente distribuídos dentro do intervalo $[0, 1; 1, 0]$ ao longo das colunas (como um gradiente), e todos os nós na mesma coluna receberam o mesmo valor. Na circular, os valores foram estabelecidos como proporcionais à distância Euclidiana de cada posição ao nó no centro da grade.

A intensidade de cada posição na grade foi utilizada para influenciar o crescimento de uma estrutura ramificada (neurônio artificial) utilizando o modelo de referência da Seção 2.2. Essas influências foram implementadas como dependências lineares entre alguns parâmetros do modelo e a expressão gênica local, definidas como as funções:

- Ângulo de bifurcação: $F_1(\theta) = \eta\theta$;
- Ângulo de torsão: $F_2(\phi) = \eta\phi$;
- Número de segmentos: $F_3(N_s) = (1 + \eta)N_s/2$;
- Probabilidade de crescimento: $F_4(\rho_g) = (1 + \eta)\rho_g/2$;
- Probabilidade de remover um ramo: $F_5(\rho_r) = (1 - \eta)\rho_r$.

Os parâmetros do modelo foram configurados como $\theta \in [0, \pi]$, $\phi \in [-\pi, \pi]$, $\sigma_\theta^2 \in [0, \pi/6]$, $\sigma_\phi^2 \in [0, \pi/9]$, $\rho_g \in [0, 1]$, $\rho_r \in [0, 1]$, e o número de segmentos $N_s = [8, 800]$.

Foram assumidos para os demais parâmetros: $\ell_{max} = 100$, $\bar{\theta} = 2\pi/3$, e $\bar{\phi} = 2\pi/9$. Além disso, as estruturas geradas possuem pelo menos uma bifurcação.

Nas Figuras 6.2 e 6.3, são exibidos neurônios gerados considerando a influência da expressão gênica de acordo com as funções apresentadas, respectivamente nas grades de padrões listrado e circular. Quanto maior o valor de intensidade da expressão gênica, mais claro é o nível de cinza. Para facilitar a visualização, os neurônios foram exibidos com uma caixa de contorno tridimensional. Desta forma, pode-se ter uma melhor percepção de suas dimensões. Assim, é possível visualizar variações morfológicas, como o aumento no tamanho de altura e profundidade e diferenças no padrão de ramificação.

6.2 Construção da rede neuronal

Os neurônios gerados pela metodologia na Seção 6.1 foram conectados entre si para criar uma rede neuronal. Cada neurônio teve seu soma posicionado num nó da rede. Os nós foram espacialmente distribuídos formando uma grade, com distância Δ entre os nós. Foi considerado que dois neurônios i e j possuem uma conexão quando ao menos uma bifurcação ou uma extremidade de um deles está próximo do soma do outro, numa distância menor que $\delta = \Delta/5$. Assim, o único parâmetro livre é o espaçamento da grade Δ . Para pequenos valores de Δ , os somas dos neurônios encontram-se bastante próximos, o que resulta num maior número de conexões, correspondendo a uma rede densa. Conforme o valor de Δ é incrementado, a probabilidade de dois neurônios se tocarem é reduzida e a rede se torna mais esparsa. Na Figura 6.4, são mostradas as redes formadas considerando o padrão de grade listrado para $\Delta = 5$ (a) e $\Delta = 10$ (b).

Figura 6.2 – Influência da concentração de expressão gênica no crescimento neuronal em uma grade de padrão listrado. Os valores de intensidade variam nas colunas de 0,1 a 1,0, igualmente espaçados, e são proporcionais aos níveis de cinza mostrados. Os neurônios artificiais na mesma coluna foram gerados considerando a mesma influência de expressão gênica (mesmo valor de concentração). Esta grade 10×10 exibe neurônios amostrados de uma grade 100×100 .

Figura 6.3 – Neurônios artificiais que cresceram numa grade de padrão circular de distribuição de valores de intensidade de expressão gênica. Esta grade 10×10 exibe neurônios amostrados de uma grade 100×100 .

Figura 6.4 – Redes formadas considerando o padrão de grade listrado para $\Delta = 5$ (a) e $\Delta = 10$ (b). São exibidas apenas as conexões entre os nós. Os neurônios não são mostrados. Como esperado, pode-se visualizar que a rede (a) é mais densa que a rede (b).

6.3 Resultados e discussão

6.3.1 Influência da expressão gênica na morfologia

A morfologia de um neurônio é determinada pelo modelo de crescimento (Seção 2.2), que é influenciada pela intensidade de expressão gênica seguindo as funções descritas na Seção 6.1. Isto significa que a morfologia neuronal é afetada pela intensidade de expressão gênica, juntamente com outras variáveis aleatórias do modelo. Essa influência pode ser detectada pela análise da variação das medidas morfológicas. As medidas morfológicas utilizadas foram (42, 60): número de bifurcações (M1), número de extremidades (M2), número de ramos (M3), largura (M4), profundidade (M5), altura (M6), alongamento (M7), comprimento médio de segmento (M8), comprimento total (M9), comprimento médio de ramo (M10), e ângulo médio de bifurcação (M11).

Cada função pode afetar uma ou mais características morfológicas medidas. Por exemplo, a medida M11 (ângulo médio de bifurcação) é evidentemente afetada somente pela

função $F_1(\theta)$ (função de influência do ângulo de bifurcação), enquanto que M4 (largura) é influenciada por todas as cinco funções. Pode-se visualizar na Tabela 6.1 quais funções de expressão gênica influenciam cada medida. As medidas M1 (número de bifurcações), M2 (número de extremidades) e M3 (número de ramos) estão diretamente associadas ao número de segmentos, regulado pela função $F_3(N_s)$. Essas medidas também são influenciadas por $F_5(\rho_r)$ (probabilidade de remover um ramo), que reduz o número de bifurcações e possibilita ramos mais longos. Nas medidas M4 (comprimento), M5 (profundidade), M6 (altura) e M7 (alongamento), as funções relacionadas com os tamanhos de segmento e ramo são influências evidentes ($F_3(N_s)$ - número de segmentos, $F_4(\rho_g)$ - probabilidade de crescimento, e $F_5(\rho_r)$ - probabilidade de remover um ramo). O efeito das funções $F_1(\theta)$ e $F_2(\phi)$ (ângulos de bifurcação e torsão) são relacionados ao fato de que os menores ângulos de bifurcação geram neurônios mais longos e finos, o que afeta o comprimento e a largura, enquanto que os ângulos de torsão criam o aspecto 3D por permitir maiores valores de profundidade quando aumenta. As medidas M8 (tamanho médio de segmento) e M11 (ângulo de bifurcação médio) possuem apenas a influência de uma única função, respectivamente $F_4(\rho_g)$ (probabilidade de crescimento) e $F_1(\theta)$ (ângulo de bifurcação). As outras duas medidas, M9 (comprimento total) e M10 (comprimento de ramo médio), são ambas afetadas pela função $F_4(\rho_g)$ (probabilidade de crescimento). Adicionalmente, M9 é também influenciada pela função $F_3(N_s)$ (número de segmentos), enquanto que M10 também é afetada por $F_5(\rho_r)$ (probabilidade de remover um ramo).

Medidas	Funções
M1) Número de bifurcações	$F_3(N_s), F_5(\rho_r)$
M2) Número de extremidades	$F_3(N_s), F_5(\rho_r)$
M3) Número de ramos	$F_3(N_s), F_5(\rho_r)$
M4) Largura	$F_1(\theta), F_2(\phi), F_3(N_s), F_4(\rho_g), F_5(\rho_r)$
M5) Profundidade	$F_1(\theta), F_2(\phi), F_3(N_s), F_4(\rho_g), F_5(\rho_r)$
M6) Altura	$F_1(\theta), F_2(\phi), F_3(N_s), F_4(\rho_g), F_5(\rho_r)$
M7) Alongamento	$F_1(\theta), F_2(\phi), F_3(N_s), F_4(\rho_g), F_5(\rho_r)$
M8) Comprimento médio de segmento	$F_4(\rho_g)$
M9) Comprimento total	$F_3(N_s), F_4(\rho_g)$
M10) Comprimento médio de ramo	$F_4(\rho_g), F_5(\rho_r)$
M11) Ângulo médio de bifurcação	$F_1(\theta)$

Tabela 6.1 – Medidas morfológicas e as respectivas funções de influência da expressão gênica (ver as funções na Seção 6.1).

A partir dos gráficos de dispersão das medidas morfológicas, podem-se organizar essas medidas em quatro grupos de padrões de dispersão similares: GM1 = {M1, M2, M3}, GM2 = {M4, M5, M6, M7}, GM3 = {M8, M9, M10}, e GM4 = {M11}. Na Figura 6.5, podem ser visualizados gráficos de dispersão representativos para cada grupo, nos dois padrões de grade, listrado e circular.

Para identificar quais medidas melhor refletem a influência da expressão gênica na

Figura 6.5 – Gráficos de dispersão das medidas morfológicas para o padrão de grade listrado (a-d), e para o padrão circular (e-h). Os pontos pretos em destaque indicam os elementos nos limites inferior e superior das distribuições.

morfologia neuronal, foi primeiro utilizada a análise de entropia condicional. Como pode ser visto na Figura 6.6, os valores da entropia condicional nos padrões listrado (ML) e circular (MC) são qualitativamente consistentes. Baixos valores indicam que as medias apresentam informações importantes. O ângulo médio de bifurcação (M11) apresentou o menor valor. De fato, esta medida exibe a distribuição mais significativa (ver gráficos de dispersão (d) e (h) na Figura 6.5). Os outros valores pequenos pertencem às medidas no grupo GM3.

Figura 6.6 – Entropia condicional das medidas morfológicas considerando os padrões de grade listrado (ML) e circular (MC).

As demais medidas apresentaram valores altos de entropia condicional. Ao analisar seus gráficos de dispersão, é possível visualizar que as medidas do grupo GM2 (Figura 6.5(b) e (f)) realmente não apresentaram informação relevante para a caracterização

da influência da expressão gênica. Por outro lado o grupo de medidas GM1 (Figura 6.5(a) e (e)) possui uma característica interessante que não foi capturada pela entropia condicional. Seus gráficos de dispersão apresentam uma inclinação bem definida no limite superior de suas distribuições. A análise do modelo e das funções de influência da expressão gênica revelaram que este comportamento é consistente e reflete a variação da intensidade de expressão gênica que neste caso atua limitando o valor máximo dessas medidas. Assim, para capturar este comportamento, foi proposta a utilização da análise de inclinação (Seção 3.10). Os pontos detectados nos limites são mostrados em preto na Figura 6.5. O resultado é mostrado na Figura 6.7. Como esperado, as medidas no grupo GM1 apresentaram os maiores valores para as inclinações dos limites superiores.

Figura 6.7 – Medida de inclinação dos limites inferior e superior dos gráficos de dispersão das medidas morfológicas para os padrões de grade listrado (a) e circular (b).

Portanto, as medidas morfológicas que mais evidentemente refletem a influência da expressão gênica é o ângulo médio de bifurcação (M11), pois apresenta o menor valor de entropia condicional e o maior fator de inclinação dos limites inferior e superior. Isto também pode ser confirmado pela análise dos gráficos de dispersão (Figura 6.5(d) e (h)), nos quais é possível verificar o crescimento de seus valores medidos em correlação com o aumento da intensidade de expressão gênica. As outras medidas que mostraram resultados relevantes foram as do grupo GM3 (em ordem crescente de entropia condicional): comprimento médio de segmento (M8), comprimento total (M9), comprimento médio de ramo (M10). Em seguida, pode-se considerar o grupo GM1: número de bifurcações (M1), número de extremidades (M2), e número de ramos (M3). Embora essas medidas tenham apresentado valores de entropia condicional altos, seus valores de fator de inclinação para o limite superior são bastante altos, refletindo a influência da expressão gênica na determi-

nação desse limite. As últimas quatro medidas (grupo GM2), largura (M4), profundidade (M5), altura (M6), e alongamento (M7) podem ser desconsideradas, pois apresentaram altos valores de entropia condicional e baixos valores de fator de inclinação. Além disso, como pode ser observado nos gráficos de dispersão (Figura 6.5(b) e (f)), as distribuições de seus pontos não variam significativamente em função da expressão gênica. Assim, pode-se constatar a eficácia em quantificar e determinar a influência da expressão gênica nas medidas morfológicas a partir da utilização e análise combinada dos resultados da entropia condicional e do fator de inclinação.

6.3.2 Influência da expressão gênica na topologia

Para caracterizar a topologia, foram obtidas medidas topológicas das redes formadas pelos neurônios artificiais utilizando-se o NetworkAnalyzer (63), um *plugin* do Cytoscape (64). Para cada nó n , foram calculadas as medidas topológicas:

1. *Neighborhood connectivity*: a conectividade de um nó é o seu número de vizinhos. Assim, a conectividade de vizinhança de um nó n é definida pela média da conectividade de todos os vizinhos de n .
2. *Stress*: número de menores caminhos passando por um nó.
3. *Degree*: número de arestas conectadas ao nó n (grau do nó).
4. *Topological coefficient*: $T_n = \langle J(n, m) \rangle / k_n$, em que k_n é o número de vizinhos. $J(n, m)$ é definido para todos os nós m que compartilham pelo menos um vizinho com n . O valor de $J(n, m)$ é o número de vizinhos compartilhados entre os nós n e m , mais um se há uma conexão direta entre eles. O coeficiente topológico é uma medida que quantifica a tendência de um nó em compartilhar vizinhos com outros nós. Nós com menos de 2 vizinhos recebem valor zero.
5. *Betweenness centrality*: $C_b(n) = \sum_{s \neq n \neq t} (\sigma_{st}(n) / \sigma_{st})$, em que s e t são nós da rede diferentes de n , σ_{st} é o número de menores caminhos entre s e t , e $\sigma_{st}(n)$ é o número de menores caminhos entre s e t que passam por n . Valor normalizado entre 0 e 1. Nesta medida, nós que conectam comunidades (sub-redes densas) apresentam valores maiores do que nós dentro de uma comunidade.
6. *Radiality*: esta medida é calculada a partir da subtração entre Average Shortest Path Length de um nó n e o diâmetro do componente conexo, mais 1. Em seguida,

este valor é dividido pelo diâmetro do componente conexo, gerando um valor entre 0 e 1.

7. *Eccentricity*: o maior valor de menor caminho entre o nó n e outro nó na rede. Se n for um nó isolado, o valor deste atributo é zero.
8. *Closeness centrality*: $C_c(n) = 1/\langle L(n, m) \rangle$, em que $L(n, m)$ é o comprimento do menor caminho entre dois nós n e m . O seu valor apresenta-se entre 0 e 1, sendo que nós isolados possuem valor 0. Esta é uma medida de quão rápido a informação se espalha a partir de um dado nó para os outros nós alcançáveis na rede.
9. *Average Shortest Path Length*: média dos comprimentos de menor caminho entre o nó n e todos os demais. Se n for um nó isolado, o valor deste atributo é zero.
10. *Clustering coefficient*: $C_n = 2w_n/(k_n(k_n - 1))$, em que k_n é o número de vizinhos de n e w_n é o número de pares conectados entre todos os vizinhos de n . Esta medida é a razão entre o número de arestas entre os vizinhos de n e o número máximo de arestas que poderiam ser formadas. Também pode-se entender esta medida como a razão entre o número de triângulos que passam pelo nó n e o número máximo de triângulos que poderiam passar por n . Os valores estão entre 0 e 1.

A topologia, além de ser afetada pela morfologia individual de cada neurônio (nó), também é influenciada pela distância entre os nós. Foram testadas duas diferentes distâncias entre os nós: 5 e 10 unidades. Como é possível visualizar na Figura 6.8, as correlações de Pearson são qualitativamente similares em relação à variação da distância, apresentando diferenças na magnitude dos valores, como se pode constatar ao comparar TL05 a TL10, e TC05 a TC10. Nessas siglas, T refere-se à medida topológica, L ou C significam que foi utilizado o padrão de grade listrado ou circular, e os números 05 e 10 indicam a distância entre os nós da rede. Como neste caso somente interessa a magnitude da correlação, foram utilizados os valores absolutos. Enquanto que nas grades listradas a maior correlação é na medida de conectividade de vizinhança, na grade circular as maiores correlações estão nas medidas de *Radiality*, *Eccentricity*, *Closeness centrality* e *Average Shortest Path Length*. Isso demonstra que a expressão gênica atua de maneira diferente sobre a topologia, dependendo de sua organização espacial. Os gráficos de dispersão dessas quatro medidas nas duas grades são mostradas na Figura 6.9. Ao examinar esses gráficos de dispersão, é possível verificar que essas medidas são de fato significativas e refletem a influência indireta da expressão gênica na formação topológica. A distribuição dos dados nos gráficos de dispersão pode ser considerada como significativa quando há uma tendência dos dados aumentar ou reduzir seu valor ao longo de um dos eixos em função do outro eixo. Ou seja, há uma correlação entre essas duas variáveis, representadas por cada eixo.

Figura 6.8 – Correlação de Pearson das medidas topológicas nas quatro configurações de grade.

Figura 6.9 – Gráficos de dispersão das medidas topológicas com os maiores valores de correlação de Pearson. Os primeiros quatro gráficos (a-d) são respectivamente as medidas radialidade, excentricidade, proximidade da centralidade e comprimento médio do menor caminho, considerando a configuração de grade TC05, enquanto que os outros quatro gráficos (e-h) são as mesmas medidas, mas para TC10.

6.3.3 Influência da expressão gênica na dinâmica

Para avaliar a influência da expressão gênica na dinâmica, foi utilizada a dinâmica integra-e-dispara (descrita na Seção 3.8). As medidas consideradas para a caracterização da dinâmica foram:

1. *Spiking rate*: taxa de disparo.

2. *Maximum interspike interval*: intervalo máximo entre disparos.
3. *Interspike entropy*: entropia dos intervalos entre disparos.

As medidas de dinâmica nas redes neuronais foram analisadas tanto no formato original quanto na escala logarítmica, e os maiores valores de correlação foram considerados. Somente os valores da medida *maximum interspike interval* foram maiores na escala logarítmica, o que indica sua natureza log. Os valores das correlações de Pearson podem ser visualizados na Figura 6.10. Assim como na topologia, na dinâmica também foram utilizados somente os valores absolutos de correlação. Valores altos de correlação foram obtidos na medida *spiking rate* da configuração de grade DL05 e em todas as medidas de DC05, indicando que essas são as medidas de dinâmica de rede que melhor refletem a influência indireta da expressão gênica. Os gráficos de dispersão dessas quatro medidas, que possuem os maiores valores de correlação, são mostrados na Figura 6.11.

Uma importante característica observada é que a medida *spiking rate* em DL05 apresentou valor de correlação bastante alto (0,76), enquanto que a medida *degree* (grau) em TL05 apresentou correlação de quase 0,45 (também significativo). É conhecido que o grau é um fator topológico importante na determinação da taxa de disparo (65). Sendo assim, as medidas conseguiram capturar essa relação, no sentido de que as altas correlações foram observadas nessas duas medidas. É também possível observar uma variação consistente ao comparar as demais configurações de rede. Pode-se observar nas medidas topológicas (Figura 6.8) uma variação da correlação do grau, que diminui para cerca de metade ao dobrar a distância, tanto no padrão de grade listrado, como no circular. Ao comparar qualitativamente com os valores de taxa de disparo nas medidas de dinâmica (Figura 6.10), pode-se observar a mesma variação em função da distância, com os valores de correlação reduzindo para cerca de metade ao dobrar a distância.

6.3.4 Quantificação da influência da expressão de dois genes

Para verificar a possibilidade de quantificar a influência de mais de um gene e diferenciar o efeito de cada gene, foi utilizado o mesmo modelo gerador de neurônios artificiais, com as mesmas funções de influência da expressão gênica. Porém, ao invés de ter apenas um padrão artificial de expressão gênica influenciando as funções, foram utilizadas duas imagens de expressão gênica reais do AMBA. Foi utilizada uma imagem de expressão do gene *Dach1* para a expressão gênica 1 (G1) e do gene *Hpca* para a expressão gênica 2

Figura 6.10 – Correlação de Pearson das três medidas de dinâmica em rede (ver legenda) em cada uma das quatro configurações de grade: DL05, DL10, DC05 and DC10. D é o identificador das medidas de dinâmica, L e C referem-se ao padrão de distribuição da expressão gênica na grade (listrado ou circular), e o número indica a distância entre os nós (em pixels). Os maiores valores de correlação indicam as medidas que melhor representam a influência da expressão gênica na dinâmica ocorrendo na rede. Assim, mostram quais características da dinâmica são as mais afetadas pela influência da expressão gênica.

(G2). Essas imagens são mostradas na Figura 6.12. G1 ficou responsável por influenciar as funções $F_3(N_s)$, $F_4(\rho_g)$ e $F_5(\rho_r)$, ou seja, as funções relacionadas com tamanho e número de ramos. As funções de ângulo, $F_1(\theta)$ e $F_2(\phi)$, ficaram sob a influência de G2. As imagens possuem tamanho 400×400 e foram divididas em quadrados de 4×4 . Foi tirada a média da intensidade de expressão gênica de cada quadrado, gerando uma grade de 100×100 . Foi gerado um neurônio para cada posição da grade, considerando os valores correspondentes de expressão em G1 e G2.

As medidas morfológicas foram analisadas com a entropia condicional, a correlação de Pearson e a análise de inclinação. Na entropia condicional (Figura 6.13), somente a medida M11 (ângulo médio de bifurcação) apresentou diferença significativa, com valor bem mais alto em G1 do que em G2, o que reflete a influência direta de G2. Na correlação de Pearson (Figura 6.14), M11 também se diferenciou significativamente. Também é possível observar que as medidas com os maiores valores de correlação em G1 são M1, M2, M3, M8 e M9 (correlações maiores que 0,2), que são medidas relacionadas ao tamanho e número de ramos, influenciadas por G1. Em G2 é evidente a influência no ângulo médio

Figura 6.11 – Gráficos de dispersão das medidas de dinâmica com os maiores valores de correlação de Pearson: (a) taxa de disparo em DL05, (b) taxa de disparo em DC05, (c) intervalo máximo entre disparos em DC05, e (d) entropia entre disparos em DC05.

Figura 6.12 – Imagens de expressão gênica utilizadas: (a) Dach1; (b) Hpca

de bifurcação (M11), que apresentou correlação aproximadamente igual a 1. As demais medidas em G2 ficaram em torno ou bem abaixo de 0,1, destacando a não influência de G2 nessas medidas.

Na análise de inclinação, a utilização de histograma com caixa fixa por número de

Figura 6.13 – Entropia condicional das medidas morfológicas considerando as duas expressões gênicas, G1 (Dach1) e G2 (Hpca).

Figura 6.14 – Correlação de Pearson das medidas morfológicas considerando as expressões gênicas G1 e G2.

pontos (no caso, 100 pontos em cada caixa) obteve melhores resultados do que com a caixa fixa por faixa de valor. Isso se deve à distribuição irregular dos pontos, na qual as regiões com poucos pontos espalhados acabavam por prejudicar a determinação da interpolação da reta para calcular a inclinação. Os resultados da análise de inclinação podem ser visualizados na Figura 6.15, enquanto que os gráficos de dispersão das medidas podem ser observados na Figura 6.16. Conforme proposto na Seção 6.3.1 as medidas estão distribuídas em quatro grupos e somente uma medida representativa de cada grupo

é mostrado na Figura 6.16. Os gráficos (a-d) mostram a distribuição das medidas em relação à expressão gênica G1, enquanto que os demais gráficos (e-f) exibem a distribuição em relação a G2. Nesses gráficos é evidente que os grupos de medidas $GM1 = \{M1, M2, M3\} = \{\text{n. de bifurcações, n. de extremidades, n. de ramos}\}$ e $GM3 = \{M8, M9, M10\} = \{\text{comprimento médio de segmento, comprimento total, comprimento médio de ramo}\}$ possuem limites superiores bem mais definidos, com pontos menos espalhados e mais alinhados quando plotados com a expressão gênica G1 do que com G2 (comparar (a) com (e) e (c) com (g)). Isso faz sentido, já que G1 afeta as funções de influência das características medidas por GM1 e GM3. Como G2 afeta os ângulos de bifurcação e torção, a medida do grupo $GM4 = \{M11\} = \{\text{ângulo médio de bifurcação}\}$ exibe correlações evidentes na Figura 6.16(h) e uma distribuição dispersa em (d). No caso do grupo $GM2 = \{M4, M5, M6, M7\} = \{\text{comprimento, profundidade, altura, alongamento}\}$, as distribuições não resultam em correlações evidentes entre as medidas e as expressões gênicas. Dessa análise, é possível constatar que as medidas se apresentam de forma mais organizada quando plotadas com a expressão gênica que a influenciou. Assim, a análise de inclinação mostrada na Figura 6.15 reflete essas características, apresentando valores bastante elevados para as medidas dos grupos GM1 e GM3 para o limite superior em G1 e para a medida M11 do grupo GM4 em ambos os limites inferior e superior em G2.

Figura 6.15 – Análise de inclinação das medidas morfológicas para G1 e G2, nos limites superior e inferior das distribuições (ver Figura 6.16)

Figura 6.16 – Gráficos de dispersão das medidas topológicas, com destaque aos pontos que compõem os limites superior e inferior das distribuições. A distribuição das medidas em relação ao G1 são apresentadas na primeira linha (a-d), enquanto que as da segunda linha (e-f) são as medidas em relação ao G2.

6.4 Considerações finais

Neste estudo, foi proposto um arcabouço para mensurar a influência da expressão gênica ao longo da escala biológica, do micro para o macro. Ou seja, a partir da escala molecular da expressão gênica, mensurar sua influência em escala celular na morfologia dos neurônios, avançando à escala topológica de uma rede, e alcançando o nível funcional dos processos dinâmicos sobre essa rede. Neste contexto, deve-se ressaltar que a influência da expressão gênica é direta sobre a morfologia e indireta sobre a topologia e a dinâmica. Essas informações são relevantes na investigação de como a expressão gênica influencia todo o processo, desde o neurônio individual até o funcionamento cerebral, que já é conhecido ser de fundamental importância para o desenvolvimento e manutenção da vida. Este arcabouço fornece uma metodologia sistemática, com um conjunto de ferramentas para tal análise. Com os resultados obtidos, o arcabouço proposto mostrou-se promissor na detecção da influência da expressão gênica na forma e na função.

7 Conclusões

O trabalho desenvolvido teve como objetivo o estudo da influência da expressão gênica no desenvolvimento da forma e da função. Para isso, foram desenvolvidas as atividades de estudo da distribuição espacial das formas neuronais no neuromorfoespaço (Seção 7.1), distribuição espacial de densidades de expressão gênica (Seção 7.2), e propagação de doenças em população de estruturas ramificadas (Seção 7.3). Assim, integrando essas atividades, foi possível desenvolver o foco do trabalho sobre a influência da expressão gênica (Seção 7.4). O sumário das principais contribuições e as referências dos trabalhos publicados são apresentados na Seção 7.5, enquanto que a lista dos trabalhos futuros encontra-se na Seção 7.6.

7.1 Hiperespaço neuromorfológico

No estudo do espaço neuromorfológico, foi proposto um procedimento computacional para o estudo da forma de neurônios utilizando o modelo de espaço morfológico de McGhee (53). Medidas obtidas dos quase 6.000 neurônios reais da base pública NeuroMorpho foram utilizadas para se caracterizar cada neurônio. Em seguida, um algoritmo gerador de estruturas ramificadas (neurônios artificiais) foi utilizado para obter variadas formas, obedecendo aos limites dos parâmetros estatísticos baseados nas medidas desses neurônios reais. A seguir, foi empregado o PCA como técnica de visualização de dados para constatar a restrita distribuição dos neurônios reais no amplo espaço morfológico possível, identificado pelos neurônios artificiais. Os dados originais das 20 medidas também foram analisados com correlação de Pearson, PCA, Análise Canônica e função de densidade radial para descobrir possíveis relações entre as medidas. A morfologia dos neurônios fornece informações potencialmente valiosas, não só sobre a função neuronal, mas também sobre a evolução e a ecologia das espécies, e sobre diferenças funcionais

entre as áreas do cérebro. No entanto, o tamanho atual do banco de dados só permite estudos globais. Questões importantes do tipo “como a morfologia neuronal evoluiu ao longo das espécies” ou “se os neurônios se tornaram mais ou menos ramificados ao longo da escala filogenética” permanecem intratáveis. Os resultados indicam uma tendência de similaridade morfológica entre neurônios de espécies próximas, como macacos e humanos, e ratos e camundongos, mas não é suficiente para prever um comportamento geral. O crescimento do banco de dados também pode futuramente ajudar a responder questões relacionadas à ecologia, como se “neurônios de espécies interrelacionadas compartilham algum traço morfológico implícito pela co-existência e compartilhamento de habitats”. Estes tópicos poderão ser considerados em trabalhos futuros, bem como a melhoria do modelo proposto, incorporando um maior número de medidas a fim de diminuir o grau de degeneração implicado por utilizar apenas algumas características morfológicas. Este trabalho envolveu a colaboração da Universidade de Cambridge e foi publicado na revista *Frontiers in Computational Neuroscience*.

7.2 Distribuição espacial das densidades de expressão gênica no cérebro

O algoritmo desenvolvido no estudo da distribuição espacial dos padrões de expressão gênica no cérebro foi uma contribuição original. Embora tenha como base um algoritmo de análise anteriormente publicado, aquele algoritmo necessitou ser modificado, dando origem a uma nova proposta, mais genérica e ampliada. O algoritmo original foi projetado para comparar dois conjuntos de pontos espacialmente distribuídos num plano. Já no algoritmo modificado no novo trabalho, foi proposta uma forma de se analisar imagens de intensidade de níveis de cinza, típicas de imagens que mostram expressão gênica. Também foi proposta uma extensão do algoritmo original para comparar não somente duas imagens, mas também dois conjuntos de imagens (fatias de dois cérebros de rato) e organizar os resultados gerados em forma de redes (grafos), com duas formas de visualização utilizando PCA. Deve-se ressaltar que este trabalho, embora focado em imagens de expressão gênica em cérebros de rato, é uma metodologia computacional genérica, que pode ser aplicado para se comparar quaisquer conjuntos de imagens em níveis de cinza em que se esteja interessado em analisar a relação da distribuição espacial de densidades. Assim, é possível quantificar a similaridade nas distribuições locais de densidades entre dois conjuntos de imagens. Foram analisados os dados públicos de expressão gênica em

cérebros de camundongo disponibilizado no Allen Mouse Brain Atlas (AMBA). Como similaridades na distribuição espacial de densidades entre genes pode indicar funcionalidade correlacionada, a identificação destas similaridades é útil para indicar potenciais direções de pesquisa para descobrir interações entre genes e suas funções correlacionadas. Neste sentido, o algoritmo proposto foi aplicado em alguns genes do AMBA, o que proporcionou medidas capazes de caracterizar a similaridade da distribuição de densidades entre esses genes, além de permitir visualizar os resultados com redes e PCA. Os resultados sugerem correlações nas densidades cerebrais entre alguns genes, que podem ser futuramente explorados. Um importante passo antes do uso deste algoritmo é o processo de registro das imagens, ou seja, o alinhamento das imagens com o atlas anatômico cerebral para possibilitar a comparação de diferentes amostras cerebrais. Este é um campo de pesquisa ainda em aberto e uma tarefa difícil de se solucionar. Como neste estudo o foco era a metodologia de análise, optou-se por realizar o registro manual em algumas imagens de expressão gênica para demonstrar o método. Trabalhos futuros podem se concentrar em investigar métodos de registro automático adequados para as imagens do AMBA. Desta forma, será possível analisar toda a base de dados. Outra possibilidade de investigação pode ser realizada em um conjunto maior de dados com funções e interações gênicas conhecidas para estudar a inter-relação dessas informações com os padrões de expressão gênica. Este trabalho foi publicado na revista *Journal of Neuroscience Methods*.

7.3 Propagação de doenças numa população de estruturas ramificadas

O estudo da influência da forma individual nas dinâmicas ocorrendo numa rede foi realizado utilizando o conceito de dinâmica de propagação de doenças (modelo SIR) numa rede de raízes reais. Tradicionalmente, estes modelos de propagação de doença são estudados desprezando-se as peculiaridades de cada indivíduo. Neste trabalho, a originalidade está na consideração da morfologia de raízes de feijão reais para a modelagem de cada indivíduo da rede e no estudo das variações na dinâmica de propagação que isto desencadeia. Este trabalho foi realizado utilizando-se de recursos e procedimentos computacionais, desde a segmentação das imagens, passando pela implementação da simulação da dinâmica e indo até a análise dos dados obtidos com o modelo de discos para possibilitar explicar os resultados. Além disso, o conjunto das rotinas utilizadas, necessário para a concretização desta investigação, constitui em si uma metodologia computacional antes

não existente. Descobriu-se que a morfologia dos indivíduos (raízes) é um importante fator para a propagação da epidemia SIR. As correlações de curta distância entre as taxas de transmissão num sistema heterogêneo, como é o caso das raízes, tornam a população menos suscetível a surtos de epidemia do que em um sistema homogêneo correspondente. Além disso, as características morfológicas de anisotropia e desordem são de particular importância nas correlações. Devido às características similares de propagação de alguns agentes patogênicos em neurônios, os resultados encontrados neste estudo também são relevantes à neurociência.

Foi assumido que as propriedades dos indivíduos, incluindo suas morfologias, não dependem do tempo. Esta suposição é verdadeira nos casos de epidemias relativamente rápidas ou alterações bastante lentas na forma das raízes. Mas, há também o caso de ocorrer mudanças morfológicas significativas enquanto a epidemia se propaga, requerendo a modelagem de transmissibilidades que variam com o tempo, o que está além do escopo deste trabalho, mas que se consiste num importante tópico para estudos futuros. Este trabalho também foi realizado em colaboração com a Universidade de Cambridge e foi publicado na revista *Journal of the Royal Society Interface* e no congresso CompleNet 2010 (*Communications in Computer and Information Science - CCIS*).

7.4 Influência da expressão gênica na morfologia, topologia e dinâmica

Na atividade sobre a influência da expressão gênica no desenvolvimento da morfologia, topologia e dinâmica, foi utilizado o modelo de estruturas ramificadas desenvolvido na atividade de investigação do hiperespaço neuromorfológico. Assim, utilizando-se desse modelo, neurônios artificiais foram crescidos considerando-se a intensidade de expressão gênica. Foram extraídos destes neurônios as suas medidas morfológicas, que foram comparadas às intensidades de expressão gênica. Em seguida, foi analisada a rede formada por esses neurônios, tanto em relação à topologia, quanto à dinâmica integra-e-dispara. A combinação dos resultados entre a entropia condicional e a análise de inclinação mostrou-se capaz de identificar as medidas morfológicas que melhor refletem a influência das variações nas intensidades de expressão gênica. Esta é uma metodologia computacional original, na qual foi utilizado um conceito de teoria da informação (entropia condicional de Shannon) para avaliar a variação morfológica relacionada à expressão gênica a partir da análise

de medidas que caracterizam a morfologia neuronal. Utilizando correlações de Pearson, também foi possível observar que as medidas conseguiram capturar a relação entre o grau (medida topológica) com a taxa de disparo (medida de dinâmica). O grau é um fator topológico importante na determinação da taxa de disparo. Também foi possível diferenciar e quantificar a influência na morfologia da expressão de dois genes. Para o futuro, é possível adicionar melhorias incrementais ao arcabouço, acrescentando novas medidas, novas ferramentas de análise e modelos neuronais cada vez mais realistas. Por exemplo, podem-se utilizar outros modelos geradores de neurônios artificiais mais sofisticados e considerar mais expressões gênicas. Assim, pode-se também investigar a possibilidade de incluir uma rede de regulação gênica.

7.5 Principais contribuições

As principais contribuições podem ser sumarizadas como:

- Proposta de uma abordagem computacional sistemática no estudo da morfologia neuronal, utilizando o modelo de McGhee.
- A grande expansão no número de neurônios reais segmentados, disponíveis na base de dados pública NeuroMorpho, possibilitou uma abrangente investigação da forma neuronal real, que ainda não tinha sido realizada por insuficiência de dados.
- Um novo algoritmo para quantificar a similaridade nas distribuições locais de densidades de expressão gênica em conjuntos de imagens.
- Uma nova metodologia computacional de investigação da influência da forma ramificada na dinâmica de propagação de doenças. Em particular, foi proposto um modelo de círculos para avaliar as correlações.
- Foi verificado que as correlações de curta distância entre as taxas de transmissão num sistema heterogêneo tornam a população menos suscetível a surtos de epidemia do que num sistema homogêneo correspondente.
- Foi constatado que a estrutura ramificada dos indivíduos (raízes) é um importante fator para a propagação da epidemia no modelo SIR. Mais especificamente, anisotropia (desvio da raiz pivotante na direção vertical) e desordem (distribuição de densidade, relacionada ao grau de ramificação) influenciam as correlações.

- Proposta de um novo arcabouço para a quantificação da influência da expressão gênica na morfologia, topologia e dinâmica, utilizando-se de métodos físicos e computacionais.

Os trabalhos desenvolvidos resultaram até o momento em quatro publicações indexadas (três em revistas e um em congresso):

- 1 MIAZAKI, M.; COSTA, L. da F. Study of cerebral gene expression densities using Voronoi analysis. *Journal of Neuroscience Methods*, v. 203, n. 1, p. 212–219, 2012.
- 2 COSTA, L. da F.; ZAWADZKI, K.; MIAZAKI, M.; VIANA, M. P.; TARASKIN, S. N. Unveiling the neuromorphological space. *Frontiers in Computational Neuroscience*, v. 4, p. 150, 2010.
- 3 HANDFORD, T. P.; PEREZ-RECHE, F. J.; TARASKIN, S. N.; COSTA, L. da F.; MIAZAKI, M.; NERI, F. M.; GILLIGAN, C. A. Epidemics in networks of spatially correlated threedimensional root-branching structures. *Journal of the Royal Society Interface*, v. 8, n. 56, p. 423–434, 2011.
- 4 HANDFORD, T. P.; PEREZ-RECHE, F. J.; TARASKIN, S. N.; COSTA, L. da F.; MIAZAKI, M.; NERI, F. M.; GILLIGAN, C. A. Epidemics in anisotropic networks of roots. In: WORKSHOP ON COMPLEX NETWORKS, 2., 2010, Rio de Janeiro. *Proceedings...* Berlin: Springer-Verlag, 2011. p. 146–153. (Communications in Computer and Information Science – CCIS, v. 116).

Os trabalhos dos dados NeuroMorpho e raízes foram desenvolvidos em colaboração com o grupo de pesquisa do Prof. Sergei N. Taraskin. Relacionados a estes trabalhos, foram realizadas duas visitas técnicas ao seu grupo na Universidade de Cambridge, Inglaterra. A primeira visita técnica contribuiu significativamente para o desenvolvimento do estudo de propagação de doenças SIR nas raízes, que gerou os artigos 3 e 4. O artigo 3 refere-se ao conteúdo apresentado no Apêndice A, menos a Subseção A.4.4 sobre anisotropia, que foi publicado no artigo 4. Já durante a segunda visita técnica nesse mesmo grupo de pesquisa, foi realizado o estudo dos dados NeuroMorpho, descrito na Seção 4 e publicado no artigo 2.

7.6 Trabalhos futuros

Como trabalhos futuros, podem ser listados:

- Quando o banco de dados NeuroMorpho se tornar suficientemente grande, seria interessante realizar estudos considerando questões mais específicas em relação a espécies, tipos de neurônios e regiões cerebrais.
- Utilização de uma metodologia de registro (alinhamento) automático das imagens do AMBA para permitir a análise de todas as imagens disponíveis.
- Obtenção de mais imagens de expressão gênica, com informações conhecidas sobre funções e interações entre genes.
- Verificar se há algum benefício em utilizar Voronoi 3D, ao invés da abordagem 2D, no algoritmo de Análise de Densidade Local por Voronoi.
- No modelo de propagação de doenças, avaliar a utilização de: neurônios, redes 3D como ocorre no cérebro, e redes irregulares.
- Considerar variações na morfologia dos indivíduos no decorrer do tempo, juntamente com a dinâmica de propagação da doença.
- Aperfeiçoamento do modelo gerador de neurônios artificiais.
- Avaliar a utilização de outras dinâmicas no arcabouço.
- Adicionar mais ferramentas de análise no arcabouço.
- Inclusão de mais dados de expressão gênica no arcabouço.
- Inclusão de uma rede de regulação gênica no arcabouço.
- Validação do arcabouço com dados biológicos reais em todos os níveis.

REFERÊNCIAS

- 1 CAJAL, S. R. *Recollections of my life*. Massachussets: MIT Press, 1989. 634 p.
- 2 DE FELIPE, J. Sesquicentenary of the birthday of Santiago Ramón y Cajal, the father of modern neuroscience. *Trends in Neurosciences*, v. 25, n. 9, p. 481–484, 2002.
- 3 ALBRIGHT, T. D.; JESSELL, T. M.; KANDEL, E. R.; POSNER, M. I. Neural science: a century of progress and the mysteries that remain. *Neuron*, v. 25, n. 1, p. S1–S55, 2000. Supplement 1. DOI:10.1016/S0896-6273(00)80912-5.
- 4 KANDEL, E. R.; SCHWARTZ, J. H.; JESSEL, T. M. *Principles of neural science*. New York: McGraw-Hill, 2000. 1414 p.
- 5 ASCOLI, G. A. (Ed.). *Computational neuroanatomy: principles and methods*. Totawa: Humana Press, 2002.
- 6 TRAPPENBERG, T. *Fundamentals of computational neuroscience*. USA: Oxford University Press, 2002. 333 p.
- 7 SPORNS, O.; TONONI, G.; KÖTTER, R. The human connectome: a structural description of the human brain. *PLoS Computational Biology*, v. 1, n. 4, p. e42, 2005. DOI:10.1371/journal.pcbi.0010042.
- 8 PELT, J. van; UYLINGS, H. B. M. Modeling neuronal growth and shape. In: LAUBICHLER, M. D.; MÜLLER, G. B. (Ed.). *Modeling biology: structures, behaviors, evolution*. Cambridge: MIT Press, 2007. p. 195–215.
- 9 HOSKING, C. R.; SCHWARTZ, J. L. The future's bright: imaging cell biology in the 21st century. *Trends in Cell Biology*, v. 9, n. 11, p. 553–554, 2009.

-
- 10 MILNER, B.; SQUIRE, L. R.; KANDEL, E. R. Cognitive neuroscience and the study of memory. *Neuron*, v. 20, n. 3, p. 445–468, 1998.
- 11 SHOLL, D. A. Dendritic organization in the neurons of the visual and motor cortices of the cat. *Journal of Anatomy*, v. 87, pt. 4, p. 387–406, 1953.
- 12 KOCH, C.; POGGIO, T.; TORRE, V. Retinal ganglion cells: a functional interpretation of dendritic morphology. *Philosophical Transactions of the Royal Society of London Series B - biological sciences*, v. 298, n. 1090, p. 227–264, 1982.
- 13 COOK, J. E. Getting to grips with neuronal diversity: what is a neuronal type? In: CHALUPA, L.; FINLAY, B. (Ed.). *Development and organization of the retina*. New York: Plenum, 1998. p. 91–120.
- 14 COSTA, L. da F.; BARBOSA, M. S. An analytical approach to connectivity in regular neuronal networks. *European Physical Journal B*, v. 42, n. 4, p. 573–580, 2004.
- 15 SCHIERWAGEN, A. Neuronal morphology: shape characteristics and model. *Neurophysiology*, v. 40, n. 4, p. 310–315, 2008.
- 16 BOTA, M.; SWANSON, L. W. The neuron classification problem. *Brain Research Reviews*, v. 56, n. 1, p. 79–88, 2007.
- 17 MASLAND, R. H. Neuronal cell types. *Current Biology*, v. 14, n. 13, p. 497–500, 2004.
- 18 COSTA, L. da F. *Morphological complex networks: can individual morphology determine the general connectivity and dynamics of networks?* Disponível em: <<http://arxiv.org/pdf/q-bio/0503041v1.pdf>>. Acesso em: 22 Abr 2012.
- 19 AHNERT, S. E.; TRAVENÇOLO, B. A. N.; COSTA, L. da F. Connectivity and dynamics of neuronal networks as defined by the shape of individual neurons. *New Journal of Physics*, v. 11, n. 10, p. 103053, 2009. DOI:10.1088/1367-2630/11/10/103053.
- 20 KREINDLER, A. *Experimental epilepsy*. Amsterdam: Elsevier Publishing Company, 1965. 213 p. (Progress in Brain Research, v. 19).
- 21 ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. *Molecular biology of the cell*. 7th ed. New York: Garland Science, 2002.

-
- 22 WEN, Q.; STEPANYANTS, A.; ELSTON, G. N.; GROSBURG, A. Y.; CHKLOVSKII, D. B. Maximization of the connectivity repertoire as a statistical principle governing the shapes of dendritic arbors. *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, n. 30, p. 12536–12541, 2009.
- 23 COSTA, L. da F.; MONTEIRO, E. T. M. A percolation approach to neural morphometry and connectivity. *Neuroinformatics*, v. 1, n. 1, p. 65–80, 2003.
- 24 AGMON-SNIR, H.; CARR, C. E.; RINZEL, J. The role of dendrites in auditory coincidence detection. *Nature*, v. 393, n. 6682, p. 268–272, 1998. DOI:10.1038/30505.
- 25 COSTA, L. da F.; MANOEL, E. T. M.; FAUCEREAU, F.; CHELLY, J.; PELT, J. van; RAMAKERS, G. A shape analysis framework for neuromorphometry. *Network: computation in neural systems*, v. 13, n. 3, p. 283–310, 2002.
- 26 JAN, Y.-N.; JAN, L. Y. The control of dendrite development. *Neuron*, v. 40, n. 2, p. 229–242, 2003.
- 27 BERKOWITZ, A.; YOSTEN, G. L. C.; BALLARD, R. M. Somato-dendritic morphology predicts physiology for neurons that contribute to several kinds of limb movements. *Journal of Neurophysiology*, v. 95, n. 5, p. 2821–2831, 2006.
- 28 PÉREZ-RECHE, F. J.; TARASKIN, S. N.; COSTA, L. da F.; NERI, M. F.; GILLIGAN, A. C. Complexity and anisotropy in host morphology make populations less susceptible to epidemic outbreaks. *Journal of the Royal Society Interface*, v. 7, n. 48, p. 1083–1092, 2010.
- 29 SEGEV, I. Sound grounds for computing dendrites. *Nature*, v. 393, n. 6682, p. 207–208, 1998.
- 30 WEN, Q.; CHKLOVSKII, D. B. A cost-benefit analysis of neuronal morphology. *Journal of Neurophysiology*, v. 99, n. 5, p. 497–500, 2008.
- 31 LEWIN, B. *Genes VII*. Tradução H. B. Ferreira. Porto Alegre: Artmed, 2001. 955 p.
- 32 STRYER, L. *Biochemistry*. 4th ed. New York: W.H. Freeman, 1995. 1064 p.

- 33 HASTY, J.; MCMILLEN, D.; ISAACS, F.; COLLINS, J. J. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nature Reviews Genetics*, v. 2, n. 4, p. 268–279, 2001.
- 34 GILBERT, S. F. *Developmental biology*. 7th ed. Sunderland: Sinauer Associates, 2003. 838 p.
- 35 KAMMERMEIER, L.; REICHERT, H. Common developmental genetic mechanisms for patterning invertebrate and vertebrate brains. *Brain Research Bulletin*, v. 55, n. 6, p. 675–682, 2001.
- 36 O’LEARY, D. D. M.; CHOU, S.-J.; SAHARA, S. Area patterning of the mammalian cortex. *Neuron*, v. 56, n. 2, p. 252–269, 2007. DOI:10.1016/j.neuron.2007.10.010.
- 37 KRUBITZER, L. In search of a unifying theory of complex brain evolution. *Annals of the New York Academy of Sciences*, v. 1156, n. 1, p. 44–67, 2009. DOI:10.1111/j.1749-6632.2009.04421.x.
- 38 BONCINELLI, E.; GULISANO, M.; BROCCOLI, V. *Emx* and *Otx* homeobox genes in the developing mouse brain. *Journal of Neurobiology*, v. 24, n. 10, p. 1356–1366, 1993.
- 39 CECCHI, C.; MALLAMACI, A.; BONCINELLI, E. *Otx* and *Emx* homeobox genes in brain development. *International Journal of Developmental Biology*, v. 44, n. 6, p. 663–668, 2000.
- 40 TRAVENÇOLO, B. A. N. *Métodos computacionais para a caracterização e análise da relação entre anatomia e expressão gênica em sistemas biológicos*. Tese (Doutorado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2007.
- 41 ASCOLI, G. A. Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nature Reviews Neuroscience*, v. 7, n. 4, p. 318–324, 2006.
- 42 ASCOLI, G. A.; DONOHUE, D. E.; HALAVI, M. NeuroMorpho.Org: a central resource for neuronal morphologies. *Journal of Neuroscience*, v. 27, n. 35, p. 9247–9251, 2007.
- 43 JONES, A. R.; OVERLY, C. C.; SUNKIN, S. M. The Allen Brain Atlas: 5 years and beyond. *Nature Reviews Neuroscience*, v. 10, n. 11, p. 821–828, 2009. DOI:10.1038/nrn2722.

- 44 HALAVI, M.; POLAVARAM, S.; DONOHUE, D. E.; HAMILTON, G.; HOYT, J.; SMITH, K. P.; ASCOLI, G. A. NeuroMorpho.Org implementation of digital neuroscience: dense coverage and integration with the NIF. *Journal of Neuroinformatics*, v. 6, n. 3, p. 241–252, 2008.
- 45 SCORCIONI, R.; POLAVARAM, S.; ASCOLI, G. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, v. 3, n. 5, p. 866–76, 2008.
- 46 LEIN, E. S.; HAWRYLYCZ, M. J.; AO, N.; AYRES, M.; BENSINGER, A.; BERNARD, A.; AL. et. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, v. 445, p. 168–176, 2007. DOI:10.1038/nature05453.
- 47 HICKS, D. G.; LONGORIA, G.; PETTAY, J.; GROGAN, T.; TARR, S.; TUBBS, R. In situ hybridization in the pathology laboratory: general principles, automation, and emerging research applications for tissue-based studies of gene expression. *Journal of Molecular Histology*, v. 35, n. 6, p. 595–601, 2004.
- 48 DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. New York: Wiley-Interscience, 2001. 654 p.
- 49 HÄRDLE, W. K.; SIMAR, L. *Applied multivariate statistical analysis*. 2nd ed. Berlin: Springer, 2007. 458 p.
- 50 FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, p. 179–188, 1936.
- 51 MCLACHLAN, G. J. *Discriminant analysis and statistical pattern recognition*. New York: Wiley-Interscience, 2004. 552 p.
- 52 COSTA, L. da F.; BONCI, D. M. O.; SAITO, C. A.; ROCHA, F. A. F.; SILVEIRA, L. C. L.; VENTURA, D. F. Voronoi analysis uncovers relationship between mosaics of normally placed and displaced amacrine cells in the thraira retina. *Neuroinformatics*, v. 5, p. 59–77, 2007. DOI:10.1385/NI:5:1:59.
- 53 MCGHEE, G. R. *The geometry of evolution: adaptive landscapes and theoretical morphospaces*. Cambridge: Cambridge University Press, 2006. 212 p.
- 54 OKABE, A.; BOOTS, B.; SUGIHARA, K.; CHIU, S. N. *Spatial tessellations: concepts and applications of Voronoi diagrams*. 2nd ed. Chichester: John Wiley, 2000. 671 p.

- 55 BERG, M.; de van KREVELD, M.; OVERMARS, M.; SCHWARZKOPF, O. *Computational geometry: algorithms and applications*. 2nd ed. Berlin: Springer, 2000. 379 p.
- 56 COSTA, L. da F.; SILVA, F. N. Hierarchical characterization of complex networks. *Journal of Statistical Physics*, v. 125, n. 4, p. 841–872, 2006. DOI:10.1007/s10955-006-9130-y.
- 57 COSTA, L. da F.; CESAR JUNIOR, R. M. *Shape analysis and classification: theory and practice*. Boca Raton: CRC Press, 2001. 680 p.
- 58 GUARDIOLA, X.; DÍAZ-GUILERA, A.; LLAS, M.; PÉREZ, C. J. Synchronization, diversity, and topology of networks of integrate and fire oscillators. *Physical Review E*, v. 62, n. 4, p. 5565–5570, 2000.
- 59 ROXIN, A.; RIECKE, H.; SOLLA, S. A. Self-sustained activity in a small-world network of excitable neurons. *Physical Review Letters*, v. 92, n. 19, p. 198101, 2004.
- 60 COSTA, L. da F.; ZAWADZKI, K.; MIAZAKI, M.; VIANA, M. P.; TARASKIN, S. N. Unveiling the neuromorphological space. *Frontiers in Computational Neuroscience*, v. 4, p. 150, 2010. DOI:10.3389/fncom.2010.00150.
- 61 BREITKREUTZ, B. J.; STARK, C.; TYERS, M. Osprey: a network visualization system. *Genome Biology*, v. 4, n. 3, p. R22, 2003. DOI:10.1186/gb-2003-4-3-r22.
- 62 MIAZAKI, M.; COSTA, L. da F. Study of cerebral gene expression densities using Voronoi analysis. *Journal of Neuroscience Methods*, v. 203, n. 1, p. 212–219, 2012. DOI:10.1016/j.jneumeth.2011.09.009.
- 63 ASSENOV, Y.; RAMÍREZ, F.; SCHELHORN, S.-E.; LENGAUER, T.; ALBRECHT, M. Computing topological parameters of biological networks. *Bioinformatics*, v. 24, n. 2, p. 282–284, 2008. DOI:10.1093/bioinformatics/btm554.
- 64 SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N. S.; WANG, J. T.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, v. 13, n. 11, p. 2498–2504, 2003. DOI:10.1101/gr.1239303.
- 65 COMIN, C. H.; BUNORO, J. B.; VIANA, M. P.; COSTA, L. da F. *Unveiling the relationship between structure and dynamics in complex networks*. Disponível em: <<http://arxiv.org/pdf/1109.2963v1.pdf>>. Acesso em: 22 Abr 2012.

- 66 BAILEY, D.; OTTEN, W.; GILLIGAN, C. A. Saprotrophic invasion by the soil-borne fungal plant pathogen *Rhizoctonia solani* and percolation thresholds. *New Phytologist*, v. 146, n. 3, p. 535–544, 2000. DOI:10.1046/j.1469-8137.2000.00660.x.
- 67 OTTEN, W.; BAILEY, D.; GILLIGAN, C. A. Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. *New Phytologist*, v. 163, n. 1, p. 125–132, 2004. DOI:10.1111/j.1469-8137.2004.01086.x.
- 68 REISS, C. S. (Ed.). *Neurotropic viral infections*. Cambridge: Cambridge University Press, 2008. 484 p.
- 69 LOEWY, A. D. Viruses as transneuronal tracers for defining neural circuits. *Neuroscience & Biobehavioral Reviews*, v. 22, n. 6, p. 679–684, 1998. DOI:10.1016/S0149-7634(98)00006-2.
- 70 OZTAS, E. Neuronal tracing. *Neuroanatomy*, v. 2, n. 1, p. 2–5, 2003.
- 71 RAMOS-VARA, J. A. Technical aspects of immunohistochemistry. *Veterinary Pathology*, v. 42, n. 4, p. 405–426, 2005. DOI:10.1354/vp.42-4-405.
- 72 HORNBY, D. (Ed.). *Biological control of soil-borne plant pathogens*. Wallingford: CAB International, 1990.
- 73 GILLIGAN, C. A. Mathematical modelling and analysis of soilborne pathogens. In: KRANZ, J. (Ed.). *Epidemics of plant diseases: mathematical analysis and modeling*. Heidelberg: Springer, 1990. p. 96–142.
- 74 GILLIGAN, C. A. An epidemiological framework for disease management. *Advances in Botanical Research*, v. 38, n. 1, p. 1–64, 2002.
- 75 GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. Tradução Roberto Marcondes Cesar Junior, Luciano da Fontoura Costa. São Paulo: Editora Edgard Blücher, 2000. 509 p.
- 76 GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. *Digital image processing using MATLAB*. Upper Saddle River: Prentice Hall, 2004. 624 p.
- 77 SCHROEDER, W.; MARTIN, K.; LORENSEN, B. *The Visualization Toolkit: an object-oriented approach to 3D graphics*. 4th ed. USA: Kitware, 2006. 528 p.

- 78 OLIVEIRA, M. C. F.; MINGHIM, R. Uma introdução à visualização computacional. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 17., 1997, Brasília. *Anais...* Brasília: SBC, 1997. p. 85–131.
- 79 PAIVA, A. C.; SEIXAS, R. B.; GATTASS, M. *Introdução à Visualização Volumétrica*. Rio de Janeiro: PUC, 1999. 107 p. Disponível em: <ftp://ftp.inf.puc-rio.br/pub/docs/techreports/99_03_paiva.pdf>. Acesso em 22 Abr 2012.
- 80 NÅSELL, I. Stochastic models of some endemic infections. *Mathematical Biosciences*, v. 179, n. 1, p. 1–19, 2002.
- 81 NEWMAN, M. E. J. *Networks: an introduction*. Oxford: Oxford University Press, 2010. 772 p.
- 82 MARRO, J.; DICKMAN, R. *Nonequilibrium phase transitions in lattice models*. Cambridge: Cambridge University Press, 1999.
- 83 HINRICHSSEN, H. Non-equilibrium critical phenomena and phase transitions into absorbing states. *Advances in Physics*, v. 49, n. 7, p. 815–958, 2000.
- 84 DYBIEC, B.; KLECZKOWSKI, A.; GILLIGAN, C. A. Controlling disease spread on networks with incomplete knowledge. *Physical Review E*, v. 70, n. 6, p. 066145, 2004. DOI:10.1103/PhysRevE.70.066145.
- 85 GRASSBERGER, P. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, v. 63, n. 2, p. 157–172, 1983. DOI:10.1016/0025-5564(82)90036-0.
- 86 NEWMAN, M. Spread of epidemic disease on networks. *Physical Review E*, v. 66, n. 1, p. 016128, 2002. DOI:10.1103/PhysRevE.66.016128.
- 87 SANDER, L.; WARREN, C. P.; SOKOLOV, I. M.; SIMON, C.; KOOPMAN, J. Percolation on heterogeneous networks as a model for epidemics. *Mathematical Biosciences*, v. 180, n. 1, p. 293–305, 2002. DOI:10.1016/S0025-5564(02)00117-7.
- 88 GILLIGAN, C. A. Modelling soil-borne plant pathogens: reaction-diffusion models. *Canadian Journal of Plant Pathology*, v. 17, n. 2, p. 96–108, 1995. DOI:10.1080/07060669509500700.

- 89 ISICHENKO, M. B. Percolation, statistical topography, and transport in random media. *Reviews of Modern Physics*, v. 64, n. 4, p. 961–1043, 1992. DOI:10.1103/RevModPhys.64.961.
- 90 HEERMANN, D. W.; STAUFFER, D. Influence of boundary conditions on square bond percolation near p_c . *Zeitschrift für Physik B: condensed matter*, v. 40, n. 1, p. 133–136, 1980. DOI:10.1007/BF01295081.
- 91 NERI, F. M.; PÉREZ-RECHE, F. J.; TARASKIN, S. N.; GILLIGAN, C. A. Heterogeneity in susceptible-infected-removed (SIR) epidemics on lattices. *Journal of the Royal Society Interface*, v. 8, n. 55, p. 201–209, 2011.
- 92 KUULASMAA, K. The spatial general epidemic and locality dependent random graphs. *Journal of Applied Probability*, v. 19, n. 4, p. 745–758, 1982.
- 93 COX, J. T.; DURRETT, R. Limit theorems for the spread of epidemics and forest fires. *Stochastic Processes and their Applications*, v. 30, n. 2, p. 171–191, 1988. DOI:10.1016/0304-4149(88)90083-X.
- 94 KENAH, E.; ROBINS, J. M. Second look at the spread of epidemics on networks. *Physical Review E*, v. 76, n. 3, p. 036113, 2007. DOI:10.1103/PhysRevE.76.036113.
- 95 MILLER, J. C. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, v. 76, n. 1, p. 010101, 2007. DOI:10.1103/PhysRevE.76.010101.
- 96 SYKES, M. F.; ESSAM, J. W. Exact critical percolation probabilities for site and bond problems in two dimensions. *Journal of Mathematical Physics*, v. 5, n. 8, p. 1117–1127, 1964. DOI:10.1063/1.1704215.
- 97 HANDFORD, T. P.; PEREZ-RECHE, F. J.; TARASKIN, S. N.; COSTA, L. da F.; MIAZAKI, M.; NERI, F. M.; GILLIGAN, C. A. Epidemics in anisotropic networks of roots. In: WORKSHOP ON COMPLEX NETWORKS, 2., 2010, Rio de Janeiro. *Proceedings...* Berlin: Springer-Verlag, 2011. p. 146–153. (Communications in Computer and Information Science – CCIS, v. 116).
- 98 HANDFORD, T. P.; PEREZ-RECHE, F. J.; TARASKIN, S. N.; COSTA, L. da F.; MIAZAKI, M.; NERI, F. M.; GILLIGAN, C. A. Epidemics in networks of spatially correlated three-dimensional root-branching structures. *Journal of the Royal Society Interface*, v. 8, n. 56, p. 423–434, 2011. DOI:10.1098/rsif.2010.0296.

Apêndice A Propagação de doenças numa população de estruturas ramificadas

O estudo de raízes teve como propósito analisar o comportamento de propagação de doenças numa rede de indivíduos biologicamente realísticos, cuja estrutura ramificada individual afeta o desenvolvimento dinâmico da epidemia na rede, ou seja, a forma afeta o modo como a doença se propaga na rede de raízes (função). Um exemplo de agente patogênico que se propaga de raiz em raiz pelo solo é o fungo *Rhizoctonia solani* (66, 67). De forma semelhante, neurônios também possuem estruturas ramificadas e são suscetíveis a doenças que se propagam de forma similar, de neurônio a neurônio. Os vírus neurotrópicos, como os que provocam a raiva e a herpes, são exemplos de agentes patogênicos que infectam e se propagam pelos neurônios.

A raiva é geralmente transmitida pela saliva de um animal infectado através de mordidas ou arranhões. O vírus da raiva entra no sistema nervoso através de um neurônio sensitivo ou motor. Em seguida, este vírus viaja de um neurônio ao seguinte ao longo da medula espinhal até o cérebro (68). Isto causa mudanças de comportamento, como o estado de fúria em cachorros ou hidrofobia em humanos. Após a invasão do cérebro, o vírus se propaga para as glândulas salivares e pode infectar outro indivíduo. Após o vírus entrar no sistema nervoso central, nenhum tratamento terapêutico pode combater a infecção e a raiva é quase sempre fatal.

Outra motivação para estudos relacionados à melhor compreensão da relação forma-função (morfologia neuronal - dinâmica na propagação de doenças) é a importante utilidade dos vírus neurotrópicos para a identificação de caminhos e interconexões numa rede neuronal. Este método, denominado *neuronal tracing* (rastreamento neuronal) (69, 70), consiste em injetar um vírus neurotrópico em um tecido periférico ou região cerebral para serem transportados pelos terminais nervosos, causando uma infecção controlada que se espalha sequencialmente dentro da rede neuronal. O vírus se desloca para o soma do neurônio e passa a controlá-lo para sua própria replicação. Em seguida, os vírus produzidos

se espalham através das sinapses, passando pelo axônio ao neurônio vizinho (transporte anterógrado) ou na direção oposta, do neurônio infectado ao axônio do vizinho (transporte retrógrado). Desta forma, o vírus necessita de tempo suficiente para se propagar pelo sistema nervoso. Porém, este tempo não pode ser excessivamente longo, pois o vírus destrói a célula neuronal. Os neurônios infectados podem ser visualizados utilizando coloração imuno-histoquímica (técnica de coloração que localiza biomarcadores num tecido biológico, usando o princípio da ligação de anticorpos a antígenos) (71). O *Herpes simplex virus 1* (HSV-1), agente patogênico que causa o herpes labial, é um exemplo de vírus neurotrópico que pode ser utilizado para *neuronal tracing*.

As raízes foram analisadas utilizando o modelo SIR (Suscetível-Infectado-Recuperado), considerando que plantas infectadas eventualmente morrem (ou seja, permanentemente vão para a classe R). O modelo SIR pode ser aplicado numa variedade de agentes patogênicos que se propagam pelo solo (72) e se espalham numa população de plantas (73, 74). Assim, foi analisada como a morfologia das raízes de feijão (*Phaseolus vulgaris*) afeta a probabilidade de invasão da epidemia de um fungo patogênico comum, o *Rhizoctonia solani*, que se propaga pelo solo através de extensões miceliais (filamentos do fungo) do indivíduo infectado para o suscetível (66).

Ao analisar imagens tridimensionais digitalizadas de estruturas ramificadas de raízes de feijão utilizando um modelo epidemiológico SIR, pode-se mostrar como as estruturas ramificadas de raízes, correlacionadas espacialmente, afetam a eficiência de transmissão e, portanto, o critério de invasão de um agente patogênico que se propaga pelo solo através de indivíduos morfologicamente complexos. Foi constatado que as transmissibilidades (taxas de transmissão) heterogêneas que surgem das correlações, definidas pelo grau de sobreposição entre raízes vizinhas, tornam a população de raízes menos suscetível a invasões de doenças do que em um sistema homogêneo correspondente. Também foram analisados vários componentes de complexidade morfológica que contribuem para a desordem e heterogeneidade na transmissibilidade da infecção. Foi verificado que a anisotropia na forma das raízes aumenta a resistência a invasões epidêmicas, enquanto que incrementar a ramificação aumenta a propagação da epidemia na população de raízes.

A.1 Raízes

Os grãos de feijão foram selecionados com base em suas medidas, sendo considerado o comprimento de 10,5 a 11,0 mm, altura de 7,0 a 7,5 mm, e largura de 5,0 a 5,5 mm. Para a germinação das sementes, os grãos foram colocados sobre algodão umedecido com água. Após dois dias, os que brotaram foram transferidos para um aquário para se desenvolverem por quatro dias. O próximo passo foi fixar cada raiz num bloco de parafina para fatiar e fotografar. Para melhorar o contraste da raiz com a parafina, as raízes foram mergulhadas numa solução de água com anilina azul (proporção de 50%) para adquirirem uma coloração escura. A parafina foi derretida e um corante branco foi misturado para deixá-la opaca. Em seguida, cada raiz foi mergulhada na parafina ainda líquida. Após secar, foi obtido um bloco sólido para ser cortado em um micrótomo (Figura A.1). O bloco foi cortado em fatias de 0,1 mm de espessura. Cada fatia foi fotografada com uma câmera digital Sony Cyber-shot DSC-H50. Em seguida, cada imagem de fatia foi segmentada automaticamente usando o método de Otsu (75, 76). Na Figura A.2, são exibidas algumas das fatias de uma raiz de feijão (a). Em (b), são mostradas suas respectivas segmentações.

Figura A.1 – Micrótomo e bloco de parafina contendo uma raiz

A reconstrução foi realizada por um software implementado pelo autor em Borland C++ Builder, utilizando a biblioteca VTK (Visualization Toolkit) (77). O VTK é uma biblioteca de código aberto (*open source*) destinada a apoiar o desenvolvimento de aplicações em computação gráfica, processamento de imagens e visualização. É utilizado por diversos pesquisadores e desenvolvedores no mundo. O VTK possui diversos algoritmos implementados, como métodos volumétricos, redução de polígonos, suavização, triangulação de Delaunay, aplicação de textura, entre outros. No programa desenvolvido, o VTK foi utilizado para gerar imagens tridimensionais a partir das sequências de imagens.

Figura A.2 – Exemplos de imagens das fatias de uma raiz de feijão (a), suas respectivas segmentações (b) e reconstrução tridimensional (c).

A técnica utilizada para realizar a reconstrução foi a renderização por isosuperfícies (superfícies de valor limiar constante). Primitivas geométricas são ajustadas automaticamente sobre a superfície detectada através do método de intersecção de voxels *marching cubes* (77). Após a definição das isosuperfícies, são utilizados os métodos tradicionais de renderização de polígonos para a geração da visualização (78, 79). Na Figura A.2(c), pode-se visualizar a reconstrução tridimensional obtida a partir das imagens segmentadas de fatias de uma raiz de feijão (Figura A.2(b)).

A.2 Modelo SIR de propagação de doenças

O estudo do processo de propagação de doenças é de grande interesse científico, pois ao compreender seus mecanismos de transmissão e disseminação numa população pode-se prever seu comportamento e desenvolver formas de prevenção e combate. Neste sentido, atualmente podem ser encontradas na literatura várias propostas de modelos de propagação de doenças (80, 81). Um modelo bastante conhecido e estudado é o SIR. Neste modelo, os indivíduos de uma população podem encontrar-se em um dos três estados: suscetível (S), infectado (I) ou recuperado (R). Os indivíduos alteram seu estado seguindo certas regras. Um indivíduo suscetível (no estado S) pode ser infectado pelo seu vizinho (passando para o estado I) e, mais tarde, é removido da epidemia devido à sua morte ou porque se recuperou tornando-se imune à doença (indo ao estado R). Uma população pode ser representada por uma rede, na qual os nós representam os indivíduos e as conexões entre os nós descrevem possíveis caminhos de transmissão da doença. A epidemia

pode encontrar-se nos regimes: não ativo (a epidemia se propaga pouco na população, não caracterizando uma epidemia) e ativo (surto de epidemia). A transição entre estes dois regimes ocorre em um ponto crítico, que pode ser descrito como uma transição de fase de percolação (82, 83). Extensões e modificações deste modelo têm sido propostas para descrever diversas características de epidemias de forma mais realística (84).

No modelo SIR, existem dois parâmetros importantes: a transmissibilidade T (probabilidade de transmissão da doença de um indivíduo a outro, alterando o estado de S para I) e o tempo de recuperação τ (que modifica o estado do indivíduo de I para R). Uma rede é considerada homogênea quando a probabilidade de transmissão T é igual em todas as arestas e o tempo de recuperação τ é igual para todos os nós. É possível demonstrar em redes homogêneas que ocorre uma epidemia quando $R > 1$, sendo $R = T/\tau$. Porém, as redes reais em geral não são homogêneas, pois seus nós e arestas costumam apresentar características individuais próprios, os quais coletivamente determinam os valores de T e τ .

Foi considerado neste trabalho que um indivíduo infectado se mantém neste estado por um período fixo de tempo $\tau = 1$ (uma iteração) e, em seguida, deixa de ser infeccioso, ou seja, vai permanentemente para o estado recuperado. Durante o período de infecção, um nó infectado pode transmitir o agente patogênico a seus vizinhos mais próximos numa taxa β . Com estas regras, a probabilidade T da doença passar de um indivíduo infectado para o seu vizinho suscetível antes que se torne recuperado é dado pela expressão (85):

$$T = 1 - e^{-\tau\beta} \equiv 1 - e^{-\beta} . \quad (\text{A.1})$$

Este valor, chamado de transmissibilidade, quantifica o processo de transmissão entre indivíduos vizinhos. No sistema como um todo, T determina a probabilidade de invasão P_{inv} de ocorrer um surto de epidemia. Para calcular a probabilidade P_{inv} , foram geradas várias realizações estocásticas da epidemia e foram contadas quantas vezes todas as bordas foram alcançadas (critério adotado para considerar que ocorreu um surto de epidemia). Assim, T é o parâmetro de controle para o modelo SIR homogêneo (que é variado para estudar a transição) e P_{inv} é o parâmetro de ordem (parâmetro no qual se observa a mudança de fase). Se a transmissibilidade é menor que um certo valor crítico, $T < T_c$, a invasão epidêmica não pode ocorrer em um sistema infinito e a probabilidade de invasão para a epidemia é zero, $P_{\text{inv}} = 0$. Se $T > T_c$, a probabilidade de invasão é $P_{\text{inv}} > 0$, aproximando-se de um quando $T \rightarrow 1$ (86).

É possível demonstrar que uma epidemia em um sistema heterogêneo não correlacionado com transmissibilidades $\{T_{ij}\}$ é equivalente a uma epidemia em um sistema homogêneo no qual todas as transmissibilidades são substituídas por um valor médio de

transmissibilidade $\langle T \rangle$ (87):

$$P_{\text{inv}}(\{T_{ij}\}) = P_{\text{inv}}(\langle T \rangle) . \quad (\text{A.2})$$

Esta afirmação é correta somente no caso de transmissibilidades independentes entre diferentes pares de indivíduos. As raízes foram utilizadas em simulações de Monte Carlo de múltiplos eventos de invasão possíveis para demonstrar que para indivíduos realísticos com morfologia complexa esta hipótese nem sempre é válida. Isto ocorre porque as correlações de curta distância em transmissibilidades entre diferentes pares de indivíduos alteram o valor limiar de transmissibilidade.

A.3 Rede de raízes

As raízes foram consideradas como potenciais hospedeiras em uma rede regular quadrada para estudar a propagação de uma epidemia SIR em uma população de plantas através de crescimento micelial de raiz para raiz. Para a criação da rede, para cada nó, foi selecionada aleatoriamente uma raiz a partir das nove disponíveis ($N_r = 9$), mostradas nas três primeiras linhas da Figura A.3(a). A raiz selecionada foi rotacionada por um ângulo $0 < \varphi < 2\pi$ em torno de seu eixo vertical e posicionada em um nó (com a posição da semente coincidindo com o nó) i ($i = 1, 2, \dots, L^2$) de uma rede quadrada de tamanho $L \times L$ e com espaçamento a . Na Figura A.3(a), as raízes A, B e C são exemplos de raízes selecionadas, posicionadas em nós da rede e rotacionadas (as versões rotacionadas são, respectivamente, A', B' e C').

Assume-se que o agente patogênico infecta uma planta inicial (no centro da rede) e cresce ao longo e ao redor do indivíduo por expansão miceliar, explorando o solo próximo. Se outra raiz se encontra nas proximidades, a colônia de fungos pode alcançá-la e eventualmente infectá-la. Desta forma, a colônia de fungos se propaga ao longo de uma população de raízes. A taxa de transmissão entre plantas no processo SIR, que é efetivada pela expansão miceliar entre raízes, é proporcional à sobreposição efetiva entre diferentes raízes 3D. Para caracterizar esta sobreposição quantitativamente, foi primeiro definida a densidade da raiz (número de voxels por volume) da seguinte maneira. As raízes são representadas por imagens digitais tridimensionais. Cada imagem i consiste em um conjunto de N_i voxels (pixels tridimensionais), com volume $V_{\text{voxel}} = 10^{-3} \text{ mm}^3$ e localização definida pelo vetor de posição $r_{ni} = (x_{ni}, y_{ni}, z_{ni})$, sendo $n = 1, \dots, N_i$. A densidade dos

Figura A.3 – As projeções horizontais de raízes de nove feijões, dispostas sobre uma rede regular quadrada com espaçamento a , são mostradas nas primeiras três linhas de (a). A última linha exibe as três raízes A', B' e C', obtidas por rotação em torno do eixo vertical das plantas aleatoriamente escolhidas A, B e C nos ângulos φ_i , φ_j e φ_k , respectivamente. Esta linha representa um padrão típico de rede estudado. A visualização tridimensional da raiz D é mostrada em (b).

voxels pode ser escrita como a soma das funções delta de Dirac:

$$\rho_{\text{scan},i}(r) = \sum_{n=1}^{N_i} \delta(r - r_{ni}) . \quad (\text{A.3})$$

Deve-se notar que a densidade deve descrever um volume adicional ao redor da raiz, correspondente à região pela qual o fungo pode se propagar pelo solo e que, portanto, intermedia a transmissão da infecção de um doador (infectado) para um recipiente (suscetível) (88). Para isto, a raiz é dilatada substituindo as funções delta de Dirac na Equação A.3 por uma representação Gaussiana:

$$\rho_i(r) \simeq \int dr' \rho_{\text{scan},i}(r') \left(\frac{1}{2\pi\sigma^2} \right)^{3/2} \exp \left[-\frac{(r - r')^2}{2\sigma^2} \right] , \quad (\text{A.4})$$

com o parâmetro de dilatação $\sigma = 1, 5$. O valor de sigma é muito menor que o espaçamento entre nós, ou seja $\sigma \ll a$, de tal forma que não há sobreposição com os vizinhos além dos mais próximos e, portanto, a transmissão do agente patogênico a eles pode ser ignorada no modelo. Assim, a doença consegue se propagar quando há sobreposição entre os ramos das raízes. Portanto, pode-se assumir que a transmissão entre indivíduos próximos i e j ocorre numa taxa β_{ij} , proporcional à sobreposição tridimensional J_{ij} entre os dois indivíduos, ou

seja:

$$\beta_{ij} = kJ_{ij} , \quad (\text{A.5})$$

em que:

$$J_{ij} = V_{\text{voxel}} \int dr \rho_i(r) \rho_j(r) \quad (\text{A.6})$$

$$= V_{\text{voxel}} \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \left(\frac{1}{4\pi\sigma^2} \right)^{3/2} \exp \left[-\frac{(r_{mi} - r_{nj})^2}{4\sigma^2} \right] . \quad (\text{A.7})$$

O valor de J_{ij} é o número de voxels da raiz i na região de sobreposição com a raiz j . Da definição, $J_{ij} = J_{ji}$. O coeficiente k (eficiência da infecção) caracteriza a taxa de transmissão por voxel da sobreposição, e é o mesmo em todas as sobreposições entre as raízes. A eficiência de transmissão representa a habilidade do agente patogênico de usar um ponto de contato entre dois indivíduos para se propagar entre eles. Desta forma, são levados em consideração no modelo tanto a morfologia particular de um indivíduo específico, quanto também condições ambientais. Assim, pode-se estudar a influência da morfologia nas taxas envolvidas entre duas raízes, além das propriedades descritas por k . Assim, a taxa de transmissão β_{ij} entre um par arbitrário de vizinhos próximos e a transmissibilidade:

$$T_{ij} = 1 - \exp\{-kJ_{ij}\} , \quad (\text{A.8})$$

(ver Equação A.1) depende da eficiência de transmissão k e da sobreposição aleatória J_{ij} , que depende da distância a entre os indivíduos. Assim, k e a são parâmetro de controle. A aleatoriedade nas sobreposições é causada pela morfologia complexa das raízes. Desta forma, toda a informação sobre uma dada configuração espacial de um sistema (uma particular disposição de raízes e ângulos em cada nó) está contida no conjunto de sobreposições $\{J_{ij}\}$, enquanto que toda a informação sobre as propriedades epidemiológicas do sistema encontra-se no conjunto de transmissibilidades $\{T_{ij}\}$.

A.4 Resultados e discussão

Inicialmente, é avaliada a probabilidade de invasão (parâmetro de ordem, ou seja, o parâmetro no qual se observa a mudança de fase) em função de dois parâmetros de controle, eficiência de transmissão e espaçamento entre os nós, destacando os valores limiares destes

parâmetros que separam os regimes epidêmicos de invasão e não-invasão. Em seguida, foi comparado o comportamento da epidemia na rede de raízes com uma rede homogênea, o que revelou um papel significativo das correlações de transmissibilidade nos pares de ligações a distâncias relativamente curtas. E finalmente, “*toy models*” foram investigados para compreender o efeito das correlações.

A.4.1 Diagrama de fases

Uma das mais importantes características da disseminação de epidemias é a probabilidade de invasão. Em termos de transição de fase, a probabilidade de invasão P_{inv} é o parâmetro de ordem que define se o estado da epidemia está no regime não-invasivo ($P_{\text{inv}}=0$) ou invasivo ($P_{\text{inv}} > 0$). No modelo, a probabilidade de invasão depende de dois parâmetros de controle – a eficiência de transmissão (relacionada especificamente ao sistema indivíduo/agente patogênico e independente da configuração espacial) e o espaçamento entre os nós da rede (que afeta o conjunto de sobreposições $\{J_{ij}\}$): $P_{\text{inv}} = P_{\text{inv}}(k, a)$.

Para calcular $P_{\text{inv}}(k, a)$ para um conjunto particular de transmissibilidades $\{T_{ij}\}$, é utilizado o mapeamento para percolação por ligações, na qual cada conexão $i-j$ é ocupada com probabilidade T_{ij} . A probabilidade de invasão é definida em um sistema infinito como a probabilidade de qualquer dado nó pertencer a um cluster infinito. Este tipo de cluster sempre existirá acima do limiar de invasão num sistema infinito e nunca existirá abaixo dele. Portanto, a probabilidade de invasão é igual à fração de nós de todo o sistema que pertence a este cluster, ou zero caso não exista tal cluster. Como é evidente que não haverá um cluster infinito em um sistema finito (que é o possível de ser modelado), pode-se definir o cluster “infinito” como aquele que se espalha pela rede e toca todas as quatro bordas. Assim, pode-se estimar P_{inv} verificando se este cluster existe e calculando a fração do sistema incorporada por ele. A média destas estimativas em várias realizações estocásticas deste sistema fornece a probabilidade de invasão para um conjunto de valores particulares T_{ij} . Como diferentes realizações do conjunto T_{ij} produzem valores levemente diferentes de probabilidades de invasão, é calculada a média de várias realizações de T_{ij} , obtendo-se $\langle P_{\text{inv}}(k, a) \rangle$.

Os resultados de simulações numéricas para a média de probabilidade de invasão para uma rede de tamanho finito $L \times L$, $\langle P_{\text{inv}}(k, a; L) \rangle$, são apresentados na Figura A.4, com $L = 200$. Como esperado, para valores grandes de espaçamento entre os nós e pequenas taxas de eficiência de transmissão, o sistema é inadequado para a invasão patogênica. Por

outro lado, espaçamentos pequenos e altos valores de k aumentam a probabilidade de invasão e, portanto, de ocorrer uma epidemia. A superfície da probabilidade de invasão $\langle P_{inv}(k, a; L) \rangle$ mostrada na Figura A.4 depende do espaçamento entre os nós L . Quando o tamanho do sistema se torna bastante grande, $L \rightarrow \infty$, os valores de probabilidade de invasão tendem a uma certa superfície limite. A linha dos pontos críticos $a_c = a_c(k_c)$ (a curva sólida no plano horizontal na Figura A.4) separando os regimes não-invasivo ($\lim_{L \rightarrow \infty} \langle P_{inv}(k, a; L) \rangle = 0$) e invasivo ($\lim_{L \rightarrow \infty} \langle P_{inv}(k, a; L) \rangle > 0$) é de particular importância, pois permite identificar a região paramétrica segura (não-invasiva).

Figura A.4 – Dependência da média da probabilidade de invasão $\langle P_{inv}(k, a) \rangle$ na eficiência de transmissão k (escala log) e no espaçamento a entre os nós da rede. A média considerou 1000 realizações numa rede de 200×200 . A curva no plano $a \times \log(k)$ representa a borda de fase entre os regimes invasivo e não-invasivo de uma epidemia SIR. Os valores críticos de k dado a foram obtidos por colapso de escala de tamanho finito, com erro relativo menor que 2%.

Os pontos da fronteira de fases ($a_c = a_c(k_c)$) foram obtidos por colapso de escala de tamanho finito (*finite-size scaling collapse*) da seguinte maneira. A dependência entre o parâmetro de ordem e o tamanho do sistema por volta do ponto crítico (k_c) para a transição de fase contínua é bem estabelecida (89):

$$\langle P_{inv}(k, a; L) \rangle = L^{-\beta/\nu} \tilde{P}_{inv}((k - k_c)L^{1/\nu}), \quad (\text{A.9})$$

em que $\tilde{P}_{inv}$ é a função de escala, e β e ν são os expoentes universais de escala. A função de escala pode ser encontrada por colapso de escala de $\langle P_{inv}(k, a; L) \rangle$ para vários tamanhos de sistema usando k_c e os expoentes de escala como parâmetros de alinhamento (90). Isto é feito plotando $\langle P_{inv}(k, a; L) \rangle L^{\beta/\nu}$ versus $(k - k_c)L^{1/\nu}$ para vários valores de L e encontrando os valores de ν , β e k_c que levam ao colapso de todas as curvas de diferentes L para uma única curva, que corresponde a $\tilde{P}_{inv}$. Esta análise foi utilizada para encontrar os valores críticos de k_c para os dados espaçamentos entre os nós e os resultados

são mostrados através da linha no plano horizontal na Figura A.4. Os colapsos de escala também fornecem os valores dos expoentes de colapso, $\beta = 0,13 \pm 0,008$ e $\nu = 1,32 \pm 0,05$, que coincidem com os da classe de universalidade de percolação isotrópica (89).

A.4.2 Correlações

As probabilidades de invasão apresentadas na Figura A.4 para uma rede heterogênea com transmissibilidades T_{ij} foram calculadas numericamente estimando a fração média de nós pertencentes ao cluster infinito no problema de percolação por ligações. Alternativamente, pode-se assumir que um sistema heterogêneo é equivalente a um homogêneo com transmissibilidade média $\langle T \rangle$ (Equação A.4), o que é verdade se não há correlações entre as transmissibilidades. Foi calculada a dependência entre probabilidade de invasão e transmissibilidade média para os sistemas heterogêneo e homogêneo (respectivamente, curvas sólida e tracejada na Figura A.5). Pode-se visualizar que o valor crítico de transmissibilidade no sistema de raízes reais é maior que no sistema homogêneo. Portanto, o sistema real com heterogeneidade inerente nas transmissibilidades é menos vulnerável a invasão epidêmica que o sistema homogêneo com a mesma transmissibilidade média. Isto implica a presença de correlações nas transmissibilidades entre indivíduos no sistema (28, 87, 91). Deve-se notar que estas correlações são de natureza diferente das espaço-temporais na densidade de indivíduos infectados/removidos observadas durante a propagação da doença. As correlações de curta distância nas transmissibilidades tornam a epidemia menos provável de ocorrer no sistema heterogêneo do que num sistema homogêneo para valores de $P_{\text{inv}} \lesssim 0,95$.

Para valores relativamente grandes de transmissibilidade média (por exemplo, $\langle T \rangle = 0,62$ na Figura A.5), a probabilidade de invasão para um sistema heterogêneo é maior que para o homogêneo. Isto é uma consequência das correlações negativas na transmissão (ver Figura A.5). Este comportamento está em contraste com o de sistemas que não possuem correlações negativas na transmissão, como um sistema de tempos de recuperação heterogêneo. Neste caso, pode-se rigorosamente comprovar que a probabilidade de invasão num sistema homogêneo não pode ser maior que a de um sistema heterogêneo para qualquer valor de $\langle T \rangle$ (92–95). No entanto, este teorema não se aplica a sistemas com correlações negativas, como o sistema de raízes estudado aqui.

A diferença nas curvas de invasão para sistemas heterogêneos e homogêneos (curvas sólida e tracejada na Figura A.5) é uma clara indicação de correlações nas transmissibi-

lidades para sistemas heterogêneos. O gráfico menor interno exibido na Figura A.5 é a função de correlação espacial para transmissibilidades:

$$g(r) = \frac{\langle T_{ij}T_{km} \rangle - \langle T_{ij} \rangle \langle T_{km} \rangle}{\langle T_{ij} \rangle \langle T_{km} \rangle} = \frac{\langle T_{ij}T_{km} \rangle}{\langle T \rangle^2} - 1, \quad (\text{A.10})$$

que mostra a presença de correlações negativas de curta distância (entre ligações próximas) em $r/a = 1$. Na Equação A.10, r é a distância entre os pontos médios de diferentes ligações $i - j$ e $k - m$, e a média é calculada considerando todos os pares de ligações de mesma distância.

Figura A.5 – Probabilidade de invasão P_{inv} versus transmissibilidade média $\langle T \rangle$ para sistemas realísticos de raízes complexas, nos quais as transmissibilidades entre dois indivíduos são calculadas de acordo com a Equação A.8 (curva sólida preta), e para o sistema homogêneo, no qual todas as transmissibilidades assumem o valor da transmissibilidade média $\langle T \rangle$ (curva tracejada vermelha). O espaçamento entre os nós da rede é $a = 120$ px e os valores de todos os outros parâmetros são os mesmos utilizados na Figura A.4. No gráfico menor interno, pode-se visualizar a função de correlação espacial $g(r)$ para transmissibilidades avaliadas no ponto crítico de eficiência de transmissão $k_c(a = 120) \cong 0,00344$.

A.4.3 Modelos baseados em discos

Para compreender a origem das correlações de curta distância nas transmissibilidades entre raízes, foram estudados três modelos simplificados (“*toy models*”) de variada

complexidade. Em todos estes modelos, cada raiz i é substituída por um disco bidimensional horizontal, com densidade $\rho_i(r)$ e centro de massa em r_i . Os centros dos discos são aleatoriamente deslocados dos nós (caracterizados pelos vetores de posição $r_i^{(0)}$), ou seja, $r_i = r_i^{(0)} + \Delta r_i$, nos quais os vetores de deslocamento são distribuições normais $\Delta r_i \sim N(0, \sigma_\Delta^2)$, com valor zero de esperança e valor σ_Δ^2 de variância. A distribuição normal de Δr_i se assemelha à anisotropia na forma das raízes reais, ou seja, o fato de que os centros de massa das raízes estão deslocados em relação à posição da semente (nó). Os modelos de disco se diferem entre si pela forma da função usada por $\rho_i(r)$ e, por consequência, pelas sobreposições entre dois discos.

Figura A.6 – Modelos baseados em discos (“toy models”): (a) discos sólidos deslocados; (b) discos dispersos deslocados; (c) discos deslocados com bordas irregulares; (d) discos não-deslocados com bordas irregulares.

No primeiro modelo (Figura A.6(a)), todos os discos possuem o mesmo raio R e densidade constante $\rho_i(r) = \rho_0$ se $|r - r_i| \leq R$, ou zero caso contrário. A sobreposição

entre dois discos i e j com centros separados pela distância $d_{ij} = |r_i - r_j|$ é:

$$J_{ij}^{(1)} = V_{\text{px}} \int dr \rho_i(r) \rho_j(r) \quad (\text{A.11})$$

$$= 2\rho_0^2 R^2 \left[\cos^{-1} \left(\frac{d_{ij}}{2R} \right) - \left(\frac{d_{ij}}{2R} \right) \sqrt{1 - \left(\frac{d_{ij}}{2R} \right)^2} \right], \quad (\text{A.12})$$

se $d_{ij}/2R \leq 1$ ou zero caso contrário (com a área em pixels $V_{\text{px}} = 1$). Este é um dos modelos mais simples, correspondente à representação de raízes reais por cilindros verticais uniformes idênticos projetados em um plano horizontal.

O segundo modelo (Figura A.6(b)) considera a propriedade das raízes reais de possuir um decaimento na densidade conforme se distancia do ramo principal. Neste modelo, a densidade segue uma distribuição normal ao redor de r_i com variância σ_R^2 , ou seja:

$$\rho_i(r) = \tilde{\rho} (2\pi\sigma_R^2)^{-1} \exp \left(-\frac{|r - r_i|^2}{2\sigma_R^2} \right), \quad (\text{A.13})$$

onde $\tilde{\rho}$ é uma constante de normalização adimensional, e a sobreposição entre dois discos dispersos é representada pela função Gaussiana:

$$J_{ij}^{(2)} = \frac{\tilde{\rho}^2}{4\pi\sigma_R^2} \exp \left(-\frac{d_{ij}^2}{4\sigma_R^2} \right). \quad (\text{A.14})$$

O terceiro modelo (Figura A.6(c)) procura reproduzir a irregularidade angular na forma das raízes reais, ou seja, o fato da sobreposição depender significativamente do ângulo φ no qual as raízes são colocadas nos nós da rede (Figura A.3(a)). Neste modelo, assume-se que a sobreposição possui duas contribuições:

$$J_{ij}^{(3)} = J_{ij}^{(1)} + \delta J_{ij}, \quad (\text{A.15})$$

em que $J_{ij}^{(1)}$ é dada pela Equação A.12 do primeiro modelo e a distribuição normal aleatória $\delta J_{ij} \sim N(0, \sigma_J^2)$, com variância σ_J^2 , que depende de $J_{ij}^{(1)}$.

Para ilustrar que somente a irregularidade na forma é insuficiente para causar o efeito observado, também foi considerado um quarto modelo, que é um caso especial do terceiro, no qual não há deslocamento dos discos (ou seja, $\sigma_\delta^2 = 0$). Neste caso, pode-se verificar que não há correlações (ver o gráfico de correlações inserido na Figura A.7) e, portanto, o sistema pode ser aproximado por um sistema homogêneo equivalente (as curvas do quarto modelo e do sistema homogêneo coincidem na Figura A.7).

Os modelos baseados em discos dependem de diversos parâmetros, como R , σ_δ^2 , σ_R^2 e σ_J^2 . As duas constantes de normalização ρ_0 e $\tilde{\rho}$ podem ser incorporadas no valor de eficiência de infecção k e são configuradas, por conveniência, em $\rho_0 = \tilde{\rho} = 1$. Os valores

Figura A.7 – Curvas de invasão para o sistema homogêneo (curva sólida com cruzeiros), sistema de raízes reais (sólido), modelo de disco sólido (tracejado), modelo de disco disperso (ponto-traço) e discos com bordas irregulares (ponto duplo - traço). São mostrados no gráfico menor interno as funções de correlação no intervalo de valores negativos (são usados os mesmos estilos de curva).

dos raios dos discos R no primeiro modelo e a largura das Gaussianas σ_R no segundo modelo podem ser estimados do valor médio (entre todas as N_r raízes) da variância da densidade das raízes reais projetadas no plano horizontal, ou seja:

$$R_2 = \sigma_R^2 = N_r^{-1} \sum_i^{N_r} \delta r_{\perp i}^2, \quad (\text{A.16})$$

em que

$$\begin{cases} \delta r_{\perp i}^2 = \int dr (r - r_i)_{\perp}^2 \rho_i(r) \\ r_i = \int dr (r - r_i^{(0)}) \rho_i(r) \end{cases}, \quad (\text{A.17})$$

com $\rho_i(r)$ dado pela Equação A.4 e o sufixo $\perp$ referindo-se ao vetor perpendicular ao eixo vertical. Este valor foi calculado como sendo $R = \sigma_R \simeq 73$ px. Similarmente, o valor de σ_{Δ}^2 foi estimado como sendo:

$$\sigma_{\Delta}^2 = N_r^{-1} \sum_i^{N_r} (r_i - r_i^{(0)})^2 \simeq 1444 \text{px}. \quad (\text{A.18})$$

O valor de σ_J^2 pode ser estimado a partir da variância da sobreposição $J_{ij}(\varphi_i, \varphi_j)$ entre duas raízes reais, que dependem estocasticamente de suas orientações, caracterizadas por φ_i e φ_j (Figura A.3(a)), e foi calculado como sendo $\sigma_J^2 / \langle J_{ij}^{(1)} \rangle \simeq 1,35$.

Com as estimações de todos os parâmetros para os modelos de discos, pode-se calcular numericamente a dependência da probabilidade de invasão na transmissibilidade média. Os resultados mostrados na Figura A.7 demonstram que, para todos os modelos que exi-

bem anisotropia (1, 2, 3 e raízes reais), o ponto crítico T_c é deslocado para a direita do ponto crítico do sistema homogêneo e a probabilidade de invasão torna-se menor na faixa $0 \leq P_{\text{inv}} \lesssim 0,95$, quando comparada ao sistema homogêneo (deve-se notar que os modelos 2 e 3 reproduzem o sistema de raízes reais satisfatoriamente bem). Isto é consequência das correlações de curta distância nas transmissibilidades presentes nos modelos de disco. O grande deslocamento no primeiro modelo quando comparado com os menos significativos deslocamentos dos modelos 2 e 3 reflete a respectiva redução na escala das correlações nestes últimos modelos. As relativamente grandes correlações no primeiro modelo são devido às bordas rígidas dos discos, que tornam as sobreposições significativamente grandes (menores) para discos espacialmente próximos (distantes) quando comparadas com o valor médio. Esta diferença é reduzida utilizando Gaussiana na densidade dos discos, tornando as correlações do segundo modelo menos acentuadas. As flutuações angulares de densidade no modelo 3 resultaram numa redução das correlações similarmente encontrada nos discos sólidos.

Pode-se concluir da análise dos modelos de disco que a anisotropia na forma das raízes é em grande parte responsável pelas correlações de curta distância nas transmissibilidades entre raízes. A anisotropia surge principalmente dos deslocamentos horizontais do centro de massa relativo à localização do nó (semente) na rede. Este efeito é em grande parte captado pelo primeiro modelo, que exhibe os valores mais significativos de correlação. Já a dispersão e a irregularidade nas formas das raízes, reproduzidas pelos modelos 2 e 3, reduzem o nível de correlações. E se os discos irregulares não forem deslocados dos nós (modelo 4), as correlações desaparecem completamente e o modelo torna-se equivalente ao sistema homogêneo.

A.4.4 Anisotropia global

Nas seções anteriores, as raízes foram dispostas na rede sem uma orientação predominante, ou seja, sem anisotropia global na orientação das raízes. Desta forma, considera-se que não há nenhum campo gradiente de recursos externo as atraindo. No entanto, este campo pode estar presente e causar um crescimento predominantemente numa certa direção. Em plantações, este efeito pode ser causado em resposta a variações de distribuições no solo de água, nutrientes ou temperatura, por exemplo.

Para estudar o efeito de um campo de recursos externo, foi assumido que o campo é aplicado ao longo do eixo x horizontal. Este campo força as raízes crescerem predomina-

temente nesta direção. As raízes são localmente anisotrópicas em relação a rotações em torno do eixo vertical passando pelo nó. Isto significa que a densidade da raiz depende do ângulo polar φ (no plano $x - y$) e possui um valor máximo numa certa direção. Para reproduzir o efeito do campo de recursos externo, foi disposta a mesma raiz (escolhida entre as nove) em cada nó da rede de forma a inicialmente alinhar a direção de densidade máxima com $\varphi = 0$, ou seja, ao longo do eixo x . Em seguida, cada raiz foi aleatoriamente rotacionada por um ângulo φ de acordo com a função densidade probabilidade:

$$f_n(\varphi) = (2n + 1 - \delta_{n,N})^{-1} \left[\sum_{k=-n}^n \delta(\varphi + k\varphi_0) - \delta_{n,N} \delta(\varphi - \pi) \right], \quad (\text{A.19})$$

em que $\varphi_0 = \pi/N$ e $0 \leq n \leq N$ (com $N, n \in \mathbb{N}$). Esta função descreve rotações discretas uniformemente distribuídas no intervalo $\varphi \in [-\Delta\varphi, \Delta\varphi]$, com $\Delta\varphi = n\varphi_0$ e φ_0 sendo o ângulo de rotação elementar que divide π em N partições. A função $f_n(\varphi)$ permite que sejam estudadas variados graus de força do campo externo. Se $\Delta\varphi = \pi$, todas as orientações das raízes são equivalentemente possíveis e, portanto, o campo externo não tem influência no sistema. Neste caso, não há anisotropia global, mas a anisotropia local, definida pela forma da raiz, continua presente. No caso oposto, $\Delta\varphi = 0$, todas as raízes estão orientadas com o eixo x , de tal forma que as anisotropias locais coerentemente induzem anisotropia global na população.

A dependência da probabilidade de invasão na transmissibilidade média para duas raízes representativas (raiz 1 e raiz 4), com variados valores de $\Delta\varphi$ é mostrada na Figura A.8. Estas curvas de invasão foram numericamente calculadas a partir da média de 10^3 realizações de epidemias num sistema de tamanho finito ($L = 200$).

As curvas de invasão dependem da força do campo de recursos externo, definido pelo parâmetro $\Delta\varphi$. Consequentemente, os valores limiares para a transmissibilidade média também dependem de $\Delta\varphi$. Esta dependência é exibida para várias raízes na Figura A.9. No caso de um campo forte, correspondente à anisotropia global máxima, $\Delta\varphi = 0$, o processo SIR pode ser mapeado num problema de percolação por ligações anisotrópico, caracterizado por duas transmissibilidades distintas T_x e T_y , correspondentes às sobreposições horizontais e verticais, respectivamente. Neste caso, o valor limiar de transmissibilidade médio pode ser exatamente encontrado (96) pela equação $T_x + T_y = 1$ (que é válida no regime crítico). Portanto, na criticalidade $\langle T \rangle = (T_x + T_y)/2 = 1/2$, o que coincide com o limiar de um sistema homogêneo (como pode ser visto na Figura A.9).

No caso oposto extremo, $\Delta\varphi = \pi$, quando não há influência de um campo de recursos externo, o limiar de invasão é tipicamente encontrado em torno ou acima do limiar do sistema homogêneo, ou seja, $T_c(\Delta\varphi = \pi) \gtrsim T_{c,\text{hom}} = 1/2$. No entanto, não se pode excluir a princípio a possibilidade do limiar de invasão estar abaixo do valor homogêneo. De

Figura A.8 – Dependência da probabilidade de invasão P_{inv} na transmissibilidade $\langle T \rangle$ para duas raízes, dispostas numa rede regular quadrada ($L = 200$). As curvas correspondem a variados graus de força do campo de recursos externo, determinado pelo parâmetro $\Delta\varphi$ (como pode ser visto na legenda). A curva sólida com cruzeiros representa um sistema homogêneo (população de raízes com transmissibilidades iguais).

Figura A.9 – Dependência entre o limiar de invasão T_c e o ângulo de abertura $\Delta\varphi$ da distribuição de rotações aleatórias das raízes, como definido pela Equação A.19.

fato, foi constatado que a curva de invasão da raiz 4 localiza-se à esquerda da curva do caso homogêneo (ver linha com duplo tracejado e ponto, e curva sólida com cruzeiros na Figura A.8). Para forças intermediárias do campo de recursos externo, o comportamento do limiar de invasão é não-monotônico e T_c pode ficar abaixo de $T_{c, \text{hom}}$ (veja as curvas das raízes 2, 4 e 7 na Figura A.9 ou permanecer acima (veja as curvas das raízes 1 e 9). Os valores de T_c na Figura A.9 foram calculados utilizando colapso de escala de tamanho

finito (*finite-size scaling collapse*).

Portanto, foi verificado que o campo de recursos externo pode fazer com que a população de raízes se torne menos resiliente a invasões de epidemias SIR, se comparado com o caso sem campo externo. Isto pode ser verificado nas dependências $T_c(\Delta\varphi)$ para as raízes 2, 4 e 7 na Figura A.9, na qual $T_c < T_{c,\text{hom}} = 1/2$ para valores intermediários de $\Delta\varphi$. Também foi observado que um campo de recursos externo bastante forte não reduz o limiar abaixo do valor para um sistema homogêneo, $T_c(\Delta\varphi = 0) = T_{c,\text{hom}}$, e somente campos menos fortes, que permitem alguma flutuação angular local das raízes, podem ser perigosos no sentido de que $T_c < T_{c,\text{hom}}$ (o ponto crítico limiar da transmissibilidade para que ocorra uma epidemia é menor do que num sistema homogêneo). O incremento na desordem das orientações locais faz com que o sistema se torne globalmente isotrópico (sem predominância em alguma direção), o que aumenta o limiar de invasão, de tal forma que $T_c(\Delta\varphi = \pi)$ é tipicamente em torno ou maior que $T_{c,\text{hom}}$.

Como a redução do limiar de invasão abaixo do valor homogêneo pode indicar a presença de correlações negativas nas transmissibilidades (92, 93), foi calculada a função de correlação dos mesmos sistemas analisados na Figura A.8. Pela Figura A.10, pode-se constatar a presença de correlações negativas em todos os casos (além de correlações positivas também). No entanto, a presença de correlações negativas não é suficiente para reduzir o limiar de invasão abaixo do valor homogêneo (que ocorre com a raiz 4) e, como resultado de uma complexa interação entre as correlações de diferentes sinais, $T_c \geq T_{c,\text{hom}}$ para a raiz 1.

Outra observação levantada pelos resultados é que populações de raízes com forte anisotropia local (raízes 1 e 9) sem um campo de recursos externo são mais resilientes à invasão epidêmica do que um sistema homogêneo. O campo de recursos externo reduz a resiliência destes sistemas, trazendo-as ao nível do sistema homogêneo para campos bastante fortes (veja as curvas correspondentes na Figura A.9).

A.5 Considerações finais

A ocorrência de transição de fase bem definida entre os estados de não-invasão e invasão de um agente patogênico (epidemia) numa rede de indivíduos homogêneos é bem conhecida (85). Após o trabalho inicial de Bailey (66) sobre a propagação do fungo *R. solani* no solo, os fenômenos de percolação crítica e transição de fase têm sido verificados

Figura A.10 – As funções de correlação para as mesmas raízes analisadas na Figura A.8. A função de correlação da raiz 1 com $\Delta\varphi = 0$ oscila entre -0.57 and 0.57 (não mostrada devido à incompatibilidade na escala).

em análise experimental. Este trabalho estendeu a análise teórica do fenômeno de percolação crítica para levar em conta as correlações espaciais que ocorrem no conjunto de indivíduos tridimensionais de morfologia complexa, representados pelas raízes.

Foram utilizadas imagens digitalizadas de raízes para prover exemplos realísticos de morfologias complexas de estruturas ramificadas para testar a metodologia. Os resultados e abordagens metodológicas, em particular os novos modelos baseados em disco, forneceram uma base para novas análises de transmissão de doenças em uma ampla gama de estruturas morfológicas. Também foi avaliado o efeito de um campo gradiente de recursos externo no limiar de invasão (anisotropia global), que tende a atrair a direção de crescimento da raiz e, portanto, influencia na direção em que há maior densidade.

As principais descobertas podem ser sumarizadas como:

1. A morfologia dos indivíduos representadas por raízes tridimensionais é um importante fator para a epidemia SIR. Duas características morfológicas, anisotropia e desordem, são de particular importância para a disseminação da epidemia.
2. A anisotropia se refere principalmente ao desvio da direção vertical na raiz pivotante, o que desloca o centro de massa da raiz para fora da posição da semente (nó). Assim, a anisotropia causa correlações de curta distância na transmissão de doenças que fazem o sistema menos vulnerável a invasões de epidemias.
3. A desordem é uma característica relacionada ao grau de ramificação da raiz, que afeta sua distribuição de densidade, o que torna as sobreposições mais suaves e reduz

as correlações. Desta forma, a desordem diminui o efeito de anisotropia e, portanto, incrementa a vulnerabilidade do sistema a um surto epidêmico.

4. A anisotropia global influencia a probabilidade de invasão e faz com que a população geralmente fique menos resistente a epidemias SIR.

O estudo da influência da anisotropia global (Seção A.4.4) foi publicado no congresso CompleNet 2010 (*Communications in Computer and Information Science* - CCIS) (97), enquanto que o restante do trabalho foi publicado na revista *Journal of the Royal Society Interface* (98).