UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

FILIPPI NASCIMENTO SILVA

Dimensão e simetria em redes complexas: uma abordagem
multiescala

São Carlos

2015

FILIPPI NASCIMENTO SILVA

Dimensão e simetria em redes complexas: uma abordagem multiescala

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

Área de concentração: Física Aplicada
Opção: Física Computacional
Orientador: Prof. Dr. Luciano da Fontoura Costa

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2015

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

Ficha catalográfica revisada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Silva, Filipi Nascimento

Dimensão e simetria em redes complexas: uma abordagem multiescala / Filipi Nascimento Silva; orientador Luciano da Fontoura Costa - versão corrigida -- São Carlos, 2015.
176 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Computacional) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2015.

1. Redes complexas. 2. Dimensão. 3. Simetria. 4. Física interdisciplinar. 5. Visualização. I. Costa, Luciano da Fontoura, orient. II. Título.

Dedico este trabalho à memória de minha querida avó Nélia, pelo amor e carinho eterno.

Agradecimentos

Agradeço ao professor Luciano da Fontoura Costa, pela oportunidade de poder realizar pesquisas científicas interessantes, por sua orientação séria e competente, pela confiança e pela possibilidade de realização deste trabalho.

À minha mãe Edna e à minha tia Eliana, pelo amor, carinho, apoio e paciência.

Ao professor Chu e à Maria Bardosova, pelas oportunidades de colaboração e discussões que permitiram enriquecer nossos trabalhos conjuntos.

Ao professor Carlos da FMUSP, pela conversas, pelos ensinamentos e pelas oportunidades de colaboração.

Aos amigos, André, Celso, Danilo, Diego, Fabio, Mariane e Rejane; pelos momentos de descontração. Almoços semanais com alguns e partidas de FIFA ou MKX com outros.

Aos colegas e ex-colegas do grupo de visão cibernética, Bruno, Cynthia, César, Henrique, Gustavo e Matheus; pelo apoio e discussões acadêmicas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo suporte em trabalhos vinculados ao projeto temático de número 11/50761-2.

Ao IFSC-USP pela infraestrutura apoio acadêmico, em especial aos funcionários da CPG: Ricardo e Patricia, pela atenção e por terem sempre uma solução para os nossos problemas acadêmicos do dia a dia.

Às bibliotecárias do IFSC, em especial à Neusa pela simpatia e ajuda nos últimos estágios do desenvolvimento deste trabalho.

*“Eu acho que se você faz algo e isso fica muito bom,
então você deve ir fazer outra coisa incrível,
não se apoie muito no sucesso anterior,
apenas descubra o próximo passo.”*
(Steve Jobs)

*“Nuvens não são esferas,
montanhas não são cones,
litorais não são círculos,
e a casca das árvores não é lisa,
tampouco a luz viaja em linha reta.”*
(Benoit Mandelbrot)

Resumo

Silva, F.N. **Dimensão e simetria em redes complexas:** uma abordagem multiescala. 2015. 176 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, 2015.

Uma nova perspectiva da ciência tem sido adotada com sucesso por diversos pesquisadores ao redor do mundo, resultando em avanços tecnológicos nas mais variadas áreas do conhecimento. A nova abordagem científica é amplamente interdisciplinar e envolve métodos para a análise, armazenamento e visualização eficiente de grandes quantidades de informação, assim como formas de entender e quantificar a própria estrutura das ciências. Entretanto, os novos problemas tratados por esta abordagem também são mais complexos e tendem a apresentar uma intrincada união entre conceitos e suas relações. Redes complexas representam naturalmente sistemas desta natureza, sendo amplamente adotadas por pesquisadores de diversas áreas científicas. Este trabalho tem como objetivo estudar e desenvolver técnicas multiescala para caracterizar a estrutura de redes complexas através de vários aspectos ligados à complexidade, como a dimensão, a regularidade e a organização de sistemas em subestruturas. Para isso, são sugeridas novas métricas baseadas no conceito de níveis e padrões concêntricos, que permitem estender, de forma natural, medidas tradicionais de redes complexas ao contexto multiescala. Desta forma, é introduzida uma nova medida de dimensão multiescala, que permite a quantificação da dimensão, tanto ao redor de vértices quanto em escalas globais de uma rede. A regularidade de redes é abordada através da ideia de simetria, redefinida em termos da simetria concêntrica, que permitiu, na prática, a quantificação e comparação da simetria entre diferentes redes. Outro conceito explorado é a caracterização do papel desempenhado por vértices dentro de comunidades, que estão diretamente relacionados às subestruturas de sistemas. Para isso, definimos uma nova medida multiescala de pertinência baseada na diversidade das comunidades presentes na vizinhança de um vértice. Também foram utilizadas e desenvolvidas técnicas e ferramentas para visualização de redes complexas, que se mostraram fundamentais durante todos os estágios deste trabalho, em especial na criação de mapeamentos científicos. A união das metodologias enriqueceu a caracterização de redes complexas, permitindo a criação de um mapa da complexidade desses sistemas. Verificou-se que a complexidade de sistemas, em geral, está associada à dimensão e simetria. Sistemas com baixa simetria e alta dimensionalidade tendem a ser aqueles de maior complexidade compartilhando a mesma região que redes de conhecimento de da infraestrutura da Internet. Sob esses aspectos, também é estudado como redes são influenciadas pelas restrições geográficas ou de planejamento. As metodologias introduzidas neste trabalho culminaram em diversas aplicações em vários segmentos como: classificação de genes em redes de transcriptoma humano, detecção de crises no mercado financeiro, visualização e sumarização automática de dados científicos,

assim como a quantificação da interdisciplinaridade de revistas científicas. No geral, os desenvolvimentos presentes aqui mostram o potencial da abordagem multiescala na análise de sistemas complexos através da representação por redes.

Palavras-chave: Redes complexas. Dimensão. Simetria. Física interdisciplinar. Visualização.

Abstract

Silva, F.N. **Dimension and symmetry in complex networks:** a multiscale approach. 2015. 176 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, 2015.

A new perspective of making science is being adopted by many researchers around the world, resulting in several technological achievements in many research fields. The new approach to science is largely interdisciplinary and wraps together many new methods used to analyze, store and visualize large amounts of data, as well as techniques to understand the structure of science itself. However, the problems usually handled by this new approach also are very complex and present an intricate collection of concepts and relationships. Complex networks can be used to naturally represent such systems, having being adopted by many researchers of different backgrounds. The objective of this work is to study and develop new multiscale techniques to characterize the structure of complex networks under three different aspects: dimension, regularity and the organization of systems in substructures; which are tightly related to the concept of complexity. To conduct such analysis, new measurements based on the idea of concentric levels and patterns are proposed, which permits traditional network properties to be extended to a multiscale approach. Therefore, we introduce a new measurement of dimension, the multiscale dimension, which can be used to quantify the dimension around the neighborhoods of a vertex, as well as over the entire network structure in a global scale. The regularity is considered in terms of symmetry, which is redefined to create the concentric symmetry, allowing the quantification and comparison of symmetry among networks of distinct sizes and connectivity. Another concept explored in this work is the characterization of the role of vertices inside a community, which represents the substructure organization of the underlying system. This is done by defining another multiscale measurement, in this case, based on the diversity of the communities of vertices in the neighborhoods of a central reference. Also, we developed new tools and techniques to visualize complex networks, which have proven essential to all the steps of the analysis conducted in this work, including the task of mapping the science structure. The combination of the proposed methodologies enriched the analysis and the characterization of complex networks, resulting in the development of a new way of mapping complexity based on dimension and symmetry. Systems presenting low symmetry and high dimensionality are more likely to also display a more complex behavior, this is the case of knowledge networks the infrastructure of the Internet. In this context, we also study the effects of spatial restrictions and system planning on networks. The methodologies introduced in this paper culminated in several other applications in many scientific topics, such as: the classification of genes in human transcriptome networks, crisis detection on the financial market network, visualization and automatic summarization of scientific data, as well as the quantification of interdisciplinary-

ity of scientific journals. In general, the contributions of this work demonstrate the potential of the multiscale analysis applied to understanding complex systems through the networks representation.

Keywords: Complex networks. Dimension. Symmetry. Interdisciplinary physics. Visualization.

Lista de Figuras

Figura 1	– Representações do problema das sete pontes da cidade de Königsberg.	30
Figura 2	– Rede de linhas de transmissão de energia da Europa (EU) (a). Cada vértice corresponde a uma estação ou usina, enquanto as arestas representam linhas de transmissão, mais informações sobre esta rede podem ser vistas no capítulo 5. Em (a) a posição no espaço original é mostrada ao longo do mapa da Europa. A visualização em (b), no entanto, é gerada apenas a partir da informação topológica da rede, isto é, desconsiderando posições e distâncias entre cada vértice. O método de visualização usado é apresentado no capítulo 3. Em ambos os painéis as cores indicam o país de origem de cada estação/usina, de acordo com a legenda.	32
Figura 3	– Entrelaçamento dos principais assuntos presentes nesta tese. Na camada mais interna é mostrada como um grafo, composto por vértices indicando as metodologias e conceitos teóricos usados ao longo do desenvolvimento deste trabalho. As arestas representam uma forte relação de dependência ou afinidade entre os conceitos. A segunda camada é composta por uma série de aplicações dos conceitos apresentados, enquanto a terceira camada fixa compõe as principais contribuições e objetivos deste trabalho. A última camada indica o relacionamento das aplicações aos conceitos de complexidade explorados nesta tese: dimensão, regularidade e participação de vértices em comunidades.	34
Figura 4	– Exemplos de redes complexas reais.	37
Figura 5	– Principais tipos de grafos.	38
Figura 6	– Exemplo de representação de uma rede não direcionada com pesos por matriz de adjacência e lista de adjacências.	40
Figura 7	– Propriedades locais de um vértice: grau (k) e coeficiente de aglomeração (C_c).	41
Figura 8	– Propriedades globais tradicionais em redes complexas. Para cada vértice é mostrado o grau e o coeficiente de aglomeração. No topo à esquerda as propriedades globais da rede são mostradas: grau médio, coeficiente de aglomeração médio e transitividade.	43
Figura 9	– Exemplo de componentes conectados em uma rede simples e em uma rede direcionada. Cada componente é destacado pela cor e destacado pelo contorno.	44
Figura 10	– Geodésicas existentes entre os vértices v_1 e v_9 mostradas em destaque laranja, $\ell_{19}^g = (v_1, v_3, v_7, v_9)$ e verde, $\ell_{19}^{g'} = (v_1, v_3, v_8, v_9)$.	45

Figura 11 – Geodésicas usadas para o cálculo da centralidade de intermediação (<i>betweenness centrality</i>) sem normalização em uma rede simples. Linhas mais finas representam as geodésicas, coloridas de acordo com a legenda, indicando a participação dos vértices A ou B. As setas indicam o início ou fim da geodésica. Neste caso a centralidade pode ser facilmente calculada pela metade da soma dos caminhos entrando no vértice (isto é, não contando os caminhos finalizando no vértice).	46
Figura 12 – Principais modelos de redes complexas usados neste trabalho.	47
Figura 13 – Duas formas de truncar a borda de uma grade 2D: (a) com efeito de bordas e (b) sem efeito de bordas. As cores destacam o grau dos vértices.	48
Figura 14 – Distribuição do grau para uma rede ER de 25 mil vértices e grau médio 12.	49
Figura 15 – Distribuição de grau para uma rede BA com 25 mil vértices e grau médio 50.	50
Figura 16 – Estágios para a geração de uma rede BA com 10 vértices e $\langle k \rangle = 4$	51
Figura 17 – Estágios para a geração de uma rede WS com 10 vértices e $\langle k \rangle = 4$	52
Figura 18 – Coeficiente de aglomeração médio (quadrados em laranja) e média das geodésicas (círculos azuis) contra a probabilidade de reconexão. A região destacada corresponde às propriedades geralmente presentes em redes do tipo pequeno mundo, com alto coeficiente de aglomeração e baixa média das geodésicas.	53
Figura 19 – Tipos de redes geográficas usadas neste trabalho. Nota-se que, diferentemente da rede geográfica aleatória (a), a rede Waxman (b) apresenta alguns poucos caminhos de longo alcance.	54
Figura 20 – Exemplo de estrutura de comunidades em uma rede complexa (a). Há preferência entre conexões de vértices dentro da mesma comunidade. O grafo reduzido (<i>coarse-grained</i>) (b) representa as relações de conectividade entre as comunidades. A largura das arestas neste grafo indicam a probabilidade estocástica de dois vértices de comunidades diferentes se conectarem.	56
Figura 21 – Exemplo de isomorfismo em uma rede simples. Os rótulos de vértices da rede à esquerda podem ser associados aos da direita de modo que a topologia das duas redes não seja alterada. Portanto, as duas redes são isomorfas entre si. Propriedades topológicas consideradas em ambas as redes são indistinguíveis a não ser pelo rótulo dos vértices.	57
Figura 22 – Isomorfismo de cores. Enquanto que A, B e C são isomorfos topologicamente, ao considerar as cores, apenas A e B se mantêm isomorfos. . . .	58

Figura 23 – Possíveis automorfismos de uma rede simples. O primeiro automorfismo (a) envolve apenas a troca de rótulos entre 2 pares de vértices. Considerando o posicionamento dos vértices este automorfismo é similar à simetria vertical. Já o segundo automorfismo (b) envolve todos os vértices da rede e se mostra conceitualmente similar a uma simetria horizontal da rede.	58
Figura 24 – Exemplo de PCA para um conjunto de dados de 2 variáveis distribuído de modo normal ao longo de uma reta.	60
Figura 25 – Conceito de caracterização multiescala em redes complexas. Enquanto medidas locais e globais caracterizam redes e vértices apenas em suas respectivas escalas, medidas multiescala podem caracterizar vértices em diferentes escalas, inclusive a intermediária.	62
Figura 26 – Anéis e padrões concêntricos obtidos para $r \leq 2$ a partir de dois vértices escolhidos como centro (α e β) em uma rede simples. Em (a), anéis são mostrados na rede por contornos ao redor dos vértices. Os respectivos padrões concêntricos extraídos da rede para os vértices α e β e $r = 2$ são mostrados em (b).	64
Figura 27 – Exemplo de anéis (ou níveis) concêntricos centrados em 1 considerando um padrão concêntrico de nível $r = 3$	64
Figura 28 – Exemplo do cálculo da acessibilidade para dois padrões concêntricos de 2 níveis. Em (a) os vértices G e F podem ser acessados, a partir de A , com mesma probabilidade $P_2(A, G) = P_2(A, F) = 1/2$, desta forma a entropia é máxima e, portanto, $a_A(r) = 2$. No segundo exemplo (b), a partir de α é muito mais provável que uma caminhada termine em θ do que em ι , resultando em uma distribuição de probabilidades menos homogênea. Neste caso a acessibilidade diminui em comparação com (a), com valor $a_\alpha(r) = 1.89$	67
Figura 29 – Acessibilidade (a) e centralidade de intermediação (b) calculadas para cada vértice de uma rede geográfica aleatória, mostradas de acordo com as cores das legendas. Verifica-se que a acessibilidade se mostra mais eficiente na detecção de bordas, além de apresentar valores mais estáveis ao longo da rede.	68
Figura 30 – Exemplos de diferentes métodos de disposição de vértices aplicados a uma rede simples.	70

Figura 31 – Metodologia FRMOD de visualização de redes por forças. À esquerda é mostrada a rede na configuração inicial, com vértices aleatoriamente distribuídos sobre o plano. As forças de atração (eq. 3.2) e repulsão (eq. 3.3) são indicadas por vetores. Através de MDS a rede chega a um estado de equilíbrio (à direita) com vértices conectados próximos um dos outros e vértices mantendo a estrutura topológica da rede. A visualização final é gerada associando a cor de cada vértice a alguma propriedade, neste caso, o grau.	74
Figura 32 – Particionamento em grade de um espaço $2D$ exemplificando as interações do método Barnes-Hut em FRMOD. Considerando a partição P_{mn} , inicialmente, todos os vértices das partições vizinhas interagem entre si de acordo com o método FRMOD. As interações de longa distância, no entanto, ocorrem apenas entre pares de partições através de aproximação por expansão multipolar, ilustrada em vermelho.	75
Figura 33 – Esquema do fluxo de dados do software Network 3D, ilustrando as principais funcionalidades e módulos do software desenvolvido.	77
Figura 34 – Captura de tela do <i>software</i> Network 3D versão para computadores. . . .	78
Figura 35 – Captura de tela do <i>software</i> Network 3D versão <i>web</i>	79
Figura 36 – Exemplos de visualizações de redes complexas geradas no Network 3D. Rede gerada pelo modelo WS bidimensional (a), a cor indica o grau de cada vértice. Estrutura da Internet (b) (107), com a cor também indicando o grau de cada vértice. Rede de correlações do mercado financeiro para 1989 (108), as cores indicam as comunidades obtidas para a rede. Rede de artigos da Wikipédia (d) em subcategorias pertencentes às grandes áreas do conhecimento, as cores indicam categoria associada a cada artigo	80
Figura 37 – Ilustração da abordagem de análise de dados de expressão gênica através de redes de similaridade.	81
Figura 38 – Exemplos das três categorias de vértices estudados em (9): Hubs seguem a definição usual, sendo aqueles com maior número de conexões na rede; Vips, por outro lado, são vértices que apresentam poucas conexões, mas em geral são conectados a vértices de alto grau k_i , desta forma apresentam alto grau concêntrico de primeiro nível ($k_i(1)$); High-Hubs são vértices que apresentam as duas características, isto é, além de apresentarem muitas conexões também apresentam alto grau de primeiro nível.	83
Figura 39 – Exemplo de visualização de redes e classificação de acordo com as categorias VIPS e HUBS obtida a partir dos dados analisados de expressão gênica	84

- Figura 40 – Exemplo da entropia de pertinência para duas configurações de redes. A rede no topo da figura (a) apresenta maior diversidade de conexão com vizinhos de diferentes comunidades, portanto, apresentando maior valor de entropia de pertinência. Já no segundo caso (b), a maioria dos vizinhos pertencem a uma única comunidade (azul), apresentando baixo valor para a medida. 87
- Figura 41 – Visualização bidimensional da rede de distribuição de energia dos EUA com cores indicando as comunidades de maior pertinência para cada vértice. A entropia de pertinência H_i^* é mostrada através do tamanho dos vértices de acordo com a legenda. 88
- Figura 42 – Representação esquemática da rede de revistas científicas. As revistas são conectadas por arestas com pesos considerando citações de entrada e saída. No exemplo, A está conectada às revistas B , C e D e também possui uma autoconexão já que alguns de seus artigos contém referências internas. Um conjunto de categorias temáticas também é associado a cada revista). 90
- Figura 43 – Evolução da média e mediana de diversas propriedades das redes com o tempo. (a) O tamanho do maior componente da rede de citação é mostrado pela curva vermelha, revelando uma relação praticamente linear com o tempo. O mínimo caminho médio da rede é mostrado em azul, apresentando uma curva monotonicamente decrescente com o tempo. (b), (c) respectivamente apresentam a força média de entrada e a mediana da entropia temática variando no tempo para as diferentes grandes áreas das ciências. Os valores globais, isto é, considerando todos os vértices das redes é destacado pela linha cinza em destaque. 93
- Figura 44 – A imagem principal mostra um mapa científico bidimensional baseado na rede para o ano de 2009 com as cores indicando as áreas do conhecimento dentro dos campos de exatas e biológicas. A tabela no topo destaca as revistas com maior entropia temática, portanto maior interdisciplinaridade. No quadro inferior à direita é mostrada a distribuição da entropia temática para diversos períodos, resultando em uma lei de potência, O coeficiente desta lei tem diminuído com o tempo (como pode ser observado no quadro interno), indicando também que a própria entropia temática tem se tornado mais diversa ao longo do tempo. 94

Figura 45 – Projeção 2D (a) da rede CN com as cores indicando as comunidades obtidas, cada uma nomeada de acordo com as respectivas palavras-chave encontradas. O grafo reduzido (<i>coarse-grained</i>) (b) correspondente à rede é apresentada, onde cada vértice é associado a uma comunidade. O dendrograma (c) foi obtido através da aglomeração hierárquica das palavras-chave.	99
Figura 46 – Resumo dos principais resultados obtidos para a rede PC. Similarmente à Figura 45, esta é composta pelo mapa científico (a), grafo reduzido (b) e pelo dendrograma de palavras-chave (c)	100
Figura 47 – Exemplo de variedade bidimensional construída parametricamente pela equação indicada. Verifica-se que mesmo estando definida sob as variáveis em coordenadas cilíndricas (r_0, θ, z) , devido à dependência de apenas duas variáveis (t, u) , o objeto tem dimensão 2D ao invés de preencher o espaço 3D por completo.	104
Figura 48 – Exemplos ilustrando o processo de obtenção da dimensão de Lebesgue para algumas formas 1D e 2D. Para cada forma, uma cobertura C é mostrada com seus conjuntos abertos identificados por cores. Após o processo de refinamento topológico, obtém-se m , contando o número de conjuntos de C que compõem a maior sobreposição de cada forma geométrica (indicada por setas).	105
Figura 49 – Exemplo do cálculo da dimensão de capacidade $(143, 20)$ em uma mesma rede complexa, entretanto com duas diferentes escolhas de vértices centrais (indicados pelas setas). Em (a) a dimensão resulta em $d \approx 1.8$, mas no caso da escolha mostrada em (b), a dimensão é substancialmente menor, com $d \approx 0.8$	107
Figura 50 – Exemplos de curvas de $V_i(r)$ para duas redes reais (143): (a) rede de estradas dos EUA (149) e (b) rede da infraestrutura da Internet.(107) Enquanto a dimensão de (a) pode ser obtida observando a inclinação da reta formada pela curva, em (b) não é possível traçar uma única reta que represente uma parcela considerável dos vértices, tornando impossível o cálculo da dimensão segundo o método existente.(143, 20)	108
Figura 51 – Exemplo de rede multidimensional apresentando uma região 2D e outra região 3D. A rede foi construída de modo que, com exceção das bordas, os seus vértices apresentam grau homogêneo $k = 6$. As cores indicam a dimensão local $d_i(4)$ para cada vértice de acordo com os valores da legenda. Mesmo com todos os vértices apresentando o mesmo grau, a medida de dimensão é capaz de diferenciar claramente as duas regiões.	109

Figura 52 – Comparação da dimensão multiescala obtida para redes ER através do método LS de segunda ordem (Equação 5.9) com o resultado analítico obtido pelas equações 5.12 e 5.13.	112
Figura 53 – Mapa da dimensão máxima obtida pelas equações 5.12 e 5.13 para diferentes combinações de parâmetros. O mapa da dimensão para $ \mathcal{V} $ e $\langle k \rangle$ é apresentado em (a) com a região em destaque detalhada em (b). Similarmente, o mapa em (c) apresenta a dimensão para valores de $ \mathcal{V} $ e p com detalhe da região em destaque em (d).	113
Figura 54 – Dimensão multiescala obtida para redes regulares sem efeitos de borda (toroidais) de dimensões 1, 2, 3 e 4 através de diferentes métodos de estimação (BK, FW CE, LS) calculados em 2ª ordem.	114
Figura 55 – Exemplo de rede multidimensional toroidal construída para verificar o comportamento da dimensão multiescala. A variedade original (a) é um toróide 3D com casca. As dimensões usadas para gerar a variedade são mostradas em vermelho. A rede (b) é gerada mantendo as características dimensionais da variedade original.	115
Figura 56 – Curvas de dimensão multiescala, considerando os diferentes métodos propostos, obtidas para a rede multidimensional construída de de acordo com a Figura 55. A área em detalhe é mostrada para melhor comparar os métodos. Verifica-se que para r grande, todos os métodos convergem em uma mesma curva. Os estágios de escala da rede são ilustrados qualitativamente através de visualizações de uma rede construída da mesma forma, mas com menor número de vértices.	116
Figura 57 – Média da dimensão multiescala para redes do modelo WS unidimensional para várias probabilidades de reconexão p com tamanho fixo $N = 5000$ e $\langle k \rangle = 4$. O gráfico interno mostra a curva de dependência do valor máximo de $D(r)$ com p , que se estabiliza em $D_{\text{estável}}(r) \approx 4$, que corresponde à dimensão máxima de redes ER com mesmo número de vértices e arestas.	117
Figura 58 – Fenômeno de <i>empacotamento</i> em uma rede 2D. No topo são apresentadas visualizações de uma rede regular com efeito de borda sob o mesmo processo gerador de redes WS para diferentes probabilidades de reconexão. Na parte inferior da figura é mostrada a curva da dimensão máxima com a probabilidade de reconexão. As cores dos vértices das visualizações indicam a dimensão local de acordo com a legenda.	118
Figura 59 – Curvas de $D(r)$ para redes geográficas 2D e 3D. Para fins de comparação, são mostrados também, os valores para redes ER e regulares com tamanhos e grau médio semelhantes.	119

Figura 60 – Dimensão local para $r = 3$ de uma rede RAGEO de 10 mil vértices e $\langle k \rangle \approx 12$	120
Figura 61 – Dimensão multiescala obtida para as redes de distribuição de energia (EU e US). Também são mostradas as curvas correspondentes aos modelos geográficos RAGEO2D e RAGEO3D, assim como redes regulares 2D e 3D (REG2D e REG3D).	122
Figura 62 – Visualização 3D da rede EU sob diferentes projeções: frente (a), lado (b) e cima (c). A dimensão local para $r = 3$ é mostrada através das cores. A projeção frontal é a mesma da Figura 2(b).	123
Figura 63 – Visualização 3D da rede US sob diferentes projeções: frente (a), lado (b) e cima (c). A dimensão local para $r = 3$ é mostrada através das cores. . .	124
Figura 64 – Efeitos da dinâmica de remoção aleatória de vértices sobre as redes de distribuição de energia (EU e US) e modelos com propriedades similares (ER e RAGEO, REG2D e REG3D). O gráfico mostra para cada rede a curva do número de componentes conectados normalizado (pelo número de vértices) na rede. Os valores apresentados foram obtidos através da média de 1000 realizações para cada rede, nesta configuração. A largura das curvas corresponde ao dobro do desvio padrão.	125
Figura 65 – Dimensão local contra acessibilidade considerando o terceiro nível concêntrico ($r = 3$) para diferentes modelos geográficos de tamanho $\mathcal{V} = 5000$. O quadro à esquerda mostra o mesmo mapeamento considerando as redes de distribuição de energia. O quadro à direita mostra a curva de crescimento da correlação entre as duas medidas de modo que apenas os vértices dentro de um limiar crescente ou decrescente na acessibilidade, são considerados. Isto é, usando um limiar da acessibilidade σ_a , apenas vértices com $a_i(3) < \sigma_a$ (crescente) ou $a_i(3) > \sigma_a$ (decrescente), são usados para o cálculo da correlação.	126
Figura 66 – Dimensão multiescala obtida para para as redes de ruas estudadas. Os principais modelos comparáveis às redes também são mostrados, com a rede real correspondente indicada entre parênteses. A rede da cidade de Londres é mostrada em sua versão com e sem o transporte subterrâneo.	127
Figura 67 – Dimensão local para $r = 3$ dos vértices das cidades de San Joaquin e Oldenburg.	127

Figura 68 – Rede da cidade de Londres com dimensão local para $r = 3$ indicada de acordo com a legenda de cores. A rede é mostrada usando a posição original dos vértices sem (a) e com (b) o transporte subterrâneo. Também são mostradas projeções da rede em um espaço 3D geradas pelo visualizador de redes complexas, tanto para a rede sem (c) e com (d) o transporte subterrâneo. Em (b) e (d) as arestas referentes ao sistema de transporte subterrâneo são destacadas por linhas mais largas. 129

Figura 69 – Dimensão local da rede de estradas dos EUA (a) com destaque para a região compreendendo as cidades de Chicago, Indianapolis e Detroit (b). Em (c) a dimensão multiescala da rede e de modelos de propriedade similares é mostrada. 130

Figura 70 – Exemplo de análise de simetria pelo tamanho do conjunto de automorfismos. Iniciando de um grafo simétrico em (a), ao remover 1 aresta (b), o tamanho do conjunto de automorfismos tradicional, A , e normalizado, $\bar{A}$, são reduzidos. Entretanto, ao considerar a generalização do grafo g_n (c) com n ramos, mesmo a estrutura sendo simétrica, ocorre um decaimento rápido de $\bar{A}$ com n . Além disso, o uso de A é proibitivo já que com o acréscimo de novos vértices, A também cresce $n!$ 133

Figura 71 – Acessibilidade de uma rede RAGEO2D. Padrões concêntricos mais simétricos tendem a aparecer nas regiões mais centrais da rede, de modo que acessam vértices mais distantes uniformemente. Nas bordas da rede a acessibilidade é menor, indicando menor simetria no acesso a outros vértices. 134

Figura 72 – Exemplo das transformações *backbone* e *merged* de um padrão concêntrico de nível $r = 2$. (a) Padrão original com os níveis concêntricos em destaque. (b) Transições possíveis em uma caminhada concêntrica. Para construir o padrão *backbone* (c), todas as arestas existentes dentro de um mesmo nível concêntrico, precisam ser removidas. Já o padrão *merged* (d), é construído unindo-se vértices que pertençam ao mesmo nível concêntrico. O número de arestas existentes entre vértices unidos que levam a um mesmo nó são somados e atribuídos como peso a aresta. Na figura, a probabilidade de transição em ambas as transformações é mostrada pelos números próximos aos vértices alcançáveis. 136

Figura 73 – Exemplo do cálculo das simetrias <i>backbone</i> (b,e) e <i>merged</i> (c,f) para dois padrões concêntricos (a) e (d). O padrão concêntrico da parte superior da figura (a) é simétrico, fato que é refletido na uniformidade da probabilidade de acesso aos vértices alcançáveis em 2 passos(b,c). Já o padrão mostrado em (d) apresenta uma terminação no vértice 3, consequentemente, menor é a simetria <i>backbone</i> (e). A simetria <i>merged</i> (f) também é sutilmente menor devido ao desbalanço no acesso aos vértices à distância 2 da referência.	137
Figura 74 – Padrões concêntricos $r = 2$ (a) e $r = 3$ (b) da rede ER distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i> . A rede apresenta $ \mathcal{V} = 5000$ vértices e grau médio $\langle k \rangle = 6$	139
Figura 75 – Padrões concêntricos $r = 2$ (a) e $r = 3$ (b) da rede RAGEO2D distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i> . A rede apresenta $ \mathcal{V} = 5000$ vértices e grau médio $\langle k \rangle = 6$	140
Figura 76 – Padrões concêntricos $r = 2$ da rede WAX distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i> . A rede apresenta $ \mathcal{V} = 5000$ vértices e grau médio $\langle k \rangle = 6$	141
Figura 77 – Padrões concêntricos $r = 3$ para as redes VOR (a) e RVOR (b) distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i>	142
Figura 78 – Padrões concêntricos $r = 2$ (a) da rede BA distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i> . A rede apresenta $ \mathcal{V} = 5000$ vértices e grau médio $\langle k \rangle = 6$. Para $r = 4$, as simetrias de BA apresentam substancial correlação com as medidas tradicionais de grau e centralidade de intermediação. Esta tendência, no entanto, se deve ao efeito de tamanho finito das redes, já que em uma rede BA poucos níveis concêntricos são necessários para cobrir integralmente a rede.	143
Figura 79 – Tabela de correlações de Pearson obtidas entre as medidas de simetria concêntrica e diversas outras propriedades tradicionais considerando os modelos BA (a) e WAX (b).	143
Figura 80 – Padrões concêntricos $r = 2$ (a) da rede AIR distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i>	144
Figura 81 – Padrões concêntricos $r = 2$ (a) da rede WIKI distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i>	145
Figura 82 – Padrões concêntricos $r = 3$ para as redes de ruas das cidades San Joaquin (a) e Oldenburg (b) distribuídos ao longo das simetrias <i>backbone</i> e <i>merged</i>	147
Figura 83 – Tabela de correlações de Pearson obtidas entre as medidas de simetria concêntrica e diversas outras propriedades tradicionais considerando as redes WIKI (a) e Oldenburg (b).	148

Figura 84 – PCA aplicada aos modelos e redes reais. Em (a) apenas as medidas de simetria foram usadas, enquanto em (b) apenas medidas concêntricas tradicionais foram consideradas (ver Seção 2.8.2).	148
Figura 85 – Modelos e redes consideradas ordenadas de acordo com a simetria geral média indicada pelo componente PCA1 da Figura 84(a). Cada quadro apresenta uma visualização em miniatura da rede.	149
Figura 86 – Simetria <i>merged</i> média para $r = 2$ calculada para a rede do mercado financeiro ao longo do tempo. As regiões em destaque indicam períodos de crises documentadas na literatura.	150
Figura 87 – Tabela de correlações médias e desvios padrão entre as medidas de simetria contra as medidas topológicas tradicionais. As médias foram tomadas para a rede do mercado financeiro ao longo de todo o período considerado. Verifica-se que a simetria <i>merged</i> de nível $r = 2$ apresenta correlação quase nula com todas as outras medidas.	151
Figura 88 – Mapeamento dos modelos e redes reais estudados neste trabalho ao longo do espaço da dimensão e da simetria. O tamanho de cada rede (em número de vértices) é indicado pelo tamanho dos símbolos, como indicado pela legenda. Algumas redes são conectadas devido ao seu forte relacionamento, como as redes WS para diferentes probabilidades de reconexão e as redes de Londres, sem e com transporte subterrâneo. As elipses de densidade (171) são mostradas em tons claros, indicando o desvio padrão obtido por ajuste de distribuição normal bivariada.	156

Lista de Símbolos

$\Gamma(\mathcal{V}, \mathcal{E})$	Rede ou grafo
$\mathcal{V}$	Conjunto de vértices de um grafo
$\mathcal{E}$	Conjunto de arestas de um grafo
v_i	Vértice indexado por i
e_{ij}	Aresta entre v_i e v_j
$\mathcal{W}$	Vetor de pesos de uma rede
G	Matriz de adjacência
ξ_v	Elementos da lista de adjacências
α_v	Pesos da representação por lista
$k(v_i)$	Grau ou conectividade
$k_{\text{out}}(v_i)$	Grau de saída
$k_{\text{in}}(v_i)$	Grau de entrada
$s(v_i)$	Força
$Cc(v_i)$	Coefficiente de aglomeração
$\langle \rangle$	Média
T	Transitividade
ℓ_{ij}	Caminho
ℓ_{ij}^g	Geodésica
C_B	centralidade de intermediação
Π	Matriz de mistura
$R_i(r)$	Anel concêntrico
$n_i(r)$	Número de vértices concêntricos
$B_i(r)$	Bola concêntrica

$V_i(r)$	Volume da bola concêntrica
$e_i(r)$	Números de arestas em anel
$k_i(r)$	Grau concêntrico
$Al_i(r)$	Grau intra-nível
$El_i(r)$	Grau entre-níveis
$Cc_i(r)$	Coeficiente de aglomeração concêntrico
$Cr_i(r)$	Taxa de convergência
$Kh_i(r)$	Grau concêntrico comum
$a_i(r)$	Accessibilidade
$P_{i,c}$	Pertinência
H_i^*	Entropia de pertinência
$I(w)$	Índice de importância de palavras-chave
$d_i(r)$	Dimensão local
$D(r)$	Dimensão multiescala
$\bar{A}(g)$	Número de automorfismos normalizado
$Sb_i(r)$	Simetria <i>backbone</i>
$Sm_i(r)$	Simetria <i>merged</i>

Sumário

1	INTRODUÇÃO	29
1.1	Motivações e objetivos	29
1.2	Organização da tese	35
2	REDES COMPLEXAS	37
2.1	Definição e tipos de redes complexas	37
2.2	Representação de redes complexas	38
2.3	Propriedades tradicionais de redes complexas	40
2.3.1	Propriedades locais	40
2.3.2	Propriedades globais	42
2.4	Modelos de redes complexas	46
2.4.1	Redes Regulares	48
2.4.2	Redes Erdős-Rényi	49
2.4.3	Modelo Barabási-Albert	49
2.4.4	Redes Watts-Strogatz	51
2.4.5	Modelos de redes geográficas	53
2.5	Estrutura de comunidades e modelo estocástico de blocos	55
2.6	Isomorfismos e automorfismos em grafos	56
2.7	Análise de redes por componentes principais	59
2.8	Propriedades multiescala em redes complexas	61
2.8.1	Níveis e padrões concêntricos	63
2.8.2	Medidas concêntricas	63
2.8.3	Acessibilidade	66
3	VISUALIZAÇÃO DE REDES COMPLEXAS	69
3.1	Visualização interativa de redes através de métodos baseados em forças físicas	71
3.2	Otimização da técnica FRMOD por meio de expansão multipolar	73
3.3	Desenvolvimento de software para visualização de redes complexas	77
3.4	Visualização e caracterização de redes de expressão gênica	81
4	MAPAS CIENTÍFICOS E CARACTERIZAÇÃO DE VÉRTICES EM COMUNIDADES	85
4.1	Uso da entropia para quantificar o papel de vértices em comunidades	86

4.2	Diversidade temática e quantificação da interdisciplinaridade . .	88
4.3	Taxonomia e sumarização de áreas científicas	95
4.3.1	Análise de redes de citações obtidas a partir do WoS	96
4.3.2	Extração de palavras-chave em comunidades e construção de taxono- mias	96
4.3.3	Visão geral da rede CN	98
4.3.4	Visão geral da rede PC	100
5	DIMENSÃO LOCAL E MULTIESCALA	103
5.1	Dimensão em redes complexas	105
5.2	Definindo uma dimensão multiescala de redes	108
5.3	Dimensão multiescala de modelos de redes complexas	111
5.3.1	Estudo analítico da dimensão multiescala em redes ER	111
5.3.2	Dimensão de grades regulares e efeito de empacotamento	114
5.3.3	Dimensão em modelos de redes geográficas	117
5.4	Dimensão multiescala de redes reais	120
5.4.1	Dimensão em redes de linhas de distribuição de energia	121
5.4.2	Dimensão em redes de ruas e estradas	126
6	SIMETRIA CONCÊNTRICA	131
6.1	Tamanho do grupo de automorfismos como medida de simetria concêntrica	132
6.2	Acessibilidade como medida de simetria concêntrica	134
6.3	Simetrias <i>Backbone</i> e <i>Merged</i>	135
6.4	Caracterização da simetria de modelos de redes complexas . .	137
6.5	Simetria de redes reais	144
6.6	PCA e classificação de redes por simetria	146
6.7	Aplicação preliminar da simetria concêntrica para identificar e caracterizar crises do mercado financeiro	150
7	CONCLUSÕES	153
7.1	Principais contribuições	157
7.2	Trabalhos publicados ou submetidos em periódicos internacionais	160
7.3	Trabalhos futuros	161
REFERÊNCIAS		163
APÊNDICE A – Recursos computacionais e complexidade dos algoritmos .		174

1 Introdução

O advento da era da informação mudou muito a forma de se “produzir” ciência. Ao longo dos últimos anos, o dilúvio de dados gerados a partir de experimentos, reconstrução digital de objetos reais ou até mesmo através de mineração de dados a partir da *web*; culminou em diversos avanços científicos e tecnológicos. Esta nova abordagem científica permitiu o surgimento de uma grande variedade de métodos que têm sido aplicados com sucesso na tarefa de organizar, sumarizar, filtrar e visualizar a crescente quantidade de informação disponível em diversas áreas do conhecimento, englobando ciências da vida (1), exatas (2), humanidades (3) e até mesmo artes.(4) Exemplos bem-sucedidos incluem: técnicas para traçar melhores rotas em cidades considerando dados em tempo real (5), busca na Web (6), reconhecimento de padrões em imagens (7), prova automatizada de teoremas (8), busca de mecanismos genéticos de doenças.(1, 9) Ao mesmo tempo, esta nova ciência, que agora também envolve os problemas de processar, caracterizar, modelar e interpretar essa enorme quantidade de dados, tem se tornado cada vez mais complexa.

O aumento da complexidade da ciência parece estar diretamente relacionado à complexidade da classe de problemas que ainda restam a serem exploradas, já que muitos dos problemas *simples* já foram devidamente estudados e formalizados. Espera-se então que novas descobertas ocorram em uma interface de problemas cada vez mais complexos. Desta forma, a comunidade científica tem passado por uma transformação, partindo de uma abordagem científica reducionista a uma nova forma de construir conhecimento, mais integrada e interdisciplinar.(10) Enquanto a abordagem reducionista é útil na caracterização individual de cada elemento de um sistema, muitas vezes, a interação entre eles é que contribuem mais significativamente para seu comportamento macroscópico. No caso de sistemas complexos, estas relações são geralmente intrincadas, dificultando o desenvolvimento de modelos e abordagens analíticas. Uma forma natural de modelar e caracterizar sistemas desta natureza é através do uso de métodos oriundos da ciência das redes (*network science*), onde os elementos e relações de um sistema podem ser modelados por meio de *grafos*, ou mais precisamente, redes complexas.(11, 12)

1.1 Motivações e objetivos

Costuma-se atribuir a *Leonhard Paul Euler* o desenvolvimento inicial da *teoria dos grafos*. Isto se deve principalmente ao fato dele ter resolvido de forma definitiva, em 1735, o problema das sete pontes de Königsberg.(13) Nesse período, Königsberg, Prússia (atualmente Kalingrad, Rússia), era uma cidade rica e um importante centro comercial, ela era dividida

em quatro regiões pelo rio Pregel, onde também haviam sete pontes que permitiam a travessia entre as regiões, como mostrado na Figura 1a. Visitantes e moradores que apreciavam caminhar pela cidade eram geralmente desafiados para um jogo, onde o objetivo era determinar um caminho, partindo-se de uma das quatro regiões da cidade, que percorresse todas as sete pontes de modo que cada ponte fosse atravessada apenas uma vez, não necessitando que a região de origem e a final sejam as mesmas. A pedido do prefeito, Euler desenvolveu uma solução, onde provou que, matematicamente, tal caminho não existe.

Figura 1 – Representações do problema das sete pontes da cidade de Königsberg.
Fonte: Elaborada pelo autor.

O primeiro passo de Euler na solução do problema de Königsberg foi notar que este não se tratava de um problema geométrico, já que métricas não contribuíam na busca de uma solução. Euler reduziu o problema ao eliminar as propriedades geográficas, e, ao introduzir os conceitos de vizinhança e conectividade. Cada região de terra do mapa foi representada por um nó, ou *vértice*, e cada ponte, uma conexão, ou *aresta*, entre pares de vértices, o sistema foi reduzido ao que é chamado hoje de grafo, como ilustrado na Figura 1b.

Euler também generalizou a solução do problema de Königsberg para qualquer outro grafo, tendo como soluções os *caminhos de Euler*. Em um *caminho de Euler*, todas as arestas de um grafo são atravessadas apenas uma vez. Euler provou que a existência desses caminhos depende da *conectividade* dos vértices, isto é, o número de arestas conectadas a eles. Para o caso da região inicial ser a mesma da final, todas as conectividades precisam ser pares, já que cada vértice precisa de uma entrada e uma saída, neste caso, o nome dado a estes caminhos de Euler é *circuito de Euler*. No caso dos vértices inicial e final diferirem, é requerido que estes apresentem uma saída e uma entrada, respectivamente, desta forma, devem apresentar conectividade ímpar, enquanto que os outros vértices continuam a apresentar conectividade par. Em resumo, a rede apresentará um *caminho de Euler* apenas se houver 0 ou 2 vértices de

conectividade ímpar.

A solução dos *caminhos de Euler*, despertou grande interesse de outros matemáticos da época, o que culminou na solução e surgimento de diversos outros problemas que deram origem ao que é chamada hoje de teoria dos grafos. Em sua maioria, eles não dependiam da geometria, posição, formas, ou quaisquer outros detalhes presentes no espaço original que não fossem transformados em relações entre seus elementos. Neste contexto, estas relações eram suficientes para representar, caracterizar e solucionar estes problemas.

Enquanto a teoria dos grafos tradicionalmente desvincula a topologia de um sistema das outras propriedades de seu espaço original, na maioria dos sistemas reais, a relação entre esses dois espaços (topológico e original) apresenta forte dependência. Estudos em *redes complexas* tentam associar fenômenos topológicos a efeitos que ocorrem no espaço original e vice-versa. Esta abordagem deu origem a diversas técnicas que tentam explicar a formação de redes em sistemas complexos.(11, 14) O planejamento urbano e as limitações geográficas em cidades, por exemplo, levam também a redes que mantêm estas restrições, raramente apresentando caminhos de longa distância.(15) Já em uma rede social, o fator geográfico nem sempre é uma restrição, sendo mais influenciada pela capacidade dos indivíduos em estabelecer novas relações sociais.(16, 17) O mesmo ocorre com a rede da *web*, onde muitas vezes a qualidade ou mérito de um site contribui para o estabelecimento de novas hiperligações direcionadas a ele.(18) Muitas dessas características foram reproduzidas na área de ciência das redes, ao longo dos últimos anos, através de modelos geradores de redes.(11, 14, 12)

A Figura 2 exemplifica a forte relação entre o espaço original e o topológico, através de uma visualização da rede de linhas de energia da Europa (EU).(19) Em (a) é mostrada a posição original dos vértices e em (b) a visualização gerada através de mapeamento topológico, apenas usando a informação de conectividade da rede. Nota-se que, apesar das diferenças locais, a estrutura geográfica global permanece muito similar ao posicionamento original. Esta é uma característica presente em muitas redes geográficas (15), que advém de sistemas onde as restrições espaciais do espaço original desempenham um papel fundamental no seu padrão de conectividade. Nestas redes, torna-se interessante identificar características do espaço topológico que também possam estar presentes no espaço original. Entre estas características há destaque para a dimensão e simetria.

Surpreendentemente, tanto a dimensão quanto a simetria de redes complexas receberam pouca atenção na literatura.(20) A dimensão tem grande importância em sistemas físicos e dinâmicos, caracterizando o comportamento de equações diferenciais, em especial, dos processos de difusão (21) e propagação de ondas.(22) A simetria, por outro lado, está relacionada à regularidade e à existência de redundâncias em um sistema, o que permite que sejam modelados e entendidos por meio de blocos que se repetem. Em redes, a simetria é tratada pelo problema de automorfismos e isomorfismos (23), entretanto, tais conceitos não são aplicáveis a redes reais devido ao alto custo computacional, além de problemas de

Figura 2 – Rede de linhas de transmissão de energia da Europa (EU) (a). Cada vértice corresponde a uma estação ou usina, enquanto as arestas representam linhas de transmissão, mais informações sobre esta rede podem ser vistas no capítulo 5. Em (a) a posição no espaço original é mostrada ao longo do mapa da Europa. A visualização em (b), no entanto, é gerada apenas a partir da informação topológica da rede, isto é, desconsiderando posições e distâncias entre cada vértice. O método de visualização usado é apresentado no capítulo 3. Em ambos os painéis as cores indicam o país de origem de cada estação/usina, de acordo com a legenda.

Fonte: Elaborada pelo autor.

normalização.(24, 25)

De modo geral, a dimensão também pode ser entendida como uma forma de quantificar os graus de liberdade, ou número de variáveis independentes de um sistema, não limitando-se a características geográficas. Desta forma, por exemplo, pode fornecer informações importantes que descrevem dinâmicas macroscópicas através do uso da mecânica estatística. Além disso, sistemas complexos podem apresentar fenômenos interessantes como aqueles revelados pela dimensão fractal.(26) Isso torna a dimensão extremamente importante para a caracterização de qualquer sistema complexo, principalmente se tratando das redes que representam tais sistemas.

Sistemas complexos também apresentam dinâmicas e propriedades que mudam com a escala, por exemplo, as interações entre células são diferentes das interações entre as moléculas que as compõem. Da mesma forma, moléculas não interagem como átomos, e nem estes como partículas subatômicas. A combinação intrincada das relações de uma escala é que geralmente contribuem para as propriedades do sistema em uma escala superior. Da mesma forma que estes sistemas apresentam características intrinsecamente multiescala, este mesmo fenômeno também é esperado nas redes complexas que os representam.

Outra característica presente em sistemas complexos é a organização de sua estrutura em subestruturas ou regiões, que podem apresentar diferentes características. Similarmente, as redes que representam estes sistemas também expressam estas características,

geralmente na forma de uma estrutura de comunidades.(27) Muitas metodologias foram desenvolvidas com a finalidade de detectar comunidades em redes (28), entretanto poucos estudos caracterizam a participação de cada vértice dentro delas. Por exemplo, cidades próximas às bordas entre nações podem apresentar características culturais muito diferentes daquelas que estão localizadas no núcleo do território. Analogamente, em uma rede de conhecimento, vértices nas fronteiras entre duas ou mais áreas do conhecimento tendem a apresentar maior interdisciplinaridade do que aqueles que estão no centro de cada área.(10) Assim como a dimensão e similaridade, o conceito da participação de cada vértice dentro de uma comunidade está relacionado à complexidade do sistema, em especial à diversidade de características presentes na rede.

Por fim, a dualidade topologia/geometria de redes naturalmente espaciais também pode ser estendida a estruturas onde apenas a topologia é conhecida. Isto é feito através de técnicas de *mergulho* de redes, onde redes podem ser projetadas em um espaço métrico n -dimensional, com as posições geradas a partir da topologia (como no caso da Figura 2b). Estes métodos são geralmente empregados na visualização de redes complexas.(29, 30)

A visualização de redes, quando interativa, permite uma análise naturalmente multiescala de redes complexas, fornecendo informações qualitativas tanto da estrutura local ou de regiões da rede, quanto da rede como um todo. Apesar disso, devido às características subjetivas de nossa percepção e do alto custo computacional dependendo do tamanho das redes, a visualização interativa de redes ainda é um problema em aberto.

A proposta deste trabalho é desenvolver métodos para a análise multiescala de redes complexas, e, simultaneamente, explorar características de redes complexas, normalmente presentes em sistemas reais, mas que foram pouco abordadas na literatura, em particular, a dimensão, a simetria e a participação de vértices em comunidades. Para isso, é proposto o desenvolvimento de um conjunto de novas medidas multiescala que quantificam cada um dos aspectos de interesse. As medidas são definidas em termos de um conjunto integrado de conceitos centrados na ideia de redes complexas e análise multiescala, como mostrado na Figura 3, que resume a abordagem escolhida para esta tese, com cada conceito sendo descrito ao longo deste trabalho.

A participação de vértices em comunidades é estudada a partir de uma nova medida, a *entropia de pertinência*, que também tem utilidade na caracterização da interdisciplinaridade de conceitos em mapas do conhecimento (10), além de poder ser definida sob uma abordagem multiescala.

A dimensão em redes é estudada através do desenvolvimento de uma medida de *dimensão multiescala* (31, 32) que permite a caracterização de redes, tanto sob diferentes escalas, quanto para diferentes regiões ou vértices (localmente). Neste estudo é necessário validar a medida proposta em termos de modelos de dimensão conhecida e de acordo com resultados existentes na literatura. Com isso é possível entender tanto sobre o comportamento da di-

os tamanhos de redes geralmente estudados no campo de redes complexas. Desta forma, evitando os problemas da abordagem usual, baseada em automorfismos.

O potencial da visualização de redes complexas também é explorado ao longo de todas as análises propostas, se mostrando uma ferramenta indispensável em diversos estágios desta pesquisa. Em particular, as técnicas de visualização também são aplicadas para entender a topologia de redes de expressão gênica (9, 34), o que ajuda a determinar as métricas multiescala que podem ser úteis na caracterização dos principais mecanismos que diferem a epilepsia febril e não febril. A visualização também é aplicada na obtenção de mapas científicos, que em conjunto com análise multiescala permitem o desenvolvimento de técnicas de sumarização e obtenção de taxonomia a partir de dados científicos representados como redes de citação.

Para conduzir os experimentos, são propostas melhorias ao software de visualização 3D interativo de redes complexas já desenvolvido pelo autor, *Network 3D* (35), de modo a permitir que redes com grande número de vértices possam ser estudadas.

Por fim, é desejável a construção de um panorama de como redes em geral são distribuídas sobre os aspectos de regularidade e dimensionalidade. Permitindo a verificação da hipótese de que a noção de complexidade em sistemas pode estar relacionada às características de dimensão e regularidade, onde espera-se que redes mais complexas tendam a apresentar maior dimensão e menor simetria e vice-versa.

1.2 Organização da tese

Esta tese está dividida em 6 capítulos. O capítulo intitulado *Redes complexas* (cap. 2) abrange os principais conceitos de redes complexas abordados ao longo da tese, incluindo a descrição de modelos (Seção 2.4) e de propriedades multiescala. O capítulo *Visualização de redes complexas* (cap. 3) apresenta a metodologia de visualização de redes complexas usada ao longo do trabalho e também implementada no software desenvolvido, incluindo detalhes das otimizações realizadas (Seção 3.2) e uma aplicação em redes de transcriptoma humano (Seção 3.4). O capítulo intitulado *Mapas científicos e caracterização de vértices em comunidades* (cap. 4) apresenta duas novas abordagens para se caracterizar vértices e comunidades em redes de conhecimento, uma delas dedicada à quantificação da interdisciplinaridade de revistas científicas (Seção 4.1). Também são analisados os mapas científicos das redes estudadas. O capítulo *Dimensão local e multiescala* (cap. 5) apresenta uma nova medida multiescala de dimensão em redes complexas, onde suas propriedades são exploradas e validadas através de modelos e redes reais. No capítulo *simetria concêntrica* (cap. 6), é desenvolvida uma nova medida de simetria em redes, baseada na acessibilidade e em propriedades concêntricas, é explorado também o potencial da nova medida em caracterizar crises do mercado finan-

ceiro (Seção 6.7). O último capítulo, *Conclusões* (cap. 7), resume as principais contribuições deste trabalho além de explorar as perspectivas de trabalhos futuros.

2 Redes complexas

Redes Complexas (14, 11, 12) são estruturas de dados abstratas que representam os elementos e as relações de sistemas complexos. Uma rede complexa constitui-se basicamente de um conjunto, possivelmente numeroso, de vértices (ou nós) e arestas (ligações). Vértices podem ser associados a diversos tipos de elementos (12), tais como pessoas (17), proteínas (36, 37), trabalhos acadêmicos (38), etc; simultaneamente, as arestas indicam as relações entre os elementos indicados pelos vértices, neste contexto, por exemplo, a amizade entre pessoas, cooperação acadêmica ou semelhança entre proteínas. A Figura 4 ilustra algumas redes complexas reais.

Figura 4 – Exemplos de redes complexas reais.
 Fonte: Elaborada pelo autor.

2.1 Definição e tipos de redes complexas

Formalmente, uma rede complexa é um *grafo* (ou rede), isto é, uma estrutura abstrata $\Gamma(\mathcal{V}, \mathcal{E})$, onde $\mathcal{V}$ é o conjunto de vértices $\mathcal{V} = (v_1, v_2, v_3, \dots, v_n)$ e $\mathcal{E}$ o conjunto de arestas entre os vértices, $\mathcal{E} = \{e_{ij} \equiv (v_i, v_j) : v_i, v_j \in \mathcal{V}\}$. Para facilitar a notação de algumas equações deste trabalho, considera-se que cada vértice v_i está associado a um índice único i . Grafos são, tradicionalmente, divididos em duas categorias: *grafos não direcionados* (Figura 5a), quando as arestas *não* apresentam direção ou sentido privilegiado; ou *direcionados* (Figura 5b), também chamados de digrafos, onde a direção das arestas é importante. Grafos também podem apresentar valores numéricos associados às suas arestas, $\mathcal{W} : \mathcal{E} \rightarrow \mathbb{R}$, indicando o peso ou força de cada conexão, estes são chamados de *grafos com pesos* (Figura 5c). Também existem generalizações do conceito de grafo, onde, por exemplo, são permitidas múltiplas

conexões entre um mesmo par de vértices ou arestas do tipo e_{ii} , isto é, arestas que conectam um vértice a ele mesmo.

Figura 5 – Principais tipos de grafos.

Fonte: Elaborada pelo autor.

Devido ao fato de compartilharem a mesma estrutura abstrata, as áreas de pesquisa em redes complexas e grafos se interceptam várias vezes, compartilhando também um conjunto de metodologias e subsidio teórico. Entretanto, pesquisas de redes complexas são, em geral, direcionadas para a finalidade de caracterizar, representar e modelar de forma efetiva estruturas do mundo real. Muitas propriedades importantes em teoria de grafos têm pouca aplicação no estudo de redes complexas, isso se deve ao fato de que redes reais apresentam estruturas topológicas que dificilmente podem ser estudadas de forma analítica, o que torna ferramentas estatísticas e modelos físicos mais interessantes para esta classe de redes.

2.2 Representação de redes complexas

Uma rede complexa ou grafo é geralmente representado por sua matriz de adjacência. Uma matriz de adjacência G de uma rede $\Gamma(\mathcal{V}, \mathcal{E})$ de N vértices, tem dimensão $N \times N$, onde seus elementos são definidos por G_{ij} tal que se $G_{ij} = 1$ então $\{(v_i, v_j) \in \mathcal{E}\}$. Quando a rede é não direcionada, esta matriz é simétrica. Similarmente, uma rede com pesos também pode ser representada pela matriz de adjacência G^w , entretanto $G_{ij}^w = W(e_{ij})$ quando a aresta e_{ij} existir e zero caso contrário.

Muitos resultados teóricos importantes da teoria de grafos advêm da representação matricial, seja através de transformações, como em propriedades espectrais da matriz de Laplace (40), ou pela interpretação da rede como um sistema dinâmico acoplado. Apesar disso, a matriz de adjacência não é uma forma eficaz, em termos computacionais, de se armazenar a estrutura de redes complexas em geral. Muitos dos modelos e sistemas reais estudados são esparsos, isto é, apresentam poucas conexões comparado ao número total de possíveis conexões entre seus vértices, implicando em uma alta população de valores nulos na matriz de adjacência. Muitas outras representações podem ser usadas de modo a amenizar este problema, entre elas, matrizes esparsas, lista de arestas, banco de dados relacional (41), lista de adjacências, assim como soluções híbridas.(35)

A lista de adjacências é formada pela organização das arestas de uma rede em um conjunto de listas que indicam a adjacência de cada vértice. Cada vértice v é associado a uma lista ξ_v contendo todos os vértices com os quais está conectado. Em uma rede com pesos também é necessário associar a cada vértice uma outra lista α_v , onde cada elemento indica o peso da conexão respectiva.

Tabela 1 – Comparação dos tempos médios de diferentes operações usando as duas representações de redes complexas consideradas.

Operação	Matriz de adjacências	Listas de adjacências
v_i é adjacente a v_j ?	$O(1)$	$O(\frac{ \mathcal{E} }{ \mathcal{V} })$
lista de vértices adjacentes a v_i	$O(\mathcal{V})$	$O(\frac{ \mathcal{E} }{ \mathcal{V} })$
lista de arestas	$O(\mathcal{V} ^2)$	$O(\mathcal{V} + \mathcal{E})$
adicionar vértice	$\leq O(\mathcal{V} ^2)$	$\leq O(\mathcal{V})$
adicionar aresta	$O(1)$	$\leq O(\frac{ \mathcal{E} }{ \mathcal{V} })$
Armazenamento	$O(\mathcal{V} ^2)$	$O(\mathcal{V} + \mathcal{E})$

Fonte: Elaborada pelo autor

Para redes esparsas, a representação através de lista de adjacências é uma alternativa adequada à matriz de adjacência. Além de simples implementação e baixo consumo de memória, diversas operações podem ser realizadas em tempo médio melhor ou igual aos normalmente possíveis através de matriz de adjacência. A Tabela 1 apresenta a complexidade média de tempo para realizar algumas das operações básicas mais usadas em grafos, assim como a complexidade de espaço para armazenamento de sua estrutura. Neste trabalho são usadas apenas as representações através de matriz de adjacência e lista de adjacências. As duas formas são melhor ilustradas na Figura 6.

Figura 6 – Exemplo de representação de uma rede não direcionada com pesos por matriz de adjacência e lista de adjacências.

Fonte: Elaborada pelo autor.

2.3 Propriedades tradicionais de redes complexas

Ao construir um modelo baseado em redes complexas ou ao inferir sua topologia a partir de dados de sistemas complexos reais, o próximo passo imediato é a caracterização destas estruturas. Atualmente, pode-se contar com um conjunto de técnicas e definições para obtenção de propriedades relevantes (42) que caracterizem redes complexas sob diversas perspectivas. Algumas destas propriedades atribuem valores numéricos considerando apenas pequenas porções da rede, por exemplo, são definidas considerando a vizinhança imediata de um vértice ou aresta, chamadas de *propriedades locais*; outras, no entanto, ajudam a caracterizar as redes por inteiro e são definidas considerando toda a informação da rede, estas são chamadas de *propriedades globais*. Além disso, outra classe de propriedades, chamadas de *multiescala*, permitem tanto a caracterização local como global de uma rede, assim como nas escalas intermediárias.

As subseções seguintes descrevem sucintamente algumas das principais propriedades locais e globais de redes, além daquelas usadas ao longo deste trabalho. As propriedades multiescala, que compõem grande parte das contribuições deste trabalho, são melhor discutidas na Seção 2.8.

2.3.1 Propriedades locais

As propriedades locais caracterizam os vértices de uma rede através de medidas que podem ser obtidas a partir do próprio vértice ou de sua vizinhança imediata (isto é, daqueles com quem está conectado). Dentre as medidas comumente usadas destacam-se o *grau* de um vértice (em inglês *vertex degree*), também chamado de *conectividade*; e o *coeficiente de aglomeração* (clustering coefficient), que quantifica o agrupamento ao redor do vértice.

Figura 7 – Propriedades locais de um vértice: grau (k) e coeficiente de aglomeração (Cc).

Fonte: Elaborada pelo autor.

O grau (ou conectividade), $k(v_i)$, de um vértice v_i , é a medida topológica mais simples que se pode obter em uma rede, sendo definido como o número de arestas conectadas a este vértice (Figura. 7a). Em termos da matriz de adjacência G , o grau de um vértice v_i é a soma de todos os elementos de sua linha ou coluna correspondente, isto é,

$$k(v_i) = \sum_{j=1}^n G_{ij}. \quad (2.1)$$

Em redes direcionadas, a medida de grau pode ser considerada tanto em relação a arestas direcionadas ao vértice (*grau de entrada*), $k_{\text{in}}(v_i)$, quanto direcionadas a partir do vértice (*grau de saída*), $k_{\text{out}}(v_i)$. Redes com pesos também necessitam de uma generalização do conceito de grau. A *força* de um vértice v_i (do inglês *vertex strength*) $s(v_i)$ é definida como a soma dos pesos das arestas conectadas a v_i . Em termos da matriz de adjacência com pesos a força de um vértice v_i pode ser definida de forma similar ao grau como

$$s(v_i) = \sum_{j=1}^n G_{ij}^w. \quad (2.2)$$

O coeficiente de aglomeração, $Cc(v_i)$, de um vértice v_i , é definido como a razão do número de conexões, $e_1(v_i)$, entre os vértices da primeira vizinhança de v_i (isto é, o conjunto de vértices diretamente conectados a v_i), pelo número de conexões, $e_{\text{max}}(v_i)$, possíveis entre todos os vértices desta vizinhança. Como $e_{\text{max}}(v_i) = n_1(v_i)(n_1(v_i) - 1)/2$, onde $n_1(v_i)$ é o número de vizinhos de v_i , pode-se definir $Cc(v_i)$ como

$$Cc(v_i) = \frac{e_1(v_i)}{e_{\text{max}}(v_i)} = 2 \frac{e_1(v_i)}{n_1(v_i)(n_1(v_i) - 1)}. \quad (2.3)$$

Quando a conectividade na vizinhança é máxima, o valor de $Cc(i)$ é 1, e 0 para o outro extremo.

2.3.2 Propriedades globais

Propriedades globais são propriedades de redes obtidas considerando todos os seus elementos. Elas são úteis na tarefa de caracterização macroscópica de uma rede, assim como na caracterização do papel de seus elementos individuais, como vértices e arestas, diante da estrutura global da rede.

Essencialmente, qualquer medida local pode ser usada para caracterizar globalmente redes. Ao considerar, por exemplo, as distribuições ou estatísticas destas medidas obtidas ao longo de toda a rede, é possível inferir propriedades gerais dos sistemas que elas representam. Nesta categoria, houve grande interesse na literatura pelas distribuições de grau e valores médios do coeficiente de aglomeração. Em particular, importantes características de redes reais, como o comportamento *livre de escala* advém do estudo da distribuição de grau destas redes. Estes desenvolvimentos culminaram em modelos que reproduzem estas características, *Barabási-Albert* (BA). (43, 18, 11) A Seção 2.4.3 explora em detalhes o modelo BA.

A conectividade ou *grau médio* de uma rede, $\langle k \rangle$, é uma das medidas mais simples usadas para caracterizar redes. Além disso, tradicionalmente, é usada como parâmetro fixo na tarefa de comparação entre diferentes redes reais e modelos (42). O grau médio pode ser calculado como

$$\langle k \rangle = \sum_{v_i \in \mathcal{V}} \frac{k(v_i)}{|\mathcal{V}|} = 2 \frac{|\mathcal{E}|}{|\mathcal{V}|} \quad (2.4)$$

para redes não direcionadas. Em redes direcionadas, assim como ocorre para o grau de vértice, é necessário definir o grau médio de entrada e de saída:

$$\langle k_{\text{in}} \rangle = \sum_{v_i \in \mathcal{V}} \frac{k_{\text{in}}}{|\mathcal{V}|} \quad (2.5)$$

$$\langle k_{\text{out}} \rangle = \sum_{v_i \in \mathcal{V}} \frac{k_{\text{out}}}{|\mathcal{V}|}. \quad (2.6)$$

Outra propriedade global importante relacionada a uma propriedade local é a *transitividade* (*transitivity*) (44, 42), T , também conhecida como coeficiente de aglomeração global. Esta medida quantifica a probabilidade de dois vértices v_B e v_C estarem conectados se ambos conectam a um vértice v_A , isto é, da tripla conectada v_A, v_B, v_C ter conexões entre todos os seus vértices. Ela pode ser obtida pela razão do número de triângulos da rede, $|\Delta_\Gamma|$, pelo número de triplas conectadas $|\tau_\Gamma|$ como:

$$T(\Gamma) = \frac{3|\Delta_\Gamma|}{|\tau_\Gamma|}. \quad (2.7)$$

É importante notar que, apesar de relacionadas, em geral, o valor da transitividade difere do coeficiente de aglomeração médio, $\langle Cc \rangle$, que é obtido considerando todos os vértices de uma rede. Este último, no entanto, também é usado na literatura para quantificar a aglomeração geral de uma rede.

A Figura 8 ilustra as medidas globais tradicionais, $\langle k \rangle$, T e $\langle Cc \rangle$, em uma rede simples.

Figura 8 – Propriedades globais tradicionais em redes complexas. Para cada vértice é mostrado o grau e o coeficiente de aglomeração. No topo à esquerda as propriedades globais da rede são mostradas: grau médio, coeficiente de aglomeração médio e transitividade.

Fonte: Elaborada pelo autor.

Outra classe de propriedades globais importantes são aquelas baseadas nos *caminhos* (ou trajetórias) possíveis entre os vértices de uma rede. Um caminho ℓ_{ij} entre dois vértices (v_i, v_j) é formalmente representado por uma sequência de vértices $\ell_{ij} = (v_i, v_\alpha, v_\beta \dots v_j)$ que formam uma cadeia conectada, isto é, para cada subsequência de pares $(v_\alpha, v_\beta) \subseteq \ell_{ij}$, existe uma aresta $e_{\alpha\beta}$ que os conecta.

Muitas das dinâmicas estudadas que ocorrem sob redes baseiam-se em características de caminhos percorridos por agentes, como por exemplo caminhadas aleatórias ou autoesquivantes. Em uma rede não direcionada, os subgrafos contendo todos os vértices que possuem ao menos um caminho entre eles, são chamados de *componentes conectados*. O componente com maior número de vértices é chamado de maior componente conectado. A generalização para redes direcionadas permite duas possibilidades: uma em que a direção das arestas não importa para a definição dos caminhos que conectam os vértices, estes são os componentes fracamente conectados, que são análogos aos componentes conectados em uma rede sem direções; e outra quando os caminhos obedecem às direções indicadas pelas arestas, neste caso são chamados de componentes fortemente conectados. A Figura 9 ilustra os tipos de componentes conectados presentes em redes.

Uma *geodésica*, ℓ_{ij}^g , entre dois vértices (v_i, v_j) , também chamada de caminho mí-

Figura 9 – Exemplo de componentes conectados em uma rede simples e em uma rede direcionada. Cada componente é destacado pela cor e destacado pelo contorno.

Fonte: Elaborada pelo autor.

nimo (45), é definida como um dos caminhos com menor quantidade de vértices, isto é, seu comprimento $g_{ij} \equiv |\ell_{ij}^g|$ (conhecido também como *distância topológica*) é menor ou igual ao de qualquer outro caminho entre os mesmos vértices. Cada par de vértices pode ter uma ou mais geodésicas e estas só podem ser definidas entre vértices que pertencem a um componente conectado (a Figura 10 ilustra uma geodésica para uma rede simples). Usualmente, considera-se que $g_{ij} = \infty$ quando v_i e v_j não compartilham o mesmo componente conectado. Em redes direcionadas g_{ij} não é uma distância formal já que pode ocorrer $g_{ij} \neq g_{ji}$. Adicionalmente, redes com pesos podem definir um custo ou distância intermediária entre vértices, de modo que a distância entre dois vértices pode ser definida pelas somas dos pesos das arestas que compõe a geodésica.

A propriedade global mais simples que pode ser obtida através das geodésicas é o caminho mínimo médio, $\langle g \rangle$, que é obtido considerando todos os pares de vértices de acordo com

$$\langle g \rangle = \frac{1}{|\mathcal{V}|(|\mathcal{V}| - 1)} \sum_{v_i \neq v_j \in \mathcal{V}} g_{ij}. \quad (2.8)$$

Isto é válido somente quando a rede considerada é composta de apenas um componente conectado, isto é, quando há caminhos entre todos os pares de vértices da rede.

Outra característica importante no campo de redes é a possibilidade de quantificar o quão central um vértice é ao considerar a estrutura completa de uma rede. Medidas de *centralidade* baseiam-se na ideia de que vértices que participam de muitos caminhos mínimos são topologicamente mais centrais na rede. Diferentemente das outras propriedades globais, onde uma única característica é atribuída à rede, a caracterização por centralidade é feita

Figura 10 – Geodésicas existentes entre os vértices v_1 e v_9 mostradas em destaque laranja, $\ell_{19}^g = (v_1, v_3, v_7, v_9)$ e verde, $\ell_{19}^{g'} = (v_1, v_3, v_8, v_9)$.

Fonte: Elaborada pelo autor.

para cada vértice, mas no contexto global da rede.

Dentre as medidas mais usadas de centralidade, há grande destaque para a *centralidade de intermediação* (*betweenness centrality*), C_B . Ela é definida para cada vértice v_i como a soma das frações das geodésicas ℓ_{st}^g entre todos os pares de vértices (v_s, v_t) que contenham v_i de modo que $v_i \neq v_s$ e $v_i \neq v_t$. Como mais de uma geodésica pode existir entre os vértices v_s e v_t , $C_B(v_i)$ pode ser obtida pela razão do número de geodésicas $\sigma_{st}(i)$ que contenham v_i pelo número total de geodésicas entre v_s e v_t . Portanto, pode-se escrever a centralidade de intermediação como

$$C_B(v_i) = \sum_{v_s \neq v_t \neq v_i} \frac{\sigma_{st}(i)}{\sigma_{st}^*}, \quad (2.9)$$

onde o termo σ_{st}^* indica o número de geodésicas entre v_s e v_t . A Figura 11 ilustra a propriedade para uma rede simples. Para comparar o valor da centralidade de intermediação entre diferentes redes ou em uma rede dinâmica (onde pode haver acréscimo ou remoção de vértices e arestas, por exemplo), é recomendado que a medida seja normalizada por $(|\mathcal{V}| - 1)(|\mathcal{V}| - 2)/2$ para redes não direcionadas ou por $(|\mathcal{V}| - 1)(|\mathcal{V}| - 2)$ para redes direcionadas. Esta normalização garante que $0 \leq C_B \leq 1$, já que correspondem ao número total de caminhos existentes entre todos os pares destas redes. Na situação de $C_B(v_i) = 1.0$, todos os caminhos da rede, exceto aqueles iniciando ou terminando em v_i , passam por v_i .

Além das medidas apresentadas, consideradas tradicionais, diversas outras características globais de redes podem ser encontradas na literatura. Entre elas algumas das mais usadas são: *assortatividade de grau* (*degree assortativity*) (46), *número de vértices no clique máximo* (*clique number*) (47), *coeficiente do clube dos ricos* (*rich-club coefficient*) (48, 49) e *índice de correspondência* (*matching index*). (42) Assim como outras definições de centralidade, como a centralidade de proximidade (*closeness centrality*) (50, 51), de auto-vetor (*eigenvector centrality*) (52) e de subgrafos

Figura 11 – Geodésicas usadas para o cálculo da centralidade de intermediação (*betweenness centrality*) sem normalização em uma rede simples. Linhas mais finas representam as geodésicas, coloridas de acordo com a legenda, indicando a participação dos vértices A ou B. As setas indicam o início ou fim da geodésica. Neste caso a centralidade pode ser facilmente calculada pela metade da soma dos caminhos entrando no vértice (isto é, não contando os caminhos finalizando no vértice).

Fonte: Elaborada pelo autor.

(*subgraph centrality*).⁽⁵³⁾

2.4 Modelos de redes complexas

Um dos principais problemas em redes complexas é encontrar redes que modelem fenômenos ou estruturas da natureza. Redes reais são obtidas através de medidas experimentais e observações da natureza, e tem como objetivo, limitar o estudo a um conjunto de sistemas e as relações entre eles. Muitas vezes, é necessário obter redes que não necessariamente correspondem a sistemas físicos reais, mas, a uma aproximação representativa de uma categoria de sistemas, que possuem propriedades em comum. Diversos modelos teóricos foram desenvolvidos com essa finalidade, dentre eles destacam-se os modelos regulares, Erdős-Rényi (aleatória) (54), Watts-Strogatz (pequeno mundo) (55, 56), Barabási-Albert (livre de escala) (43) e geográficos (15), como redes geográficas aleatórias (57) e o modelo de Waxman.⁽⁵⁸⁾

Os modelos teóricos permitem a criação de redes com parâmetros controlados, fixando o número de vértices e a conectividade, por exemplo; também permitem a obtenção de resultados analíticos de algumas propriedades que, muitas vezes, podem ser estendidas

às redes reais.

A seguir os principais modelos de redes complexas usados nesta tese serão apresentados, destacando suas principais características.

Modelos Regulares (REG)

(a) Grade 2D (REG2D)

(b) Grade 3D (REG3D)

(c) Grade 2D sem bordas (REG2DTOR)

Modelos aleatórios tradicionais

(d) Erdős-Rényi (ER)

(e) Barabási-Albert (BA)

(f) Watts-Strogatz (WS)

Modelos geográficos

(g) Aleatório Geográfico 2D (RAGEO2D)

(h) Aleatório Geográfico 3D (RAGEO3D)

(i) Waxman (WAX)

Figura 12 – Principais modelos de redes complexas usados neste trabalho.

Fonte: Elaborada pelo autor.

2.4.1 Redes Regulares

Redes onde todos os vértices apresentam a mesma conectividade são chamadas de *redes regulares*. Apesar de não representarem muitos sistemas reais, estas redes são úteis na tarefa de comparação e obtenção de resultados teóricos, que podem ser generalizados para outros tipos de redes. Redes geradas a partir de grades regulares estão entre as redes regulares mais usadas na literatura.

Uma grade regular n -dimensional ($REGnD$) pode ser obtida dispondo vértices igualmente espaçados ao longo do espaço $\mathbb{Z}^n$, normalmente usa-se a disposição quadrada, entretanto, outras disposições, como a hexagonal também resultam em grades regulares. A conexão entre dois vértices (v_i, v_j) é feita considerando uma distância de vizinhança $\delta_k(v_i, v_j)$ de modo que cada vértice conecte-se com até k vizinhos. Geralmente, usa-se a distância metropolis neste espaço já que determinar $\delta_k^{\text{euclidiano}}$ (distância euclidiana) que leve a k vizinhos não é uma tarefa trivial em uma malha de dimensão arbitrária.(59)

Figura 13 – Duas formas de truncar a borda de uma grade 2D: (a) com efeito de bordas e (b) sem efeito de bordas. As cores destacam o grau dos vértices.

Fonte: Elaborada pelo autor.

A obtenção de redes finitas a partir de grades exige que elas sejam truncadas, isto é, escolhendo-se um subconjunto finito da grade. Isto pode ser feito usando diferentes formas topológicas, mas convém limitar a rede por um quadrado, cubo ou hipercubo n -dimensional dependendo da sua dimensão característica. Além da forma, também é necessário determinar como as conexões das bordas devem ser tratadas. A borda pode ser preservada eliminando as conexões que não pertencem ao componente removido, como na rede regular bi-dimensional da Figura 13a, apresentando vértices com conectividade diferente dos demais. Também é possível eliminar a borda por completo reconectando os elementos ciclicamente como se a rede fosse repetida infinitamente ao longo do espaço original, como a Figura 13b, neste caso, os vértices voltam a apresentar a regularidade da grade infinita, com todos os

vértices virtualmente indistinguíveis ao considerar propriedades topológicas. As Figuras 12a, 12b e 12c mostram alguns exemplos de redes regulares geradas a partir de grades.

2.4.2 Redes Erdős-Rényi

O modelo de redes aleatórias homogêneo, inicialmente desenvolvido por Anatol Rapoport (60) e independentemente por Paul Erdős e Alfréd Rényi (54), foi uma das primeiras tentativas em estabelecer uma representação de redes reais através de poucos parâmetros. Também conhecidas como redes Erdős-Rényi (*ER*), estas redes são estocasticamente regulares, isto é, a maioria dos vértices apresenta grau próximo à média.

Redes ER representam poucos sistemas reais, no entanto, são muito importantes na tarefa de análise e comparação do comportamento de propriedades topológicas entre diferentes redes reais. As redes aleatórias também são úteis no estudo analítico de redes complexas, podendo ser usada como a base para outros modelos mais sofisticados.

Em uma rede ER (fig. 12d), os $|\mathcal{V}|$ vértices, inicialmente desconectados, são ligados ou não dependendo de um parâmetro de probabilidade p . Para redes com grande número de vértices a distribuição do grau é binomial (distribuição de Poisson) centrada no valor de conectividade média $\langle k \rangle = p|\mathcal{V}|$, como mostrado na Figura 14.

Figura 14 – Distribuição do grau para uma rede ER de 25 mil vértices e grau médio 12.

Fonte: Elaborada pelo autor.

2.4.3 Modelo Barabási-Albert

Desenvolvido por Albert-László Barabási e Réka Albert (43), o modelo Barabási-Albert (BA) foi usado na modelagem de diversos sistemas reais, como Internet (43), WWW (61)

e redes de citação.(38) O modelo resulta em redes *livre de escala*, um fenômeno presente na maioria das redes reais. Uma rede livre de escala apresenta uma distribuição de grau que obedece uma lei de potência (Figura 15). Em particular, para o BA, a lei de potência tem a forma $P(k) \propto k^{-3}$.

Diferentemente das redes ER, onde os vértices apresentam conectividade semelhante, em uma rede livre de escala é comum o surgimento de vértices com número de conexões muito maior do que a média, a Figura 12e ilustra esse fato.

Figura 15 – Distribuição de grau para uma rede BA com 25 mil vértices e grau médio 50.

Fonte: Elaborada pelo autor.

O termo *hub* costuma ser usado para denominar os vértices com maior número de conexões em uma rede livre de escala, no entanto, não é possível defini-lo formalmente, pois nessas redes, não há um valor característico de conectividade que permita a existência de um limiar intrínseco. Neste trabalho o termo hubs é usado para definir uma pequena fração dos vértices de maior conectividade de uma rede ou vértices com número de conexões de ordens de grandeza superior à média.

O modelo BA baseia-se no conceito de que redes reais possuem tanto uma dinâmica de crescimento quanto um mecanismo preferencial de ligação. Esses 2 fatores combinados são essenciais para a geração de redes livre de escala.(43, 11)

Uma rede BA é gerada a partir um conjunto inicial de $m_0 (\geq 2)$ vértices, que podem estar desconectados ou aleatoriamente conectados. Progressivamente, t novos vértices são adicionados e ligados a $m (\geq m_0)$ vértices dos $|\mathcal{V}_t|$ já presentes na rede. A escolha dos vértices é aleatória e obedece um mecanismo preferencial de ligação onde a probabilidade depende linearmente da conectividade do vértice destino v_τ . A probabilidade $P_t(v_\tau)$ de se conectar o novo vértice do instante t a ao vértice v_τ já existente na rede ($v_\tau \in \mathcal{V}_t$) pode ser escrita

Figura 16 – Estágios para a geração de uma rede BA com 10 vértices e $\langle k \rangle = 4$.
Fonte: Elaborada pelo autor.

como:

$$P(v_\tau) = \frac{k(v_\tau)}{\sum_{v_i \in \mathcal{V}_t} k(v_i)}. \quad (2.10)$$

O processo resulta em uma rede de tamanho $\mathcal{V} = m_0 + t$ com $\langle k \rangle = 2m$. Como novos vértices são progressivamente ligados, preferencialmente, aos mais conectados, estes tendem a adquirir ainda mais conexões, tornando-se hubs. A Figura 16 ilustra a construção de uma rede BA.

2.4.4 Redes Watts-Strogatz

Pequeno-mundo é o nome dado a redes caracterizadas por terem geodésicas curtas, isto é, seus vértices podem, em geral, ser acessados com poucos passos partindo de outro vér-

tice qualquer. Esse fenômeno costuma ocorrer em diversas redes reais, em especial redes sociais.⁽¹⁷⁾ Outra característica importante destas redes é a existência de aglomeração local de vértices, isto é, vértices conectados através de um terceiro vértice têm grande chance de também estarem conectados entre si, formando triângulos. Em uma rede social, por exemplo, a existência dessas conexões é comum pois amigos de uma pessoa tem grandes chances de serem amigos entre si. Portanto, estas redes apresentam coeficiente de aglomeração médio alto, comparado a redes do tipo ER.

Figura 17 – Estágios para a geração de uma rede WS com 10 vértices e $\langle k \rangle = 4$.

Fonte: Elaborada pelo autor.

Um modelo que representa bem uma rede pequeno mundo foi desenvolvido por Duncan J. Watts e Steven Strogatz.⁽⁵⁵⁾ Neste modelo, chamado de Watts-Strogatz (WS), usa-se como base uma rede regular que progressivamente tem suas arestas reconectadas aleatoriamente.

A construção de uma rede WS começa a partir de uma rede regular (Figura 17a) n -dimensional. Através de ensaios de Bernoulli com probabilidade p algumas arestas são reconectadas aleatoriamente. A Figura 17 ilustra o processo de construção destas redes.

Dependendo da quantidade de arestas reconectadas definida pelo parâmetro p , o modelo WS pode resultar em redes que apresentam características variando de redes

regulares a redes ER. Uma rede WS começa a apresentar o fenômeno de pequeno-mundo já com valores bem baixos de p , isto é, com poucos atalhos de longe alcance, ao mesmo tempo, ela mantém a estrutura local original da rede regular. Este comportamento pode ser entendido ao observar o coeficiente de aglomeração e a média das geodésicas com o valor de p , como visto na Figura 18.

Figura 18 – Coeficiente de aglomeração médio (quadrados em laranja) e média das geodésicas (círculos azuis) contra a probabilidade de reconexão. A região destacada corresponde às propriedades geralmente presentes em redes do tipo pequeno mundo, com alto coeficiente de aglomeração e baixa média das geodésicas.

Fonte: Elaborada pelo autor.

2.4.5 Modelos de redes geográficas

Na maioria das redes reais, o espaço original onde seus elementos residem pode influenciar diretamente na forma como eles são conectados. Por exemplo, estradas e rodovias são construídas conectando progressivamente cidades próximas. Devido ao custo, dificilmente observa-se conexões de longo alcance que atravessam grandes regiões nestas redes. Em redes de relações sociais, no entanto, o efeito geográfico é diminuído já que é possível manter relações de amizade mesmo com pessoas geograficamente distantes. Redes onde as limitações do espaço original fortemente influenciam as redes geradas são chamadas de redes geográficas.

Neste trabalho, usamos dois modelos simples para gerar redes geográficas: o modelo de *redes geográficas aleatórias* (RAGEO) (57) e o modelo de *Waxman* (WAX). (58)

Figura 19 – Tipos de redes geográficas usadas neste trabalho. Nota-se que, diferentemente da rede geográfica aleatória (a), a rede Waxman (b) apresenta alguns poucos caminhos de longo alcance.

Fonte: Elaborada pelo autor.

Redes geográficas aleatórias (57) são construídas inicialmente espalhando, uniformemente, vértices sobre um espaço n -dimensional. Em seguida, pares de vértices distantes de até um parâmetro δ_{geo} são conectados. A Figura 19a ilustra o processo de criação de uma rede geográfica aleatória. Estas redes podem apresentar coeficiente de aglomeração bem alto, entretanto não há caminhos de longa distância. As Figuras 12g e 12h mostram duas redes geradas por este processo, uma 2D e outra 3D.

O modelo de Waxman (58) é similar às RAGEOs, entretanto, ao invés de conectar todos os vértices distantes por até um certo valor, há uma probabilidade $P_{wax}(D(v_i, v_j))$ de dois vértices v_i e v_j serem conectados. Esta probabilidade, que depende da distância euclidiana $D(v_i, v_j)$ entre os dois vértices, é usualmente definida como:

$$P_{wax}(D(v_i, v_j)) = \alpha e^{-D(v_i, v_j)/\beta}, \quad (2.11)$$

onde α e β são parâmetros do modelo. Como resultado, as redes WAX, assim como as RAGEO, preservam a topologia do espaço original. Ainda assim, para o modelo WAX há uma baixa probabilidade (que decai exponencialmente) de haver conexões de longo alcance. A Figura 19b ilustra o processo de criação de redes WAX e a Figura 12i mostra uma rede final do modelo.

2.5 Estrutura de comunidades e modelo estocástico de blocos

Muitas redes reais apresentam um particionamento natural de seus vértices em subestruturas, chamadas de *comunidades*. Em redes sociais, espera-se que pessoas em grupos (ou *comunidade*) com interesses em comum, ou ainda que compartilhem o mesmo ambiente de trabalho, são propícias a estabelecerem mais relações de amizade entre si do que com pessoas de outros grupos. Similarmente, em redes de conhecimento, onde conceitos são ligados quando há uma relação direta entre eles, formam comunidades que geralmente indicam assuntos em comum.

Apesar de não existir uma definição formal, a estrutura de comunidades pode ser entendida como um particionamento de uma rede onde vértices de cada conjunto (comunidade) compartilham muitas conexões entre si, mas, comparativamente, poucas conexões entre vértices de conjuntos diferentes. Como a conectividade é uma propriedade relativa e depende da natureza da rede, a relação entre a conectividade interna e externa de comunidades em uma rede é melhor quantificada pela medida de modularidade.⁽⁶²⁾ Quanto maior o valor da modularidade, maior é a distinção entre as comunidades de uma rede comparada com o que ocorreria em uma rede aleatória de mesmas características.

A maioria dos métodos de detecção de comunidades tem como objetivo encontrar particionamentos da rede que maximizem o valor da modularidade.⁽²⁸⁾ Esta abordagem apresenta limitações. Em particular, métodos baseados na maximização da modularidade podem ser incapazes de detectar comunidades em certas escalas, mesmo que estas estejam bem definidas topologicamente.⁽⁶³⁾ Apesar disso, métodos baseados na maximização de modularidade estão entre os mais usados para se inferir a estrutura de comunidades de uma rede. Neste trabalho, optamos pelo uso do método multinível de detecção de comunidades ⁽⁶⁴⁾, onde vértices são progressivamente unidos com a finalidade de maximizar a modularidade. A principal vantagem deste método é a possibilidade de ser aplicado a redes grandes (com mais de 10 mil vértices) e também por construir configurações de particionamento que resultem em uma modularidade similar à dos outros métodos. A Figura 20 ilustra uma rede e sua estrutura de comunidades.

No *modelo estocástico de blocos* ⁽⁶⁵⁾ compreende-se que a estrutura de comunidades é uma característica estocástica de redes, além disso, propõe-se uma forma de gerar redes aleatórias que apresentem comunidades definidas através de uma matriz de mistura estocástica (*mixing matrix*).⁽⁶⁵⁾

A matriz de mistura estocástica Π está associada a um grafo reduzido (*coarse-grained graph*) ⁽⁶⁶⁾, onde cada vértice representa uma comunidade de uma rede maior, a conexão entre duas comunidades (α, β) indica a probabilidade de dois vértices, escolhidos

Figura 20 – Exemplo de estrutura de comunidades em uma rede complexa (a). Há preferência entre conexões de vértices dentro da mesma comunidade. O grafo reduzido (*coarse-grained*) (b) representa as relações de conectividade entre as comunidades. A largura das arestas neste grafo indicam a probabilidade estocástica de dois vértices de comunidades diferentes se conectarem.

Fonte: Elaborada pelo autor.

aleatoriamente dentro destas comunidades, estarem conectados. Desta forma, P_i é uma matriz de adjacência com pesos e é definida por:

$$\Pi_{\alpha\beta} = \frac{K_{\alpha\beta}}{|\alpha||\beta|}, \quad (2.12)$$

onde $K_{\alpha\beta}$ é o número total de conexões entre vértices das comunidades α e β .

O grafo reduzido obtido a partir de uma rede, ilustrado na Figura 20b, é uma forma adequada de representar as relações entre suas comunidades. A visualização desta estrutura pode ser usada para sumarizar e entender a rede original em uma escala global.

A geração de novas redes a partir da matriz de mistura Π , pode ser feita através de um processo estocástico simples. Inicialmente inicializa-se os vértices da nova rede com o rótulo de uma comunidade escolhida aleatoriamente, mas respeitando a proporção dos tamanhos de comunidade da rede original, $(|\alpha|, |\beta|, \dots)$. Em seguida, através de ensaios de Bernoulli, cada par de vértices (v_i, v_j) com $v_i \in \alpha$ e $v_j \in \beta$ é conectado ou não de acordo com a probabilidade $\Pi_{\alpha\beta}$. Como resultado a rede gerada apresenta mesma estrutura de comunidades, e portanto, mesma modularidade, mas sem, necessariamente, preservar outras propriedades, como a distribuição de grau e transitividade.

2.6 Isomorfismos e automorfismos em grafos

A representação da topologia de uma rede através de matrizes de adjacência, listas de adjacências ou outras formas similares, não é única. Por exemplo, ao permutar simultane-

amente um conjunto de colunas e linhas de uma matriz de adjacência, a topologia da rede resultante não é alterada, da mesma forma, alterar a ordem das entradas de uma lista de adjacências apenas altera a rotulagem dos vértices. Usa-se o termo isomorfismo (23, 67, 68) quando há, entre duas redes, uma transformação de rótulos dos vértices que mantém a estrutura topológica da rede. Desta forma, uma rede é isomorfa a outra se topologicamente elas são indistinguíveis. A Figura 21 ilustra um isomorfismo entre duas redes.

Figura 21 – Exemplo de isomorfismo em uma rede simples. Os rótulos de vértices da rede à esquerda podem ser associados aos da direita de modo que a topologia das duas redes não seja alterada. Portanto, as duas redes são isomorfas entre si. Propriedades topológicas consideradas em ambas as redes são indistinguíveis a não ser pelo rótulo dos vértices.

Fonte: Elaborada pelo autor.

Determinar se duas redes são isomorfas é ainda uma área de pesquisa desafiadora em teoria dos grafos. Geralmente, baseia-se em determinar uma forma canônica de rotulação de vértices que seja sempre equivalente para uma dada topologia independentemente de sua rotulação original. Não se conhece no momento algoritmo em tempo polinomial para resolver o problema de isomorfismo. Os melhores algoritmos encontrados na literatura apresentam alta complexidade computacional: $O(2^{|V|})$ (69) ou, com algumas limitações, $O(2^{\sqrt{|V|} \log_2 |V|})$ (70); portanto o uso de análise de isomorfismos ainda é proibitivo ao considerar redes com grande quantidade de vértices.

Vértices em uma rede também podem ser associados a valores ou rótulos secundários, por exemplo, indicando a sua participação em uma comunidade ou particionamento, desta forma, pode-se definir o que é chamado isomorfismo de cores. Neste contexto, a cor significa um rótulo secundário de cada vértice, de modo que só existe isomorfismo quando além da topologia, os rótulos secundários (conhecidos como cores) são preservados. A Figura 22 ilustra como o isomorfismo por cores pode ser diferente do isomorfismo topológico.

Outro conceito relacionado ao isomorfismo é o automorfismo em redes, que difere do primeiro por ser uma transformação de rótulos dentro da própria rede que preserva sua topologia. O grupo de automorfismos é o conjunto de todas as possíveis transformações de

Figura 22 – Isomorfismo de cores. Enquanto que A, B e C são isomorfos topologicamente, ao considerar as cores, apenas A e B se mantêm isomorfos.

Fonte: Elaborada pelo autor.

rótulos que não alteram a topologia da rede. Este conceito está diretamente relacionado à simetria em grafos já que regiões com muitos automorfismos indicam alta redundância e repetição da estrutura topológica da rede. A Figura 23 ilustra dois automorfismos em uma rede simples. Por fim, o custo computacional para se encontrar o grupo de automorfismo de um grafo é similar ao problema de isomorfismos.

Figura 23 – Possíveis automorfismos de uma rede simples. O primeiro automorfismo (a) envolve apenas a troca de rótulos entre 2 pares de vértices. Considerando o posicionamento dos vértices este automorfismo é similar à simetria vertical. Já o segundo automorfismo (b) envolve todos os vértices da rede e se mostra conceitualmente similar a uma simetria horizontal da rede.

Fonte: Elaborada pelo autor.

Na maioria dos casos, tanto a teoria de isomorfismos quanto a de automorfismos não podem ser diretamente aplicadas a redes reais. Um dos motivos principais é claramente o seu alto custo computacional. Entretanto, também há o fato de que redes reais geralmente apresentam ruído ou são construídas a partir de amostragem, enquanto que isomorfismos e automorfismos são instáveis a pequenas mudanças na rede .(71) Na Seção 6 deste trabalho propomos uma alternativa ao uso de automorfismos para se determinar e quantificar sime-

trias em redes. Similarmente, o problema de isomorfismos para fins de comparação de redes pode ser tratado de forma mais estável como um problema de proximidade de propriedades topológicas.(72)

2.7 Análise de redes por componentes principais

Diferentes métricas revelam diferentes propriedades em redes complexas. No entanto, a composição dessas diferentes métricas pode apresentar redundância ou informações não relevantes para a caracterização de seus vértices ou das próprias redes. Outro problema é determinar quais propriedades apresentam maior quantidade de informação, de forma a reduzir o tamanho do espaço de dados, permitindo a aplicação de métodos de visualização e classificação de dados. A análise multivariável baseada nos *componentes principais*, ou *PCA* (73, 59), pode ajudar na solução deste problema.

A metodologia PCA tem como objetivo obter bases ordenadas por relevância compostas por combinações lineares do conjunto de dados e, simultaneamente, eliminar as possíveis redundâncias reduzindo a dimensionalidade. A aplicação do método é simples, e baseia-se na obtenção da base dos auto-vetores mais importantes da matriz de covariância; e na projeção dos dados nessa nova base.

Supondo um conjunto de dados, $\mathcal{X}$, composto por N observações de M diferentes variáveis, em notação vetorial, $\vec{X}_k = \{X_{1k}, X_{2k}, \dots, X_{Mk}\}$, com $k = \{1, 2, 3 \dots N\}$; deseja-se reduzir a dimensionalidade de $\mathcal{X}$ para L , com $L < M$, mantendo a maior quantidade da informação.

Os elementos da matriz de covariância, Σ_{ij} , de tamanho $M \times M$; são definidos pelos valores das covariância dos respectivos vetores observados, X_{ki} , onde X é a matriz composta pelos N vetores $\vec{X}_k$ agrupados. A Equação 2.13 define a matriz de covariância:

$$\Sigma_{ij} = \text{cov}(X_i, X_j) = \langle (X_i - \mu_i)(X_j - \mu_j) \rangle \quad (2.13)$$

onde $\langle \dots \rangle$ significa o valor esperado, ou valor médio ao longo das N observações e

$$\mu_i = \langle X_i \rangle = \frac{1}{N} \sum_{k=1}^N X_{ik} \quad (2.14)$$

Uma forma mais elegante para se definir a matriz de covariância é através do *produto externo* (ou produto tensorial) dos vetores de desvio da média, $\vec{D}_k = \{D_{1k}, D_{2k}, \dots, D_{Mk}\}$ definidos pelos elementos:

$$D_{kj} = X_{ki} - \mu_i \quad (2.15)$$

A matriz de covariância pode ser calculada pela média em k do produto externo de $\vec{D}_k$ por ele mesmo:

$$\Sigma = \langle \vec{D} \otimes \vec{D} \rangle = \langle \vec{D} \cdot \vec{D}^* \rangle \quad (2.16)$$

onde $\vec{D}^*$ é a matriz adjunta do vetor $\vec{D}$. Como o vetor é composto por números reais, $\vec{D}^* = \vec{D}^T$.

A matriz de covariância é simétrica, portanto Hermitiana, o que garante que todos os autovalores são reais e que os auto-vetores são ortogonais. Com o objetivo de eliminar a covariância entre as diferentes variáveis é desejável que a matriz de covariância para os dados transformados seja diagonal. Para isso, projetam-se os vetores dos dados originais na base ortonormal composta pelos auto-vetores de Σ , $\vec{v}_i$ com $i = \{1, 2, 3, \dots M\}$.

A cada auto-vetor $\vec{v}_i$ está associado um autovalor λ_i , que corresponde ao elemento $\Sigma_{ii}^{(\text{transformada})}$ da matriz de covariância dos vetores projetados, $\vec{P}_k$. Os elementos da diagonal da matriz de covariância equivalem à variância:

$$\lambda_i = \Sigma_{ii}^{(\text{transformada})} = \text{var}(P_i) = \text{var}(\vec{v}_i \cdot \vec{X}) \quad (2.17)$$

portanto, os vetores projetados mais significativos são aqueles compostos pelos auto-vetores de maiores autovalores. Pode-se, então, ordenar os autovalores e os correspondentes auto-vetores em ordem decrescente tal que $\lambda_1 > \lambda_2 > \lambda_3 > \dots \lambda_M$.

A projeção das variáveis na nova base pode ser feita membro a membro como indicado pela Equação 2.18

$$P_{ik} = \vec{v}_i \cdot \vec{X}_k \quad (2.18)$$

(a) Dados originais e auto-vetores da matriz de covariância.

(b) Projeção usando PCA.

Figura 24 – Exemplo de PCA para um conjunto de dados de 2 variáveis distribuído de modo normal ao longo de uma reta.

Fonte: Elaborada pelo autor.

Geometricamente, os valores originais foram submetidos a uma rotação n -dimensional.

(74) A Figura 24 ilustra um exemplo clássico onde os valores do conjunto de dados estão

distribuídos ao redor de uma reta. Verifica-se que as variáveis originais, “A” e “B” são altamente correlacionadas. O auto-vetor de maior autovalor, equivalente a nova variável chamada de “PCA1” ajusta-se perfeitamente na direção de maior variância, enquanto que para o outro auto-vetor, “PCA2”, é ortogonal a primeira e apresenta muito menos variância, e conseqüentemente menos informação. Neste caso, a redução de dimensionalidade para L poderia ser feita simplesmente descartando-se as variáveis menos importantes (menores autovalores). Nota-se que após a transformação, como esperado, as variáveis não estarão correlacionadas.

É possível aplicar PCA tanto na visualização do espaço gerado por redes (cada ponto indica uma rede) ou para cada vértice de um conjunto de redes. Neste último caso, espera-se que os vértices de uma mesma rede ou tipo de rede compartilhem uma mesma região.

2.8 Propriedades multiescala em redes complexas

Ao representar sistemas reais por redes complexas, algumas de suas características intrínsecas também podem emergir de suas redes. Uma delas é a formação de diferentes estruturas que coexistem no mesmo espaço, mas em diferentes escalas, um fenômeno presente em diversos sistemas reais. Por exemplo, a estrutura local de redes de vias públicas, em cidades, geralmente é formada por vias acesso de baixa velocidade (que formam as quadras) e vias de acesso rápido (como avenidas), entretanto em uma escala maior, como em um estado ou província, os detalhes internos de uma cidade já não contribuem para a caracterização do sistema, que agora é formado principalmente por algumas estradas e rodovias que conectam cidades entre si. *Medidas multiescala* (Figura 25) em redes complexas devem ser capazes de, através do ajuste de um ou mais parâmetros, caracterizar redes sob a perspectiva de diferentes escalas.

Neste trabalho o termo multiescala, apesar de relacionado, difere daquele usualmente empregado em processamento de sinais e formalizado em análise *multiresolução*.^(75, 76, 77) Neste tipo de análise, um sinal ou imagem, pode ser estudado sobre a perspectiva de diversas escalas, entretanto as medidas que garantem a possibilidade desta abordagem precisam obedecer uma série de formalismos matemáticos.⁽⁷⁵⁾ Um exemplo deste tipo de análise é a aplicação do filtro de Gabor baseado em *wavelets* (que difere do filtro de Gabor tradicional) ⁽⁷⁸⁾ em imagens. Alterando-se um parâmetro de escala do filtro é possível detectar a orientação de retas em uma imagem sob diferentes escalas. Por exemplo, em uma imagem de alta resolução obtida por satélite, é possível identificar a orientação de formações rochosas ou contornos do litoral de um país em uma escala macroscópica, da mesma forma, é possível observar a orientação de estradas e divisões rurais sob outra escala.

Um método para se analisar uma rede sobre o ponto de vista multiescala é atra-

Figura 25 – Conceito de caracterização multiescala em redes complexas. Enquanto medidas locais e globais caracterizam redes e vértices apenas em suas respectivas escalas, medidas multiescala podem caracterizar vértices em diferentes escalas, inclusive a intermediária.

Fonte: Elaborada pelo autor.

vés de técnicas de renormalização.(79). Elas se baseiam em, progressivamente, simplificar redes unindo vértices próximos de até uma certa distância topológica, que representa a escala, formando novos vértices, ao mesmo tempo em que se mantém a conectividade entre eles. Um problema desta abordagem se deve ao fato da necessidade de uma heurística para agregar os vértices que culmina em um problema de otimização. Isto torna a extensão de métodos tradicionais para a obtenção da dimensão em fractal, como a contagem de caixas (*box counting*) (79, 80), complicado de ser tratado em redes. Outro problema se deve à perda do contexto de cada vértice (que agora formam grupos de vértices), não permitindo que estes sejam caracterizados de acordo com sua participação na rede ao longo de diferentes escalas. Por fim, o custo computacional destas técnicas geralmente é proibitivo em redes com grande número de vértices e arestas.

Uma alternativa computacionalmente viável e simples para a análise multiescala em redes é a caracterização de vértices de acordo com suas sucessivas vizinhanças, também conhecida por análise *concêntrica*.(81, 82, 83) Nesta abordagem, propriedades são obtidas de subgrafos construídos a partir das vizinhanças de um vértice, chamados de *padrões concêntricos* (*concentric patterns*). Desta forma é possível caracterizar os vértices em diferentes escalas, que são expressas pelo número de vizinhanças consideradas nos padrões concêntricos. As próximas seções deste capítulo definem padrões e níveis concêntricos e algumas propriedades que podem ser consideradas nestas estruturas, como as propriedades concêntricas e a

medida de acessibilidade em redes. Nos capítulos 5 e 6, as medidas que constituem as principais contribuições deste trabalho, dimensão multiescala e simetria concêntrica, também são consideradas propriedades multiescala de redes e baseiam-se nos formalismos apresentados neste capítulo.

2.8.1 Níveis e padrões concêntricos

Propriedades básicas de redes complexas, como a conectividade e o coeficiente de aglomeração, fornecem informações importantes para a caracterização dos vértices e da própria rede. Estas propriedades, no entanto, consideram apenas a estrutura da vizinhança imediata de cada vértice. As propriedades topológicas locais de um vértice podem não ser suficientes para descrever seu papel na rede, necessitando de informação adicional, não presente na estrutura de sua vizinhança imediata. Uma alternativa para completar essas propriedades é estendê-las ao considerar não somente os vizinhos mais próximos, como também os vizinhos mais distantes, através de uma análise multiescala. Esta abordagem pode ser realizada através do uso dos conceitos de *níveis concêntricos* (ou hierárquicos) e *padrões concêntricos*, formalizados em (82) e (25, 33).

O nível concêntrico r de um vértice v_j corresponde à região da rede definida pelos vértices situados a uma distância topológica $g_{ij} = r$ a partir de v_j . Desta forma, é possível reorganizar os vértices de uma rede de acordo com os níveis concêntricos de cada um em relação a um vértice central v_i de interesse, formando estruturas chamadas de *anéis concêntricos* e *padrões concêntricos*. A Figura 26 ilustra o conceito de anéis e padrões concêntricos.

Cada anel concêntrico, $R_i(r)$, é definido pelo nível concêntrico r e por um vértice central v_i ; e contém todos os vértices $v_j \in \mathcal{V}$ tal que $g_{ij} = r$, isto é, o anel de nível 1, $R_i(1)$, contém todos os vértices que são vizinhos imediatos ao vértice central v_i , o anel de nível 2, $R_i(2)$, é formado por todos os vértices vizinhos aos do nível 1 que não pertençam a $R_i(1)$. O mesmo raciocínio pode ser usado para determinar os vértices de um anel $R_i(r+1)$ conhecendo-se $R_i(r)$. De forma similar, o padrão concêntrico, também conhecido como *bola concêntrica*, $B_i(r)$, é definido como a sub-rede construída a partir dos vértices presentes nos níveis concêntricos $\leq r$, portanto, constituído pela união de todos os anéis concêntricos do nível 0 ao nível r . A Figura 27 exemplifica a definição dos níveis concêntricos de um padrão concêntrico de nível $r = 3$.

2.8.2 Medidas concêntricas

A caracterização de níveis e padrões concêntricos pode ser feita a partir das *medidas concêntricas* (previamente chamadas de medidas hierárquicas) (82, 83), que são definidas para cada anel ao longo dos níveis concêntricos de um padrão concêntrico. Apesar de estende-

Figura 26 – Anéis e padrões concêntricos obtidos para $r \leq 2$ a partir de dois vértices escolhidos como centro (α e β) em uma rede simples. Em (a), anéis são mostrados na rede por contornos ao redor dos vértices. Os respectivos padrões concêntricos extraídos da rede para os vértices α e β e $r = 2$ são mostrados em (b).

Fonte: Elaborada pelo autor.

Figura 27 – Exemplo de anéis (ou níveis) concêntricos centrados em 1 considerando um padrão concêntrico de nível $r = 3$.

Fonte: Elaborada pelo autor.

rem as medidas tradicionais, as medidas concêntricas não se derivam trivialmente delas, por exemplo, a conectividade se estende em 4 novas propriedades e 2 outras não possuem análogo tradicional. As propriedades estão definidas para cada anel centrado em um vértice da

rede e dependem exclusivamente dos vértices que estão no anel e das conexões internas ou entre os anéis vizinhos. Os próximos parágrafos detalham as medidas concêntricas tendo como base a Figura 27 para exemplificar seus cálculos.

Número de vértices concêntricos (number of concentric nodes), $n_i(r) \equiv |R_i(r)|$, é a mais simples das medidas concêntricas e é definida como o número de vértices que pertencem ao anel $R_i(r)$. Por exemplo, na Figura 27 o número de vértices dos 4 níveis concêntricos são respectivamente: $n_1(0) = 1$, $n_1(1) = 3$, $n_1(2) = 4$ e $n_1(3) = 6$.

Volume da bola concêntrica (concentric volume), $V_i(r) \equiv |B_i(r)|$, indica o número de vértices na sub-rede $B_i(r)$. Também pode ser obtido pela soma de $n_i(r)$ ao longo dos níveis concêntricos, isto é, $V_i(r) = \sum_{d=0}^r n_i(d)$. Considerando o exemplo da figura, $V_1(0) = 1$, $V_1(1) = 4$, $V_1(2) = 8$ e $V_1(3) = 14$.

Números de arestas em anel (number of edges on a ring), $e_i(r)$, consiste no número de arestas exclusivamente contidas no anel $R_i(r)$, isto é, todas as arestas que conectam vértices compartilhando o mesmo nível concêntrico r . No exemplo da figura: $e_1(0) = 0$, $e_1(1) = 1$, $e_1(2) = 0$ e $e_1(3) = 3$.

Grau concêntrico (concentric degree), $k_i(r)$, é a extensão direta do grau clássico aplicada aos níveis concêntricos, sendo definido como o número de conexões que vértices de um anel $R_i(r)$ possuem com o do nível seguinte, $R_i(r+1)$. Para a rede da Figura 27, $k_1(0) = 3$, $k_1(1) = 5$ e $k_1(2) = 7$.

Grau intranível (intra-level degree), $Al_i(r)$, definida como o valor médio do grau dos vértices do anel considerando apenas as conexões internas, isto é, ignorando conexões com os vértices dos anéis: anterior, $R_i(r-1)$, e subsequente, $R_i(r+1)$. Também pode-se defini-la em termos das outras propriedades concêntricas pela equação como

$$Al_i(r) = \frac{2e_i(r)}{n_i(r)}. \quad (2.19)$$

Nota-se que para o anel de nível 0, seu valor é sempre $Al_i(0) = 0$. Na figura de exemplo tem-se: $Al_1(1) = 2/3$, $Al_1(2) = 0$ e $Al_1(3) = 1$.

Grau entreníveis (inter-level degree), $El_i(r)$, obtida pelo número médio de arestas que conectam o anel $R_i(r)$ ao seguinte, $R_i(r+1)$. Em termos das outras propriedades é definida por

$$El_i(r) = \frac{k_i(r)}{n_i(r)}. \quad (2.20)$$

Por exemplo, para a rede da Figura 27, tem-se: $El_1(0) = 3$, $El_1(1) = 5/3$ e $El_1(2) = 7/4$.

Coeficiente de aglomeração concêntrico (concentric clustering coefficient), $Cc_i(r)$, estende a medida do coeficiente de aglomeração tradicional. É obtido através de

$$Cc_i(r) = 2 \frac{e_i(r)}{n_i(r)(n_i(r) - 1)}. \quad (2.21)$$

Nota-se que o coeficiente de aglomeração concêntrico preserva as mesmas propriedades da medida tradicional, variando entre 0 e 1. Por exemplo, para a rede da figura, tem-se: $Cc_1(1) = 0$, $Cc_1(2) = 1/3$, $Cc_1(3) = 0$ e $Cc_1(4) = 1/5$.

Taxa de convergência (convergence ratio), $Cr_i(r)$, quantifica a razão do número de novas conexões ao longo dos níveis pelo número de arestas dos níveis seguintes. É definida como

$$Cr_i(r) = \frac{k_i(r)}{n_i(r+1)}. \quad (2.22)$$

Para a rede exemplificada, $Cr_1(0) = 1$, $Cr_1(1) = 5/4$ e $Cr_1(2) = 7/6$.

Grau concêntrico comum (concentric common degree), $Kh_i(r)$, é a média dos graus tomada para todos os vértices pertencentes ao anel $R_i(r)$, considerando, inclusive, as conexões que não compartilham mesmo nível concêntrico. Em termos das propriedades concêntricas pode ser definido pela Equação 2.23:

$$Kh_i(r) = \frac{k_i(r) + k_i(r-1) + 2e_i(r)}{n_i(r)}. \quad (2.23)$$

Para a rede da figura de exemplo, $Kh_1(0) = 1$, $Kh_1(1) = 10/3$, $Kh_1(2) = 3$ e $Kh_1(3) = 13/6$.

Em (82) e (35) as propriedades concêntricas são estudadas detalhadamente para os principais modelos de redes e algumas redes reais.

2.8.3 Acessibilidade

Assim como as medidas concêntricas, a *acessibilidade* (84, 85) também é uma medida multiescala em redes complexas e também pode ser considerada como uma extensão multi-escala do grau e caracteriza a diversidade estatística de acesso a vértices na rede a partir de uma referência. Ela é formalmente definida em termos de uma normalização da entropia de diversidade.(86)

A entropia de diversidade em redes complexas, $E_i(r)$ é definida como a entropia das probabilidades de transição, $P_r(v_i, v_j)$, de, a partir do vértice de referência v_i , alcançar um vértice v_j por meio de um caminho de comprimento r . Apesar da definição da medida não depender do tipo de caminhada escolhida, costuma-se utilizar as dinâmicas de caminhada aleatória ou autoesquivante (*self-avoiding random walk*). (87) Em uma caminhada aleatória, as probabilidades de acesso a vértices vizinhos são uniformemente distribuídas, isto é, para cada estágio da caminhada a probabilidade de se escolher algum vértice vizinho ao atual é sempre igual à dos outros vizinhos. Já em uma caminhada autoesquivante, a probabilidade em se visitar um vértice já percorrido na rede é nula, desta forma, nunca um vértice pode ser percorrido duas ou mais vezes. Neste trabalho optamos pela utilização da caminhada autoesquivante no cálculo da acessibilidade, isso se dá pelo fato de que o uso da caminhada

aleatória se torna uma métrica dependente da paridade dos tamanhos dos caminhos em grande parte das redes.(88) No capítulo 6 uma dinâmica autoesquivante modificada é usada para adaptar a acessibilidade como medida de simetria em redes.

Antes de definir a acessibilidade, é necessário definir a entropia de diversidade, $E_i(r)$, que pode ser calculada diretamente através de todas as probabilidades de transição, de acordo com uma caminhada autoesquivante de tamanho r , pela equação:

$$E_i(r) = - \sum_j \begin{cases} P_r(v_i, v_j) \ln(P_r(v_i, v_j)) & \text{se } P_r(v_i, v_j) \neq 0 \\ 0 & \text{se } P_r(v_i, v_j) = 0 \end{cases}. \quad (2.24)$$

A acessibilidade $a_i(r)$, é então definida diretamente pela exponencial da diversidade não normalizada:

$$a_i(r) = e^{E_i(r)}. \quad (2.25)$$

Através desta normalização a acessibilidade apresenta a mesma unidade de número de vértices, podendo variar entre 0 e o número de vértices alcançáveis pelas caminhadas autoesquivantes de tamanho r . Em termos práticos, a acessibilidade mede o número efetivo de vértices acessíveis em r passos para o tipo de caminhada considerada. O cálculo de $a_i(r)$ para um vértice central v_i em uma rede equivale ao do padrão concêntrico de r níveis, já que a distância máxima do vértice de referência, ao longo da caminhada é sempre menor ou igual a r . A Figura 28 ilustra como a heterogeneidade das probabilidades de transição influenciam no valor da acessibilidade.

Figura 28 – Exemplo do cálculo da acessibilidade para dois padrões concêntricos de 2 níveis. Em (a) os vértices G e F podem ser acessados, a partir de A , com mesma probabilidade $P_2(A, G) = P_2(A, F) = 1/2$, desta forma a entropia é máxima e, portanto, $a_A(r) = 2$. No segundo exemplo (b), a partir de α é muito mais provável que uma caminhada termine em θ do que em ι , resultando em uma distribuição de probabilidades menos homogênea. Neste caso a acessibilidade diminui em comparação com (a), com valor $a_\alpha(r) = 1.89$.

Fonte: Elaborada pelo autor.

A acessibilidade se mostrou eficiente para a determinação de bordas em diversas redes reais e modelos, apresentando melhor desempenho do que as medidas que usam informações globais, como as de centralidade.(86, 85) A Figura 29 ilustra a performance da

acessibilidade, comparada à centralidade de intermediação (C_B), em detectar bordas em uma rede geográfica aleatória.

Figura 29 – Acessibilidade (a) e centralidade de intermediação (b) calculadas para cada vértice de uma rede geográfica aleatória, mostradas de acordo com as cores das legendas. Verifica-se que a acessibilidade se mostra mais eficiente na detecção de bordas, além de apresentar valores mais estáveis ao longo da rede.

Fonte: Elaborada pelo autor.

3 Visualização de redes complexas

Com o crescente aumento de interesse na área de redes complexas, também surgiu a necessidade de se desenvolver e implementar métodos para visualizar e interagir com estruturas desta natureza. A visualização de redes complexas compartilha uma grande interseção com a visualização de grafos, uma subárea importantíssima da visualização científica. O principal objetivo da visualização científica é sumarizar dados científicos por meio de figuras ou ferramentas visuais interativas, que podem servir tanto para divulgação de resultados quanto para a exploração e interpretação dos mesmos. A visualização, quando interativa, se torna uma forma naturalmente multiescala de análise de redes complexas, permitindo a exploração de diferentes aspectos da rede tanto local quanto globalmente. Devido a esta característica, a visualização se torna uma ferramenta essencial para entender o comportamento das propriedades multiescala estudadas e desenvolvidas ao longo deste trabalho.

A área de visualização de grafos já foi bem explorada na literatura (89), entretanto mesmo compartilhando a mesma estrutura abstrata de representação de dados, o problema da visualização de redes complexas apresenta diversas particularidades, que em geral dependem da natureza de cada sistema estudado e de suas características de interesse. Por exemplo, propriedades topológicas, como grau e coeficiente de aglomeração, podem ser importantes no estudo de um sistema em particular, de forma que possa ser incorporado em técnicas de visualização. Isto torna a área de visualização de redes complexas um problema ainda em aberto.

Em geral, tanto em teoria de grafos quanto em redes complexas, o problema de visualização baseia-se em determinar um diagrama de símbolos (usados para representar os vértices em uma rede) e linhas, que os conectam (indicando as arestas), distribuídos ao longo de uma variedade, geralmente um espaço métrico 2D ou 3D. Tanto o problema da escolha dos símbolos quanto de suas posições, também conhecido como *problema de disposição de vértices* (*network layout*), são áreas de pesquisa em visualização científica. Na maioria dos casos, o problema da escolha dos símbolos pode ser solucionado de forma simples, por exemplo através do uso de formas geométricas para representar vértices, como círculos, quadrados, esferas, etc. Sua forma, cores e tamanho, através de um mapeamento simples, podem indicar diretamente uma ou mais propriedades particulares de cada vértice. Já o problema de disposição de vértices, que exige um mapeamento do tipo $M : \mathcal{V} \rightarrow \mathbb{R}^d$, em geral, é mais complicado e pode exigir o uso de técnicas sofisticadas para a obtenção de resultados satisfatórios.

Originalmente, o problema de associar a cada vértice $v_i \in \mathcal{V}$ uma posição em um espaço é equivalente ao problema de mergulho de grafos. Em um mapeamento de um grafo

em um espaço métrico, $M : \mathcal{V} \rightarrow \mathbb{R}^n$, é desejável que sua topologia seja preservada, ou seja, a vizinhança de cada vértice projetado deve, preferencialmente, ser mantida. Isto só é possível quando a dimensão da rede é compatível com a dimensão do espaço de mergulho. Quando uma rede é projetada em um espaço $\mathbb{R}^n$ com n menor que sua dimensão característica, não é garantido que sua topologia será preservada, entretanto, na maioria dos casos este tipo de visualização é suficiente para interpreta-las, da mesma forma que é possível interpretar qualitativamente a forma de um tesseracto (um hipercubo de dimensão 4) em um espaço 3D ou até mesmo 2D. Desta forma, na maioria dos casos, o mergulho de redes também pode ser interpretado como uma técnica de redução de dimensionalidade.

Figura 30 – Exemplos de diferentes métodos de disposição de vértices aplicados a uma rede simples.
Fonte: Elaborada pelo autor.

Diversas técnicas com a finalidade de dispor vértices ao longo de espaços métricos existem, como métodos espectrais (92), *multidimensional scaling* (94, 95), circular (90), hierárquico (91), métodos baseados em forças (96, 93), etc. A Figura 30 ilustra alguns dos métodos mais usados de visualização de redes complexas. Métodos baseados em forças estão entre os mais versáteis, principalmente pela existência de técnicas que podem facilmente ser adaptadas para a exploração de diversas propriedades de uma rede. Neste trabalho, optamos pelo uso de métodos baseado em forças para o estudo de redes complexas em geral. Esta técnica

foi escolhida devido à sua versatilidade, pois não exige quaisquer informações extras sobre as redes, e na maioria dos casos gera visualizações compreensíveis.(93) Outra vantagem é que as leis de força podem ser facilmente modificadas, gerando uma família de técnicas de visualização. Para isso, desenvolvemos e implementamos melhorias ao software já desenvolvido pelo autor ao longo de seu projeto de mestrado.(35) As próximas seções deste capítulo descrevem os métodos usados para a visualização de redes, com destaque para as melhorias implementadas neste projeto.

3.1 Visualização interativa de redes através de métodos baseados em forças físicas

Métodos dirigidos por forças (96, 93) utilizam uma analogia física, onde cada vértice é associado a uma partícula e cada aresta por uma interação de forças entre elas. A técnica, se resume em encontrar a melhor, ou uma das melhores, formas de dispor os vértices de modo que a energia total das interações deste sistema físico seja mínima. Simultaneamente, informações extras da rede podem ser incorporadas às interações deste método, de modo a refletir alguma característica de interesse, presente no sistema original.

A técnica incorporada neste trabalho é baseada no algoritmo de forças Fruchterman-Riengold (*FR*) (93), que abrange interações do tipo forças repulsivas e atrativas. A técnica inicia dispondo aleatoriamente os vértices da rede ao longo de um espaço 2D ou 3D. Nesta técnica, uma força atrativa existe entre dois vértices sempre que estes estão conectados. A força repulsiva, no entanto, está presente em todos os pares de vértices, estejam estes conectados ou não. O sistema é similar ao problema físico de N partículas carregadas de mesma carga, com alguns pares conectados por molas. Entretanto, originalmente, o conjunto de forças do método *FR* não apresenta equivalente físico na natureza, sendo a força de atração entre dois vértices conectados v_i e v_j , $\vec{f}_a^{\text{FR}}(v_i, v_j) = (\vec{R}_i - \vec{R}_j)^2 / K \hat{r}_{ij}$ e a de repulsão entre todos os pares de vértices, $\vec{f}_r^{\text{FR}}(v_i, v_j) = -K^2 / |\vec{R}_i - \vec{R}_j| \hat{r}_{ij}$, onde $\vec{R}_i$ é a posição no espaço do vértice v_i , $\hat{r}_{ij}$ é o versor na direção $(\vec{R}_j - \vec{R}_i)$ e K uma constante. Através de arrefecimento simulado (*simulated annealing*), as posições dos vértices são otimizadas para minimizar a energia do sistema físico equivalente. Como resultado, vértices conectados tendem a permanecer a uma distância K no espaço alvo, enquanto as distâncias entre vértices desconectados tendem a ser a máxima possível.

A visualização de redes através de métodos dirigidos por forças pode revelar informações qualitativas importantes acerca da estrutura topológica de redes complexas. Características como estrutura de comunidades, existência de hubs, restrições geográficas, existência de aglomerações, etc; se tornam evidentes através desta técnica.

Em (35), propomos algumas modificações ao algoritmo *FR* (*FRMOD*). Ao invés do

uso de técnicas de arrefecimento simulado, que necessitam de uma função de esfriamento adequada para que o sistema não migre para estados de mínimos locais de energia, permitimos que o sistema evoluísse da configuração original (com vértices distribuídos aleatoriamente) através de simulação molecular (*MDS*).⁽⁹⁷⁾ Nesta abordagem, os vértices se comportam exatamente como partículas que interagem por forças reais, e portanto, apresentam posição, velocidade e aceleração que variam ao longo do tempo. Também é necessário acrescentar uma força do tipo viscosa que diminui progressivamente a energia do sistema. A técnica *MDS* foi escolhida por apresentar ao usuário uma visualização com movimentos naturais em tempo real, diferentemente da abordagem original ⁽⁹³⁾, onde é necessário definir uma função de esfriamento. Com a modificação sugerida, o usuário pode alterar dinamicamente os parâmetros da visualização, que correspondem àqueles usados na simulação molecular, tornando a visualização mais agradável com movimentos naturais e previsíveis.

O sistema dinâmico simulado pelo FRMOD pode ser descrito através de um conjunto de equações diferenciais definidas ao longo do tempo para cada vértice v_i :

$$\frac{d^2 \vec{R}_i}{dt^2} = \vec{F}_a(v_i) + \vec{F}_r(v_i) - \mu \frac{d\vec{R}_i}{dt}, \quad (3.1)$$

onde $\vec{F}_a(v_i)$ e $\vec{F}_r(v_i)$ são as forças atrativas e repulsivas totais atuando sobre v_i , e μ é uma constante de viscosidade.

Ao testar outras combinações de forças, verificamos que uma força de decaimento quadrático com a distância apresentou melhor performance e mais estabilidade durante simulações ⁽³⁵⁾ de diversas redes. Desta forma optamos pelas seguintes relações de forças totais:

$$\vec{F}_a(v_i) = \sum_{(v_i, v_j) \in \mathcal{E}} a_{ij} (\vec{R}_i - \vec{R}_j)^2 \hat{r}_{ij} \quad (3.2)$$

$$\vec{F}_r(v_i) = \sum_{v_j \in \mathcal{V}} \frac{-b}{(\vec{R}_i - \vec{R}_j)^2} \hat{r}_{ij}. \quad (3.3)$$

onde a_{ij} e b são respectivamente as constantes de atração e de repulsão. A constante de atração pode também depender do par de vértices conectados, por exemplo do peso da conexão $a_{ij} \propto w_{ij}$. Com estas modificações, a distância preferencial $|r_{ij(ideal)}| = |\vec{R}_{i(ideal)} - \vec{R}_{j(ideal)}|$ ocorre quando a força total atuando sobre um vértice é nula, isto é $|r_{ij(ideal)}| = \left(\frac{b}{a_{ij}}\right)^{\frac{1}{4}}$.

Para resolver a equação diferencial 3.1, usamos o método de integração de Verlet.⁽⁹⁸⁾ Desta forma, a discretização da aceleração $d^2 \vec{R}_i(t)/dt^2$ para um dado instante t com intervalo de integração Δt fica:

$$\frac{d^2 \vec{R}_i(t)}{dt^2} \approx \frac{\Delta^2 \vec{R}_i}{\Delta^2 t} = \frac{\frac{\vec{R}_i(t+\Delta t) - \vec{R}_i(t)}{\Delta t} - \frac{\vec{R}_i(t) - \vec{R}_i(t-\Delta t)}{\Delta t}}{\Delta t} \quad (3.4)$$

$$= \frac{\vec{R}_i(t + \Delta t) - 2\vec{R}_i(t) + \vec{R}_i(t - \Delta t)}{\Delta t^2}. \quad (3.5)$$

A equação diferencial pode ser então escrita como:

$$\frac{\vec{R}_i(t + \Delta t) - 2\vec{R}_i(t) + \vec{R}_i(t - \Delta t)}{\Delta t^2} = \vec{F}_{\text{total}}(v_i) \quad (3.6)$$

$$\vec{R}_i(t + \Delta t) = 2\vec{R}_i(t) - \vec{R}_i(t - \Delta t) + \vec{F}_{\text{total}}(v_i)\Delta t^2 \quad (3.7)$$

O erro associado a esta aproximação é $\vec{O}(\Delta t^4)$. Isto pode ser demonstrando facilmente ao somar $\vec{R}(t + \Delta t)$ com $\vec{R}(t - \Delta t)$:

$$\vec{R}(t + \Delta t) = \vec{R}(t) + \vec{V}(t)\Delta t + \frac{\vec{A}(t)\Delta t^2}{2} + \frac{\vec{B}(t)\Delta t^3}{6} + \vec{O}(\Delta t^4) \quad (3.8)$$

$$+ \vec{R}(t - \Delta t) = \vec{R}(t) - \vec{V}(t)\Delta t + \frac{\vec{A}(t)\Delta t^2}{2} - \frac{\vec{B}(t)\Delta t^3}{6} + \vec{O}(\Delta t^4) \quad (3.9)$$

$$\vec{R}(t + \Delta t) = 2\vec{R}(t) - \vec{R}(t - \Delta t) + \vec{A}(t)\Delta t^2 + \vec{O}(\Delta t^4), \quad (3.10)$$

que tem a mesma forma da Equação 3.7.

Como não é requerido que o sistema se comporte perfeitamente como um sistema físico real, o cálculo da velocidade $d\vec{R}_i(t)/dt$, usado para determinar a força viscosa, é feito através do método de Euler, portanto:

$$\frac{d\vec{R}_i(t)}{dt} \approx \frac{\Delta \vec{R}_i}{\Delta t} = \frac{\vec{R}_i(t + \Delta t) - \vec{R}_i(t)}{\Delta t}. \quad (3.11)$$

Como resultado, o módulo da velocidade apresenta um erro de ordem $\vec{O}(\Delta t^2)$.(99)

A equação completa para o cálculo da nova posição $\vec{R}_i(t + \Delta t)$ de um vértice v_i a partir de um instante t é então dada por:

$$\vec{R}_i(t + \Delta t) = 2\vec{R}_i(t) - \vec{R}_i(t - \Delta t) + \left(\vec{F}_a(v_i) + \vec{F}_b(v_i) - \mu \frac{\vec{R}_i(t + \Delta t) - \vec{R}_i(t)}{\Delta t} \right) \Delta t^2. \quad (3.12)$$

A Figura 31 sumariza a metodologia FRMOD de visualização de redes complexas.

3.2 Otimização da técnica FRMOD por meio de expansão multipolar

É importante notar que o cálculo das forças atrativas entre todos os vértices do método FRMOD tem complexidade computacional $O(|\mathcal{E}|)$ e o cálculo das forças repulsivas, $O(|\mathcal{V}|^2)$. Portanto a complexidade do algoritmo completo é $O(|\mathcal{V}|^2 + |\mathcal{E}|)$. Em redes esparsas, que englobam grande parte das redes complexas de sistemas reais, como $|\mathcal{E}| \ll |\mathcal{V}|^2$, então $O(|\mathcal{V}|^2)$ é o custo mínimo esperado através de otimizações simples para estas redes.

Com o custo computacional alto das técnicas FR e FRMOD, a visualização de redes com mais de 5000 vértices se torna proibitiva, especialmente quanto a possibilidade do uso

Figura 31 – Metodologia FRMOD de visualização de redes por forças. À esquerda é mostrada a rede na configuração inicial, com vértices aleatoriamente distribuídos sobre o plano. As forças de atração (eq. 3.2) e repulsão (eq. 3.3) são indicadas por vetores. Através de MDS a rede chega a um estado de equilíbrio (à direita) com vértices conectados próximos um dos outros e vértices mantendo a estrutura topológica da rede. A visualização final é gerada associando a cor de cada vértice a alguma propriedade, neste caso, o grau.

Fonte: Elaborada pelo autor.

de visualização interativa nessas redes. Entretanto, diversos métodos foram sugeridos na literatura para reduzir o custo computacional de técnicas baseadas no algoritmo FR.⁽⁹³⁾ Um dos que apresenta melhor performance, é o método introduzido por Yifan Hu.⁽²⁹⁾ Neste algoritmo a rede é progressivamente simplificada em camadas e a energia do sistema é otimizada para cada camada isoladamente, reduzindo o custo computacional do cálculo das forças repulsivas para até $O(|V| \log |V|)$. Entretanto, pela dificuldade em adaptá-lo para o contexto de simulação molecular presente no FRMOD, um problema também observado com outras técnicas de otimizações comumente empregadas ao método FR, optamos por desenvolver um método de otimização baseado nas técnicas já existentes em MDS. Entretanto, foge do escopo deste trabalho o desenvolvimento e validação de um algoritmo de performance superior aos já existentes na literatura, tendo em vista que o objetivo principal é desenvolver um algoritmo que se mostre suficientemente eficiente para a obtenção de visualizações interativas das redes complexas estudadas ao longo deste trabalho.

Uma das técnicas mais simples e mais usadas para se melhorar a performance em MDS é o método de Barnes-Hut.⁽¹⁰⁰⁾ O método baseia-se em particionar o espaço alvo em regiões de modo que cada uma contenha uma parcela pequena das partículas (vértices) do sistema. O particionamento pode ser feito através de diversas técnicas, como pelo do uso de *quadtrees* para espaços 2D, *octrees* no caso 3D, ou ainda de forma mais simples, dividindo o espaço como uma grade regular. Como a força de repulsão entre dois vértices do método FRMOD decai com o quadrado da distância entre eles, a força total entre aqueles que es-

tão muito distantes é pequena. Desta forma, pode-se aproximar cada partição do sistema por uma distribuição de cargas que interagem entre si. Como resultado, as interações de repulsão não mais ocorrem entre todos os pares de vértices, mas sim entre pares de vértices dentro de cada partição do espaço e também entre as próprias partições. Através desta técnica a complexidade computacional é definida então pelo número de partições existentes no sistema, $|\mathcal{P}|$, e pelo número médio de vértices existentes em cada partição, $\langle |p| \rangle$, como $O(|\mathcal{P}|^2 \langle |p| \rangle^2)$. Com um particionamento por *quadtrees* ou *octrees*, por exemplo, é possível implementar um algoritmo de complexidade $O(|\mathcal{V}| \log |\mathcal{V}|)$.(100)

Figura 32 – Particionamento em grade de um espaço 2D exemplificando as interações do método Barnes-Hut em FRMOD. Considerando a partição P_{mn} , inicialmente, todos os vértices das partições vizinhas interagem entre si de acordo com o método FRMOD. As interações de longa distância, no entanto, ocorrem apenas entre pares de partições através de aproximação por expansão multipolar, ilustrada em vermelho.

Fonte: Elaborada pelo autor.

Para simplificar a descrição do método Barnes-Hut aplicado em FRMOD, considere o particionamento em grade mostrado na Figura 32, com cada partição P_{mn} indexada por um par ordenado (m, n) . Inicialmente, determina-se as forças de repulsão locais existentes entre os vértices de cada partição P_{mn} , também incluindo as partições vizinhas $(P_{(m-1)(n-1)}, P_{(m-1)n}, P_{(m-1)(n+1)}, P_{m(n-1)}, P_{m(n+1)}, P_{(m+1)(n-1)}, P_{(m+1)n}, P_{(m+1)(n+1)})$, em destaque na figura; usando a Equação 3.12. Em seguida, é necessário calcular as interações entre cada par de partições, que podem ser obtidas através de uma técnica similar à expansão multipolar do potencial elétrico proveniente de uma distribuição de cargas.(22)

A força de repulsão entre dois vértices no FRMOD se comporta como o campo elétrico existente entre duas partículas carregadas, portanto, para a força repulsiva total, $\vec{F}_{(r)mn}(\vec{R}_i)$, gerada pela partição P_{mn} , centrada em $\vec{R}_{mn}$, atuando em um vértice v_i na posi-

ção $\vec{R}_i$, pode-se definir um potencial $\phi_{mn}(\vec{R}_i)$:

$$\vec{F}_{(r)mn}(\vec{R}_i) = b \sum_{v_j \in P_{mn}} \frac{1}{(\vec{R}_i - \vec{R}_j)^2} \hat{r}_{ji} = -b \vec{\nabla} \sum_{v_j \in P_{mn}} \frac{1}{|\vec{R}_i - \vec{R}_j|} = -b \vec{\nabla} \phi_{mn}(\vec{R}_i). \quad (3.13)$$

Pode-se então expandir $\phi_{mn}(\vec{R}_i)$ em termos de multipolos centrados em $\vec{R}_{mn}$, resultando em:

$$\begin{aligned} \phi_{mn}(\vec{R}_i) = & \frac{|P_{mn}|}{|\vec{R}_i - \vec{R}_{mn}|} + \frac{\vec{p}_{mn} \cdot (\vec{R}_i - \vec{R}_{mn})}{|\vec{R}_i - \vec{R}_{mn}|^3} + \\ & + \frac{1}{2} \sum_k^3 \sum_l^3 \frac{(\mathbf{Q}_{mn})_{kl} (\vec{R}_i - \vec{R}_{mn})_k (\vec{R}_i - \vec{R}_{mn})_l}{|\vec{R}_i - \vec{R}_{mn}|^5} + \\ & + O\left(\frac{1}{|\vec{R}_i - \vec{R}_{mn}|^7}\right), \end{aligned} \quad (3.14)$$

onde $|P_{mn}|$, $\vec{p}_{mn}$ e $\mathbf{Q}_{mn}$ correspondem respectivamente aos momentos monopolo, vetor dipolo e tensor quadrupolo da partição P_{mn} .

O termo quadrupolo da Equação 3.14 pouco contribui para o cálculo da força resultante $\vec{F}_{(r)mn}(\vec{R}_i)$, portanto, é possível aproximar a interação entre partições apenas pelos termos de monopolo e dipolo. Substituindo a Equação 3.14 em 3.13 e eliminando o termo de quadrupolo, obtém-se:

$$\vec{F}_{(r)mn}(\vec{R}_i) \approx b \left(\frac{|P_{mn}|}{(\vec{R}_i - \vec{R}_{mn})^2} \hat{r}_{mn(i)} + 3 \frac{(\vec{p}_{mn} \cdot \hat{r}_{mn(i)}) \hat{r}_{mn(i)} - \vec{p}_{mn}}{|\vec{R}_i - \vec{R}_{mn}|^3} \right), \quad (3.15)$$

onde $\hat{r}_{mn(i)}$ é o versor na direção $(\vec{R}_i - \vec{R}_{mn})$, e o vetor de dipolo $\vec{p}_{mn}$ pode ser escrito em função das posições dos vértices em P_{mn} relativas ao centro:

$$\vec{p}_{mn} = \sum_{v_j \in P_{mn}} (\vec{R}_j - \vec{R}_{mn}). \quad (3.16)$$

Verificamos qualitativamente que as modificações propostas nesta seção para a técnica FRMOD contribuíram para um considerável ganho de performance na visualização interativa de redes. Apesar da possível dependência quanto ao método de particionamento do espaço escolhido, tanto o uso de particionamento através de *quadtrees/octrees* e grade regular resultaram em melhorias sem perda visível de qualidade das visualizações. Apesar disso, as modificações sugeridas não podem ser adaptadas diretamente para o cálculo das interações do tipo atrativas, que continuam a apresentar complexidade de $O(\mathcal{E})$, tornando proibitiva a visualização de redes não esparsas. Em estudos futuros, deverão ser investigadas outras técnicas que podem ser usadas para melhorar o desempenho do cálculo das forças atrativas, por exemplo, através do uso de Sammon Mapping (101) ou técnicas que utilizem pontos de controle ou *pivots*.(102, 103)

3.3 Desenvolvimento de software para visualização de redes complexas

Apesar da existência de ótimos softwares dedicados à visualização de redes complexas (104, 105, 106), a flexibilidade de se ter uma ferramenta para esta finalidade desenvolvida pelo próprio pesquisador pode ajudar a integrá-la e adaptá-la aos seus próprios interesses. Por este motivo, durante o projeto de mestrado do autor (35) foi desenvolvido um *software* de visualização interativa de redes complexas, o *Network 3D*. Ao longo deste trabalho, além de pequenas alterações e correções, também foi incorporada ao *software* a técnica FRMOD (descrita na seção anterior), a geração de visualizações interativas de redes com dezenas de milhares de vértices. Foi desenvolvido também uma versão mais simples do visualizador de redes, voltada para o uso através da *web*. Para isso, foram usadas tecnologias recentes como *HTML5* e *WebGL*, permitindo que pesquisadores possam explorar redes grandes de maneira interativa mesmo que não estejam em suas estações preferenciais de trabalho, necessitando apenas de um navegador compatível.

Figura 33 – Esquema do fluxo de dados do software Network 3D, ilustrando as principais funcionalidades e módulos do software desenvolvido.

Fonte: Elaborada pelo autor.

A Figura 33 apresenta um esquema simplificado do fluxo de dados no *software* desenvolvido. O processo de visualização é iniciado ao abrir um novo arquivo a partir do disco (a) formatado como “.net” (104) ou “.xnet”. Ambos os formatos de arquivos podem conter não somente a estrutura da rede, como vértices, arestas e pesos, como também propriedades (c),

incluindo cores, vetores de posição, valores escalares (como medidas obtidas para vértices), classes (por exemplo, comunidades), etc. O *software* apresenta ao usuário um painel com uma projeção 3D da rede (d). Inicialmente, se o arquivo formatado contém informações de posições, estas são usadas para a apresentação da rede, caso contrário, os vértices são distribuídos aleatoriamente ao longo do espaço 3D. Também é possível limitar o posicionamento de vértices a um plano, permitindo também que o *software* seja usado para visualização 2D de redes.

Figura 34 – Captura de tela do *software* Network 3D versão para computadores.

Fonte: Elaborada pelo autor.

Após a inicialização da rede, o usuário do *software* (e) pode interagir de diversas formas com o painel de visualização, podendo alterar o mapa de cores usado para apresentar propriedades escalares e classes de vértices (f). Outro tipo de interação permitida é a filtragem de vértices e arestas através de propriedades (g). O usuário pode usar uma propriedade para definir um intervalo onde vértices serão ou não mostrados, similarmente, também é possível escolher um intervalo de corte (*threshold*) referente ao peso de arestas. O usuário também pode interagir com o painel através do movimento de câmera (h), que en-

Figura 35 – Captura de tela do software Network 3D versão *web*.

Fonte: Elaborada pelo autor.

volve rotação, movimento panorâmico e zoom, assim como alterando propriedades gerais de visualização da rede (cor de fundo, largura das arestas, etc.). O software permite também que o usuário ative o algoritmo de disposição de vértices FRMOD (i), resultando em uma animação em tempo real da disposição de vértices (j). Cada parâmetro usado na simulação MDS também pode ser alterado em tempo de execução (k), com retorno imediato para o usuário. Por fim, o usuário pode exportar arquivos de imagens (l) ou visualizações interativas para a *web* (m), que pode ser aberta em qualquer navegador moderno. As mudanças realizadas na rede também podem ser salvas de volta para os formatos originais (b). Além da interface padrão para o usuário, também é possível automatizar tarefas através do uso de roteiros escritos em linguagem *Python*, através de um terminal embutido no software.

As Figuras 34 e 35 mostram respectivamente a captura de tela obtida a partir da versão *desktop* e da versão *web* do software desenvolvido. Na Figura 36 são mostradas diversas visualizações de redes, algumas delas, em particular, as redes com mais de 10 mil vértices, só puderam ser estudadas após a implementação das mudanças no algoritmo FRMOD. Grande parte das figuras presentes neste trabalho envolvendo visualizações de redes foram geradas pelo software.

Figura 36 – Exemplos de visualizações de redes complexas geradas no Network 3D. Rede gerada pelo modelo WS bidimensional (a), a cor indica o grau de cada vértice. Estrutura da Internet (b) (107), com a cor também indicando o grau de cada vértice. Rede de correlações do mercado financeiro para 1989 (108), as cores indicam as comunidades obtidas para a rede. Rede de artigos da Wikipédia (d) em subcategorias pertencentes às grandes áreas do conhecimento, as cores indicam categoria associada a cada artigo

Fonte: Elaborada pelo autor.

3.4 Visualização e caracterização de redes de expressão gênica

Sistemas complexos, em geral, apresentam vínculos entre seus elementos, o que impede que suas propriedades variem individualmente. A similaridade entre propriedades de elementos diferentes pode indicar um grau de dependência ou proximidade entre eles. Desta forma, é possível que as propriedades dos elementos do sistema sejam inferidas considerando a adjacência entre eles, e vice-versa. Como resultado, também é possível obter uma rede complexa que represente estes vínculos ou a similaridade entre as propriedades dos elementos. Redes geradas com esta finalidade, chamadas de *redes de similaridade*, têm sido usadas com sucesso em diversos problemas reais.(9, 109, 110, 111) Nesta abordagem, supondo um sistema composto por N elementos caracterizados por M propriedades $V = v_i : i \in U, v_i \in \mathbb{R}^M$, uma rede composta destes elementos, totalmente conectada e com pesos, pode ser construída tomando como valores para os pesos uma medida de similaridade $S : V^2 \rightarrow \mathbb{R}$, que é obtida para todos os pares de propriedades dos elementos $(u_i, u_j) : u_i \in U, u_j \in U$. Como a rede completa apresenta pouca informação sobre a estrutura topológica do sistema, normalmente, uma nova rede é gerada podando-se conexões de baixa similaridade, por exemplo através de um limiar.(9, 1, 112) Neste trabalho, o limiar foi escolhido manualmente buscando valores que apresentassem estabilidade de propriedades da rede, como distribuição de grau, e através de análise visual (através de visualização). Entretanto, na literatura, já é possível encontrar metodologias mais robustas para a definição do limiar, entre eles destaca-se ambiente WGCNA (112) que apresenta um método para eliminar arestas pouco significativas de modo a obter uma rede livre de escala. Esta e outras técnicas para podar, ou definir um limiar de correlação, deverão ser estudadas e comparadas em trabalhos futuros.

Figura 37 – Ilustração da abordagem de análise de dados de expressão gênica através de redes de similaridade.

Fonte: Elaborada pelo autor.

A análise de redes de expressão gênica (1) é uma nova forma de se obter, indiretamente, informações sobre mecanismos genéticos através de dados de *microarrays* de DNA. Estas redes são, em geral, obtidas por meio da similaridade dos padrões de expressão gênica provenientes de tecidos de diversos indivíduos selecionados de modo a cobrirem um fenômeno biológico em particular, como uma doença. Desta forma, estes sistemas podem ser representados através de redes de similaridade. Em particular, nessas redes, emprega-se o coeficiente de correlação de Pearson (73) como medida de similaridade entre genes. A Figura 37 exemplifica o método para geração de redes de similaridade a partir de dados de expressão gênica. O processo inicia-se com a construção de um vetor de características para cada gene composto pelas expressões obtidas de diversos *microarrays* (em geral uma para cada indivíduo). No segundo passo é computada a correlação de Pearson entre os vetores de características de todos os pares de genes. A rede é construída conectando genes cuja correlação de Pearson é maior que um limiar C_{Limiar} , fornecido como parâmetro. O passo 3 compreende a análise dessa rede, que pela nossa abordagem é baseada na classificação de vértices em Vips ou Hubs, e em visualização.

A determinação do limiar para a geração de redes de similaridade a partir de correlações ainda é um problema em aberto na área. Apesar disso, novas técnicas têm sido desenvolvidas como a *Weighted correlation network analysis* (WGCNA) (112), onde busca-se manter conexões que resultem em uma rede livre de escala. Esta abordagem, apesar de promissora, pode embutir um viés às redes que naturalmente não são livres de escala, ou que apresentem estruturas híbridas. Desta forma, para este trabalho, por não conhecermos *a priori* a estrutura das redes estudadas, optamos por estimar o limiar inicial, ϕ_{inicial} pela significância estatística, considerando um $p\text{-value} < 0.01$.(171) Considerando apenas valores de threshold $\phi > \phi_{\text{inicial}}$, obtivemos buscamos, através de incrementos pequenos de ϕ a região onde a rede apresentava ao menos 90% dos vértices ainda conectados ao componente principal e que também se mantivesse qualitativamente estável através de visualização das redes, isto é, encontrando valores de ϕ onde, nas proximidades, a rede resultante varie pouco. Para esta tarefa usamos o software de visualização interativa de redes, que possibilitou que alterações manuais dos valores de ϕ gerassem visualizações da rede correspondente em tempo real. Em trabalhos mais recentes, também buscamos por valores de ϕ que correspondem a regiões onde outras propriedades das redes são estáveis, como por exemplo, a modularidade.(172, 173, 174)

Em colaboração com pesquisadores do grupo do professor Carlos A. Moreira da Faculdade de Medicina da Universidade de São Paulo, foram analisados dados de transcriptoma humano provenientes de pacientes portadores de epilepsia.(9) Esta doença atinge mais de 50 milhões de pessoas no mundo¹. Outro agravante é o fato de que aproximadamente 30% dos pacientes apresentam quadros crônicos que não respondem ao tratamento por medicamentos antiepilépticos disponíveis atualmente.(113)

¹ Disponível em: <http://www.folhaweib.com.br/?id_folha=2-1--669-20131007> Acesso em: 20 jan. 2015.

Figura 38 – Exemplos das três categorias de vértices estudados em (9): Hubs seguem a definição usual, sendo aqueles com maior número de conexões na rede; Vips, por outro lado, são vértices que apresentam poucas conexões, mas em geral são conectados a vértices de alto grau k_i , desta forma apresentam alto grau concêntrico de primeiro nível ($k_i(1)$); High-Hubs são vértices que apresentam as duas características, isto é, além de apresentarem muitas conexões também apresentam alto grau de primeiro nível.

Fonte: Elaborada pelo autor.

A base de dados é constituída de tabelas de transcriptoma humano obtidas por meio de *microarrays* de DNA considerando tecidos de três fontes. Um conjunto de dados engloba tecidos obtidos da região CA3 do hipocampo de indivíduos portadores de epilepsia de diferentes categorias, como doença de característica febril ou não-febril além de classes obtidas por histologia e histopatologia.(34, 9)

A análise baseia-se inicialmente no uso de visualização interativa de redes. A caracterização de vértices vem em seguida e é direcionada segundo os indicativos obtidos através das visualizações. Em (9), ao explorar as redes através de visualização, verificou-se que além de Hubs, outra categoria de vértices se destacava: os Vips. A Figura 38 ilustra o conceito de Vips e Hubs.

No trabalho conduzido (9), a visualização de redes combinada com o uso das medi-

Figura 39 – Exemplo de visualização de redes e classificação de acordo com as categorias VIPS e HUBs obtida a partir dos dados analisados de expressão gênica

Fonte: Elaborada pelo autor.

das concêntricas revelou a existência de vias metabólicas responsáveis por efeitos compensatórios na região CA3 do hipocampo, fato antes apenas sugerido na literatura especializada. (114) A Figura 39 exemplifica uma das redes geradas a partir da metodologia proposta (Figura 39a) para dados obtidos da região CA3 de indivíduos portadores de epilepsia. A classificação no espaço Vips e Hubs é apresentada ao lado (Figura 39b). Além disso, diversas classes de genes participando em diferentes vias metabólicas foram revelados pela análise, como mecanismos de diferenciação neuronal e transmissão sináptica. Avanços na compreensão destes sistemas são importantes para o entendimento dos mecanismos da doença, assim como a identificação de vias metabólicas alteradas, podendo futuramente ajudar na formulação de novos medicamentos ou técnicas de tratamento.

Também foram desenvolvidas diversas ferramentas computacionais associadas ao visualizador, dedicadas à análise de redes de expressão gênica, estas foram disponibilizadas junto ao grupo de pesquisa do professor Carlos A. Moreira, para que seus membros também pudessem conduzir as análises com maior independência. Futuramente, estas ferramentas deverão também ser disponibilizadas para a comunidade científica através do site: <<http://cyvision.ifsc.usp.br/>> (Acesso em 10.07.2015).

4 Mapas científicos e caracterização de vértices em comunidades

Seja através de propriedades concêntricas ou de visualização interativa, a análise multiescala de redes complexas permite caracterizar a participação de vértices ao longo de suas vizinhanças, assim como no contexto global da rede. Muitos sistemas, no entanto, são organizados em subestruturas ou hierarquias (62) que por sua vez formam módulos, que no contexto de redes são usualmente chamados de comunidades.(28, 62) Desta forma, uma nova escala de interesse (além da local e da global) emerge naturalmente em sistemas com subestruturas. Torna-se, então, desejável determinar não somente a participação de vértices nas escalas usuais da rede, como também, sua participação na escala destas subestruturas.

O uso inteligente de mapeamento de propriedades em visualização de redes complexas, por exemplo, através das cores, pode ser útil para, qualitativamente, caracterizar tanto a relação de vértices em comunidades quanto as relações entre as próprias comunidades. Quando estas redes envolvem dados cobrindo tópicos de conhecimento científico, suas visualizações são chamadas de mapas científicos.(115, 116, 117)

O uso de mapas científicos tem grande importância na compreensão da estrutura geral das relações intrincadas existentes entre os tópicos e subtópicos de uma área científica. Além disso, técnicas desta natureza podem ser tratadas como uma forma rudimentar de sumarização. Neste contexto também é importante identificar a participação de cada vértice em relação aos tópicos gerais, representados pelas comunidades. Por exemplo, vértices situados nas interfaces entre diferentes assuntos científicos tendem a refletir conceitos mais interdisciplinares, enquanto vértices localizados nas regiões mais centrais, em geral compõem as bases ou núcleo daquela área científica.

A caracterização de vértices em comunidades e em mapas científicos é também um problema de análise multiescala, já que é necessário contextualizar as características de um vértice no nível de escala das comunidades da rede. Neste capítulo propomos uma metodologia multiescala para caracterizar vértices de acordo com sua participação em comunidades. esta medida é baseada no conceito de diversidade e apresenta similaridades com a medida de acessibilidade. Com pequenas alterações na metodologia, foi possível desenvolver uma nova medida para quantificar a interdisciplinaridade de revistas científicas em uma rede de citações (118, 119) ao longo do tempo.(10) Similarmente, propomos também um método para caracterizar as comunidades de uma rede de citações baseada nas informações contextuais dos vértices. Esta última análise permite criar taxonomias informativas sobre uma dada área científica.

4.1 Uso da entropia para quantificar o papel de vértices em comunidades

A quantificação da participação de vértices, tanto em comunidades, determinadas pela própria estrutura da rede, quanto em categorias, conhecidas *a priori*, foi pouco explorada na literatura. Guimera e Amaral (120) propuseram que o papel de um vértice em sua comunidade pode ser determinado por sua taxa de participação P_i que é definida em termos da razão do grau interno (considerando apenas conexões dentro da mesma comunidade) pelo grau externo (que considera apenas conexões com outras comunidades). Entretanto, apesar de fornecer um forte indicativo de como os vértices estão descritos, a medida é de difícil interpretação, além de depender explicitamente do grau dos vértices, tornando difícil a comparação entre vértices de comunidades com diferentes padrões de conectividade.

Podemos entender que, em uma rede modular, podemos encontrar basicamente três tipos de vértices: vértices centrais dentro de suas respectivas comunidades, vértices que estabelecem pontes entre comunidades diferentes e vértices localizados nas regiões mais afastadas das redes, que também correspondem às bordas externas das comunidades. Já sabemos que este último grupo de vértices pode ser caracterizado pela acessibilidade (ver Seção 2.8.3). Para identificar os outros dois grupos, sugerimos uma abordagem semelhante, baseada na quantificação da diversidade de conexões através da entropia, entretanto, considerando as categorias ou comunidades associadas aos vértices.

No problema geral, vértices podem apresentar mais do que uma categoria (ou comunidade), neste caso, para cada vértice v_i , define-se um valor de *pertinência* $P_{i,c}$ a cada categoria c existente em na rede. No caso específico de vértices pertencendo a apenas uma categoria, considera-se que $P_{i,c} = 1$ se v_i pertence à categoria c e $P_{i,c} = 0$ caso contrário.

A partir do conjunto de valores referentes às pertinências de vértices em categorias, podemos definir uma medida local que quantifique objetivamente a diversidade das categorias com que um vértice v_i se conecta, a *entropia de pertinência* H_i . A Figura 40 ilustra a ideia geral da medida. Considerando esta rede organizada em N_C categorias (ou comunidades) com cada vértice v_i apresentando um grau de pertinência a uma comunidade c denotado por $P_{i,c}$, define-se a entropia de pertinência H_i^* como a entropia da distribuição de comunidades considerando o vértice v_i e seus vizinhos imediatos, $(P_{i,c}^*)$, isto é

$$P_{i,c}^* = \frac{1}{k_i + 1} \left(P_{i,c} + \sum_{(v_i, v_j) \in \mathcal{E}} P_{j,c} \right) \quad (4.1)$$

$$H_i^* = - \sum_c^{N_C} P_{i,c}^* \ln(P_{i,c}^*). \quad (4.2)$$

Em redes com pesos, os termos $P_{j,c}$ podem ser ponderados pelo peso w_{ij} das cone-

Figura 40 – Exemplo da entropia de pertinência para duas configurações de redes. A rede no topo da figura (a) apresenta maior diversidade de conexão com vizinhos de diferentes comunidades, portanto, apresentando maior valor de entropia de pertinência. Já no segundo caso (b), a maioria dos vizinhos pertencem a uma única comunidade (azul), apresentando baixo valor para a medida.

Fonte: Elaborada pelo autor.

xões, de modo que $P_{i,c}^*$ pode ser redefinido como:

$$P_{i,c}^{\text{weighted}} = \frac{1}{k_i + 1} \left(P_{i,c} + \frac{1}{s_i} \sum_{(v_i, v_j) \in \mathcal{E}} w_{ij} P_{j,c} \right), \quad (4.3)$$

onde s_i é a força do vértice i , isto é, $s_i = \sum_{(v_i, v_j) \in \mathcal{E}} w_{ij}$.

A Figura 41 mostra a entropia de pertinência aplicada em uma rede de distribuição de energia dos EUA (ver Seção 5.1 para mais informações sobre a rede). As comunidades encontradas foram obtidas através do método CONGA (121) que não somente detecta as comunidades, como também atribui pertinências.

Também é possível redefinir a medida da entropia de pertinência de modo multiescala, permitindo a caracterização de vértices através do conceito de níveis concêntricos, ao invés de simplesmente considerar apenas a vizinhança imediata dos vértices. Desta forma,

Figura 41 – Visualização bidimensional da rede de distribuição de energia dos EUA com cores indicando as comunidades de maior pertinência para cada vértice. A entropia de pertinência H_i^* é mostrada através do tamanho dos vértices de acordo com a legenda.

Fonte: Elaborada pelo autor.

define-se a *pertinência concêntrica*, $P_{i,c}(r)$ e a entropia de pertinência concêntrica $H_i(r)$ como:

$$P_{i,c}(r) = \frac{1}{V_i(r)} \left(\sum_{h=0}^r \sum_{j \in R_i(h)} P_{j,c} \right) \quad (4.4)$$

$$H_i(r) = - \sum_c^{N_C} P_{i,c}(r) \ln(P_{i,c}(r)), \quad (4.5)$$

onde $V_i(r)$ e $R_i(r)$ correspondem respectivamente ao volume e ao anel concêntricos de r níveis centrados em v_i . Nota-se que a entropia de pertinência definida anteriormente é um caso especial da pertinência concêntrica: $H_i^* = H_i(1)$.

4.2 Diversidade temática e quantificação da interdisciplinaridade

Uma nova perspectiva da ciência tem sido adotada por diversos pesquisadores ao redor do mundo. Com problemas cada vez mais complexos e intrincados, a filosofia reducionista vem perdendo espaço no campo científico, sendo complementada por análises mais completas e robustas que, em geral, só podem ser conduzidas envolvendo pesquisadores de diversas áreas científicas. Desta forma há uma percepção geral de que as ciências têm se tornado mais interdisciplinares ao longo das últimas décadas. Recentemente, muitos estudos

também vêm sendo conduzidos para se entender a evolução e a estrutura das ciências, uma área de pesquisa que compõe o que é hoje conhecido como *e-Science*.^(122, 123, 124) Apesar disso, estudos que exploram modelos e a caracterização da própria interdisciplinaridade nas ciências ainda são escassos na literatura.

A maior dificuldade em se quantificar uma qualidade como a interdisciplinaridade está no fato deste ser um fenômeno subjetivo. Entretanto, através de algumas inferências e análise de dados é possível identificar alguns padrões que podem permitir a sua caracterização. Como discutido anteriormente, espera-se que conceitos mais interdisciplinares se relacionem com outros de áreas de pesquisa diferentes. Assim, uma definição intuitiva fortemente relacionada ao conceito de interdisciplinaridade é a diversidade.

Considerando a estrutura do conhecimento como uma rede complexa, é então possível quantificar a interdisciplinaridade através da entropia de pertinência. Para ilustrar a capacidade do método, foi usada a rede de citações entre revistas científicas indexadas no *Web of Science* (WoS)¹ que também pode ser entendida como uma forma de representação do conhecimento.

Através de mineração de dados foi obtido todo o conjunto de 7387 revistas científicas das áreas de exatas e biológicas indexadas no *Journal Citation Reports*(JCR)², incluindo informações do número de citações entre os artigos de cada revista, o fator de impacto, assim como a sua classificação temática (ou *Subject Category*). Os dados foram obtidos para um período de 11 anos, cobrindo o período de 1999 a 2009. Para cada ciclo anual foi gerada uma rede direcionada com pesos, de modo que vértices representam periódicos e o peso de cada aresta o número de citações totais (considerando todos os artigos) entre cada par de revistas. Ao longo do tempo, a rede foi construída de modo incremental, ou seja, a rede referente a uma época T contém informação de todos os instantes anteriores $T - 1, T - 2, \dots 0$. Por exemplo, a rede correspondente ao ano de 2005 contém a rede de 2004, tal como a de 2003 e assim por diante. Diferentemente das análises normalmente empregadas em dados dessa natureza, onde são usados apenas dados referentes a um subtópico ou amostras da rede, aqui, a rede é usada de forma integral.

A Figura 42 ilustra a estrutura geral das redes obtidas. É importante notar que como um artigo pode citar outro dentro de uma mesma revista, também são permitidas autoconexões nestas redes. Entretanto, foi analisado o efeito da presença ou não de autocitações (dentro da mesma revista). As medidas não se alteraram com ou sem a presença dessas conexões, assim, conexões deste tipo foram mantidas.

O método para quantificar a interdisciplinaridade de vértices está diretamente relacionado à diversidade com que revistas se conectam a diferentes categorias temáticas.

¹ Disponível em: <<http://webofknowledge.com>> Acesso em: 09 abril 2015.

² Disponível em: <http://thomsonreuters.com/products_services/science_products/a-z/journal_citation_reports/> Acesso em: 09 abril 2015.

Figura 42 – Representação esquemática da rede de revistas científicas. As revistas são conectadas por arestas com pesos considerando citações de entrada e saída. No exemplo, *A* está conectada às revistas *B*, *C* e *D* e também possui uma autoconexão já que alguns de seus artigos contém referências internas. Um conjunto de categorias temáticas também é associado a cada revista).

Fonte: Elaborada pelo autor.

Portanto, é esperado que nesta rede, os vértices interdisciplinares estejam localizados nas fronteiras das regiões com diferentes conjuntos de categorias temáticas. Essa ideia equivale a de determinar a participação dos vértices nas comunidades através da entropia de pertinência.

Como a quantidade de vértices pertencentes a cada categoria temática pode variar substancialmente, é necessário normalizar a pertinência pelo tamanho total de cada categoria, N_c . A pertinência temática, $P_{i,c}^{\text{subject}}$, de um vértice v_i apresentando um conjunto de categorias C_i é então definida como:

$$P_{i,c}^{\text{subject}} = \begin{cases} \frac{N_c}{N} & \text{se } c \in C_i \\ 0 & \text{caso contrário} \end{cases} \quad (4.6)$$

No contexto específico deste problema, a entropia de pertinência é chamada de *entropia temática* quando calculada usando a Equação 4.3 sobre as pertinência temáticas dos vértices. Esta medida é a que caracteriza a interdisciplinaridade das revistas científicas e também das áreas gerais do conhecimento.

A entropia temática foi obtida para todas as revistas da JCR dentro dos intervalos de tempo considerados. Outras medidas de redes complexas como o mínimo caminho médio, força média e tamanho do maior componente também foram obtidas para a rede. A Figura 43 apresenta as curvas de evolução dessas propriedades ao longo do tempo para as diferentes áreas do conhecimento. Como esperado o tamanho do maior componente conectado cresceu com o aumento do número de revistas, entretanto é interessante notar que a curva apresenta uma lei linear de crescimento (com exceção de 2009), indicando que neste

período houve em média 90 novas revistas por ano.

O número de citações média de cada revista aumentou consideravelmente nos últimos anos, como mostrado na Figura 43b. Este é o caso de química, biologia e física. A evolução global das citações da rede segue um padrão muito similar ao de medicina, muito provavelmente porque corresponde a 34% de todos os vértices da rede. Entretanto, curiosamente, medicina não é uma das áreas mais citadas, indicando que o número de revistas em uma área não necessariamente indica maior quantidade de citações.

Tabela 2 – Correlação entre as diversas medidas obtidas para as revistas científicas

	Força de entrada	Centralidade de intermediação	Entropia temática	Fator de impacto	Número de artigos
Força de entrada	-	0.67	0.92	0.38	0.65
Centralidade de intermediação	0.67	-	0.69	0.18	0.43
Entropia temática	0.92	0.69	-	0.42	0.62
Fator de impacto	0.38	0.18	0.42	-	0.13
Número de artigos	0.65	0.43	0.62	0.13	-

Fonte: Elaborada pelo autor

A Figura 43c mostra efetivamente que as áreas gerais das ciências têm se tornado cada vez mais interdisciplinares, confirmando então as perspectivas subjetivas mencionadas anteriormente. Outro resultado importante, mostrado na tabela 2, é que, dentre as medidas consideradas, a entropia temática foi a que apresentou maior correlação com o fator de impacto, apesar de moderada. Isto indica que muitas revistas de excelência também apresentam perfil interdisciplinar.

O mapa científico da rede correspondente ao ano de 2009 é mostrado na Figura 44. Essa projeção fornece informações sobre a proximidade topológica entre as revistas na rede. Medicina aparece bem conectada entre si e com biologia, assim como ciências veterinárias. Contrariamente, matemática pura aparece isolada, conectada ao maior componente conectado através de engenharia e física. A biologia funciona como uma ponte entre medicina e a maioria das áreas exatas. Na mesma figura também é mostrada a classificação das revistas com maiores índices de interdisciplinaridade (entropia temática).

Uma projeção 3D da rede também foi obtida usando o software de visualização desenvolvido neste trabalho, uma animação desta rede pode ser vista no website do grupo de pesquisa³.

Outro resultado interessante aparece ao observar as projeções das revistas *Nature* e *Science*, que compartilham uma mesma posição tanto no mapa 2D (Figura 44) quanto no 3D.

³ Disponível em: <<http://cyvision.ifsc.usp.br/Cyvision/?page=INTERDISCIPLINARITY>>, Acesso em: 09 abril 2015.

Isso indica que os padrões de citação são muito similares, além disso, são as únicas revistas que apresentam esta característica na projeção 3D.

Figura 43 – Evolução da média e mediana de diversas propriedades das redes com o tempo. (a) O tamanho do maior componente da rede de citação é mostrado pela curva vermelha, revelando uma relação praticamente linear com o tempo. O mínimo caminho médio da rede é mostrado em azul, apresentando uma curva monotonicamente decrescente com o tempo. (b), (c) respectivamente apresentam a força média de entrada e a mediana da entropia temática variando no tempo para as diferentes grandes áreas das ciências. Os valores globais, isto é, considerando todos os vértices das redes é destacado pela linha cinza em destaque.

Fonte: Elaborada pelo autor.

Figura 44 – A imagem principal mostra um mapa científico bidimensional baseado na rede para o ano de 2009 com as cores indicando as áreas do conhecimento dentro dos campos de exatas e biológicas. A tabela no topo destaca as revistas com maior entropia temática, portanto maior interdisciplinaridade. No quadro inferior à direita é mostrada a distribuição da entropia temática para diversos períodos, resultando em uma lei de potência. O coeficiente desta lei tem diminuído com o tempo (como pode ser observado no quadro interno), indicando também que a própria entropia temática tem se tornado mais diversa ao longo do tempo.

Fonte: Elaborada pelo autor.

4.3 Taxonomia e sumarização de áreas científicas

O trabalho de explorar e sumarizar a produção bibliográfica de uma área científica é normalmente delegada a especialistas e pesquisadores familiarizados com técnicas tradicionais para organizar e extrair informações relevantes da literatura relacionada. Entretanto, a tarefa de sumarização bibliográfica tem se tornado cada vez mais difícil e demorada. Uma das razões principais é a limitação inerente da capacidade humana em lidar com grandes quantidades de informação de forma integrada, como é o caso da crescente quantidade de novos conteúdos científicos que vem sendo disponibilizados ao longo dos últimos anos.⁽¹⁰⁾ Adicionalmente, o efeito causado, mesmo que não intencional, pela polarização das preferências do pesquisador durante o processo de revisão, pode prejudicar a visibilidade de algumas descobertas científicas potencialmente importantes. Uma forma de atenuar estes problemas, é através do uso de ferramentas computacionais e métodos sistemáticos para organização e sumarização de dados desta natureza.

A sumarização de textos é uma área tradicional de processamento de língua natural (PLN).⁽¹²⁵⁾ Métodos desta área baseiam-se no uso de técnicas de aprendizado de máquina supervisionado, onde métricas são usadas para quantificar a saliência de uma sentença. As medidas usadas geralmente expressam pouco conteúdo contextual, por exemplo, o número médio de palavras em uma sentença, número de palavras iniciadas por letra maiúscula, frequência de palavras-chave (previamente obtidas), etc.⁽¹²⁶⁾ Apesar de ter se mostrado eficaz ao considerar apenas as sentenças em textos usuais, muitas bases de dados científicos não disponibilizam acesso fácil e padronizado a trabalhos de forma integral. Normalmente, estas bases cobrem dados referentes a citações, resumo e dados gerais como nome dos autores, instituição de origem, etc. Uma forma natural de considerar todas estas características disponíveis é através do uso de redes complexas.

Recentemente, redes complexas têm sido empregadas em PLN com sucesso na caracterização de textos, apresentando vantagens por ser independente da língua e principalmente do sentido das palavras no texto. Essa nova metodologia se mostrou importante na caracterização de textos de conteúdo desconhecido ⁽¹²⁷⁾ por exemplo. No contexto do problema de sumarização, redes complexas construídas a partir de informações usualmente disponíveis em bases públicas, como redes de citações, podem ser usadas lado a lado com técnicas de PLN que permitam também a análise de conteúdo.

Para ilustrar a abordagem sugerida desenvolvemos um novo método preliminar ⁽¹¹⁷⁾ que permite gerar taxonomias a partir de dados envolvendo resumos e citações de artigos científicos. A técnica se baseia em utilizar a estrutura de comunidades, usualmente presente nessas redes, juntamente com estatísticas simples, oriundas de PLN, para enfim obter um conjunto de palavras-chave relevantes de cada comunidade, assim como da própria base por completo. Em seguida essas palavras-chave são usadas para enriquecer a visualização

de mapas científicos, permitindo nomear as comunidades de acordo com palavras chave, além de permitir a construção de uma árvore de relações entre as palavras. Para esta análise adotamos dois conjuntos de dados referentes a artigos científicos da base do *Web of Science* para áreas: “Redes Complexas” (CN) e “Cristais Fotônicos” (PC). Para cada área considerada, uma rede de citações foi obtida, assim como o conteúdo parcial de cada artigo na forma de resumos (*abstracts*).

4.3.1 Análise de redes de citações obtidas a partir do WoS

As redes usadas para esta análise foram extraídas através de mineração direta do Web of Science (WoS)⁴. Os termos usados para a busca no banco de dados foram “*Complex Networks*” (para a rede CN) e “*Photonic Crystals*” (para a rede PC). A rede resultante CN abrange 11063 artigos com $\langle k_{out}^{CN} \rangle \approx 8.5$, enquanto a rede PC apresentou 20230 artigos e $\langle k_{out}^{PC} \rangle \approx 6.6$. Foram considerados artigos desde 1991 a 2014.

As comunidades das redes foram obtidas usando o método multinível (64), introduzido por Blondel e colaboradores. Este método foi escolhido por resultar em valores de modularidade comparáveis a outros métodos, entretanto com tempo computacional reduzido. A rede CN apresentou 22 comunidades e modularidade $q_{CN} \approx 0.53$. A rede PC apresentou 20 comunidades e modularidade $q_{PC} \approx 0.63$.

4.3.2 Extração de palavras-chave em comunidades e construção de taxonomias

Para identificar as palavras-chave referentes às comunidades, permitindo que um rótulo contextual seja atribuído a cada uma, consideramos os resumos dos artigos, que em geral compõem um *corpus* de algumas poucas centenas de palavras. Devido ao seu pequeno tamanho, métodos tradicionais para a obtenção de palavras-chave PLN não podem ser aplicados diretamente aos resumos.(128, 129, 130) A solução encontrada foi adaptar estatísticas simples presentes na área de PLN, em particular, a contagem de unigramas e bigramas. Também foram removidas dos resumos, palavras funcionais, que apresentam pouco sentido semântico, e pontuação. O texto de cada resumo também passou pelo processo de lematização, onde termos são reduzidos a formas canônicas, removendo-se inflexões como gênero, número ou tempo de verbos.(131, 128). A tabela 3 mostra os estágios da metodologia aplicada a um exemplo de texto⁵.

⁴ Disponível em: <<http://thomsonreuters.com/thomson-reuters-web-of-science/>> Acesso em: 09 abril 2015

⁵ Texto retirado da Wikipédia: <https://pt.wikipedia.org/wiki/Ciência_das_redes> Acesso em: 20 nov 2015.

Tabela 3 – Exemplo dos procedimentos de PLN utilizados para a obtenção dos unigramas e bigramas a partir dos resumos.

Texto original	Remoção de palavras funcionais	Lematização
“O estudo das redes surgiu em diversas disciplinas como um meio de análise de dados relacionais complexos. O livro mais antigo conhecido neste campo é o famoso Sete pontes de Königsberg escrito por Leonhard Euler em 1736.”	estudo redes surgiu diversas disciplinas meio análise dados relacionais complexos livro mais antigo conhecido campo famoso Sete pontes Königsberg escrito Leonhard Euler 1736	estudar, rede, surgir, diverso, disciplina, meio, análise, dado, relacional, complexo, livro, mais, antigo, conhecer, campo, famoso, 7, ponte, königsberg, escrever, leonhard, euler, 1736

Fonte: Elaborada pelo autor

Entende-se que termos de alta relevância, ou palavras-chave, de um subconjunto de textos tendem a estar presentes de forma substancialmente maior dentro do próprio subconjunto do que no conjunto que o engloba. Com a finalidade de quantificar este conceito de forma simples em conjuntos de resumos e da respectiva rede de citações, é proposto, um *índice de importância* $I(w)$ que pode ser atribuído para cada unigrama ou bigrama w existente em algum dos resumos. Este índice é definido pela relação da frequência relativa $F_\alpha^{in}(w)$ de w aparecendo dentro de uma comunidade α , com a frequência dela no restante da rede $F_\alpha^{out}(w)$. Para determinar $F_\alpha^{in}(w)$ e $F_\alpha^{out}(w)$ é necessário contar o número de vezes $n_\alpha(w)$ que um artigo apresentando a palavra w aparece na comunidade α . Note que para esta análise, mesmo que uma palavra w apareça mais de uma vez no mesmo resumo, ela é contada apenas uma vez. Podemos então calcular as frequências em termos de $n_c(w)$ como:

$$F_\alpha^{in}(w) = \frac{n_\alpha(w)}{|\alpha|}, \quad (4.7)$$

onde $|\alpha|$ é o número de artigos pertencentes a comunidade α , e

$$F_\alpha^{out}(w) = \sum_{\gamma \neq \alpha} \frac{n_\gamma(w)}{|\mathcal{V}| - |\alpha|}. \quad (4.8)$$

O índice de importância $I(w)$ é enfim definido pela maior diferença entre $F_\alpha^{in}(w)$ e $F_\alpha^{out}(w)$ ao longo de todas as comunidades.

$$I(w) = \max_\alpha [F_\alpha^{in}(w) - F_\alpha^{out}(w)]. \quad (4.9)$$

Tanto unigramas quanto bigramas são ordenados de acordo com o índice de importância. As palavras com maiores valores de $I(w)$ correspondem às palavras chave. Uma taxonomia da rede pode então ser obtida através de aglomeração hierárquica (73, 59) das palavras-chave usando como métrica a distância topológica média entre os artigos onde elas estão presentes. Como consequência, as palavras podem ser aglomeradas, gerando um dendrograma através da própria topologia da rede. Em resumo, estes dendrogramas incorporam tanto o conceito da rede de citação (através das comunidades e da aglomeração por distância topológica), assim como o contexto presente nos resumos. Este detalhe é crucial já que permite que a estrutura de citações da rede do conhecimento seja usada para inferir a estrutura de organização das palavras presentes nos resumos. Esta análise não poderia ser conduzida de outra forma, como por exemplo, somente considerando os resumos (devido ao seu tamanho limitado).

Por fim, como unigramas e bigramas são classificados de acordo com a mesma métrica, se um bigrama apresenta valores altos para $I(w)$, os unigramas que o compõem também deverão aparecer entre as mais bem classificadas palavras-chave. Para resolver este problema, removemos unigramas do conjunto de palavras-chave que também sejam parte de algum outro bigrama do conjunto.

4.3.3 Visão geral da rede CN

A Figura 45 mostra os principais resultados obtidos para a rede CN. O mapa científico da rede (a) mostra que muitas comunidades obtidas preenchem diferentes regiões do espaço projetado, com exceção das comunidades B e D que parecem compartilhar uma mesma região. Entretanto, através de visualizações 3D desta mesma rede, foi observada uma clara separação entre as duas, portanto trata-se de um fator limitante da projeção 2D. Nesta rede, a maioria das comunidades se estendem a partir da região mais central da projeção. Os unigramas e bigramas que apresentaram maior valor de $I(w)$ são usados para rotular as comunidades.

Na mesma figura, o grafo reduzido (*coarse grained*) também é mostrado (b). Este revela que as comunidades B , C e D são altamente conectadas entre si. A comunidade B (*epidemic, spreading, dynamics*) mantém, conectadas, diversas outras comunidades, se tornando um dos núcleos da rede, assim como H (*fractal, self-similar*). Isso se deve ao fato de que dinâmicas de epidemia representadas pela comunidade B são altamente presentes no campo de redes complexas, sendo uma das dinâmicas mais usadas e aplicadas.(12) Enquanto A (*synchronization and coupling*) é a maior comunidade, ela se conecta fortemente apenas com G (*brain and cortical networks*), fato conhecido por especialistas da área já que dinâmicas de sincronização são úteis na modelagem de sistemas neurais. Surpreendentemente, a comunidade E (*gene regulatory networks, protein interaction, etc*) está entre as mais fracamente conectadas à rede. Isto

Figura 45 – Projeção 2D (a) da rede CN com as cores indicando as comunidades obtidas, cada uma nomeada de acordo com as respectivas palavras-chave encontradas. O grafo reduzido (*coarse-grained*) (b) correspondente à rede é apresentada, onde cada vértice é associado a uma comunidade. O dendrograma (c) foi obtido através da aglomeração hierárquica das palavras-chave.

Fonte: Elaborada pelo autor.

pode indicar que artigos representados por esta comunidade ainda não se aproveitaram por completo das metodologias presentes em redes complexas.

O dendrograma obtido através das palavras-chave na rede CN, , mostrado na Figura 45c, também apresenta detalhes interessantes. Por exemplo, palavras chave relacionadas às redes da área de ecologia se apresentam próximas entre si. Exceto por alguns exemplos contra intuitivos como a alta proximidade entre as palavras-chave *promote* e *player*, que devem ser melhor investigadas no futuro. No geral, a taxonomia é consistente com o que se esperaria de um especialista na área.

4.3.4 Visão geral da rede PC

Os principais resultados obtidos para a rede PC são sumarizados na Figura 46. A projeção da rede (a) já torna evidente que a rede PC apresenta dois grandes grupos de comunidades aglomeradas. Através de uma rápida análise das palavras-chave obtidas para as comunidades que os compõem, verifica-se que cada grupo corresponde a trabalhos produzidos por pesquisadores de áreas bem diferentes. O maior grupo é composto por tópicos usualmente tratados na física experimental e teórica de cristais fotônicos. Já a menor comunidade, refere-se a assuntos tratados por engenheiros que utilizam fibras em telecomunicações. Observa-se também que a interface entre os dois grupos apresenta poucas conexões, indicando que há pouca interação científica entre os dois grupos principais.

Figura 46 – Resumo dos principais resultados obtidos para a rede PC. Similarmente à Figura 45, esta é composta pelo mapa científico (a), grafo reduzido (b) e pelo dendrograma de palavras-chave (c)

Fonte: Elaborada pelo autor.

A interface entre os dois grupos principais pode ser melhor observada no grafo reduzido (Figura 46b). Conexões entre as comunidades *E* (*one dimensional, transfer matrix, matrix*

method, omnidirectional), *G* (*negative refraction, self collimation*), *I* (*detection, biosensor, label free*) e *L* (*vertical cavity, cavity surface, vcsel, surface emit*) compõem a interface entre os dois grupos.

Em (117) as análises apresentadas nesta seção são expandidas e melhor detalhadas. Adicionalmente, através da consultoria com uma pesquisadora especialista na área, foi possível confirmar diversas inferências observadas nesta análise. Nesse trabalho também é discutida a evolução das palavras-chave ao longo do tempo, que pode ser útil para determinar as principais tendências de pesquisa de cada período.

5 Dimensão local e multiescala

A noção de dimensão está diretamente relacionada aos graus de liberdade existentes em um sistema, conceito que também pode ser entendido como o número mínimo de coordenadas, ou variáveis, necessárias para descrevê-lo. Sob esta perspectiva, a dimensão também está ligada à dificuldade em se descrever os estados de um sistema, e, portanto à sua complexidade. Por exemplo, traçar melhores rotas em um espaço 3D pode se tornar uma tarefa muito mais complexa do que o mesmo problema definido em um espaço 2D, assim como o problema de segmentação de objetos se torna muito mais complexo em 3D.⁽¹³²⁾ Similarmente, a tarefa se torna trivial quando executada sobre uma linha, i.e., no espaço 1D. A dimensão, então, se torna uma das características mais fundamentais de sistemas complexos. Como muitos destes sistemas podem ser representados por redes complexas, há grande interesse em também definir uma medida de dimensão para estas estruturas.

Um sistema definido em um espaço de n variáveis geralmente pode ter dimensão até n . Entretanto, nem sempre é garantido que a dimensão do espaço onde o sistema se encontra seja igual à quantidade de variáveis do sistema, isto é, nem sempre a existência de n variáveis implica em um sistema n -dimensional. A Figura 47 ilustra um objeto 3D que apresenta dimensão 2D (uma variedade bidimensional) devido às relações de dependência entre suas variáveis. Uma condição para que isso se verifique é, por exemplo, que as variáveis deste sistema sejam independentes. Há ainda sistemas onde as variáveis e suas relações de dependência não podem ser obtidas diretamente. Desta forma, enquanto é possível determinar o número de variáveis independentes (e, portanto a dimensão) de algumas estruturas matemáticas bem definidas, como variedades e subespaços gerados por funções (26); outra abordagem é necessária para generalizar a dimensão de estruturas onde o conceito de variáveis não é bem estabelecido, como em espaços topológicos ou grafos.

Uma definição mais geral da dimensão é a *dimensão de cobertura de Lebesgue* (133), geralmente estudada em espaços topológicos. Ela é definida a partir da informação do conjunto de vizinhanças (topologia) do espaço e, portanto, não necessita de uma definição de métrica ou de número de variáveis. Intuitivamente, com base na estrutura do sistema formando um espaço topológico $\mathcal{T}$, a dimensão de Lebesgue d_ℓ é definida em termos do número mínimo m de conjuntos de uma cobertura aberta C de $\mathcal{T}^1$ de modo que cada elemento de $\mathcal{T}$ está contido em no máximo m conjuntos. O número de conjuntos na cobertura C precisa ser suficientemente grande para acomodar a possível dimensão do espaço $\mathcal{T}$. O processo de obtenção de m pode ser feito iniciando a partir de uma cobertura C que passa por um processo de refinamento topológico (133), que ajusta a cobertura de modo a minimizar a sobreposição de seus conjuntos. A dimensão de Lebesgue é então definida por $d_\ell = m - 1$.

¹ Uma cobertura C de $\mathcal{T}$ é um conjunto de conjuntos abertos que contém $\mathcal{T}$.⁽¹³³⁾

Figura 47 – Exemplo de variedade bidimensional construída parametricamente pela equação indicada. Verifica-se que mesmo estando definida sob as variáveis em coordenadas cilíndricas (r_0, θ, z) , devido à dependência de apenas duas variáveis (t, u) , o objeto tem dimensão 2D ao invés de preencher o espaço 3D por completo.

Fonte: Elaborada pelo autor.

A Figura 48 ilustra o processo para se obter a d_ℓ de algumas formas geométricas.

Outra alternativa que generaliza o conceito de dimensão é a dimensão de Hausdorff (134, 135, 136) d_H , que já se mostrou eficiente no estudo de uma grande variedade de problemas matemáticos e sistemas reais.(31, 32) Ela pode ser definida de forma intuitiva em termos da dimensão de capacidade $d_H = d_C$, exceto para alguns espaços com estruturas notáveis, construídos matematicamente e geralmente raros na natureza.(136) A dimensão de capacidade baseia-se na ideia de que o volume generalizado de uma bola $V(r)$ em um espaço n -dimensional varia com o raio r através de uma lei de potência do tipo $V^*(r) \propto r^{d_C}$. Por exemplo, em uma linha, o volume generalizado corresponde ao comprimento do segmento de tamanho r , logo $V_{1D}^*(r) \propto r$. Similarmente, para um espaço 2D, obtém-se a área $V_{2D}^*(r) \propto r^2$ e para o 3D, o volume $V_{3D}^*(r) \propto r^3$. Por esta definição, um espaço de dimensão nula é aquele onde o volume generalizado é constante, independentemente de r , isto inclui espaços contendo um ponto ou o espaço sem elementos. Costuma-se, então, definir

Figura 48 – Exemplos ilustrando o processo de obtenção da dimensão de Lebesgue para algumas formas 1D e 2D. Para cada forma, uma cobertura C é mostrada com seus conjuntos abertos identificados por cores. Após o processo de refinamento topológico, obtém-se m , contando o número de conjuntos de C que compõem a maior sobreposição de cada forma geométrica (indicada por setas).

Fonte: Elaborada pelo autor.

a dimensão de capacidade como:

$$d_C = \lim_{\epsilon \rightarrow 0} \frac{\ln V(\epsilon)}{\ln \epsilon} \quad (5.1)$$

Diferentemente da dimensão tradicional e de Lebesgue, a dimensão de Hausdorff permite valores reais não necessariamente inteiros. Quando $d_C \neq d_L$, a estrutura deste sistema é chamada de fractal. Desta forma, é possível encontrar ou construir estruturas que apresentem dimensão fractal maior do que a natural. Por exemplo, um segmento da curva de Koch (137) não possui comprimento finito, da mesma forma que também não apresenta área definida, estando entre as dimensões 1 e 2. Este fenômeno pode ser observado em diversas estruturas da natureza (138, 137), assim como em construções matemáticas (133) e sistemas complexos da física.(134)

5.1 Dimensão em redes complexas

O conceito de dimensão em teoria de grafos esteve inicialmente associado à noção de planaridade de grafos.(139) Um grafo planar é aquele que pode ser mergulhado em um espaço 2D sem que ocorram cruzamentos entre suas arestas. Enquanto essa definição restringe o conjunto de grafos planares, a extensão para dimensões maiores não é imediata, já que qualquer grafo pode ser mergulhado em um espaço 3D ou superior, sem que ocorram

cruzamentos de arestas. Posteriormente, outro conceito de dimensão foi introduzido por Erdős em (140), onde uma medida de dimensão de grafos d_e foi definida para um grafo G como a dimensão do menor espaço métrico em que é possível mergulhar G de modo que as distâncias entre todos os seus vértices conectados seja 1. Ele também provou a relação $d_e(G) \leq 2\chi(G)$ que liga o conceito de números cromáticos de grafos χ à dimensão. O número cromático $\chi(G)$ indica o mínimo número de cores que preenchem todos os vértices de G de modo que nenhum vizinho apresente a mesma cor. Apesar disso, o cálculo tanto de $\chi(G)$ quanto de $d_e(G)$ são problemas NP-Completo (141), tornando o método, na prática, útil apenas para estudos teóricos de grafos pequenos ou regulares.

Anos depois, a caracterização da dimensão de grafos foi revisitada, no contexto de redes complexas, por Csányi (142) e consolidada por Gastner e Newman.(143) Recentemente, Havlin e colaboradores (20) analisaram três diferentes métodos para calcular a dimensionalidade de redes, baseados em diferentes dinâmicas físicas: difusão, caminhada aleatória e percolação; resultando em alta concordância entre as medidas de dimensão usadas. Também foi descoberto que a dimensão finita em redes não depende do tamanho nem do grau médio de muitas redes reais. Nesta nova abordagem, a dimensão de grafos é entendida pela formulação da dimensão de Hausdorff, onde é possível estimar o fator de escala, isto é, a dimensão de capacidade d_C , baseada no volume de vértices, $V(r)$, definido pela bola concêntrica $B(r)$ de distância topológica r . Supõe-se que, analogamente ao caso de estruturas contínuas, $V(r) \propto r^{d_C}$. A dimensão de redes definida em (143) e também usada como dimensão de difusão em (20) é então:

$$d_C = \lim_{r \rightarrow \infty} \frac{\ln V(r)}{\ln r}. \quad (5.2)$$

É importante notar que diferentemente da Equação 5.1, devido ao caráter discreto de redes, quanto maior r , melhor é a estimativa da dimensão de Hausdorff, comparada àquela obtida no caso contínuo. Entretanto, como redes reais são finitas considera-se a dimensão como a inclinação da curva gerada no gráfico log-log de $\ln V_i(r) \times \ln r$, considerando como referência um vértice central da rede v_i . A Figura 49a ilustra a curva $\ln V_i(r) \times \ln r$ para uma rede simples tomando o vértice mais central da rede como referência. A dimensão é estimada através do coeficiente angular da melhor reta que se ajusta aos pontos.

Em outra abordagem, Shanker (144) generalizou o conceito de dimensionalidade ao desenvolver uma nova forma, matematicamente coerente, de se determinar a dimensão baseada na função zeta de Riemann. Ele também provou que redes ER têm dimensão infinita.(145) Apesar disso, com exceção da possibilidade de estudos analíticos da dimensão em redes, o uso desta metodologia não apresenta benefícios práticos para o cálculo da dimensão de redes reais.

As medidas de dimensionalidade (143, 20) propostas na literatura apresentam bons indicativos sobre a estrutura global das redes, entretanto, não caracterizam individualmente

Figura 49 – Exemplo do cálculo da dimensão de capacidade (143, 20) em uma mesma rede complexa, entretanto com duas diferentes escolhas de vértices centrais (indicados pelas setas). Em (a) a dimensão resulta em $d \approx 1.8$, mas no caso da escolha mostrada em (b), a dimensão é substancialmente menor, com $d \approx 0.8$.

Fonte: Elaborada pelo autor.

os vértices, nem permitem quantificar a dimensão sobre o ponto de vista multiescala. Redes interdependentes (146), construídas pela união de diversas redes, por exemplo, podem apresentar regiões com dimensões bem diferentes. Em seu artigo, Gastner comenta que a dimensionalidade pode mudar consideravelmente entre os vértices da rede e, portanto, uma medida local de dimensão se torna necessária para caracterizar completamente este aspecto em redes. Apesar disso, nenhum trabalho subsequente estudou ou sugeriu novas formas de se caracterizar a dimensão local, ao redor de diferentes vértices de referência.

A influência da escolha do vértice de referência é crucial para o cálculo da dimensão em redes finitas, como pode-se ver na Figura 49b comparada à Figura 49a.

É conhecido que algumas redes geográficas reais que não são pequeno mundo, assim como grafos regulares, apresentam distribuição de distâncias que seguem uma lei de potência (147, 148) em concordância com o que ocorre em sistemas contínuos. Entretanto, há redes onde $V(r)$ não cresce respeitando uma lei de potência do tipo r^d , nestes casos, a

rede é considerada como não tendo uma dimensão estável ou dimensão bem definida.(149) A Figura 50 ilustra este problema ao comparar as curvas de $V_i(r)$ obtidas para a rede de estradas dos EUA com aquela obtida para a infraestrutura da internet. Desta forma, para a análise destas redes, torna-se necessário generalizar a medida de dimensão para que possa caracterizar mesmo uma estrutura onde a dimensão varia com a escala.

Figura 50 – Exemplos de curvas de $V_i(r)$ para duas redes reais (143): (a) rede de estradas dos EUA (149) e (b) rede da infraestrutura da Internet.(107) Enquanto a dimensão de (a) pode ser obtida observando a inclinação da reta formada pela curva, em (b) não é possível traçar uma única reta que represente uma parcela considerável dos vértices, tornando impossível o cálculo da dimensão segundo o método existente.(143, 20)

Fonte: Elaborada pelo autor.

5.2 Definindo uma dimensão multiescala de redes

Para melhor definir a dimensão de capacidade em redes complexas e no contexto multiescala, consideramos uma rede, em determinada escala, como uma amostra de uma variedade d_m -dimensional. Isso é necessário, pois, originalmente, a dimensão de qualquer espaço discreto é zero e, portanto, apresenta estritamente dimensão $d = 0$. Supõe-se, portanto, que a dimensão de uma rede é uma estimativa da dimensão de sua variedade característica, de modo que a rede possa ser mergulhada nesta variedade mantendo sua estrutura topológica de forma similar à dimensão de Erdős. Entretanto, assim como em diversos sistemas complexos, é interessante que seja permitido que redes complexas apresentem dimensão fractal d , de modo que podem não preencher completamente a sua variedade intrínseca $d_m - 1 \leq d \leq d_m$. Pelo teorema de imersão de Whitney (134), isso implica que a rede pode ser mergulhada em um espaço métrico de dimensão $(2d_m - 1)$, já que devido à sua natureza discreta, não há a necessidade de que este mapeamento seja injetivo. Com isso, é possível relacionar a dimensão topológica d à dimensão do espaço métrico e de sua variedade intrínseca que também pode ser verificada através do uso de métodos de mergulho de redes, como o apresentado neste trabalho para visualização de redes (ver capítulo 3).

Devido à alta variedade de padrões concêntricos, $V_i(r)$, pode variar substancialmente de acordo com a escolha do vértice referência v_i (83), ou mesmo pode não seguir estritamente uma lei de potência. Nestas redes, uma única dimensão não define sua estrutura e, portanto, podem ser consideradas como multidimensionais. Por exemplo, ao considerar apenas rodovias e estradas em um sistema de transporte a rede deste sistema muito provavelmente reflete uma estrutura planar devido à geográfica, entretanto, ao adicionar outros meios de transporte, como metrô ou mesmo o sistema de transporte aéreo, o sistema não pode mais ser representado por uma estrutura unicamente bidimensional. Nestes sistemas, aparece então a necessidade de uma medida *local* e *multiescala* da dimensão. A Figura 51 ilustra uma rede multidimensional gerada artificialmente e com vértices de mesma conectividade, com exceção das bordas.

Figura 51 – Exemplo de rede multidimensional apresentando uma região 2D e outra região 3D. A rede foi construída de modo que, com exceção das bordas, os seus vértices apresentam grau homogêneo $k = 6$. As cores indicam a dimensão local $d_i(4)$ para cada vértice de acordo com os valores da legenda. Mesmo com todos os vértices apresentando o mesmo grau, a medida de dimensão é capaz de diferenciar claramente as duas regiões.

Fonte: Elaborada pelo autor.

Partindo-se da dimensão de Hausdorff (ou de capacidade) já definida em redes complexas. (143, 20) A definição da dimensionalidade pode ser estendida de modo que esta possa variar não somente com a escolha do vértice de referência v_i , como também da escala r de interesse. A dimensão local multiescala, $d_i(r)$ pode então ser definida como a inclinação da curva log-log de $V_i(r)$ contra r , logo:

$$V_i(r) \propto r^{d_i(r)} \quad (5.3)$$

$$d_i(r) \approx \frac{d}{d \ln r} \ln V_i(r). \quad (5.4)$$

A definição da dimensionalidade local multiescala está relacionada à dimensão fractal multiescala (150), obtida geralmente pelo uso do método de Bouligand-Minkowsk (80),

empregada na caracterização de formas 2D em imagens. Entretanto, aqui a dimensionalidade está medindo os limites do espaço ao redor de cada vértice e não medindo a dimensão fractal de algum conjunto especial ou forma de pontos. É importante notar que a dimensão multiescala em redes depende apenas da topologia da rede, em contraste com outras metodologias (151), que necessitam que a rede esteja mergulhada em um espaço métrico.

A equação 5.4 também pode ser expressa em termos de r e $V_i(r)$ diretamente, como:

$$d_i(r) = \frac{r}{V_i(r)} \frac{d}{dr} V_i(r). \quad (5.5)$$

Também é possível obter uma medida de dimensão multiescala geral $D(r)$, (sem considerar um vértice de referência), para isso, usa-se a média $\langle V(r) \rangle$ de $V_i(r)$ sobre todos os vértices v_i , logo

$$D(r) = \frac{r}{\langle V(r) \rangle} \frac{d}{dr} \langle V(r) \rangle. \quad (5.6)$$

Também é possível expressar $D(r)$ por meio da distribuição de geodésicas da rede $\rho(r)$, que indica o número de geodésicas com tamanho r na rede, portanto:

$$\langle V(r) \rangle = \sum_{v_i \in \mathcal{V}} \frac{V_i(r)}{|\mathcal{V}|} = \rho(r)(|\mathcal{V}| - 1) \quad (5.7)$$

$$D(r) = \frac{r}{\rho(r)} \frac{d}{dr} \rho(r). \quad (5.8)$$

Por simplicidade, usaremos os termos *dimensão local* para $d_i(r)$ e *dimensão multiescala* para $D(r)$. Verificamos experimentalmente que para todos os modelos de redes reais tratados neste trabalho $D(r) \approx \langle d(r) \rangle$, permitindo que ambas sejam intercambiáveis. O uso desta última definição permite que também possa ser estimado o erro $\sigma(d(r))$ associado à medida da dimensão multiescala.

De acordo com as equações 5.5 e 5.6, os valores de $d_i(r)$ e $D(r)$ podem ser estimados através do uso de métodos de diferenças finitas.(99) Outra alternativa é também estimar diretamente os coeficientes $d_i(r)$ e $D(r)$ a partir de ajuste local, por exemplo através do uso de mínimos quadrados.(152) Enquanto o método escolhido não deve influenciar os valores das medidas para r alto, na região de interesse local, com $r < 5$, a escolha do método de diferenciação não pode ser arbitrária já que influencia diretamente. Desta forma, optamos por avaliar 4 métodos para se estimar os valores das dimensões, sendo eles: diferenças progressivas (*dFW*), diferenças regressivas (*dBK*), diferenças centradas (*dCe*) e através de mínimos quadrados locais (*LS*).

Os métodos baseados em diferenças finitas (*dFW*, *dBK* e *dCE*) podem ser aplicados sob diferentes ordens de acordo com os coeficientes descritos em (99). Já o método *LS* precisa de modificações para permitir a estimação local das funções de dimensão e também sob diferentes ordens de precisão. Baseando-se na definição tradicional de mínimos quadrados

em leis de potência (152) obtemos sua versão local considerando apenas valores em uma janela de n valores anteriores e posteriores a r , onde n é a ordem do método. Portanto, para o cálculo da dimensão $d_i^{n\text{-LS}}(r)$ para o método LS com ordem n , usa-se a equação:

$$d_i^{n\text{-LS}}(r) = \frac{(2n+1) \sum_{h=r-n}^{r+n} (\ln V_i(h) \ln h) - \sum_{h=r-n}^{r+n} (\ln V_i(h)) \sum_{h=r-n}^{r+n} (\ln h)}{(2n+1) \sum_{h=r-n}^{r+n} (\ln h)^2 - \left(\sum_{h=r-n}^{r+n} \ln h \right)^2}. \quad (5.9)$$

A ordem de precisão dos métodos de estimação pode ser entendida como a resolução, ou nível de detalhes da dimensão, de modo que, para ordens maiores, as curvas de dimensão ao longo de r são mais suaves, entretanto, simultaneamente revelam menos detalhes. Na Figura 51 a dimensão local foi obtida através do método LS de ordem $n = 2$ e é mostrada de acordo com as cores dos vértices. Como será visto na próxima seção, o método que melhor adere ao conceito de dimensão em redes regulares foi o LS.

5.3 Dimensão multiescala de modelos de redes complexas

Nesta seção será explorado o comportamento da dimensão local e multiescala para alguns modelos de redes complexas, incluindo o modelo ER, redes baseadas em grades regulares, modelos de redes pequeno mundo e modelos de redes geográficas.

5.3.1 Estudo analítico da dimensão multiescala em redes ER

Apesar de redes ER não representarem muitos sistemas reais, elas usualmente são consideradas modelos nulos para diversos problemas reais, sendo um ótimo modelo para estabelecer comparações de medidas topológicas. Desta forma, é importante investigar o comportamento da dimensão multiescala nestas redes, além de também permitir a validação da medida de acordo com resultados já conhecidos da dimensão para redes ER.

Sabe-se que para uma probabilidade p de conexão fixa e tamanho infinito a dimensão de uma rede ER também é infinita. (144, 145) Entretanto, em geral, redes ER são geradas com tamanho e grau médio finito para que possam ser comparadas a outros modelos ou redes reais, portanto, além dos limites extremos também é importante caracterizar a dimensão dessas redes para redes finitas geradas a partir dos parâmetros: número de vértices $|\mathcal{V}|$ e probabilidade de conexão p ou grau médio esperado $\langle k \rangle$.

Em (153), Blondel e colaboradores determinaram uma aproximação analítica e fechada para a distribuição de caminhos mínimos $\rho^{\text{ER}}(r)$ em uma rede ER, obtida diretamente a partir dos parâmetros do modelo, p e $|\mathcal{V}|$, como:

$$\rho^{\text{ER}}(r) = 1 - \exp \left(-\frac{(|\mathcal{V}|p)^r}{|\mathcal{V}|} \right), \quad (5.10)$$

que também pode ser escrita em termos de $\langle k \rangle$ já que $p = \frac{\langle k \rangle}{|\mathcal{V}|}$, logo

$$\rho^{\text{ER}}(r) = 1 - \exp\left(-\frac{\langle k \rangle^r}{|\mathcal{V}|}\right). \quad (5.11)$$

Substituindo a Equação 5.10 em 5.8 obtém-se a dimensão multiescala de redes ER:

$$D^{\text{ER}}(r) = \frac{r \ln(|\mathcal{V}|p)}{|\mathcal{V}|} \frac{(|\mathcal{V}|p)^r}{\left(e^{\frac{(|\mathcal{V}|p)^r}{|\mathcal{V}|}} - 1\right)}, \quad (5.12)$$

que também pode ser expressa em termos de $\langle k \rangle$:

$$D^{\text{ER}}(r) = \frac{r \ln(\langle k \rangle)}{|\mathcal{V}|} \frac{\langle k \rangle^r}{\left(e^{\frac{\langle k \rangle^r}{|\mathcal{V}|}} - 1\right)}. \quad (5.13)$$

Para checar a validade das equações 5.12 e 5.13, comparamos as curvas da dimensão multiescala analíticas com as obtidas a partir de redes ER considerando diversos valores de grau médio. Para $\langle k \rangle > 2$ os resultados analíticos representam bem as curvas obtidas pelo método LS de segunda ordem, como mostrado na Figura 52. O tamanho da rede, $|\mathcal{V}| = 10000$, também não parece influenciar substancialmente a correspondência com o resultado analítico.

(a) $|\mathcal{V}| = 10000$

(b) $|\mathcal{V}| = 100000$

Figura 52 – Comparação da dimensão multiescala obtida para redes ER através do método LS de segunda ordem (Equação 5.9) com o resultado analítico obtido pelas equações 5.12 e 5.13.

Fonte: Elaborada pelo autor.

Outra verificação importante, é checar se o comportamento assintótico de $D^{\text{ER}}(r)$ quando $\mathcal{V} \rightarrow \infty$ faz sentido, similarmente ao estudo já realizado usando a função zeta de Riemann.⁽¹⁴⁴⁾ Considerando $D_p^{\text{ER}*}(r)$, a probabilidade de conexão p é mantida fixa, portanto equivale ao limite de $\mathcal{V} \rightarrow \infty$ para a Equação 5.12:

$$D_p^{\text{ER}*}(r) = \lim_{\mathcal{V} \rightarrow \infty} \frac{r \ln(|\mathcal{V}|p)}{|\mathcal{V}|} \frac{(|\mathcal{V}|p)^r}{\left(e^{\frac{(|\mathcal{V}|p)^r}{|\mathcal{V}|}} - 1\right)}. \quad (5.14)$$

Para $r > 1$ é fácil verificar que este limite é nulo $D_p^{\text{ER}*}(r > 1) \rightarrow 0$, assim como para $r = 0$, entretanto, para $r = 1$, o limite é infinito, $D_p^{\text{ER}*}(r = 1) \rightarrow \infty$, concordando com o resultado obtido em (144).

Figura 53 – Mapa da dimensão máxima obtida pelas equações 5.12 e 5.13 para diferentes combinações de parâmetros. O mapa da dimensão para $|\mathcal{V}|$ e $\langle k \rangle$ é apresentado em (a) com a região em destaque detalhada em (b). Similarmente, o mapa em (c) apresenta a dimensão para valores de $|\mathcal{V}|$ e p com detalhe da região em destaque em (d).

Fonte: Elaborada pelo autor.

Como será visto nas seções seguintes, a dimensão multiescala máxima $D_{\max}$ aparenta estar relacionada à dimensão aparente do espaço de melhor mergulho da rede. A Figura 53 apresenta o mapa de dimensão máxima para o modelo ER considerando diversos parâmetros: $|\mathcal{V}|$, p e $\langle k \rangle$. Observa-se que, mesmo para uma rede estatisticamente homogênea como ER, a dimensão apresenta comportamento complexo com os parâmetros. Através dos ma-

pas é possível, por exemplo construir redes ER que apresentem uma dimensão específica que pode ser usada para fins de comparação com outras redes e modelos.

5.3.2 Dimensão de grades regulares e efeito de empacotamento

Outra condição necessária para validar a dimensão multiescala é verificar o seu comportamento em grades bem definidas, isto é, em redes regulares. Devido ao caráter discreto de grades regulares, sua dimensão pode diferir daquela esperada considerando a variedade equivalente. Entretanto, espera-se que o valor da dimensão multiescala se aproxime rapidamente (com o aumento de r) do valor da variedade equivalente, podendo variar sutilmente devido ao número de vizinhanças conectadas e tamanho da rede. Como a dimensão de um vértice é nula, espera-se também que a função tenha um início crescente, após um período, é esperado que a rede alcance uma dimensão máxima e posteriormente caia para zero em redes finitas. Isso ocorre pois em uma escala muito maior do que a da rede, ela apresentará características de um único ponto e portanto, dimensão nula.

Figura 54 – Dimensão multiescala obtida para redes regulares sem efeitos de borda (toroidais) de dimensões 1, 2, 3 e 4 através de diferentes métodos de estimativa (BK, FW CE, LS) calculados em 2ª ordem.

Fonte: Elaborada pelo autor.

Para verificar qual é o método de estimativa da dimensão que melhor se ajusta às condições mencionadas, foram obtidas as curvas da dimensão multiescala para diversas redes regulares infinitas, mostradas na Figura 54. Observa-se que, com o aumento de r , os valores das dimensões multiescala se aproximam da dimensão original das redes regulares. O método que melhor reproduz as condições de rápida aproximação para a dimensão do espaço e de valores crescentes no início é o LS. Verificamos também em outras combinações de redes e para outras dimensões, em todos os casos o método LS foi superior aos outros.

Portanto, adotamos o método LS como padrão para o cálculo da dimensão nas seções subsequentes.

Figura 55 – Exemplo de rede multidimensional toroidal construída para verificar o comportamento da dimensão multiescala. A variedade original (a) é um toróide 3D com casca. As dimensões usadas para gerar a variedade são mostradas em vermelho. A rede (b) é gerada mantendo as características dimensionais da variedade original.

Fonte: Elaborada pelo autor.

Podemos também verificar se a rede captura as mudanças de dimensão em uma rede multidimensional. Para isso construímos uma nova rede multidimensional. Ela é construída a partir de um toróide 3D (dimensão da variedade) com diferentes números de vértices para cada dimensão. Como resultado obtém-se uma rede que tem aspecto de tubo com 3 principais dimensões: comprimento, diâmetro interno e largura da casca externa, como mostrado na Figura 55. Espera-se então que localmente ele seja indistinguível de um objeto 3D, para escalas maiores ela deve passar pelos estágios 2D (superfície do toróide) e posteriormente para um estado 1D (comprimento do toróide). Por fim, a dimensão deve cair para zero devido ao caráter finito da rede.

A Figura 56 mostra as curvas de dimensão para a rede multidimensional construída considerando diferentes métodos de estimação. Todos os métodos apresentaram os resultados esperados, com a rede passando pelos 4 estágios de escala, com destaque para o método LS, que novamente se apresentou superior às outras técnicas. Desta forma, verificamos o potencial da medida para detectar diferentes escalas em redes complexas.

Outra interpretação da dimensão multiescala é que ela pode medir a complexidade ou falta de regularidade de uma rede quando observada em uma certa escala já que se trata de uma medida dos graus de liberdade dessas redes. Por exemplo, redes complexas grandes

Figura 56 – Curvas de dimensão multiescala, considerando os diferentes métodos propostos, obtidas para a rede multidimensional construída de acordo com a Figura 55. A área em detalhe é mostrada para melhor comparar os métodos. Verifica-se que para r grande, todos os métodos convergem em uma mesma curva. Os estágios de escala da rede são ilustrados qualitativamente através de visualizações de uma rede construída da mesma forma, mas com menor número de vértices.

Fonte: Elaborada pelo autor.

podem ser formadas por pequenas e simples estruturas de modo que quando observada por propriedades imediatamente locais não revelem suas características complexas. Um agente observando a este sistema nesta escala facilmente entenderia o funcionamento destas estruturas simples, mas não poderia explicar o funcionamento global do sistema. Por isso esses sistemas são melhores descritos quando as observações ocorrem dentro de uma escala onde os detalhes são mais ricos, que pode ser entendida como aquela que corresponde à maior dimensão multiescala.

A ideia de que as escalas de complexidade estão relacionadas às de dimensão é ilustrada em redes do tipo WS mostradas na Figura 57. Estas redes são geradas partindo-se de uma grade unidimensional que se torna cada vez mais interconectada reconectando arestas aleatoriamente com uma probabilidade p . Com o aumento de p as redes geradas são menores e apresentam maior riqueza em detalhes, portanto a dimensão máxima cresce da mesma forma que sua escala característica se torna menor. Para este fenômeno, damos o nome de *empacotamento* (*wrapping effect*) de redes.

O efeito de *empacotamento* de redes é melhor observado em redes 2D. A Figura 58 mostra uma grade regular onde atua progressivamente um processo de reconexão aleatória de arestas, de modo análogo ao modelo WS. Inicialmente é gerada uma grade 2D com efeitos de borda que em seguida tem uma parcela p_r de suas arestas reconectadas aleatoria-

Figura 57 – Média da dimensão multiescala para redes do modelo WS unidimensional para várias probabilidades de reconexão p com tamanho fixo $N = 5000$ e $\langle k \rangle = 4$. O gráfico interno mostra a curva de dependência do valor máximo de $D(r)$ com p , que se estabiliza em $D_{\text{estável}}(r) \approx 4$, que corresponde à dimensão máxima de redes ER com mesmo número de vértices e arestas.

Fonte: Elaborada pelo autor.

mente. O fenômeno de empacotamento pode ser observado pelo aumento visual da dimensão dessas redes. Para p_r muito pequeno, a rede ainda pode ser mergulhada em um espaço bidimensional, entretanto, após algumas poucas reconexões (por exemplo para $p_r = 0.001$) a rede se torna praticamente 3D. O aumento de p_r aumenta ainda mais a sua dimensão máxima até alcançar a dimensão máxima característica do modelo ER de mesmo tamanho e conectividade.

Em redes reais, o efeito de empacotamento está diretamente ligado à criação de novos atalhos, que não ocorrem naturalmente em redes com características puramente geográficas. Por exemplo, ao considerar o sistema de transporte rápido subterrâneo em uma cidade, a dimensão da rede de transporte pode crescer substancialmente (esta aplicação é explorada na subseção 5.4.2). Devido a capacidade abrupta de acessar novos vértices, em regiões próximas a um atalho, em geral, os vértices apresentam dimensão local maior do que nas regiões de características geográficas (como pode ser visto nas visualizações da Figura 58).

5.3.3 Dimensão em modelos de redes geográficas

Após o modelo ER e a grade n -dimensional, o modelo de redes geográficas aleatórias (RAGEO) é a escolha imediata para se entender o comportamento da dimensão em

Figura 58 – Fenômeno de *empacotamento* em uma rede 2D. No topo são apresentadas visualizações de uma rede regular com efeito de borda sob o mesmo processo gerador de redes WS para diferentes probabilidades de reconexão. Na parte inferior da figura é mostrada a curva da dimensão máxima com a probabilidade de reconexão. As cores dos vértices das visualizações indicam a dimensão local de acordo com a legenda.

Fonte: Elaborada pelo autor.

redes complexas, já que é o modelo mais simples com características geográficas e também por permitir o controle de sua dimensão característica. Assim como observado para redes regulares, espera-se que a dimensão multiescala também reflita as características do espaço original dessas redes. Para isso, a dimensão multiescala e local foram obtidas para redes RAGEO 2D e 3D com diferentes conectividades: $\langle k \rangle = 6$, $\langle k \rangle = 12$, e $\langle k \rangle = 24$.

Figura 59 – Curvas de $D(r)$ para redes geográficas 2D e 3D. Para fins de comparação, são mostrados também, os valores para redes ER e regulares com tamanhos e grau médio semelhantes. Fonte: Elaborada pelo autor.

A Figura 59 mostra as curvas da dimensão multiescala para as redes geográficas consideradas comparadas a outros modelos. As curvas apresentam um pico que indica a escala de maior dimensionalidade da rede, seguido por uma região côncava de decaimento. Nas redes 3D o efeito de tamanho finito da rede já é evidente, como pode ser observado pelo início do decaimento antes que haja estabilidade da dimensão. Mesmo assim, as redes geográficas parecem refletir melhor a dimensão do espaço original quando comparadas às redes regulares, principalmente no caso 3D. Esta característica pode estar ligada ao fato de que redes geográficas são altamente aglomeradas (154) fazendo com que as curvas de dimensão alcancem o valor máximo de dimensionalidade mais rapidamente. Verifica-se também que as restrições de conectividade locais presentes no modelo RAGEO são refletidas pela considerável diferença quando comparada à dimensão obtida para redes ER.

Ao observar a dimensão local de redes RAGEO, foi observado que há variações substanciais de dimensão ao longo de suas estruturas. A Figura 60 ilustra este fenômeno apresentando a dimensão local para $r = 3$, $d_i(3)$, de uma rede geográfica. A rede apresenta regiões de baixa dimensão próximas às bordas, enquanto que algumas regiões centrais apresentam dimensão maior que 2D. Nas bordas, devido ao tamanho finito da rede, é esperado

Figura 60 – Dimensão local para $r = 3$ de uma rede RAGEO de 10 mil vértices e $\langle k \rangle \approx 12$.

Fonte: Elaborada pelo autor.

que a dimensão seja menor, pois, no caso de uma rede $2D$, localmente, a borda apresenta comportamento que varia entre $1.5D$ e $2D$, nos casos das regiões mais afastadas do centro, a dimensão pode cair ainda mais, chegando perto de uma estrutura $1D$. No centro, as regiões de maior dimensionalidade parecem constituir lacunas com poucos vértices funcionando como pontes, desta forma apresentam conexões de maior alcance, o que lhes garante maior dimensionalidade.

5.4 Dimensão multiescala de redes reais

Como visto na seção anterior, as medidas de dimensão propostas correspondem às definições esperadas de uma medida de dimensão em redes, além de se apresentarem robustas na caracterização do espaço original de redes RAGEO. Nesta seção, verificaremos o potencial das medidas de dimensão local e multiescala na tarefa de caracterizar a estrutura topológica de redes complexas reais. Para isso, serão consideradas redes de linhas de distribuição de energia, assim como redes de ruas e estradas. Ambos os conjuntos de dados representam redes geográficas de grande interesse na literatura, principalmente sob o ponto

de vista do problema de robustez (de linhas de transmissão de energia) e de otimização do transporte (redes de ruas e estradas).

5.4.1 Dimensão em redes de linhas de distribuição de energia

Redes de distribuição de energia são estruturas geográficas naturalmente originárias de uma variedade bidimensional, entretanto devido à existência de algumas conexões de longo alcance, a presença de comunidades e densidade heterogênea de vértices e outras características, sua topologia pode não refletir exatamente o seu espaço original. Como a dimensão de um sistema está relacionada também à sua robustez.⁽²⁰⁾ Por exemplo, para separar uma rede composta por uma linha, isto é 1D, apenas é necessário quebrar uma pequena região de pontos em qualquer posição ao longo de seu comprimento. No caso 2D, para que ocorra separação, é necessário quebrar uma linha inteira de vértices, e no 3D um plano é necessário. Sobre este ponto de vista, a análise da dimensão de redes de distribuição de energia tem grande importância. Nestas redes, estações e usinas são representadas por vértices e as linhas de distribuição por arestas.

Para este estudo foram consideradas duas redes de distribuição de energia. Uma correspondendo à rede de distribuição da Europa continental (*EU*) (19) no ano 2003, cobrindo a maioria das estações controladas pela UCTE (*Union for the Coordination of Transmission of Electricity*), com 2783 vértices e conectividade $\langle k \rangle = 2.8$. Como discutido na introdução, verifica-se que sua estrutura topológica da rede EU apresenta uma semelhança notável com sua geografia original, como pode ser observado na Figura 2. A outra rede estudada corresponde às linhas de transmissão dos estados do oeste dos EUA (*US*) (55) disponível no banco de dados de redes do software *Pajek*². Esta rede apresenta um total de 4941 vértices e $\langle k \rangle = 2.66$.

Primeiramente, foram obtidas as curvas de dimensão multiescala para as redes consideradas, mostradas na Figura 61. Todas as curvas são caracterizadas por um pico seguidas por um decaimento. A rede EU apresenta dimensão máxima mais próxima à geográfica, com $D_{max} = 2.4$, diferentemente da rede americana que apresentou $D_{max} = 3$. Isso reforça a ideia de que a dimensão topológica nem sempre reflete a dimensão do espaço original de uma rede. Estes resultados corroboram aqueles previamente obtidos por Havlin e colaboradores (20), obtidos usando as medidas de dimensão clássicas.

Apesar de manterem um comportamento semelhante ao longo dos primeiros níveis concêntricos, as curvas de $D(r)$ para as redes de distribuição de energia começam a diferir substancialmente, com a EU apresentando um decaimento mais lento após o pico. Este comportamento está relacionado ao tamanho finito das redes e ao fato de que a rede US apresenta maior aglomeração local (possivelmente devido à maioria dos centros urbanos do

² Disponível em: <<http://vlado.fmf.uni-lj.si/pub/networks/data/mix/mixed.htm>> Acesso em: 20 ago. 2015.

Figura 61 – Dimensão multiescala obtida para as redes de distribuição de energia (EU e US). Também são mostradas as curvas correspondentes aos modelos geográficos RAGEO2D e RAGEO3D, assim como redes regulares 2D e 3D (REG2D e REG3D).

oeste dos EUA estarem na região litorânea). Desta forma, a rede US apresenta alta dimensão na escala local, mas que logo decai para escalas maiores. Pode-se entender também que a estrutura de dimensão da rede EU é mais homogênea. Este comportamento fica claro ao observar as curvas para as redes regulares também apresentadas na Figura 61, que tendem a manter a mesma dimensão em todas as escalas.

Para verificar o comportamento da dimensão local nas redes de distribuição de energia, foi realizando um mergulho das redes em um espaço 3-dimensional usando o algoritmo do visualizador de redes (Seção 3). Desta forma é possível relacionar a dimensão apropriada para o mergulho da rede com a sua dimensão topológica. As Figuras 62 e 63 mostram as redes EU e US mergulhadas no espaço 3D sob diferentes ângulos de visão. Primeiramente, verifica-se que a rede US se espalha muito mais ao longo do espaço 3D, concordando com a dimensão multiescala maior obtida para esta rede. A rede EU adere muito bem a um espaço 2D, apresentando apenas algumas regiões aglomeradas que constituem uma estrutura 3D, como na região da Espanha e França.

A dimensão local de cada vértice também é mostrada nas Figuras 62 e 63. Em ambos os casos a dimensão local próxima de 3 está presente principalmente em vértices localizados nos centros de aglomerações que constroem subestruturas 3D nestas redes. Também é possível verificar que em regiões construídas por cadeias de vértices (como a presente nas proximidades da Itália), a dimensão local fica muito próxima de 1.

Os resultados apresentados nesta seção fortalecem a hipótese de que as medidas de dimensão local e multiescala apresentam alto potencial na tarefa de caracterização de redes reais. Também é interessante verificar como esta dimensão está relacionada a características importantes de redes geográficas, como a robustez a ataques (155) e dinâmicas de caminhadas aleatórias.

Figura 62 – Visualização 3D da rede EU sob diferentes projeções: frente (a), lado (b) e cima (c). A dimensão local para $r = 3$ é mostrada através das cores. A projeção frontal é a mesma da Figura 2(b).

Fonte: Elaborada pelo autor.

Para entender como a dimensão pode influenciar a robustez de uma rede, realizamos um ataque aleatório (155) às redes de distribuição de energia e em alguns modelos com propriedades similares. Para isso, removemos progressivamente arestas destas redes e obtivemos o número de componentes conectados presentes após as remoções. Este processo permite entender como a remoção de arestas afeta a rede, e portanto sua robustez. Os resultados desta análise são mostrados na Figura 64. Observa-se que as redes de distribuição de energia são aquelas que apresentam menor robustez a ataques quando comparadas aos modelos ER, RAGEO2D e redes regulares. A rede regular 3D, mesmo sendo construída com 6 vizinhanças é muito mais robusta do que a rede 2D construída com 8 vizinhança, corroborando a ideia de que redes de dimensão maior são mais robustas. Entretanto, o mesmo pensamento não é válido para as redes EU e US ao analisar apenas a dimensão multiescala máxima. A rede EU se apresenta sutilmente mais robusta do que a rede US, mesmo apresentando menor dimensão. Uma explicação para este fato é que a rede US, apesar de apresentar dimensão máxima maior, ela não se mantém ao longo das escalas e para todos os vértices (como visto na Figura 63), desta forma, a robustez não depende somente da dimensão máxima, mas também da estabilidade de seu valor ao longo das escalas. A verificação e estudo deste fenômeno é algo que deverá ser futuramente investigado.

Figura 63 – Visualização 3D da rede US sob diferentes projeções: frente (a), lado (b) e cima (c). A dimensão local para $r = 3$ é mostrada através das cores.

Fonte: Elaborada pelo autor.

Em todas as análises das redes estudadas foi observado que vértices localizados nas bordas das redes tendem a apresentar menor valor de dimensão local. Em geral, nas bordas de uma rede, os graus de liberdade, isto é, os movimentos permitidos de um agente localizado nestes vértices, são reduzidos devido a caminhos com terminações inesperadas. Como a dimensão também é uma forma de caracterizar os graus de liberdade de um sistema, é interessante entender como se dá a relação dela com dinâmicas de caminhada nestas regiões.

Para compreender melhor a relação entre bordas, graus de liberdade e dimensão, obtivemos a medida de acessibilidade (Seção 2.8.3) de vértices em redes geográficas e nas redes de distribuição de energia estudadas. A acessibilidade é uma medida dinâmica de diversidade de caminhos e que já se mostrou adequada para identificar bordas em redes geográficas.^(86, 85) Desta forma é interessante comparar a acessibilidade com a dimensão local de vértices e verificar se há alguma relação forte entre elas.

A Figura 65 mostra gráficos da acessibilidade contra a dimensão local para algumas

Figura 64 – Efeitos da dinâmica de remoção aleatória de de vértices sobre as redes de distribuição de energia (EU e US) e modelos com propriedades similares (ER e RAGEO, REG2D e REG3D). O gráfico mostra para cada rede a curva do número de componentes conectados normalizado (pelo número de vértices) na rede. Os valores apresentados foram obtidos através da média de 1000 realizações para cada rede, nesta configuração. A largura das curvas corresponde ao dobro do desvio padrão.

Fonte: Elaborada pelo autor.

redes RAGEO2D e para as redes de distribuição de energia. Em todos os casos há alguma correlação entre as medidas, como mostrado também na tabela 4.

Tabela 4 – Coeficiente de correlação de Pearson entre a dimensão local e acessibilidade para a escala $r = 3$.

Rede	GEO2D $\langle k \rangle = 4$	GEO2D $\langle k \rangle = 6$	GEO2D $\langle k \rangle = 12$	US Powergrid	EU Powergrid
Correlação	0.67	0.68	0.77	0.50	0.74

Fonte: Elaborada pelo autor

Foi verificado, também, que os vértices que mais contribuem na correlação são aqueles de baixa acessibilidade (isto é, os vértices localizados nas bordas das redes). Isto pode ser observado no quadro à direita na Figura 65, onde a seleção progressiva dos vértices de menor acessibilidade aos de maior (crescente) aumenta abruptamente a correlação entre as medidas. A inclusão progressiva na outra direção (decrescente) segue um padrão bem mais lento de crescimento. Outro resultado importante é que a correlação foi consideravelmente

Figura 65 – Dimensão local contra acessibilidade considerando o terceiro nível concêntrico ($r = 3$) para diferentes modelos geográficos de tamanho $\mathcal{V} = 5000$. O quadro à esquerda mostra o mesmo mapeamento considerando as redes de distribuição de energia. O quadro à direita mostra a curva de crescimento da correlação entre as duas medidas de modo que apenas os vértices dentro de um limiar crescente ou decrescente na acessibilidade, são considerados. Isto é, usando um limiar da acessibilidade σ_a , apenas vértices com $a_i(3) < \sigma_a$ (crescente) ou $a_i(3) > \sigma_a$ (decrescente), são usados para o cálculo da correlação.

Fonte: Elaborada pelo autor.

maior para a rede EU do que para a rede US. Isso pode ser explicado ao comparar as projeções 3D de ambas as redes. Enquanto a rede EU, por ser mais próxima de um espaço bidimensional, apresenta bordas bem definidas, é difícil identificar as bordas e o centro da rede US, já que esta apresenta características de dimensão fractal entre o 2.5D e 3D.

5.4.2 Dimensão em redes de ruas e estradas

Como discutido anteriormente, a dimensão revela características fundamentais sobre a navegabilidade e sobre os caminhos em uma rede. Estas características tendem a ser importantíssimas nos estudos de problemas de transporte, como, por exemplo no acesso através de ruas e estradas em uma cidade ou estado/província. Torna-se então, importante o estudo da dimensão em redes desta natureza.

Para verificar o potencial das dimensões propostas na caracterização de redes de transporte, conduzimos um estudo semelhante ao realizado nas redes de distribuição de energia, agora para redes de ruas e estradas. Para esta análise foram usadas 4 redes. Três redes de regiões urbanas, construídas a partir das ruas da cidade de *San Joaquín* ($\mathcal{V} = 18263$, $\langle k \rangle \approx 2.6$) (149, 156), *Oldenburg* ($\mathcal{V} = 6105$, $\langle k \rangle \approx 2.3$) (149, 156), *Londres* ($\mathcal{V} = 6539$, $\langle k \rangle \approx$

2.7) (157); e outra construída a partir de todo o sistema de *estradas dos EUA* ($\mathcal{V} = 175813$, $\langle k \rangle \approx 2.0$)³.(149) Em especial, a rede da cidade de Londres também apresenta informação sobre o sistema de transporte rápido subterrâneo, podendo ser acoplada à rede de ruas, assim como feito em (157).

Figura 66 – Dimensão multiescala obtida para as redes de ruas estudadas. Os principais modelos comparáveis às redes também são mostrados, com a rede real correspondente indicada entre parênteses. A rede da cidade de Londres é mostrada em sua versão com e sem o transporte subterrâneo.

Fonte: Elaborada pelo autor.

Figura 67 – Dimensão local para $r = 3$ dos vértices das cidades de San Joaquin e Oldenburg.

Fonte: Elaborada pelo autor.

³ Disponível em: <<http://www.cs.utah.edu/~lifeifei/SpatialDataset.htm>> Acesso 20 ago 2015.

A Figura 66 mostra a dimensão multiescala obtida para as redes de cidades. Enquanto as redes de Londres (sem transporte subterrâneo) e San Joaquin apresentam dimensão estável e próxima de 2 (relativamente próximas aos padrões obtidos para os modelos RAGEO2D), a rede de Oldenburg apresenta uma dimensão mais próxima de 3 e é menos estável ao longo da escala. Esta observação já foi feita em (149), entretanto as metodologias empregadas até então não permitiam calcular uma dimensão característica desta rede, que na verdade apresenta comportamento multiescala, assim como observado, por exemplo, na rede US da seção anterior. A explicação para este fenômeno é que a cidade de Oldenburg, por ser mais velha e não planejada, apresenta uma estrutura que diverge daquela esperada em uma grade regular, apresentando, portanto uma estrutura mais complexa. Este efeito corresponde bem às visualizações mostradas na Figura 67, onde as redes de San Joaquin e Oldenburg são mostradas. Enquanto a rede de San Joaquin apresenta uma estrutura mais organizada em grade, a de Oldenburg é composta por caminhos sinuosos e sem orientação preferencial. Desta forma a dimensão também parece estar relacionada à regularidade nestas redes, fenômeno que também deverá ser investigado no futuro.

Os resultados obtidos para a rede de Londres mostram uma manifestação real do efeito de empacotamento de redes. Sem o sistema de transporte subterrâneo, a rede de Londres apresenta dimensão similar ao do modelo RAGEO2D correspondente, entretanto. Ao conectar o sistema subterrâneo à rede, sua dimensão característica sobe para $D_{\max} \approx 3$ e começa a apresentar características de dimensão instável. Este efeito é melhor investigado na Figura 68. Com o sistema subterrâneo (d) os caminhos da rede colapsam e a rede é deformada, tornando próximas regiões que antes eram distantes, em comparação com (c). Também é possível verificar que a dimensão local das regiões ao redor das linhas de transporte subterrâneo ficaram mais altas, com máximos justamente para vértices dessas linhas.

Para verificar a viabilidade do uso da dimensão em redes grandes, também analisamos a rede de estradas dos EUA através das metodologias propostas.

A Figura 69 mostra os principais resultados obtidos para a rede de estradas dos EUA. Em (a) é mostrada uma visão geral da rede, que é composta por diversas cadeias que formam as estradas ao longo do percurso. Em uma escala local, a estrutura destas estradas se assemelha a uma linha e portanto, em geral apresentam dimensão local próxima de 1. Em regiões metropolitanas como na mostrada em detalhes (b), mesmo em uma escala local vértices já se conectam formando grades e estruturas bidimensionais, resultando em uma dimensão próxima de 2. Já em regiões com comutação de muitas estradas, como nas cidades, a dimensão tende a ser ainda maior. O comportamento multiescala desta rede é observado também na dimensão multiescala (c), que, comparativamente à rede regular equivalente, apresenta crescimento muito mais lento até alcançar a dimensão estável 2. Em trabalhos anteriores, a dimensão desta rede foi considerada como 1.5 (149), entretanto fica claro que a rede apresenta um gradiente da dimensão ao longo das escalas. Esta propriedade só pôde

Figura 68 – Rede da cidade de Londres com dimensão local para $r = 3$ indicada de acordo com a legenda de cores. A rede é mostrada usando a posição original dos vértices sem (a) e com (b) o transporte subterrâneo. Também são mostradas projeções da rede em um espaço 3D geradas pelo visualizador de redes complexas, tanto para a rede sem (c) e com (d) o transporte subterrâneo. Em (b) e (d) as arestas referentes ao sistema de transporte subterrâneo são destacadas por linhas mais largas.

Fonte: Elaborada pelo autor.

ser observada através das medidas de dimensão propostas neste trabalho.

Figura 69 – Dimensão local da rede de estradas dos EUA (a) com destaque para a região compreendendo as cidades de Chicago, Indianapolis e Detroit (b). Em (c) a dimensão multiescala da rede e de modelos de propriedade similares é mostrada.

Fonte: Elaborada pelo autor.

6 Simetria concêntrica

Estudos de simetria de grafos são tradicionalmente conduzidos com base na teoria algébrica de grafos (23, 67, 68), onde ela é definida em termos de isomorfismos e do grupo de automorfismos de um grafo (ver Seção 2.6). Classicamente, uma rede é considerada simétrica se ela apresenta ao menos um automorfismo, caso contrário ela é considerada assimétrica.

Em geral, a presença de simetrias dentro de um sistema pode indicar características importantes relacionadas à regularidade e redundância de sua estrutura. Apesar disso, pouca atenção foi dada ao estudo da simetria em redes complexas (24, 158, 159, 160, 161), as quais representam muitos sistemas reais. Em teoria de grafos, a análise de simetria só pode ser realizada para grafos pequenos ou aqueles bem caracterizados matematicamente, portanto, não pode ser aplicada em sua forma original na caracterização de redes complexas.

Recentemente, Xiao e colaboradores (158) adaptaram técnicas baseadas em teoria de grafos para quantificar a simetria global de redes complexas. Especificamente, através do uso de normalizações do tamanho do grupo de automorfismos (159) e do conceito de órbita de grafos¹. No trabalho de Xiao, foi mostrado também que a simetria parece emergir de padrões similares de conectividade.

Em outro estudo, Holme (24) introduziu uma nova medida para quantificar a simetria, que diverge das técnicas tradicionais existentes em teoria de grafos. Ela é definida de acordo com as vizinhanças de vértices, isto é, em termos dos padrões concêntricos de uma rede. A caracterização desta simetria é feita através da comparação das sequências dos graus, obtidos de vértices em caminhos com origem em um vértice de referência. Uma alta correlação das sequências de grau entre diversos caminhos indica uma alta simetria ao redor do vértice de referência. Entretanto, a medida apresenta diversos problemas, como o fato de não ter fronteiras de valores bem definidos, permitindo inclusive valores negativos, além de perturbações causadas pela existência de vértices com grau elevado. Isto torna difícil a tarefa de comparação da simetria entre redes com tamanhos e conectividade diferentes.

Outra abordagem estudada para quantificar a simetria em redes é baseada em caminhadas quânticas (162), uma analogia em redes da propagação de ondas quânticas de sistemas físicos reais. Esta técnica se baseia na ideia de que os padrões de interferência destrutiva de ondas caminhando em uma rede, estão relacionadas à simetria de sua estrutura. Esta técnica, no entanto, requer a simulação de diversas caminhadas quânticas em uma

¹ A órbita de um vértice é o conjunto de vértices que podem ser trocados com ele sob permutações do grupo de automorfismos.(23, 67, 68)

rede, resultando em altíssimo custo computacional.

Como já mostrado nos capítulos anteriores, o uso de medidas baseadas no conceito de níveis concêntricos permite a caracterização de redes em diferentes escalas, enriquecendo substancialmente a análise de redes complexas. Esta característica também é desejável para a análise da simetria de redes, que deve ser capaz de capturar a simetria tanto local quanto global. Muitas redes reais e modelos são geradas partindo-se de características locais de simetria. Por exemplo, redes de ruas obtidas de cidades planejadas, apresentam estruturas que relembram grades regulares, resultando em uma alta simetria local. No entanto, em redes com grande diversidade de grau como redes livres de escala, espera-se que a simetria local seja perturbada pela presença frequente dos *hubs*. A presença ou não de simetria local pode impactar drasticamente no comportamento de dinâmicas sobre estas estruturas.

Ao longo das próximas seções, é apresentado o desenvolvimento de uma nova medida de simetria, a *simetria concêntrica* (25), que é baseada no conceito de acessibilidade, sendo também definida em termos de níveis e padrões concêntricos. A nova medida é verificada em termos de modelos teóricos e aplicada na caracterização local e global de redes reais. Por fim também apresentamos uma análise preliminar do uso da simetria concêntrica para estudar redes dinâmicas, em especial, uma rede construída a partir de dados da bolsa de valores de Nova Iorque.

6.1 Tamanho do grupo de automorfismos como medida de simetria concêntrica

Em busca de uma nova medida de simetria que possa ser definida tanto localmente quanto globalmente, optamos inicialmente por estender as técnicas baseadas no tamanho do grupo de automorfismos (67, 158), de teoria dos grafos, para que pudessem ser definidas em termos de padrões concêntricos. Entretanto, o número de automorfismos tende a crescer como $V_g(r)!$, isto tende a ocorrer mesmo usando a definição do problema de automorfismo com cores. Neste caso, o conjunto de automorfismos, de tamanho $A(g)$ para um padrão concêntrico g , é determinado com as cores associadas a cada nível concêntrico dos vértices. Desta forma a estrutura geral dos níveis concêntricos é mantida em cada elemento do grupo de automorfismos. Uma abordagem geralmente empregada para normalizar $A(g)$ é determinar o valor máximo possível da medida para um certo conjunto de vértices e cores. (67, 158) O número de automorfismos normalizados $\bar{A}(g)$ é definido então como:

$$\bar{A}(g) = \left(\frac{A(g)}{n_g(0)!n_g(1)!n_g(2)! \cdots n_g(r)!} \right)^{\frac{1}{V_g(r)}}, \quad (6.1)$$

onde $n_g(h)$ corresponde ao número de vértices no nível concêntrico h do padrão concêntrico g e $V_g(r)$, seu total de vértices.

Figura 70 – Exemplo de análise de simetria pelo tamanho do conjunto de automorfismos. Iniciando de um grafo simétrico em (a), ao remover 1 aresta (b), o tamanho do conjunto de automorfismos tradicional, A , e normalizado, $\bar{A}$, são reduzidos. Entretanto, ao considerar a generalização do grafo g_n (c) com n ramos, mesmo a estrutura sendo simétrica, ocorre um decaimento rápido de $\bar{A}$ com n . Além disso, o uso de A é proibitivo já que com o acréscimo de novos vértices, A também cresce $n!$.

Fonte: Elaborada pelo autor.

Apesar de quantificarem apropriadamente a simetria concêntrica ao redor de vértices, ambas as medidas baseadas em automorfismo falham ao se tentar comparar padrões concêntricos de diferentes tamanhos ou diferentes organizações dos níveis. A Figura 70 ilustra o problema. Enquanto é possível comparar a simetria dos grafos g_2 (a) e g'_2 (b), onde há quebra de simetria; ao adicionar novos ramos (c), já não é possível comparar com o caso anterior já que $\bar{A}(g_n)$ tende a decair rapidamente com o número de ramos. Esperava-se, no entanto, que a simetria concêntrica se mantivesse constante, já que a estrutura continua sendo simétrica, independentemente do número de ramos. Desta forma, o uso do número de automorfismos exige uma normalização especial que pode ser difícil de ser encontrada, tornando problemático o uso desta medida em um caso geral de comparação da simetria entre diferentes padrões concêntricos, e, portanto, imprópria para caracterizar vértices em uma rede.

6.2 Acessibilidade como medida de simetria concêntrica

Devido às dificuldades em se determinar uma medida de simetria concêntrica baseada em automorfismos, propomos uma outra abordagem baseada na medida de acessibilidade. Ao verificar os padrões concêntricos de alta acessibilidade, verificamos que grande parte deles apresenta uma estrutura visualmente similar ao que se espera de uma estrutura simétrica. Similarmente, ao examinar vértices de mesmo grau, mas que apresentam baixa acessibilidade em uma rede, estes tendem a apresentar estruturas menos simétricas. A Figura 71 ilustra esta característica, que também pode estar relacionada às bordas das redes.

Figura 71 – Acessibilidade de uma rede RAGEO2D. Padrões concêntricos mais simétricos tendem a aparecer nas regiões mais centrais da rede, de modo que acessam vértices mais distantes uniformemente. Nas bordas da rede a acessibilidade é menor, indicando menor simetria no acesso a outros vértices.

Fonte: Elaborada pelo autor.

Apesar disso, a medida de acessibilidade não pode ser usada diretamente para quantificar a simetria de vértices, isso porque ela apresenta grande dependência do número de nós alcançáveis. Outro problema é quanto à escolha da dinâmica usada para o seu cálculo. Ao considerar a caminhada aleatória autoesquivante, usualmente empregada em sua definição tradicional (veja a Seção 2.8.3), a acessibilidade é calculada também sobre caminhos que podem retornar às proximidades da origem. Isto leva a uma redundância de caminhos que pode levar a perturbações em uma possível medida de simetria. Desta forma, a

acessibilidade ainda necessita de modificações para que seja possível quantificar a simetria concêntrica em redes.

6.3 Simetrias *Backbone* e *Merged*

O primeiro problema da acessibilidade $a_i(r)$, referente à sua dependência com o número de vértices alcançáveis em um padrão concêntrico, é facilmente resolvido normalizando-se $a_i(r)$ pelo número de vértices alcançáveis em r passos a partir do vértice de origem v_r . Desta forma, a medida fica condicionada a valores que variam de 0 a 1.0. Já o segundo problema, referente à perda de informação devido a caminhos com retorno, necessita da adoção de uma nova dinâmica de caminhada. Esta não deve permitir caminhos com retorno a vértices de níveis concêntricos anteriores. Entretanto, ainda é necessário tratar o caso de arestas existentes dentro de um mesmo nível.

Para solucionar a degenerescência causada por caminhos dentro de um mesmo nível, propomos o uso de duas dinâmicas, e como consequência duas simetrias. Uma delas refere-se à dinâmica onde o custo para se caminhar dentro de um mesmo nível é infinito, este comportamento pode ser obtido através de uma transformação no padrão concêntrico, a qual damos o nome de padrão *backbone* (ou padrão de esqueleto), onde arestas dentro de um mesmo nível são removidas. A outra dinâmica baseia-se no outro extremo, onde o custo de uma caminhada dentro de um mesmo nível concêntrico é nula (se dois vértices estiverem conectados). Neste caso a transformação inerente que mantém essa dinâmica, chamada aqui de padrão *merged* (ou padrão incorporado), baseia-se em unir todos os vértices conectados dentro de um mesmo nível. Desta forma, obtém-se as simetrias *backbone* e *merged*, associadas respectivamente aos padrões transformados.

A Figura 72 ilustra a construção dos padrões *backbone* (c) e *merged* (d) a partir de um padrão concêntrico (a). A figura também mostra as transições possíveis de acordo com a dinâmica sugerida, assim como as probabilidades de acesso.

Por fim, também é necessário tratar a existência de vértices com terminações anteriores ao nível (ou número de passos) máximo do padrão concêntrico de interesse. Como é esperado que a presença destas terminações diminua a simetria do padrão concêntrico, optamos por reduzir os valores da medida através de normalização. Para isso, consideramos não só o número de vértices alcançáveis $n_i(r)$ a partir de v_i , assim como, também, o número de terminações, $\Xi_i(h)$, de cada nível concêntrico h com $h < r$. Formalmente, uma terminação no nível h é qualquer vértice sem arestas conectando-se a algum vértice do nível $h + 1$. As terminações, no entanto, não são consideradas no cálculo das probabilidades de transição, isto é, a probabilidade de transição $P_r(v_i \rightarrow v_T)$ partindo do vértice de referência v_i para alguma terminação v_T de nível $h < r$ é sempre nula, $P_r(v_i \rightarrow v_T) = 0$.

Figura 72 – Exemplo das transformações *backbone* e *merged* de um padrão concêntrico de nível $r = 2$. (a) Padrão original com os níveis concêntricos em destaque. (b) Transições possíveis em uma caminhada concêntrica. Para construir o padrão *backbone* (c), todas as arestas existentes dentro de um mesmo nível concêntrico, precisam ser removidas. Já o padrão *merged* (d), é construído unindo-se vértices que pertençam ao mesmo nível concêntrico. O número de arestas existentes entre vértices unidos que levam a um mesmo nó são somados e atribuídos como peso a aresta. Na figura, a probabilidade de transição em ambas as transformações é mostrada pelos números próximos aos vértices alcançáveis.

Fonte: Elaborada pelo autor.

Tanto a simetria *backbone* $Sb_i(r)$, quanto a simetria *merged*, $Sm_i(r)$, podem ser definidas em termos de uma única simetria $S_i(r)$, simplesmente considerando o seu cálculo nos padrões concêntricos respectivamente transformados. Podemos então definir $S_i(r)$, tendo como vértice de referência v_i e para o nível r , como:

$$S_i(r) = \frac{\exp \left\{ - \sum_{v_j \in R_i(r)} P_r(v_i \rightarrow v_j) \ln[P_r(v_i \rightarrow v_j)] \right\}}{n_i(r) + \sum_{h=0}^{r-1} \Xi_i(h)}, \quad (6.2)$$

onde $P_r(v_i \rightarrow v_j)$ indica a probabilidade de transição partindo-se de v_i até o vértice v_j presente no anel concêntrico $R_i(r)$. A Figura 73 apresenta dois exemplos de padrões concêntricos, (a) e (d), para exemplificar o cálculo das simetrias. O primeiro (a) apresenta simetria máxima tanto de *backbone*(b) quanto de *merged* (c). O segundo padrão mostrado (d) apresenta uma estrutura *backbone* (e) com a presença de uma terminação (vértice 3), portanto não apresenta Sb máximo.

Figura 73 – Exemplo do cálculo das simetrias *backbone* (b,e) e *merged* (c,f) para dois padrões concêntricos (a) e (d). O padrão concêntrico da parte superior da figura (a) é simétrico, fato que é refletido na uniformidade da probabilidade de acesso aos vértices alcançáveis em 2 passos(b,c). Já o padrão mostrado em (d) apresenta uma terminação no vértice 3, consequentemente, menor é a simetria *backbone* (e). A simetria *merged* (f) também é sutilmente menor devido ao desbalanço no acesso aos vértices à distância 2 da referência.

Fonte: Elaborada pelo autor.

Intuitivamente, a simetria *backbone* está relacionada ao conceito de simetria radial, enquanto a simetria *merged* é similar à simetria angular presente em estruturas com simetria por rotação.

Um software de código livre foi especialmente desenvolvido para calcular as medidas de simetria e encontra-se disponível em <<http://github.com/filipinascimento/CVSymmetry>> (Acesso em: 20 ago. 2015).

6.4 Caracterização da simetria de modelos de redes complexas

Nesta seção será explorado o comportamento das medidas de simetria propostas para alguns modelos de redes. Para isso foram considerados diversos modelos, começando

com o modelo ER. A Figura 74 mostra os padrões concêntricos obtidos a partir de uma rede ER espalhados ao longo do plano das medidas de simetria, $S_b \times S_m$ para $r = 2$ e $r = 3$. Em ambos os casos, praticamente todos os padrões concêntricos obtidos estão situados em regiões de alta simetria (tanto *backbone* quanto *merged*). Como a presença de triângulos em redes ER é baixa, a maioria dos padrões concêntricos para $r = 2$ são do tipo árvore, levando a uma alta correlação entre os dois tipos de simetria. Este comportamento começa a mudar ao considerar $r = 3$, onde já não é rara a existência de arestas conectando vértices dentro do mesmo nível, tornando as medidas de simetria sutilmente diferentes. Há também uma migração dos padrões concêntricos para uma região de menor simetria. A explicação deste fenômeno está no fato de que redes ER são pequeno mundo, portanto, mesmo para valores pequenos de r , uma parte considerável da rede pode já estar sendo considerada no padrão concêntrico respectivo, o que pode levar a efeitos causados pelo tamanho finito da rede. Apesar disso alta simetria local em redes ER é uma consequência de sua regularidade estocástica, onde seus padrões base tendem a não ser muito diversos, levando à simetria.

Também é interessante entender os efeitos que restrições espaciais, como as presentes em redes geográficas, têm sob as medidas de simetria. Para isso, consideramos 3 modelos geográficos: redes geográficas aleatórias 2D (RAGEO2D), modelo de Waxman (WAX) e redes geradas por tesselação de Voronoi (VOR).^(15, 163, 164)

A Figura 75 mostra os padrões de simetria obtidos para a rede RAGEO2D. Diferentemente das redes ER, os padrões concêntricos de redes geográficas não apresentam estrutura do tipo árvore, mas sim uma alta densidade de triângulos e ciclos de ordem maior, isso é refletido diretamente na relação entre as duas medidas de simetria, que divergem substancialmente entre si. Além disso, a simetria *merged* tende a apresentar valores ainda mais divergentes (e baixos) para $r = 3$.

Os resultados das simetrias obtidas para o modelo WAX são apresentados na Figura 76. Comparada com a rede RAGEO2D, as simetrias obtidas para o modelo WAX são menores, além disso, a análise visual dos padrões indica que estes apresentam estrutura sutilmente mais complexa do que as reveladas no modelo RAGEO2D. Isso se deve ao fato das redes WAX permitirem algumas poucas conexões de longa distância, gerando maior diversidade dos padrões concêntricos, e como consequência, menor simetria.

Também foram analisadas redes geradas através de tesselação de Voronoi (VOR). Redes deste modelo são construídas espalhando-se vértices uniformemente sobre o plano e conectando aqueles pares cujas células de Voronoi se tocam. Este problema é equivalente à tesselação por triangularização de Delaunay.⁽¹⁶⁵⁾ Neste modelo, as condições espaciais são mais estritas, desta forma, também espera-se maior simetria. A Figura 77a, mostra os resultados da simetria para o modelo VOR considerando $r = 3$. A grande maioria dos vértices apresenta simetria *merged* máxima, vértices que divergem deste valor são aqueles pertencen-

(a) $r = 2$ (b) $r = 3$

Figura 74 – Padrões concêntricos $r = 2$ (a) e $r = 3$ (b) da rede ER distribuídos ao longo das simetrias *backbone* e *merged*. A rede apresenta $|\mathcal{V}| = 5000$ vértices e grau médio $\langle k \rangle = 6$.

Fonte: Elaborada pelo autor.

(a) $r = 2$ (b) $r = 3$

Figura 75 – Padrões concêntricos $r = 2$ (a) e $r = 3$ (b) da rede RAGEO2D distribuídos ao longo das simetrias *backbone* e *merged*. A rede apresenta $|\mathcal{V}| = 5000$ vértices e grau médio $\langle k \rangle = 6$.
Fonte: Elaborada pelo autor.

Figura 76 – Padrões concêntricos $r = 2$ da rede WAX distribuídos ao longo das simetrias *backbone* e *merged*. A rede apresenta $|\mathcal{V}| = 5000$ vértices e grau médio $\langle k \rangle = 6$.

Fonte: Elaborada pelo autor.

tes às bordas da rede. A simetria *backbone*, no entanto, varia substancialmente, indicando a presença de alta diversidade de padrões concêntricos. Para entender melhor como a quebra das restrições geográficas pode influenciar a rede, foi também obtida uma variação da rede VOR, onde algumas arestas sofreram processo de reconexão aleatória (similarmente ao que ocorre no modelo WS). A rede RVOR foi obtida a partir da VOR reconectando 0.1% das arestas. Os resultados presentes na Figura 77, indicam que este processo reduz consideravelmente a simetria da rede, ilustrando o papel fundamental das restrições espaciais nestas redes.

O modelo BA também foi analisado de acordo com sua simetria, como mostrado na Figura 78. Verifica-se que dentre todos os modelos, o BA é aquele com maior diversidade de padrões de simetria, cobrindo uma região substancial do mapa das medidas. Como esperado, há uma forte relação entre as medidas de simetria, já que isto reflete o fato de que nestas redes, assim como no modelo ER, a densidade de triângulos, e portanto, o coeficiente de aglomeração são baixos. Observa-se também, que nestas redes, os padrões de maior simetria são aqueles de menor grau.

Por fim, investigamos as correlações entre as medidas de simetria e medidas topológicas tradicionais. Obtivemos as correlações de Pearson para todos os modelos estudados, entretanto, os resultados para as redes aleatórias BA e ER ficaram muito parecidos, e o

(a) VOR $r = 3$ (b) RVOR $r = 3$

Figura 77 – Padrões concêntricos $r = 3$ para as redes VOR (a) e RVOR (b) distribuídos ao longo das simetrias *backbone* e *merged*.

Fonte: Elaborada pelo autor.

Figura 78 – Padrões concêntricos $r = 2$ (a) da rede BA distribuídos ao longo das simetrias *backbone* e *merged*. A rede apresenta $|\mathcal{V}| = 5000$ vértices e grau médio $\langle k \rangle = 6$. Para $r = 4$, as simetrias de BA apresentam substancial correlação com as medidas tradicionais de grau e centralidade de intermediação. Esta tendência, no entanto, se deve ao efeito de tamanho finito das redes, já que em uma rede BA poucos níveis concêntricos são necessários para cobrir integralmente a rede.

Fonte: Elaborada pelo autor.

Figura 79 – Tabela de correlações de Pearson obtidas entre as medidas de simetria concêntrica e diversas outras propriedades tradicionais considerando os modelos BA (a) e WAX (b).

Fonte: Elaborada pelo autor.

mesmo ocorreu entre RAGEO2D e WAX, por este motivo apenas os referentes aos modelos BA e WAX são mostrados na Figura 79. As correlações entre as medidas de simetria são altas para os modelos aleatórios (BA). Como mencionado antes, isso se deve ao fato desses modelos apresentarem baixa densidade de triângulos. As redes geográficas, no entanto, mostram menor correlação entre as medidas de simetria, isso se deve ao fato dessas redes apresentarem maior coeficiente de aglomeração.⁽¹⁵⁾ Em ambos os casos, a simetria correlaciona fracamente com outras medidas tradicionais como o grau e a centralidade de intermediação. Isto indica que estas propriedades revelam informações complementares sobre a estrutura dessas redes.

6.5 Simetria de redes reais

Similarmente à metodologia empregada nos modelos de redes complexas, nesta seção, algumas redes reais são estudadas através das medidas de simetria. Em particular consideramos quatro redes reais para esta análise, Aeroportos (AIR), Wikipédia (WIKI), e as redes das cidades de San Joaquin e Oldenburg.

Figura 80 – Padrões concêntricos $r = 2$ (a) da rede AIR distribuídos ao longo das simetrias *backbone* e *merged*.

Fonte: Elaborada pelo autor.

A rede de aeroportos (AIR) abrange pares de aeroportos internacionais que são co-

nectados se eles compartilham uma rota entre si². A rede AIR contém $|\mathcal{V}| = 2940$ vértices e tem conectividade média $\langle k \rangle \approx 21$. Os resultados da simetria para a rede AIR são mostrados na Figura 80. Verifica-se que os padrões concêntricos são muito mais diversos do que aqueles obtidos para os modelos de redes. Outra tendência é que vários padrões concêntricos apresentam algum tipo de simetria máxima. Aqueles com simetria *backbone* máxima correspondem a padrões centrados em vértices de baixa conectividade, mas que são conectados a hubs, levando a uma construção do tipo árvore. Estes vértices representam aeroportos pequenos que são conectados a aeroportos de maior movimento. No caso de simetria *merged* máxima, observam-se padrões altamente aglomerados, onde praticamente todos os vértices de um nível são conectados entre si.

Figura 81 – Padrões concêntricos $r = 2$ (a) da rede WIKI distribuídos ao longo das simetrias *backbone* e *merged*.

Fonte: Elaborada pelo autor.

Outra rede estudada envolve artigos da área científica presentes na Wikipédia (WIKI). Cada vértice representa um artigo conectado através de citações. Neste caso, a direção das arestas é ignorada, de forma que uma conexão indique apenas se dois conceitos científicos estão relacionados e não exatamente a hierarquia dessa relação. A rede apresenta $|\mathcal{V}| = 45876$ vértices e $\langle k \rangle \approx 5.8$. Os resultados da simetria para a rede WIKI são mostrados na Figura 81. Novamente, observa-se grande dispersão no espaço das simetrias, preenchendo

² A rede foi construída por César Henrique Comin a partir dos dados de rotas disponíveis em: <<http://sourceforge.net/p/openflights/code/HEAD/tree/openflights/data/>> Acesso em: 20 ago. 2015.

praticamente metade deste espaço. Isso indica que estas redes apresentam grande diversidade de padrões concêntricos. Este fato pode estar relacionado à alta complexidade da organização destas estruturas, o que torna cada vértice único.

Diferentemente do observado para as redes AIR e WIKI, as redes de cidades (descritas em detalhes na Seção 5.4.2) apresentaram simetrias com comportamentos similares àqueles observados nos modelos RAGEO e ER, como mostrado na Figura 82. Isso possivelmente se deve às restrições espaciais e de planejamento presentes nestas redes, limitando o escopo de possíveis padrões concêntricos. Outro resultado interessante é que mesmo considerando que ambas as cidades apresentam estruturas diferentes, a dependência entre as simetrias é parecida.

Também foram obtidas correlações das simetrias com outras propriedades tradicionais de redes complexas para as redes reais estudadas. Na Figura 83 é mostrada a tabela de correlações de Pearson para as redes WIKI, que também apresentou resultados similares à rede AIR, e da cidade de Oldenburg, que foram similares aos obtidos para a rede de San Joaquin. Ambas as redes, apresentaram baixa correlação entre as medidas de simetria e as tradicionais. As correlações entre os dois tipos de simetria são altas na rede de Oldenburg. Entretanto, para as duas redes reais, a correlação entre as simetrias de diferentes níveis é baixa, indicando que informação da simetria para cada nível pode ser usada de forma complementar.

6.6 PCA e classificação de redes por simetria

Para melhor entender a relação geral entre as redes estudadas (tanto modelos quanto redes reais), aplicamos PCA aos padrões concêntricos da rede. Como nem todas elas possuem mesmo número de vértices, a análise foi conduzida por amostragem, selecionando 1000 vértices de cada rede para que não houvessem perturbações causadas pela representatividade de cada grupo de redes.

A Figura 84 mostra as projeções obtidas através dos dois primeiros componentes principais para as redes estudadas considerando apenas as medidas de simetria (*backbone* e *merged*) para $r = 2, 3$ e 4 (a). Para fins de comparação foi obtida uma projeção via PCA das redes, mas através do uso de medidas concêntricas para as mesmas escolhas de níveis (b). É importante notar que as propriedades concêntricas também envolvem muitas propriedades tradicionais, como o coeficiente de aglomeração e grau.

O PCA da Figura 84a sumariza muitos dos resultados deste capítulo e destaca as principais características de simetria das redes estudadas. Os dois primeiros componentes correspondem a $\approx 86\%$ da total variância. Notavelmente, o componente *PCA1* corresponde a praticamente ao oposto da média dos valores de simetria obtidos para os níveis considera-

(a) San Joaquin $r = 3$ (b) Oldenburg $r = 3$

Figura 82 – Padrões concêntricos $r = 3$ para as redes de ruas das cidades San Joaquin (a) e Oldenburg (b) distribuídos ao longo das simetrias *backbone* e *merged*.

Fonte: Elaborada pelo autor.

Figura 83 – Tabela de correlações de Pearson obtidas entre as medidas de simetria concêntrica e diversas outras propriedades tradicionais considerando as redes WIKI (a) e Oldenburg (b).

Fonte: Elaborada pelo autor.

Figura 84 – PCA aplicada aos modelos e redes reais. Em (a) apenas as medidas de simetria foram usadas, enquanto em (b) apenas medidas concêntricas tradicionais foram consideradas (ver Seção 2.8.2).

Fonte: Elaborada pelo autor.

dos, sendo definido como: $PCA1 = -0.37Sb_i(2) - 0.43Sm_i(2) - 0.44Sb_i(3) - 0.43Sm_i(3) - 0.43Sb_i(4) - 0.34Sm_i(4)$. Desta forma, quanto menor o valor de $PCA1$, maior é a simetria do padrão concêntrico. Padrões encontrados nas redes WIKI, BA e AIR estão entre

aqueles com maior $PCA1$, portanto, menor simetria, possivelmente refletindo a natureza livre de escala presente nestas redes. Surpreendentemente, os modelos RAGEO2D e WAX apresentam baixa sobreposição entre eles, mesmo sendo modelos geográficos. Isso pode ser causado pela pequena quantidade de conexões de longo alcance permitidas em modelos WAX. As redes que apresentaram maior sobreposição foram aquelas obtidas a partir de cidades, além disso, sob o ponto de vista de simetria, estão mais próximos de modelos ER do que dos modelos geográficos.

Comparativamente, a projeção obtida através de PCA para as medidas de simetria (Figura 84a), é claramente uma representação mais discriminativa das redes estudadas em relação à obtida somente através de propriedades concêntricas (Figura 84b). Isso é refletido em menor sobreposição entre os padrões das redes e melhor aproveitamento do espaço de medidas. Estas qualidades de medidas obtidas a partir de padrões concêntricos são melhor discutidas e analisadas em (166).

Figura 85 – Modelos e redes consideradas ordenadas de acordo com a simetria geral média indicada pelo componente $PCA1$ da Figura 84(a). Cada quadro apresenta uma visualização em miniatura da rede.

Fonte: Elaborada pelo autor.

Considerando que o eixo $PCA1$ mantém 71% da variância original das medidas de simetria, é possível classificar cada rede de acordo com sua simetria média, isto é $\langle PCA1 \rangle$, como mostrado na Figura 85. As redes de cidades são aquelas que apresentam maior simetria, fato que possivelmente está relacionado com o planejamento urbano que deve permitir que estas redes sejam simples de se navegar. Redes apresentando distribuição de grau heterogênea (BA, WIKI e AIR) são aquelas de menor simetria já que a presença ou não de

hubs em padrões concêntricos é responsável por causar um desbalanço nas probabilidades de transição. Redes *ER* e *VOR* são os modelos que apresentaram maior simetria. Curiosamente, as restrições espaciais empregadas nos modelos *RAGEO2D* e *WAX* não contribuem para o aumento da simetria, pelo contrário, tendem a aumentar a diversidade de padrões concêntricos se comparada ao modelo *ER*.

6.7 Aplicação preliminar da simetria concêntrica para identificar e caracterizar crises do mercado financeiro

Para ilustrar o potencial do uso das medidas de simetria introduzidas neste trabalho em uma aplicação real, analisamos de forma preliminar a simetria de uma rede do mercado financeiro ao longo do tempo. A rede, obtida de (167), foi construída a partir de dados do *Yahoo! Finance*³. Este banco de dados contém a informação dos preços de fechamento diários das ações de 347 empresas negociadas na bolsa de Nova Iorque ao longo do período de janeiro de 1986 a fevereiro de 2002. Cada vértice da rede gerada corresponde a uma das 347 empresas, onde pares de vértices são conectados ou não de acordo com a correlação do preço das ações seguindo o método descrito em (167). Neste trabalho a rede foi gerada com 5% das arestas mais representativas (de maior correlação) usando uma janela deslizante de 28 dias.

Figura 86 – Simetria *merged* média para $r = 2$ calculada para a rede do mercado financeiro ao longo do tempo. As regiões em destaque indicam períodos de crises documentadas na literatura.

Fonte: Elaborada pelo autor.

A Figura 86 mostra a simetria *merged* para $r = 2$ obtida ao longo do tempo para a rede do mercado financeiro considerada. As regiões em destaque indicam períodos conheci-

³ Disponível em: <<http://finance.yahoo.com>> Acesso em: 09 abril 2015

dos de crises⁴.(168) Verifica-se que a rede tende a ser altamente simétrica durante períodos de estabilidade econômica, entretanto, durante crises e períodos de pessimismo econômico ocorre um decaimento rápido da simetria, indicando que a topologia da rede se torna mais heterogênea. Isto sugere que durante crises as empresas tendem adotar estratégias distintas, já que seus padrões concêntricos tendem a ser mais diversos, resultado que corrobora estudos recentes na área.(169)

Figura 87 – Tabela de correlações médias e desvios padrão entre as medidas de simetria contra as medidas topológicas tradicionais. As médias foram tomadas para a rede do mercado financeiro ao longo de todo o período considerado. Verifica-se que a simetria merged de nível $r = 2$ apresenta correlação quase nula com todas as outras medidas.

Fonte: Elaborada pelo autor.

Como verificado para as outras redes estudadas, também obtivemos as correlações das medidas de simetria com as tradicionais, neste caso, a tabela de correlações foi tomada ao longo do tempo através da média e desvio padrão, como mostrado na Figura 87. Surpreendentemente, a correlação entre a simetria merged para $r = 2$, empregada na análise

⁴ Disponível em: <<http://www.wsj.com/articles/SB119239926667758592>> Acesso em: 09 abril 2015
Disponível em: <<http://www.telegraph.co.uk/finance/2773265/Billionaire-who-broke-the-Bank-of-England.html>> Acesso em: 09 abril 2015
Disponível em: <<http://www.guardian.co.uk/technology/2005/may/16/media.business>> Acesso em: 09 abril 2015
Disponível em: <http://news.bbc.co.uk/onthisday/hi/dates/stories/april/26/newsid_2503000/2503271.stm> Acesso em: 09 abril 2015
Disponível em: <<http://www.theguardian.com/world/2003/apr/09/russia.artsandhumanities>> Acesso em: 09 abril 2015

anterior, e as outras medidas, inclusive outras simetrias, é praticamente inexistente. Isso ocorre devido à capacidade notável da medida em capturar os dois comportamentos do mercado financeiro de forma estável. Esta característica não está presente ao usar outras propriedades de redes complexas, como mostrado em (108). No geral, mesmo preliminares, estes resultados mostram a aplicabilidade das medidas de simetria concêntrica na caracterização de redes.

7 Conclusões

Nesta tese foram desenvolvidas novas metodologias para a análise multiescala de redes complexas sob a perspectiva de três diferentes aspectos importantes de sistemas complexos: a dimensão, a simetria e a caracterização multiescala de vértices em comunidades. Tanto os conceitos de simetria quanto de dimensão foram pouco explorados na literatura dentro do contexto de redes complexas. Enquanto a organização em comunidades apesar de bem explorada, tem como foco a determinação da estrutura de comunidades e não a identificação e caracterização de seus elementos.

De acordo com a proposta inicial, foi possível definir as novas medidas em termos de níveis e padrões concêntricos de modo que adquirissem características multiescala. As dimensões local e multiescala foram construídas através da extensão da dimensão de capacidade, agora tomadas ao longo dos níveis concêntricos, podendo também estar centradas em um vértice específico.

A dimensão multiescala foi verificada de acordo com os modelos teóricos, como no caso ER, grades regulares e redes geográficas. Nas redes ER foi possível desenvolver uma solução analítica, enquanto que nas redes regulares e geográficas a dimensão encontrada correspondeu à esperada, validando, portanto, a metodologia. Também foi estudado um caso específico de rede multidimensional, construída de modo a apresentar três estágios de dimensionalidade de acordo com a escala observada, onde a medida proposta foi bem-sucedida em determinar todos os três estágios presentes na rede.

Após validada, a medida de dimensão foi usada para caracterizar redes reais, onde observou-se a correspondência entre a dimensão topológica e aquela obtida através de mergulho em espaços métricos. Além disso, foi descoberto que a dimensão local também apresenta uma forte relação com as bordas em redes geográficas, e como consequência com a acessibilidade, indicando que os graus de liberdade de agentes caminhando ao longo das bordas são comprometidos.

Também foram estudados os efeitos de dinâmicas de reconexão de vértices, onde algumas arestas da rede são reconectadas aleatoriamente, de modo análogo ao que ocorre no modelo WS. Ao aplicar essa dinâmica, as redes são empacotadas, apresentando aumento considerável da dimensão máxima. Este efeito se tornou claro ao comparar a rede da cidade de Londres com mesma rede incluindo conexões referentes ao transporte subterrâneo, de longa distância. A rede que antes apresentava características 2D, adquire dimensão próxima de 3, fenômeno que pode ser observado tanto através de visualização da rede, quanto pelo uso da dimensão multiescala.

As análises conduzidas para entender o comportamento das dimensões propostas

validam e demonstram o potencial das medidas para a caracterização topológica de redes, o que permite também que sistemas não necessariamente geográficos possam ser caracterizados pela dimensão.

Diferentemente do desenvolvimento da dimensão multiescala, onde foi possível entender diretamente as definições já presentes na literatura, a abordagem direta para se definir uma medida de simetria em redes, estendendo o conceito de automorfismos, não foi bem sucedida. Isto ocorreu principalmente devido à dificuldade em se encontrar normalizações adequadas para o número de automorfismos. Portanto, propomos como alternativa, o uso do conceito de diversidade de caminhos, que estendem a medida de acessibilidade, para definir a simetria concêntrica em redes. Foi necessário definir uma nova dinâmica baseada em níveis concêntricos, onde apenas as transições entre níveis consecutivos são permitidas. Devido à degenerescência desta dinâmica, também foi necessário dividir o espaço de simetria em dois tipos de medidas, a *backbone* e a *merged*. As medidas de simetria são definidas para todos os subgrafos que constituem padrões concêntricos em redes.

A análise de redes reais e modelos, através das medidas de simetria, ajudaram a estabelecer o perfil dos padrões concêntricos dessas redes. Redes geradas a partir de grades regulares ou homogêneas, como aquelas do modelo ER, apresentaram padrões concêntricos de alta simetria, já redes mais heterogêneas como as obtidas através do modelo BA e redes reais apresentaram maior diversidade no espaço das simetrias. Redes com restrições espaciais, como as de cidades e modelos geográficos também apresentaram alta simetria. Foi possível identificar, que, por exemplo, o modelo Waxman, por permitir algumas poucas conexões de longo alcance apresenta padrões ligeiramente mais diversos e de menor simetria que outros modelos geográficos. As medidas sugeridas também correlacionaram fracamente com as medidas tradicionais, sugerindo que apresentam informação extra sobre esses sistemas.

Ao aplicar PCA para reduzir a dimensionalidade do espaço de simetrias obtido, observou-se que os padrões concêntricos de cada rede tendem a apresentar baixa sobreposição entre aquelas de tipos diferentes, além de preencherem o espaço transformado de modo uniforme. Essa característica é desejável na caracterização de vértices pois permite identificar as redes de origem apenas observando os padrões concêntricos. Esse resultado não é reproduzido ao considerar as medidas concêntricas tradicionais, onde ocorre maior sobreposição e menor uniformidade no preenchimento do espaço. Essas características são melhor exploradas em (166).

Surpreendentemente, o exemplo de aplicação em redes do mercado financeiro mostra uma relação forte entre a simetria *merged* e a história de crises e períodos de pessimismo no mercado. Este vínculo deverá ser aprofundado no futuro, seguindo as análises presentes em (108).

Devido ao baixo custo computacional (ver apêndice A), as medidas de simetria con-

cêntrica e dimensão multiescala (assim como a local), podem ser aplicadas mesmo em redes com grande número de vértices, tornando-se descritores de interesse para toda a comunidade científica da área de redes complexas.

Ambos os aspectos considerados neste trabalho: simetria (ou regularidade) e dimensão; também são conhecidos por estarem relacionados à complexidade de um sistema.⁽¹⁷⁰⁾ Quanto mais regular um sistema, menos elementos são necessários para descreve-lo, e, portanto, menor é a sua diversidade. Similarmente, quanto menor a dimensão, menos variáveis independentes são necessárias para descrever os estados desse sistema. Portanto, quanto mais complexo o sistema, espera-se maior dimensionalidade ou menor regularidade em sua estrutura e vice-versa.

Com as medidas definidas aqui, podemos sumarizar o conceito descrito acima através da Figura 88, que exemplifica um mapeamento das redes e modelos estudados neste trabalho distribuídos ao longo dos eixos que representam a dimensão (dimensão máxima) e a simetria (média da simetria).

A Figura 88 ilustra muitos dos resultados obtidos nesta tese. Por exemplo, redes ER apresentam alta simetria, mas também tendem a apresentar alta dimensão, dependendo dos parâmetros de grau e número de vértices. Redes regulares, modelos geográficos e cidades tendem a apresentar suas dimensões características além de apresentarem alta simetria. Contrariamente, a Wiki e outras redes sem restrições espaciais como redes de conhecimento (JCR2009, CN, PC) tendem a apresentar baixa simetria e alta dimensionalidade, juntamente com modelos BA e com a rede que representa a infraestrutura da Internet.

No mapeamento, também é possível observar o efeito de empacotamento da rede WS, tanto no espaço de simetria, como no espaço da dimensão. A dimensão desta rede aumenta com p até alcançar uma dimensão de saturação correspondente ao modelo ER, entretanto, ao longo do caminho a simetria desta rede cai drasticamente, principalmente entre $p \approx 5\%$ e $p \approx 10\%$. Este efeito pode indicar uma segunda transição de fase nestas redes que ainda não foi explorada na literatura. Como observado nos resultados do capítulo 5, a rede da cidade de Londres apresenta um aumento significativo de dimensão. Apesar disso, diferentemente do que ocorre para as redes WS, não há uma queda substancial no valor médio de simetria com a adição de conexões de longo alcance. Este fenômeno pode estar relacionado às características da dinâmica que estabelece as conexões de longa distância em cada uma dessas redes. Enquanto o estudo dos efeitos na dimensão e simetria, de diferentes dinâmicas topológicas em redes, não faz parte do escopo deste trabalho, este conceito deverá ser melhor explorado em trabalhos futuros.

As elipses de densidade ⁽¹⁷¹⁾ mostradas na figura representam os desvios da média considerando uma função normal bivariada. A redes mapeadas no quadrante de alta dimensão e baixa simetria, apresentam vértices com grande diversidade de valores de simetria e principalmente dimensão. Outro fato observado é que as orientações das elipses são

Figura 88 – Mapeamento dos modelos e redes reais estudados neste trabalho ao longo do espaço da dimensão e da simetria. O tamanho de cada rede (em número de vértices) é indicado pelo tamanho dos símbolos, como indicado pela legenda. Algumas redes são conectadas devido ao seu forte relacionamento, como as redes WS para diferentes probabilidades de reconexão e as redes de Londres, sem e com transporte subterrâneo. As elipses de densidade (171) são mostradas em tons claros, indicando o desvio padrão obtido por ajuste de distribuição normal bivariada.

Fonte: Elaborada pelo autor.

similares entre as redes reais (i.e., desconsiderando o modelo BA) que se encontram nesta região, isso pode indicar um padrão interno comum, presente em redes reais, que relaciona a dimensão e a simetria dos vértices.

Como previsto, redes consideradas mais complexas apresentaram menor simetria e maior dimensão. Redes de conhecimento, por exemplo, são altamente complexas, revelando pouca redundância em sua estrutura, além de não apresentar nenhuma restrição espacial, ou se ela existir, está presa a um espaço de altíssima dimensão, que ainda é desconhecido.

Interessantemente, não encontramos redes que preenchem a região de baixa dimensão e baixa simetria, o que pode indicar que restrições espaciais tendem a aumentar naturalmente a simetria de sistemas. O modelo que mais se desvia disso é o WAX, possivelmente por manter parte das restrições espaciais do seu espaço original e, ao mesmo tempo, permitir a existência de algumas conexões de longo alcance.

A análise de complexidade e dimensão mostra também que muitos sistemas reais apresentam alta dimensão, o que pode dificultar a visualização efetiva de suas redes. Isso indica que mesmo considerando um espaço 3D que permita interatividade, alguns detalhes são perdidos devido à projeção, necessitando de novas técnicas para reduzir a dimensionalidade de modo que as características da estrutura topológica de interesse sejam mantidas. Apesar disso, no geral, as visualizações geradas ao longo deste trabalho, somente possíveis devido às otimizações realizadas no algoritmo foram cruciais para a interpretação dos resultados, assim como para a análise inicial das estruturas.

Em especial, a abordagem multiescala permitida pelo visualizador interativo teve um papel fundamental na obtenção de mapas científicos, que, em conjunção com a caracterização de comunidades e de seus elementos, deram origem a novos métodos para quantificar a interdisciplinaridade nas ciências, assim como, também estabelecer taxonomias de uma área científica que podem ser usadas no processo de sumarização. Além disso, as ferramentas de visualização desenvolvidas permitiram que, através de colaboração extensiva com especialistas da área médica, novas metodologias específicas para a análise de redes de expressão gênica fossem desenvolvidas, fornecendo resultados importantes na área médica.

Em geral, as metodologias exploradas até aqui contribuem para a literatura, fornecendo novas observações e ideias sobre os conceitos de dimensão, da estrutura multiescala e nas direções de como devem ser exploradas novas métricas. Estas contribuições deverão ajudar a entender melhor a dinâmica e organização de redes complexas.

7.1 Principais contribuições

- *Visualização de redes* – Neste trabalho, as modificações sugeridas ao algoritmo FR usual melhoraram o custo computacional vinculado à simulação molecular usada para o

posicionamento de vértices no software desenvolvido para visualização de redes complexas, *Network3D*. Desta forma, a complexidade passou de $O(|\mathcal{V}|^2 + |\mathcal{E}|)$ para apenas $O(|\mathcal{V}| \log |\mathcal{V}| + |\mathcal{E}|)$, permitindo a visualização de redes com grande número de vértices. Embora existam outras técnicas mais sofisticadas para a visualização de redes (29), a metodologia empregada foi adequada para a análise de todas as redes estudadas neste trabalho. Isso, no entanto, não seria possível através do uso de outros softwares já disponíveis publicamente, já que nenhum deles apresenta a integração requerida para esta abordagem, como, por exemplo, interatividade, interface amigável e alta velocidade de plotagem. A visualização interativa em três dimensões foi crucial para o desenvolvimento de todos os tópicos deste trabalho, em especial na análise de mapas científicos e no estudo de dimensão em redes. Futuramente, o software deverá ser distribuído ao público.

- *Caracterização de redes de expressão gênica* – Com a ajuda do visualizador e de técnicas multiescala, foi possível desenvolver um método para classificar e analisar vértices em redes de expressão gênica com o objetivo de encontrar genes que participam de forma notável em redes. Estes desenvolvimentos culminaram na criação de um conjunto de ferramentas computacionais, que incluem o visualizador, programas para gerar redes a partir de dados de transcriptoma e rotinas prontas para o cálculo de propriedades. Atualmente estes recursos estão disponíveis ao grupo de pesquisa do professor Carlos A. Moreira da Faculdade de Medicina da Universidade de São Paulo, que por meio de colaboração, já resultaram em duas publicações (9, 172) e dois capítulos de livro.(173, 174)
- *Dimensão local e multiescala* – Apesar de já existirem algumas definições de dimensão em redes complexas, a nova medida introduzida nesta tese é vinculada à estrutura de níveis concêntricos, portanto, naturalmente multiescala, permitindo a caracterização tanto global quanto dos vértices de uma rede. As análises apresentadas aqui pavimentaram o caminho para o uso dessa propriedade na caracterização de redes complexas em geral. Outros pesquisadores já estão adotando esta definição de dimensão, inicialmente disponibilizada em (175). Em (176) a dimensão local é aplicada em redes geradas a partir de imagens com o objetivo de realçar bordas. A nova técnica apresenta vantagens sobre os métodos de detecção de bordas usuais, principalmente em imagens naturais, onde ocorrem formas complexas além do aparecimento de ruído. Em outra abordagem (177), a dimensão local foi comparada com outras medidas na tarefa de identificar vértices influentes em diversas redes. Como resultado, a performance da dimensão local se mostrou superior a todas as outras medidas, principalmente em redes com estrutura de comunidades.
- *Simetria concêntrica* – A simetria introduzida neste trabalho difere das definições usuais, baseadas em automorfismos. A nova abordagem permitiu que a simetria de redes gran-

des pudesse ser quantificada de forma multiescala, caracterizando também a simetria ao redor de vértices. A análise de correlação e PCA revelou que simetria apresenta informações novas quando comparada às medidas tradicionais e até mesmo com relação às medidas concêntricas. A análise preliminar da simetria concêntrica aplicada no mercado financeiro mostrou o potencial da medida ao identificar de modo singular as crises, simultaneamente apresentando correlações próximas de zero com relação a outras medidas. Em (166) a simetria apresentou performance amplamente superior na tarefa de caracterização de vértices em modelos e redes reais, quando comparada às medidas concêntricas, que embutem várias medidas tradicionais de redes. A simetria também foi usada com sucesso, em (33), na tarefa de detecção de autoria de obras literárias usando redes complexas.

- *Quantificação da interdisciplinaridade* – Foi desenvolvida uma nova metodologia para quantificar a interdisciplinaridade de áreas científicas (10), baseada na ideia de que conceitos interdisciplinares tendem a estar relacionados a um conjunto mais diverso de tópicos. Para isso, foi construída uma rede de citações entre revistas científicas ao longo do tempo e uma medida de entropia temática foi desenvolvida para quantificar a diversidade de assuntos com que uma revista recebe citações. Através dessa análise, foi possível identificar que de fato a ciência está se tornando interdisciplinar, além de traçar um perfil da evolução da interdisciplinaridade ao longo do tempo das principais áreas científicas. Tanto a metodologia, quanto os resultados podem ser de grande interesse para agências e instituições responsáveis por novas políticas que visem aumentar a integração da pesquisa científica em um setor.
- *Sumarização e taxonomia de áreas científicas* – Através da combinação de mapas científicos, caracterização de comunidades e PLN foi desenvolvida uma nova metodologia para extrair taxonomias a partir de redes de citações e resumos de artigos.(117) A metodologia baseia-se em obter palavras-chave que descrevam as comunidades encontradas em uma rede de citação, de modo que também possam ser relacionadas para formar dendrogramas informativos que sumerizem um tópico científico. A técnica foi aplicada com sucesso em duas redes geradas a partir das áreas de “Cristais fotônicos” e “redes complexas”. Esta abordagem pode ajudar tanto pesquisadores experientes na tarefa de organizar trabalhos de levantamento (surveys) sobre uma certa área de pesquisa, quanto pesquisadores novatos que desejem iniciar os estudos nesta área.
- *Softwares* – Além do software de visualização, foram desenvolvidas diversas rotinas para o cálculo das propriedades introduzidas nesta tese, incluindo a dimensão multiescala e a simetria concêntrica. Algumas delas já estão disponíveis publicamente. Os ponteiros para os softwares desenvolvidos podem ser encontrados no apêndice A.
- *Banco de dados de redes* – As diversas redes obtidas ao longo deste projeto também cons-

tituem uma das principais contribuições deste trabalho, já que estas podem ser usadas para outras análises nos desenvolvimentos futuros daquelas presentes aqui. As redes desenvolvidas para este trabalho foram: citações entre revistas científicas, as redes de citações das áreas de “Cristais fotônicos” e “Redes complexas” e redes da Wikipédia.

7.2 Trabalhos publicados ou submetidos em periódicos internacionais

- SILVA, F. N.; VIANA, M. P.; TRAVENÇOLO, B. A. N.; COSTA, L. da F. Investigating relationships within and between category networks in wikipedia. *Journal of Informetrics*, v. 5, n. 3, p. 431 – 438, 2011. doi: 10.1016/j.joi.2011.03.003.
- SILVA, F. N.; COSTA, L. da F. Local dimension of complex networks. 2012. (Submetido: Physical Review E) Disponível em: <<http://arxiv.org/abs/1209.2476>>. Acesso em: 10.6.2015.
- SILVA, F. N.; RODRIGUES, F. A.; OLIVEIRA JUNIOR, O. N.; COSTA, L. da F. Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics*, v. 7, n. 2, p. 469 – 477, 2013. doi: 10.1016/j.joi.2013.01.007.
- BANDO, S. Y.; SILVA, F. N.; COSTA, L. da F.; SILVA, A. V.; PIMENTEL-SILVA, L. R.; CASTRO, L. H. M.; WEN, H.-T.; AMARO JUNIOR, E.; MOREIRA-FILHO, C. A. Complex network analysis of CA3 transcriptome reveals pathogenic and compensatory pathways in refractory temporal lobe epilepsy. *PLoS ONE*, v. 8, n. 11, p. e79913, 2013. doi: 10.1371/journal.pone.0079913.
- SILVA, F. N.; COMIN, C. H.; PERON, T. K. D.; RODRIGUES, F. A.; YE, C.; WILSON, R. C.; HANCOCK, E.; COSTA, L. da F. Concentric network symmetry. *Information Sciences*, v. 333, p. 61 – 80, 2015. doi: 10.1016/j.ins.2015.11.014.
- MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; SILVA, F. N.; COSTA, L. da F. Methods for gene coexpression network visualization and analysis. In: PASSOS, G. A. (Ed.). *Transcriptomics in health and disease*. Berlin: Springer, 2014. cap. 4, p. 79–94. Disponível em: <http://dx.doi.org/10.1007/978-3-319-11985-4_4>.
- MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; SILVA, F. N.; COSTA, L. da F.; CARNEIRO-SAMPAIO, M. Thymus gene coexpression networks: a comparative study in children with and without down syndrome. In: PASSOS, G. A. (Ed.). *Transcriptomics in health and disease*. Berlin: Springer, 2014. cap. 7, p. 123–136.

- AMANCIO, D. R.; SILVA, F. N.; COSTA, L. da F. Concentric network symmetry grasps authors' styles in word adjacency networks. *Europhysics Letters*, v. 110, n. 6, p. 68001, 2015. doi: 10.1209/0295-5075/110/68001.
- COMIN, C. H.; SILVA, F. N.; COSTA, L. da F. A framework for evaluating complex networks measurements. *Europhysics Letters*, v. 110, n. 6, p. 68002, 2015. doi: 10.1209/0295-5075/110/68002.
- SILVA, F. N.; COMIN, C. H.; PERON, T. K. D.; RODRIGUES, F. A.; YE, C.; WILSON, R. C.; HANCOCK, E.; COSTA, L. da F. On the modular dynamics of financial market networks. 2015. (Submetido: Physical Review E) Disponível em: <<http://arxiv.org/abs/1501.05040>>. Acesso em: 10.6.2015.
- SILVA, F. N.; AMANCIO, D. R.; BARDOSOVA, M.; OLIVEIRA JUNIOR, O. N.; COSTA, L. da F. Using network science and text analytics to produce surveys in a scientific topic. 2015. (Submetido: Journal of Informetrics) Disponível em: <<http://arxiv.org/abs/1506.05690>>. Acesso em: 10.6.2015.
- MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; IAMASHITA, P.; SILVA, F. N.; COSTA, L. da F.; SILVA, A. V.; CASTRO, L. H. M.; WEN, H.-T. Community structure analysis of transcriptional networks reveals distinct molecular pathways for early- and late-onset temporal lobe epilepsy with childhood febrile seizures. *PLoS ONE*, v. 10, n. 5, p. e0128174, 2015. doi: 10.1371/journal.pone.0128174.
- MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; SILVA, F. N.; COSTA, L. da F.; FERREIRA, L.; FURLANETTO, G.; CHACUR, P.; ZERBINI, M. C.; CARNEIRO-SAMPAIO, M. Modular Transcriptional Repertoire and MicroRNA target Analyses Characterize Genomic Dysregulation in the Thymus of Down Syndrome Infants. 2015. (Submetido: Oncotarget)

7.3 Trabalhos futuros

As metodologias aqui introduzidas foram aplicadas em um conjunto diverso, porém limitado de redes reais e modelos. Uma extensão direta deste trabalho seria reproduzir as metodologias aqui presentes, como os mapas de dimensão e de simetria, entretanto, sob o contexto de uma análise comparativa e exaustiva, envolvendo um conjunto grande de redes reais e modelos. Desta forma será possível estabelecer um mapa ainda mais completo das características presentes nesta tese.

Espera-se que trabalhos futuros também explorem de forma mais sistemática a relação entre os conceitos de graus de liberdade, diversidade e dimensão; expandindo a análise apresentada na Figura 65. É possível que estes aspectos possam ser conectados à dimensão

de informação (178), onde as probabilidades de transição também são envolvidas no cálculo da dimensão efetiva do sistema.

É interessante verificar também, o quão robustas as medidas introduzidas neste trabalho são diante de perturbações em redes. Em especial, é importante verificar como a simetria e a dimensão se comportam quando as redes naturalmente apresentam alguma condição não ideal. Este efeito pode ser analisado através de simulação, por exemplo, através de amostragem ou da criação e eliminação de uma pequena quantidade de vértices ou arestas. Desta forma, será possível entender melhor as limitações desta abordagem e estabelecer formas para torná-las mais robustas. Dentro dessa mesma abordagem, também é importante estudar como as medidas aqui introduzidas são influenciadas por dinâmicas topológicas (i.e., quando sua estrutura muda com o tempo) que podem ocorrer em uma rede.

Uma limitação da abordagem da quantificação de interdisciplinaridade está na dependência do método com a categorização dos vértices (em nossa abordagem, pelas categorias temáticas). Uma forma de tratar esta limitação é combinar os métodos de obtenção de taxonomia automática com o cálculo da interdisciplinaridade. Além de eliminar a possível polarização dos resultados devido ao particionamento arbitrário das áreas científicas, a metodologia poderia ser aplicada mesmo em banco de dados onde não há uma categorização inerente. Quando já existe uma estrutura de categorias inerente do sistema, os resultados obtidos considerando a categorização preconcebida e a obtida automaticamente poderiam ser comparados, verificando-se os efeitos reais da polarização.

Referências

- 1 BARABASI, A.-L.; GULBAHCE, N.; LOSCALZO, J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, v. 12, n. 1, p. 56–68, 2011. doi: 10.1038/nrg2918.
- 2 DOYE, J. P. K. Network topology of a potential energy landscape: a static scale-free network. *Physical Review Letters*, v. 88, n. 23, p. 238701, 2002. doi: 10.1103/PhysRevLett.88.238701.
- 3 PORTER, M. A.; MUCHA, P. J.; NEWMAN, M. E. J.; WARMBRAND, C. M. A network analysis of committees in the u.s. house of representatives. *Proceedings of the National Academy of Sciences*, v. 102, n. 20, p. 7057–7062, 2005. doi: 10.1073/pnas.0500191102.
- 4 M ZENG XHT, A. L. W. Cross-evaluation of metrics to estimate the significance of creative works. *Proceedings of the National Academy of Sciences*, v. 112, n. 5, p. 1281–1286, 2015. doi: 10.1073/pnas.1412198112.
- 5 ZHAO, W.; CHELLAPPA, R.; PHILLIPS, P. J.; ROSENFELD, A. Face recognition: a literature survey. *ACM Computing Surveys*, v. 35, n. 4, p. 399–458, 2003. doi: 10.1145/954339.954342.
- 6 LAWRENCE, S.; GILES, C. L. Searching the world wide web. *Science*, v. 280, n. 5360, p. 98–100, 1998. doi: 10.1126/science.280.5360.98.
- 7 DESERNO, T. M.; LOTUFO, R. A.; RITTNER, L.; AUDIGIER, R.; MACHADO, R. C.; SAÚDE, A. V. Biological and medical physics, biomedical engineering. In: _____. *Biomedical image processing*. Berlin: Springer, 2011. p. 107–129.
- 8 DETLEFS, D.; NELSON, G.; SAXE, J. Simplify: a theorem prover for program checking. *Journal of the ACM*, v. 52, n. 3, p. 365–473, 2005. doi: 10.1145/1066100.1066102.
- 9 BANDO, S. Y.; SILVA, F. N.; COSTA, L. da F.; SILVA, A. V.; PIMENTEL-SILVA, L. R.; CASTRO, L. H. M.; WEN, H.-T.; AMARO JUNIOR, E.; MOREIRA-FILHO, C. A. Complex network analysis of CA3 transcriptome reveals pathogenic and compensatory pathways in refractory temporal lobe epilepsy. *PLoS ONE*, v. 8, n. 11, p. e79913, 2013. doi: 10.1371/journal.pone.0079913.
- 10 SILVA, F. N.; RODRIGUES, F. A.; OLIVEIRA JUNIOR, O. N.; COSTA, L. da F. Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics*, v. 7, n. 2, p. 469 – 477, 2013. doi: 10.1016/j.joi.2013.01.007.
- 11 ALBERT, R.; BARABASI, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, v. 74, n. 1, p. 47–97, 2002. doi: 10.1103/RevModPhys.74.47.
- 12 COSTA, L. da F.; OLIVEIRA JUNIOR, O. N.; TRAVIESO, G.; RODRIGUES, F. A.; BOAS, P. R. V.; ANTIQUEIRA, L.; VIANA, M. P.; ROCHA, L. E. C. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, v. 60, n. 3, p. 329–412, 2011. doi: 10.1080/00018732.2011.572452.

- 13 SANDIFER, C. E. Solution of a problem relating to the geometry of position. In: _____. *The early mathematics of Leonhard Euler*. USA: MAA, 2007. p. 195–200.
- 14 NEWMAN, M. E. J. The structure and function of complex networks. *Siam Review*, v. 45, n. 2, p. 167–256, 2003. doi: 10.1137/S003614450342480.
- 15 BARTHÉLEMY, M. Spatial networks. *Physics Reports*, v. 499, n. 1–3, p. 1 – 101, 2011. doi: 10.1016/j.physrep.2010.11.002.
- 16 BIANCONI, G.; BARABÁSI, A.-L. Competition and multiscaling in evolving networks. *Europhysics Letters*, v. 54, n. 4, p. 436, 2001. doi: 10.1209/epl/i2001-00260-6.
- 17 NEWMAN, M. E. J.; PARK, J. Why social networks are different from other types of networks. *Physical Review E*, v. 68, n. 3, p. 036122, 2003. doi: 10.1103/PhysRevE.68.036122.
- 18 BARABASI, A.-L.; ALBERT, R.; JEONG, H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, v. 281, n. 1-4, p. 69–77, 2000. doi: 10.1016/S0378-4371(00)00018-2.
- 19 ROSAS-CASALS, M.; SOLÉ, R. Analysis of major failures in europe’s power grid. *International Journal of Electrical Power & Energy Systems*, v. 33, n. 3, p. 805 – 808, 2011. doi: 10.1016/j.ijepes.2010.11.014.
- 20 DAQING, L.; KOSMIDIS, K.; BUNDE, A.; HAVLIN, S. Dimension of spatially embedded networks. *Nature Physics*, v. 7, n. 6, p. 481–484, 2011. doi: 10.1038/NPHYS1932.
- 21 BEN-AVRAHAM, D.; HAVLIN, S. *Diffusion and reactions in fractals and disordered systems*. Cambridge: Cambridge University Press, 2005.
- 22 JACKSON, J. D. *Classical electrodynamics*. 3rd. ed. New York: Wiley, 1998.
- 23 SCAPELLATO, R.; LAURI, J. *Topics in graph automorphisms and reconstruction*. Cambridge: Cambridge University Press, 2003. v. 54. (London mathematical society student text, v. 54).
- 24 HOLME, P. Detecting degree symmetries in networks. *Physical Review E*, v. 74, n. 3, p. 036107, 2006. doi: 10.1103/PhysRevE.74.036107.
- 25 SILVA, F. N.; COMIN, C. H.; PERON, T. K. D.; RODRIGUES, F. A.; YE, C.; WILSON, R. C.; HANCOCK, E.; COSTA, L. da F. Concentric network symmetry. *Information Sciences*, v. 333, p. 61 – 80, 2015. doi: 10.1016/j.ins.2015.11.014.
- 26 BUNDE, A.; HAVLIN, S. *Fractals in science*. New York: Springer-Verlag, 2013.
- 27 GIRVAN, M.; NEWMAN, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, v. 99, n. 12, p. 7821–7826, 2002. doi: 10.1073/pnas.122653799.
- 28 FORTUNATO, S. Community detection in graphs. *Physics Reports*, v. 486, n. 3–5, p. 75 – 174, 2010. doi: 10.1016/j.physrep.2009.11.002.
- 29 HU, Y. Efficient and high quality force-directed graph drawing. *The Mathematica Journal*, v. 10, n. 1, p. 37–71, 2005.

- 30 WALSHAW, C. A multilevel algorithm for force-directed graph drawing. In: MARKS, J. (Ed.). *Graph drawing*. Berlin: Springer, 2001, (Lectures notes in computer science, v. 1984, v. 1984/2001). p. 31–55.
- 31 TORRES, R. da S.; FALCÃO, A.; COSTA, L. da F. A graph-based approach for multiscale shape analysis. *Pattern Recognition Letters*, v. 37, n. 6, p. 1163 – 1174, 2004. doi: 10.1016/j.patcog.2003.10.007.
- 32 BRUNO, O. M.; PLOTZE, R. de O.; FALVO, M.; CASTRO, M. de. Fractal dimension applied to plant identification. *Information Sciences*, v. 178, n. 12, p. 2722 – 2733, 2008. doi: 10.1016/j.ins.2008.01.023.
- 33 AMANCIO, D. R.; SILVA, F. N.; COSTA, L. da F. Concentric network symmetry grasps authors' styles in word adjacency networks. *Europhysics Letters*, v. 110, n. 6, p. 68001, 2015. doi: 10.1209/0295-5075/110/68001.
- 34 BANDO, S. Y.; ALEGRO, M. C.; AMARO JUNIOR, E.; SILVA, A. V.; CASTRO, L. H. M.; WEN, H.-T.; LIMA, L. d. A.; BRENTANI, H.; MOREIRA-FILHO, C. A. Hippocampal CA3 transcriptome signature correlates with initial precipitating injury in refractory mesial temporal lobe epilepsy. *PLoS ONE*, v. 6, n. 10, p. e26268, 2011. doi: 10.1371/journal.pone.0026268.
- 35 SILVA, F. N. *Redes complexas: novas metodologias e modelagem de aquisição de conhecimento*. 2009. 185 p. Dissertação (Mestrado em Ciências) — Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2009.
- 36 BILKE, S.; PETERSON, C. Topological properties of citation and metabolic networks. *Physical Review E*, v. 64, n. 3, p. 036106, 2001. doi: 10.1103/PhysRevE.64.036106.
- 37 BARABASI, A.-L.; OLTVAI, Z. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, v. 5, n. 2, p. 101–113, 2004. doi: 10.1038/nrg1272.
- 38 VAZQUEZ, A. Statistics of citation networks. 2001. Disponível em: <<http://arxiv.org/abs/0105031v1>>. Acesso em: 10.6.2015.
- 39 COLIZZA, V.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, v. 3, n. 4, p. 276–282, 2007. doi: 10.1038/nphys560.
- 40 SAMUKHIN, A.; DOROGOVTSSEV, S.; MENDES, J. Laplacian spectra of, and random walks on, complex networks: are scale-free architectures really important? *Physical Review E*, v. 77, n. 3, p. 036115, 2008. doi: 10.1103/PhysRevE.77.036115.
- 41 ANGLES, R.; GUTIERREZ, C. Survey of graph database models. *ACM Computing Surveys*, v. 40, n. 1, p. 1, 2008. doi: 10.1145/1322432.1322433.
- 42 COSTA, L. da F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: a survey of measurements. *Advances in Physics*, v. 56, n. 1, p. 167–242, 2007. doi: 10.1080/00018730601170527.
- 43 BARABASI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 509–512, 1999. doi: 10.1126/science.286.5439.509.

- 44 NEWMAN, M. E. J. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, v. 98, n. 2, p. 404–409, 2001. doi: 10.1073/pnas.021544898.
- 45 LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. *Physical Review Letters*, v. 87, n. 19, p. 198701, 2001. doi: 10.1103/PhysRevLett.87.198701.
- 46 NEWMAN, M. E. Assortative mixing in networks. *Physical Review Letters*, v. 89, n. 20, p. 208701, 2002. doi: 10.1103/PhysRevLett.89.208701.
- 47 BOMZE, I. M.; BUDINICH, M.; PARDALOS, P. M.; PELILLO, M. The maximum clique problem. In: DU, D.-Z. (Ed.). *Handbook of combinatorial optimization*. New York: Springer, 1999. p. 1–74.
- 48 ZHOU, S.; MONDRAGÓN, R. J. The rich-club phenomenon in the internet topology. *IEEE Communications Letters*, v. 8, n. 3, p. 180–182, 2004. doi: 10.1109/LCOMM.2004.823426.
- 49 COLIZZA, V.; FLAMMINI, A.; SERRANO, M. A.; VESPIGNANI, A. Detecting rich-club ordering in complex networks. *Nature Physics*, v. 2, n. 2, p. 110–115, 2006. doi: 10.1038/nphys209.
- 50 SABIDUSSI, G. Centrality index of a graph. *Psychometrika*, v. 31, n. 4, p. 581–581, 1966. doi: 10.1007/BF02289527.
- 51 BRANDES, U. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, v. 25, n. 2, p. 163–177, 2001. doi: 10.1080/0022250X.2001.9990249.
- 52 NEWMAN, M. E. J. The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, v. 8, n. 2008, p. 1–12, 2008.
- 53 ESTRADA, E.; RODRIGUEZ-VELAZQUEZ, J. Subgraph centrality in complex networks. *Physical Review E*, v. 71, n. 5, p. 056103, 2005. doi: 10.1103/PhysRevE.71.056103.
- 54 ERDÖS, P.; RÉNYI, A. On random graphs. i. *Publicationes Mathematicae Debrecen*, v. 6, n. 1, p. 290–297, 1959.
- 55 WATTS, D. J.; STROGATZ, S. H. Collective dynamics of "small-world" networks. *Nature*, v. 393, n. 6684, p. 440–442, 1998. doi: 10.1038/30918.
- 56 AMARAL, L.; SCALA, A.; BARTHELEMY, M.; STANLEY, H. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, v. 97, n. 21, p. 11149–11152, 2000. doi: 10.1073/pnas.200327197.
- 57 HAYASHI, Y. A review of recent studies of geographical scale-free networks. *IPSJ Digital Courier*, v. 47, n. 3, p. 776–785, 2006. doi: 10.2197/ipsjdc.2.155.
- 58 WAXMAN, B. M. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, v. 6, n. 9, p. 1617–1622, 1988. doi: 10.1109/49.12889.
- 59 COSTA, L. da F.; CESAR JUNIOR, R. M. *Shape classification and analysis: theory and practice*. 2nd. ed. Boca Raton: CRC Press, 2009. v. 10. (Image Processing Series, v. 10).
- 60 RAPOPORT, A. Contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics*, v. 19, n. 4, p. 257–277, 1957. doi: 10.1007/BF02478417.

- 61 ALBERT, R.; JEONG, H.; BARABASI, A.-L. Internet - diameter of the world-wide web. *Nature*, v. 401, n. 6749, p. 130–131, 1999. doi: 10.1038/43601.
- 62 NEWMAN, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, v. 103, n. 23, p. 8577–8582, 2006. doi: 10.1073/pnas.0601602103.
- 63 FORTUNATO, S.; BARTHÉLEMY, M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, v. 104, n. 1, p. 36–41, 2007. doi: 10.1073/pnas.0605965104.
- 64 BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: theory and experiment*, v. 2008, n. 10, p. P10008, 2008. doi: 10.1088/1742-5468/2008/10/P10008.
- 65 KARRER, B.; NEWMAN, M. E. J. Stochastic blockmodels and community structure in networks. *Physical Review E*, v. 83, n. 1, p. 016107, 2011. doi: 10.1103/PhysRevE.83.016107.
- 66 ROSVALL, M.; BERGSTROM, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, v. 105, n. 4, p. 1118–1123, 2008. doi: 10.1073/pnas.0706851105.
- 67 BIGGS, N. *Algebraic graph theory*. Cambridge: Cambridge University Press, 1994.
- 68 BEINEKE, L. W.; WILSON, R. J. Topics in algebraic graph theory. In: BEINEKE, L. W.; WILSON, R. J. (Ed.). *Encyclopedia of mathematics and its applications*. Cambridge: Cambridge University Press, 2004. v. 102.
- 69 MCKAY, B. D. *Practical graph isomorphism*. Tennessee: Department of Computer Science, Vanderbilt University Tennessee, 1981.
- 70 BABAI, L.; CODENOTTI, P. Isomorphism of hypergraphs of low rank in moderately exponential time. 2008. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4690999>>. Acesso em: 10.6.2015.
- 71 GAO, X.; XIAO, B.; TAO, D.; LI, X. A survey of graph edit distance. *Pattern Analysis and Applications*, v. 13, n. 1, p. 113–129, 2010. doi: 10.1007/s10044-008-0141-y.
- 72 COSTA, L. da F.; BOAS, P. R. V.; SILVA, F. N.; RODRIGUES, F. A. A pattern recognition approach to complex networks. *Journal of Statistical Mechanics: theory and experiment*, v. 2010, n. 11, p. P11015, 2010. doi: 10.1088/1742-5468/2010/11/P11015.
- 73 DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. New York: Wiley, 2001.
- 74 CAMPBELL, N.; ATCHLEY, W. The geometry of canonical variate analysis. *Systematic Zoology*, v. 30, n. 3, p. 268–280, 1981. doi: 10.2307/2413249.
- 75 MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 11, n. 7, p. 674–693, 1989. doi: 10.1109/34.192463.
- 76 LOUNSBERY, M.; DEROSE, T. D.; WARREN, J. Multiresolution analysis for surfaces of arbitrary topological type. *ACM Transactions on Graphics*, v. 16, n. 1, p. 34–73, 1997. doi: 10.1145/237748.237750.

- 77 HADDAD, P. A.; AKANSU, A. N. *Multiresolution signal decomposition: transforms, subbands, and wavelets*. Orlando: Academic Press, 2012.
- 78 LYONS, M.; AKAMATSU, S.; KAMACHI, M.; GYOBA, J. Coding facial expressions with gabor wavelets. p. 200–205, 1998. doi: 10.1109/AFGR.1998.670949.
- 79 SONG, C.; HAVLIN, S.; MAKSE, H. A. Self-similarity of complex networks. *Nature*, v. 433, n. 7024, p. 392–395, 2005. doi: 10.1038/nature03248.
- 80 BACKES, A. R.; BRUNO, O. M. Fractal and multi-scale fractal dimension analysis: a comparative study of Bouligand-Minkowski method. *INFOCOMP Journal of Computer Science*, v. 7, n. 2, p. 74–83, 2008.
- 81 COSTA, L. da F. The hierarchical backbone of complex networks. *Physical Review Letters*, v. 93, n. 9, p. 098702, 2004. doi: 10.1103/PhysRevLett.93.098702.
- 82 COSTA, L. da F.; SILVA, F. N. Hierarchical characterization of complex networks. *Journal of Statistical Physics*, v. 125, n. 4, p. 845–876, 2006. doi: 10.1007/s10955-006-9130-y.
- 83 COSTA, L. da F.; TOGNETTI, M. A. R.; SILVA, F. N. Concentric characterization and classification of complex network nodes: application to an institutional collaboration network. *Physica A: statistical mechanics and its applications*, v. 387, n. 24, p. 6201–6214, 2008. doi: 10.1016/j.physa.2008.06.034.
- 84 RODRIGUES, F. A.; COSTA, L. da F. A structure–dynamic approach to cortical organization: number of paths and accessibility. *Journal of Neuroscience Methods*, v. 183, n. 1, p. 57–62, 2009. doi: 10.1016/j.jneumeth.2009.06.038.
- 85 VIANA, M. P.; BATISTA, J. a. L. B.; COSTA, L. da F. Effective number of accessed nodes in complex networks. *Physical Review E*, v. 85, n. 3, p. 036105, 2012. doi: 10.1103/PhysRevE.85.036105.
- 86 TRAVENÇOLO, B. A. N.; VIANA, M. P.; COSTA, L. da F. Border detection in complex networks. *New Journal of Physics*, v. 11, n. 6, p. 063019, 2009. doi: 10.1088/1367-2630/11/6/063019.
- 87 HERRERO, C. P. Self-avoiding walks on scale-free networks. *Physical Review E*, v. 71, n. 1, p. 016103, 2005. doi: 10.1103/PhysRevE.71.016103.
- 88 ARRUDA, G. F. de; BARBIERI, A. L.; RODRÍGUEZ, P. M.; RODRIGUES, F. A.; MORENO, Y.; COSTA, L. da F. Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E*, v. 90, n. 3, p. 032812, 2014. doi: 10.1103/PhysRevE.90.032812.
- 89 HERMAN, I.; MELANCON, G.; MARSHALL, M. S. Graph visualization and navigation in information visualization: a survey. *IEEE Transactions on Visualization and Computer Graphics*, v. 6, n. 1, p. 24–43, 2000. doi: 10.1109/2945.841119.
- 90 BAUR, M.; BRANDES, U. Crossing reduction in circular layouts. In: HROMKOVIČ, J.; NAGL, M.; WESTFECHTEL, B. (Ed.). *Graph-theoretic concepts in computer science*. Berlin: Springer, 2005, (Lecture notes in computer science, v. 3353). p. 332–343.
- 91 HEALY, N. Hierarchical drawing algorithms. In: TAMASSIA, R. (Ed.). *Handbook of graph drawing and visualization*. Boca Raton: CRC Press, 2014.

- 92 HAREL, D.; KOREN, Y. Graph drawing by high-dimensional embedding. *Journal of Graph Algorithms and Applications*, v. 8, n. 2, p. 195–214, 2004. doi: 10.1007/3-540-36151-0_20.
- 93 FRUCHTERMAN, T. M. J.; REINGOLD, E. M. Graph drawing by force-directed placement. *Software-Practice & Experience*, v. 21, n. 11, p. 1129–1164, 1991. doi: 10.1002/spe.4380211102.
- 94 KRUSKAL, J. Multidimensional-scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, v. 29, n. 2, p. 115–129, 1964. doi: 10.1007/BF02289565.
- 95 VIANA, M. P.; COSTA, L. da F. Fast long-range connections in transportation networks. *Physics Letters A*, v. 375, n. 15, p. 1626–1629, 2011. doi: 10.1016/j.physleta.2011.03.006.
- 96 KAMADA, T.; KAWAI, S. An algorithm for drawing general undirected graphs. *Information Processing Letters*, v. 31, n. 1, p. 7–15, 1989. doi: 10.1016/0020-0190(89)90102-6.
- 97 GRIEBEL, M.; KNAPEK, S.; ZUMBUSCH, G. *Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications*. Berlin: Springer, 2007.
- 98 PRESS, W. H. *Numerical recipes: the art of scientific computing*. Cambridge: Cambridge University Press, 2007.
- 99 FORNBERG, B. Generation of finite difference formulas on arbitrarily spaced grids. *Mathematics of computation*, v. 51, n. 184, p. 699–706, 1988. doi: 10.1002/nme.1620230205.
- 100 BARNES, J.; HUT, P. A hierarchical $o(n \log n)$ force-calculation algorithm. *Nature*, v. 324, n. 6096, p. 446–449, 1986. doi: 10.1038/324446a0.
- 101 SAMMON, J. W. A nonlinear mapping for data structure analysis. *IEEE Transactions on computers*, n. 5, p. 401–409, 1969. doi: 10.1109/T-C.1969.222678.
- 102 MORRISON, A.; CHALMERS, M. A pivot-based routine for improved parent-finding in hybrid MDS†. *Information Visualization*, v. 3, n. 2, p. 109–122, 2004. doi: 10.1057/palgrave.ivs.9500069.
- 103 PAULOVICH, F. V.; NONATO, L. G.; MINGHIM, R.; LEVKOWITZ, H. Least square projection: a fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics*, v. 14, n. 3, p. 564–575, 2008. doi: 10.1109/TVCG.2007.70443.
- 104 BATAGELJ, V.; MRVAR, A. Pajek - analysis and visualization of large networks. In: MUTZEL, P.; JÜNGER, M.; LEIPERT, S. (Ed.). *Graph drawing*. Berlin: Springer, 2002, (Lecture Notes in Computer Science, v. 2265). p. 477–478.
- 105 BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. Disponível em: <<https://gephi.org/publications/gephi-bastian-feb09.pdf>>. Acesso em: 10.6.2015.
- 106 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.

- 107 ZHANG, B.; LIU, R.; MASSEY, D.; ZHANG, L. Collecting the internet AS-level topology. *ACM SIGCOMM Computer Communication Review*, v. 35, n. 1, p. 53–61, 2005. doi: 10.1145/1052812.1052825.
- 108 SILVA, F. N.; COMIN, C. H.; PERON, T. K. D.; RODRIGUES, F. A.; YE, C.; WILSON, R. C.; HANCOCK, E.; COSTA, L. da F. On the modular dynamics of financial market networks. 2015. Disponível em: <<http://arxiv.org/abs/1501.05040>>. Acesso em: 10.6.2015.
- 109 ZHANG, J.; SMALL, M. Complex network from pseudoperiodic time series: topology versus dynamics. *Physical Review Letters*, v. 96, n. 23, p. 238701, 2006. doi: 10.1103/PhysRevLett.96.238701.
- 110 PERON, T. K. D.; COMIN, C. H.; AMANCIO, D. R.; COSTA, L. da F. da; RODRIGUES, F. A.; KURTHS, J. Correlations between climate network and relief data. *Nonlinear Processes in Geophysics*, v. 21, n. 6, p. 1127–1132, 2014. doi: 10.5194/npg-21-1127-2014.
- 111 MANTEGNA, R. N. Hierarchical structure in financial markets. *The European Physical Journal B*, v. 11, n. 1, p. 193–197, 1999. doi: 10.1007/s100510050929.
- 112 ZHANG, B.; HORVATH, S. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, v. 4, n. 1, p. 1544–6115, 2005. doi: 10.2202/1544-6115.1128.
- 113 HEMMING, K.; MAGUIRE, M. J.; HUTTON, J. L.; MARSON, A. G. Vigabatrin for refractory partial epilepsy. *The Cochrane Library*, n. 1, 2013. doi: 10.1002/14651858.CD007302.pub2.
- 114 SAHASRABUDHE, S.; MOTTER, A. E. Rescuing ecosystems from extinction cascades through compensatory perturbations. *Nature Communications*, v. 2, p. 170, 2011. doi: 10.1038/ncomms1163.
- 115 SILVA, F. N.; TRAVENÇOLO, B. A. N.; VIANA, M. P.; COSTA, L. da F. Identifying the borders of mathematical knowledge. *Journal of Physics A: mathematical and theoretical*, v. 43, n. 32, p. 325202, 2010. doi: 10.1088/1751-8113/43/32/325202.
- 116 SILVA, F. N.; VIANA, M. P.; TRAVENÇOLO, B. A. N.; COSTA, L. da F. Investigating relationships within and between category networks in wikipedia. *Journal of Informetrics*, v. 5, n. 3, p. 431 – 438, 2011. doi: 10.1016/j.joi.2011.03.003.
- 117 SILVA, F. N.; AMANCIO, D. R.; BARDOSOVA, M.; OLIVEIRA JUNIOR, O. N.; COSTA, L. da F. Using network science and text analytics to produce surveys in a scientific topic. 2015. Disponível em: <<http://arxiv.org/abs/1506.05690>>. Acesso em: 10.6.2015.
- 118 CHEN, C.; HICKS, D. Tracing knowledge diffusion. *Scientometrics*, v. 59, n. 2, p. 199–211, 2004. doi: 10.1023/B:SCIE.0000018528.59913.48.
- 119 MENCZER, F. Correlated topologies in citation networks and the web. *The European Physical Journal B*, v. 38, n. 2, p. 211–221, 2004. doi: 10.1140/epjb/e2004-00114-1.
- 120 GUIMERA, R.; AMARAL, L. A. N. Functional cartography of complex metabolic networks. *Nature*, v. 433, n. 7028, p. 895–900, 2005. doi: 10.1038/nature03288.

- 121 GREGORY, S. Fuzzy overlapping communities in networks. *Journal of Statistical Mechanics: theory and experiment*, v. 2011, n. 02, p. 02017, 2011. doi: 10.1088/1742-5468/2011/02/P02017.
- 122 CRADDOCK, T.; HARWOOD, C. R.; HALLINAN, J.; WIPAT, A. e-science: relieving bottlenecks in large-scale genome analyses. *Nature Reviews Microbiology*, v. 6, n. 12, p. 948–954, 12 2008. doi: 10.1038/nrmicro2031.
- 123 DONOVAN, S. Big data: teaching must evolve to keep up with advances. *Nature*, v. 455, n. 7212, p. 461–461, 09 2008. doi: 10.1038/455461d.
- 124 BELL, G.; HEY, T.; SZALAY, A. Beyond the data deluge. *Science*, v. 323, n. 5919, p. 1297–1298, 2009. doi: 10.1126/science.1170411.
- 125 COHEN, A. M.; HERSH, W. R. A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, v. 6, n. 1, p. 57–71, 2005. doi: 10.1093/bib/6.1.57.
- 126 NENKOVA, A.; MCKEOWN, K. A survey of text summarization techniques. In: *Mining text data*. Berlin: Springer, 2012. p. 43–76.
- 127 AMANCIO, D. R.; ALTMANN, E. G.; RYBSKI, D.; OLIVEIRA JUNIOR, O. N.; COSTA, L. da F. Probing the statistical properties of unknown texts: application to the voynich manuscript. *PLoS ONE*, v. 8, n. 7, p. e67310, 2013. doi: 10.1371/journal.pone.0067310.
- 128 ANDRADE, M. A.; VALENCIA, A. Automatic extraction of keywords from scientific text: application to the knowledge domain of protein families. *Bioinformatics*, v. 14, n. 7, p. 600–607, 1998. doi: 10.1093/bioinformatics/14.7.600.
- 129 MANNING, C. D.; SCHÜTZE, H. *Foundations of statistical natural language processing*. Cambridge: MIT Press, 1999.
- 130 CARRETERO-CAMPOS, C.; BERNAOLA-GALVÁN, P.; CORONADO, A.; CARPENA, P. Improving statistical keyword detection in short texts: entropic and clustering approaches. *Physica A: statistical mechanics and its applications*, v. 392, n. 6, p. 1481 – 1492, 2013. doi: 10.1016/j.physa.2012.11.052.
- 131 AMANCIO, D. R.; ALUISIO, S. M.; OLIVEIRA JUNIOR, O. N.; COSTA, L. da. F. Complex networks analysis of language complexity. *Europhysics Letters*, v. 100, n. 5, p. 58002, 2012. doi: 10.1209/0295-5075/100/58002.
- 132 RIGO, G. V. *Modelagem de grãos confinados em invólucros utilizando redes complexas e métodos de imagem*. 2015. 155 p. Tese (Doutorado em Ciências) — Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.
- 133 MUNKRES, J. *Topology*. 2nd. ed. Upper Saddle River: Prentice-Hall, 2000.
- 134 TAKENS, F. Detecting strange attractors in turbulence. In: RAND, D. A.; YOUNG, L.-S. (Ed.). *Dynamical systems and turbulence*. Berlin: Springer, 1981, (Lecture notes in mathematics, v. 898). p. 366–381.
- 135 HARRIS, J. W.; STOCKER, H. Hausdorff dimension. In: _____. *Handbook of mathematics and computational science*. New York: Springer-Verlag, 1998. p. 113–114.

- 136 ESSEX, C.; NERENBERG, M. A. H. Fractal dimension: limit capacity or Hausdorff dimension? *American Journal of Physics*, v. 58, n. 10, p. 986–988, 1990. doi: 10.1119/1.16262.
- 137 MANDELBROT, B. *The fractal geometry of nature*. San Francisco: W.H. Freeman, 1982.
- 138 FLAKE, G. *The computational beauty of nature: computer explorations of fractals, chaos, complex systems, and adaptation*. Cambridge: MIT Press, 2000.
- 139 TRUDEAU, R. *Introduction to graph theory*. New York: Dover, 1993.
- 140 ERDOS, P.; HARARY, F.; TUTTE, W. On the dimension of a graph. *Mathematika*, v. 12, n. 2, p. 118–122, 1965.
- 141 DAILEY, D. P. Uniqueness of colorability and colorability of planar 4-regular graphs are np-complete. *Discrete Mathematics*, v. 30, n. 3, p. 289 – 293, 1980. doi: 10.1016/0012-365X(80)90236-8.
- 142 CSÁNYI, G.; SZENDROI, B. Fractal–small-world dichotomy in real-world networks. *Physical Review E*, v. 70, n. 1, p. 016122, 2004. doi: 10.1103/PhysRevE.70.016122.
- 143 GASTNER, M. T.; NEWMAN, M. E. The spatial structure of networks. *The European Physical Journal B*, v. 49, n. 2, p. 247–252, 2006. doi: 10.1140/epjb/e2006-00046-8.
- 144 SHANKER, O. Graph zeta function and dimension of complex network. *Modern Physics Letters B*, v. 21, n. 11, p. 639–644, 2007. doi: 10.1142/S0217984907013146.
- 145 SHANKER, O. Sharp dimension transition in a shortcut model. *Journal of Physics A: mathematical and theoretical*, v. 41, n. 28, p. 285001, 2008. doi: 10.1088/1751-8113/41/28/285001.
- 146 BULDYREV, S. V.; PARSHANI, R.; PAUL, G.; STANLEY, H. E.; HAVLIN, S. Catastrophic cascade of failures in interdependent networks. *Nature*, v. 464, n. 7291, p. 1025–1028, 2010. doi: 10.1038/nature08932.
- 147 LAMBIOTTE, R.; BLONDEL, V. D.; KERCHOVE, C. de; HUENS, E.; PRIEUR, C.; SMOREDA, Z.; DOOREN, P. V. Geographical dispersal of mobile communication networks. *Physica A: statistical mechanics and its applications*, v. 387, n. 21, p. 5317 – 5325, 2008. doi: 10.1016/j.physa.2008.05.014.
- 148 KOSMIDIS, K.; HAVLIN, S.; BUNDE, A. Structural properties of spatially embedded networks. *Europhysics Letters*, v. 82, n. 4, p. 48005, 2008. doi: 10.1209/0295-5075/82/48005.
- 149 LACASA, L.; GÓMEZ-GARDEÑES, J. Correlation dimension of complex networks. *Physical Review Letters*, v. 110, n. 16, p. 168703, 2013. doi: 10.1103/PhysRevLett.110.168703.
- 150 BACKES, A. R.; BRUNO, O. M. Shape classification using complex network and multi-scale fractal dimension. *Pattern Recognition Letters*, v. 31, n. 1, p. 44 – 51, 2010. doi: 10.1016/j.patrec.2009.08.007.
- 151 KONDOR, D.; MÁTRAY, P.; CSABAI, I.; VATTAY, G. Measuring the dimension of partially embedded networks. *Physica A: statistical mechanics and its applications*, v. 392, n. 18, p. 4160 – 4171, 2013. doi: 10.1016/j.physa.2013.04.046.

- 152 CHATTERJEE, S. *Regression analysis by example*. New York: Wiley, 2000.
- 153 BLONDEL, V. D.; GUILLAUME, J.-L.; HENDRICKX, J. M.; JUNGERS, R. M. Distance distribution in random graphs and application to network exploration. *Physical Review E*, v. 76, n. 6, p. 066101, 2007. doi: 10.1103/PhysRevE.76.066101.
- 154 DALL, J.; CHRISTENSEN, M. Random geometric graphs. *Physical Review E*, v. 66, n. 1, p. 016121, 2002. doi: 10.1103/PhysRevE.66.016121.
- 155 ALBERT, R.; JEONG, H.; BARABASI, A.-L. Error and attack tolerance of complex networks. *Nature*, v. 406, n. 6794, p. 378–382, 2000. doi: 10.1038/35019019.
- 156 BRINKHOFF, T. A framework for generating network-based moving objects. *Geoinformatica*, v. 6, n. 2, p. 153–180, 2002. doi: 10.1023/A:1015231126594.
- 157 COSTA, L. da F.; TRAVENCOLO, B. A. N.; VIANA, M. P.; STRANO, E. On the efficiency of transportation systems in large cities. *Europhysics Letters*, v. 91, n. 1, p. 18003, 2010. doi: 10.1209/0295-5075/91/18003.
- 158 XIAO, Y.; XIONG, M.; WANG, W.; WANG, H. Emergence of symmetry in complex networks. *Physical Review E*, v. 77, n. 6, p. 066108, 2008. doi: 10.1103/PhysRevE.77.066108.
- 159 MACARTHUR, B. D.; SÁNCHEZ-GARCÍA, R. J.; ANDERSON, J. W. Symmetry in complex networks. *Discrete Applied Mathematics*, v. 156, n. 18, p. 3525–3531, 2008. doi: 10.1016/j.dam.2008.04.008.
- 160 ESTRADA, E.; PEÑA, J. de la; HATANO, N. Walk entropies in graphs. *Linear Algebra and its Applications*, v. 443, n. 15, p. 235–244, 2014. doi: 10.1016/j.laa.2013.11.009.
- 161 ZHAO, C.; WANG, W.-X.; LIU, Y.-Y.; SLOTINE, J.-J. Intrinsic dynamics induce global symmetry in network controllability. *Scientific Reports*, v. 5, p. 8422, 2015. doi: 10.1038/srep08422.
- 162 ROSSI, L.; TORSELLO, A.; HANCOCK, E. R.; WILSON, R. C. Characterizing graph symmetries through quantum jensen-shannon divergence. *Physical Review E*, v. 88, n. 3, p. 032806, 2013. doi: 10.1103/PhysRevE.88.032806.
- 163 ESTRADA, E. *The structure of complex networks: theory and applications*. New York: Oxford University Press, 2012.
- 164 COSTA, L. da F. Voronoi and fractal complex networks and their characterization. *International Journal of Modern Physics C*, v. 15, n. 1, p. 175–183, 2004. doi: 10.1142/S0129183104005619.
- 165 OKABE, A.; BOOTS, B.; SUGIHARA, K.; CHIU, S. N. *Spatial tessellations: concepts and applications of Voronoi diagrams*. New York: Wiley, 2009.
- 166 COMIN, C. H.; SILVA, F. N.; COSTA, L. da F. A framework for evaluating complex networks measurements. *Europhysics Letters*, v. 110, n. 6, p. 68002, 2015. doi: 10.1209/0295-5075/110/68002.
- 167 PERON, T. K. D.; COSTA, L. da F.; RODRIGUES, F. A. The structure and resilience of financial market networks. *Chaos*, v. 22, n. 1, p. 013117, 2012. doi: 10.1063/1.3683467.

- 168 GUHATHAKURTA, K. Nonlinear dynamics of stock markets during critical periods. In: ABERGEL, F.; AOYAMA, H.; CHAKRABARTI, B. K.; CHAKRABORTI, A.; GHOSH, A. (Ed.). *Econophysics and data driven modelling of market dynamics*. Berlin: Springer, 2015. p. 143–165.
- 169 SANDHU, R.; GEORGIOU, T.; TANNENBAUM, A. Market fragility, systemic risk, and ricci curvature. 2015. Disponível em: <<http://arxiv.org/abs/1505.05182>>. Acesso em: 10.6.2015.
- 170 WEAVER, W. Science and complexity. *American Scientist*, v. 36, n. 4, p. 536–544, 1948. Disponível em: <<http://www.jstor.org/stable/27826254>>. Acesso em: 10.6.2015.
- 171 BERTSEKAS, D. P.; TSITSIKLIS, J. N. *Introduction to probability*. Belmont: Athena Scientific, 2002.
- 172 MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; IAMASHITA, P.; SILVA, F. N.; COSTA, L. da F.; SILVA, A. V.; CASTRO, L. H. M.; WEN, H.-T. Community structure analysis of transcriptional networks reveals distinct molecular pathways for early- and late-onset temporal lobe epilepsy with childhood febrile seizures. *PLoS ONE*, v. 10, n. 5, p. e0128174, 2015. doi: 10.1371/journal.pone.0128174.
- 173 MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; SILVA, F. N.; COSTA, L. da F. Methods for gene coexpression network visualization and analysis. In: PASSOS, G. A. (Ed.). *Transcriptomics in health and disease*. Berlin: Springer, 2014. cap. 4, p. 79–94. Disponível em: <http://dx.doi.org/10.1007/978-3-319-11985-4_4>.
- 174 MOREIRA-FILHO, C. A.; BANDO, S. Y.; BERTONHA, F. B.; SILVA, F. N.; COSTA, L. da F.; CARNEIRO-SAMPAIO, M. Thymus gene coexpression networks: a comparative study in children with and without down syndrome. In: PASSOS, G. A. (Ed.). *Transcriptomics in health and disease*. Berlin: Springer, 2014. cap. 7, p. 123–136.
- 175 SILVA, F. N.; COSTA, L. da F. Local dimension of complex networks. 2012. Disponível em: <<http://arxiv.org/abs/1209.2476>>. Acesso em: 10.6.2015.
- 176 WU, Z.; LU, X.; DENG, Y. Image edge detection based on local dimension: A complex networks approach. *Physica A: statistical mechanics and its applications*, v. 440, n. 1, p. 9 – 18, 2015. doi: 10.1016/j.physa.2015.07.020.
- 177 PU, J.; CHEN, X.; WEI, D.; LIU, Q.; DENG, Y. Identifying influential nodes based on local dimension. *Europhysics Letters*, v. 107, n. 1, p. 10010, 2014. doi: 10.1209/0295-5075/107/10010.
- 178 BAKER, G.; GOLLUB, J. *Chaotic dynamics: an introduction*. Cambridge: Cambridge University Press, 1996.

APÊNDICE A – Recursos computacionais e complexidade dos algoritmos

Neste apêndice são apresentadas a complexidade e tempos de execução dos principais algoritmos implementados (tabela 5) neste trabalho e os softwares utilizados e desenvolvidos neste trabalho (tabela 6).

Tabela 5 – Complexidade e tempos de cálculo das metodologias usadas neste trabalho. A análise de tempo de execução foi conduzida em um processador Intel Core i7 2.3 Ghz com 8GB de RAM. Foram consideradas 4 redes ER de conectividade média $\langle k \rangle = 3$ e $\langle k \rangle = 6$ assim como de tamanhos $|\mathcal{V}| = 1000$ $|\mathcal{V}| = 10000$. Os tempos para o algoritmo FRMOD foram considerados a partir do início da simulação até a estabilização da rede, enquanto aqueles baseados em níveis concêntricos foram considerados até $r = 5$.

Operação	Complexidade	Tempo $ \mathcal{V} = 1000$ $\langle k \rangle = 3$	Tempo $ \mathcal{V} = 1000$ $\langle k \rangle = 6$	Tempo $ \mathcal{V} = 10000$ $\langle k \rangle = 3$	Tempo $ \mathcal{V} = 10000$ $\langle k \rangle = 6$
Dimensão local e multiescala para nível r	$O(\langle k \rangle^r)$	0.04s	0.06s	1.3s	3.3s
Simetria concêntrica para nível r	$O(\langle k \rangle^r)$	0.04s	0.11s	0.71s	4.8s
Algoritmo de visualização FRMOD	$O(\mathcal{V} \log \mathcal{V} + \mathcal{E})$	5.3s	8.1s	13.9s	21.3s

Fonte: Elaborada pelo autor

Tabela 6 – Softwares usados ou desenvolvidos durante a execução deste trabalho.

Ferramentas de Desenvolvimento		
Clang/LLVM	v6.1/602	Conjunto de compiladores C e Objective-C. http://clang.llvm.org/
Python	v2.7	Interpretador da linguagem <i>script</i> de mesmo nome. http://www.python.org/
Xcode	v6.4	Ambiente de desenvolvimento em C e Objective-C http://developer.apple.com/technology/xcode.html
Ferramentas de Publicação		
Latex	-	Software gerador de documentos para publicação. http://www.latex-project.org/
TexShop	-	Editor de documentos Latex. http://www.uoregon.edu/~koch/texshop/
Ambientes Operacionais		
Mac OS X	v10.10	Sistema Operacional UNIX baseado em Darwin. http://www.apple.com/macosx/
Windows	8	Sistema Operacional Comercial. http://www.microsoft.com/windows/
Softwares de Análise de Redes Complexas		
Cytoscape	-	Ferramenta de obtenção de propriedades e visualização de redes complexas. http://www.cytoscape.org/
Pajek (104)	v1.24	Ferramenta de obtenção de propriedades de redes complexas. http://pajek.imfm.si/
Concentric	v0.8b	Software para extrair propriedades concêntricas de redes complexas. http://cyvision.ifsc.usp.br/~bant/hierarchical/
Softwares desenvolvidos		
ComNetKit	v1.0	Biblioteca de códigos para redes complexas em Objective-C/Cocoa. não disponível no momento.
Network 3D	v1.0	Software gerador de visualização interativas de redes complexas. http://cyvision.ifsc.usp.br/Cyvision/?page=SOFTWARE&subpage=NETWORKS3D
CVSymmetry	v0.2	Software para calcular a simetria de redes complexas. https://github.com/filipinascimento/CVSymmetry
CVAccessibility	v0.8	Software para calcular a acessibilidade de redes complexas. https://github.com/filipinascimento/CVAccessibility
CVDimension	v0.2	Software para calcular a dimensão local e multiescala de redes complexas. https://github.com/filipinascimento/CVDimension
Network 3D Web	v0.1	Visualizador de redes voltado para a web (versão preliminar). http://cyvision.ifsc.usp.br/networkstools/

Fonte: Elaborada pelo autor