

Mateus Tomazini Kinasz

CRITICALIDADE NO CONECTOMA

Dissertação submetida ao Programa
de Pós-Graduação em Física para a
obtenção do Grau de Mestre em Física.
Orientador: Prof. Dr. Marcelo H. R.
Tragtenberg

Florianópolis

2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Kinasz, Mateus Tomazini
Criticalidade no Conectoma / Mateus Tomazini
Kinasz ; orientador, Marcelo Henrique Romano
Tragtenberg, 2018.
94 p.

Dissertação (mestrado) - Universidade Federal de
Santa Catarina, Centro de Ciências Físicas e
Matemáticas, Programa de Pós-Graduação em Física,
Florianópolis, 2018.

Inclui referências.

1. Física. 2. Conectoma. 3. Transição de Fase. 4.
Criticalidade. I. Tragtenberg, Marcelo Henrique
Romano. II. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Física. III. Título.

Mateus Tomazini Kinasz

CRITICALIDADE NO CONECTOMA

Esta Dissertação foi julgada aprovada para a obtenção do Título de “Mestre em Física”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Física.

Florianópolis, 11 de setembro 2018.

Prof. Dr. Ivan
Coordenador do Curso

Banca Examinadora:

Prof. Dr. Marcelo H. R. Tragtenberg
Presidente

Prof. Dr. Nilton da Silva Branco
Universidade Federal de Santa Catarina

Prof. Dr. Nuno Crokidakis
Universidade Federal Fluminense

Dedico este trabalho ao sistema operacional Linux e ao meu computador Aspire-5733Z que fez a maior parte do trabalho.

AGRADECIMENTOS

Agradeço a CAPES pelo apoio financeiro, ao Marcelo Tragtenberg por me orientar, ao Maurício Girardi-Schappo pela sugestão do tema deste trabalho, ao Dr. Dante R. Chialvo por ceder a rede utilizada aqui, a minha família, meus amigos e minha companheira Júlia pelo apoio e suporte dado.

I should have to believe that the logic of thought is quite arbitrary—that it is a purely and strictly human invention without any basis in the physical universe. While it is true, as I have already shown, that we do project logical patterns (nets, grids, and other types of calculus) upon the wiggly physical world—which can be confusing if we do not realize what we are doing—nevertheless, these patterns do not come from *outside* the world. They have something to do with the design of the human nervous system, which is definitely in and of the world.

(Alan Watts, 1966)

RESUMO

Este trabalho descreve brevemente certas ferramentas da teoria dos grafos, duas técnicas de imagem usadas na aquisição de dados para a rede chamada de conectoma, as transições de fase em particular as de segunda ordem com seus expoentes críticos e como medi-los em sistemas finitos. Também simula a percolação e ataque no conectoma, obtendo os mesmos expoentes críticos de uma rede livre de escala. Outra simulação é a do modelo de limiar linear, que também apresentou transição de fase de segunda ordem.

Palavras-chave: Conectoma. Transição de Fase. Criticalidade.

ABSTRACT

This work briefly describes some tools of graph theory, two image techniques used in data acquisition for the network called connectome, the phase transitions especially the second order with the critical exponents and how to measure them in finite systems. It also simulates the percolation and attack in the connectome, obtaining the same critical exponents of a free-scale network. Another simulation in the conectoma is with the linear threshold model, which also presented second order phase transition.

Keywords: Connectome. Phase Transition. Criticality.

LISTA DE FIGURAS

| | | |
|-----------|---|----|
| Figura 1 | Diagrama das pontes de Königsberg feito por Euler no artigo original de 1736..... | 28 |
| Figura 2 | Representação de uma rede ou grafo não direcionado... | 29 |
| Figura 3 | Componentes de uma rede..... | 32 |
| Figura 4 | Representação de uma árvore..... | 32 |
| Figura 5 | Grafo mostrando o grau de cada vértice..... | 33 |
| Figura 6 | Representação do conectoma..... | 41 |
| Figura 7 | Localização das 998 ROIs do conectoma recobrando o córtex cerebral..... | 42 |
| Figura 8 | Redução de dados de um voxel em 6 dimensões obtido por DSI..... | 44 |
| Figura 9 | Matriz de adjacência binária do conectoma..... | 45 |
| Figura 10 | Passos para a criação do conectoma..... | 46 |
| Figura 11 | Distribuição de graus do conectoma junto com as distribuições binomial, de Poisson e uma lei de potência para cauda.... | 47 |
| Figura 12 | Distribuição de graus do conectoma em log-log e melhor ajuste de lei de potência..... | 48 |
| Figura 13 | Distribuição de pesos das arestas do conectoma..... | 49 |
| Figura 14 | Distribuição de graus ponderados do conectoma..... | 50 |
| Figura 15 | Exemplo de uma transição de fase de segunda ordem na percolação..... | 56 |
| Figura 16 | Processo de percolação em uma rede..... | 57 |
| Figura 17 | Retirada de um vértice numa árvore formando <i>clusters</i> isolados..... | 63 |
| Figura 18 | Cálculo numérico da percolação de sítios para o conectoma..... | 72 |
| Figura 19 | Cálculo numérico da percolação de ligação para o conectoma..... | 72 |
| Figura 20 | Corte no eixo y do conectoma obtendo duas sub-redes com 30% de vértices do original..... | 73 |
| Figura 21 | Parâmetro de ordem ϕ e susceptibilidade em função de p para percolação de sítios..... | 76 |
| Figura 22 | Parâmetro de ordem ϕ e susceptibilidade em função de p para percolação de ligação..... | 77 |

| | |
|---|----|
| Figura 23 Ajuste de funções gaussianas para as susceptibilidades de percolação de sítios..... | 78 |
| Figura 24 Ajuste de funções gaussianas para as susceptibilidades de percolação de ligação..... | 78 |
| Figura 25 Gráfico log-log $p_c(N) - p_c$ versus tamanho da rede N para percolação de sítios..... | 79 |
| Figura 26 Gráfico log-log $p_c(N) - p_c$ versus tamanho da rede N para percolação de ligação..... | 80 |
| Figura 27 Gráfico log-log do parâmetro de ordem no pseudo-ponto crítico $S(p_c(N))$ versus o tamanho da rede N | 80 |
| Figura 28 Gráfico log-log do parâmetro de ordem no pseudo-ponto crítico $B(p_c(N))$ versus o tamanho da rede N | 81 |
| Figura 29 Relação entre as parâmetro de ordem da percolação de sítio e de ligação..... | 82 |
| Figura 30 Ataque no conectoma..... | 83 |
| Figura 31 Ataque de arestas no conectoma..... | 84 |
| Figura 32 Parâmetro de ordem do modelo LTM em função de θ .. | 87 |
| Figura 33 Susceptibilidade do modelo LTM em função de θ e ajuste de uma função gaussiana..... | 88 |
| Figura 34 Gráfico log-log $\theta_c(N) - \theta_c$ versus tamanho da rede N para o modelo LTM..... | 89 |
| Figura 35 Gráfico log-log do parâmetro de ordem no pseudo-ponto crítico $A(\theta_c(N))$ versus o tamanho da rede N | 89 |

LISTA DE ABREVIATURAS E SIGLAS

| | | |
|------|--|----|
| DSI | Imagem por difusão por ressonância magnética, do inglês, <i>diffusion spectrum imaging</i> | 25 |
| ROIs | Regiões de interesse do córtex cerebral, do inglês, <i>regions of interest</i> | 25 |
| LTM | Modelo de limiar linear, do inglês <i>Linear Threshold Model</i> | 25 |
| ODF | função de distribuição de orientação, do inglês <i>orientation distribution function</i> | 43 |
| FSS | Escala de tamanho finito, do inglês <i>finite size scaling</i> . . | 53 |

LISTA DE SÍMBOLOS

| | | |
|---------------------------|---|----|
| n | Número de vértices de um grafo | 28 |
| \mathbb{A} | Matriz de adjacência de uma rede | 29 |
| A_{ij} | Elemento de matriz da matriz de adjacência | 29 |
| w_{ij} | Peso associado a aresta que liga os vértice i e j | 30 |
| k_i | Grau de um vértice i | 30 |
| m | Número de arestas em um grafo | 30 |
| $\langle k \rangle$ | Grau médio de uma rede | 31 |
| d | Diâmetro de uma rede | 31 |
| p_k | Distribuição de probabilidade de graus de uma rede | 32 |
| $G(n, m)$ | Ensemble de redes aleatórias com n vértices e m arestas . | 35 |
| b | Probabilidade de dois vértices estarem conectados numa rede aleatória | 35 |
| $G(n, b)$ | Ensemble de redes aleatórias com n vértices e probabilidade b de conexão entre vértices | 35 |
| $\binom{n}{k}$ | Coeficiente binomial | 35 |
| $\langle m \rangle$ | Número médio de arestas no ensemble $G(n, b)$ | 36 |
| $\zeta(\lambda)$ | Função zeta de Riemann | 39 |
| k_{min} | Valor mínimo de grau onde vale uma lei de potência para distribuição de graus | 39 |
| $\zeta(\lambda, k_{min})$ | Função zeta de Riemann generalizada | 39 |
| C_i | Coeficiente de agrupamento local de um vértice i | 47 |
| C | Coeficiente de agrupamento global de uma rede | 47 |
| σ_C | Desvio padrão do coeficiente de agrupamento do conectoma | 47 |
| l | Comprimento médio de uma rede | 48 |
| σ_{SW} | Coeficiente de mundo pequeno | 49 |
| $P(w_{ij})$ | Distribuição de pesos de arestas | 49 |
| k_w | Grau ponderado de um vértice | 49 |
| ξ | Comprimento de correlação de um sistema | 51 |
| t | Temperatura reduzida | 51 |
| T_c | Temperatura crítica de um sistema numa transição de fase | 51 |
| ρ | Parâmetro de ordem de um sistema | 52 |
| κ_T | Compressibilidade isotérmica, proporcional à segunda derivada de um potencial termodinâmico | 52 |

| | | |
|---------------------|--|----|
| L | Comprimento linear de um sistema | 53 |
| $\tilde{\chi}(x)$ | Função de escala da susceptibilidade | 53 |
| p | Probabilidade de ocupação | 55 |
| p_c | Ponto crítico de uma percolação | 55 |
| $g_0(z)$ | Função geradora de uma distribuição de graus p_k | 57 |
| $g'_0(z)$ | Derivada da função geradora $g_0(z)$ | 58 |
| $\delta(a, b)$ | Delta de Kronecker | 59 |
| σ_k^2 | Variância de uma distribuição de grau | 60 |
| q_k | Distribuição de excesso de graus de uma rede | 61 |
| $g_1(z)$ | Função geradora do excesso de graus q_k | 61 |
| π_s | Probabilidade de um vértice escolhido aleatoriamente fazer parte de um <i>cluster</i> não gigante de tamanho s | 62 |
| $h_0(z)$ | Função geradora de π_s | 62 |
| ρ_s | Probabilidade de que um vértice no extremo de uma aresta pertença a um <i>cluster</i> pequeno de tamanho s depois que essa aresta é removida. | 62 |
| $P(s k)$ | Probabilidade após um vértice i ser retirado, seus k vizinhos pertençam a pequenos <i>clusters</i> de tamanho s_i | 63 |
| $P(s-1 k)$ | Probabilidade de um vértice i pertencer a um pequeno <i>cluster</i> de tamanho s dado que ele possui grau k | 63 |
| $h_1(z)$ | Função geradora de ρ_s | 64 |
| $\langle r \rangle$ | Tamanho médio do <i>cluster</i> o qual um vértice escolhido aleatoriamente faz parte | 66 |
| $\langle s \rangle$ | Tamanho médio de um pequeno <i>cluster</i> | 68 |
| u | Probabilidade de um vértice não pertencer ao <i>cluster</i> gigante | 69 |
| S | Parâmetro de ordem da percolação de sítios | 69 |
| B | Parâmetro de ordem da percolação de ligação | 69 |
| $Z(p)$ | Maior <i>cluster</i> da rede em função da probabilidade de ocupação p | 74 |
| $\chi_S(p)$ | Susceptibilidade da percolação de sítios em função da probabilidade de ocupação | 74 |
| Zr | Maior <i>cluster</i> da rede em função da quantidade r de arestas ocupadas | 74 |
| χ_B | Susceptibilidade da percolação de ligação | 75 |
| $p_c(N)$ | Pseudo ponto crítico na percolação para uma rede com tamanho N | 77 |

| | | |
|----------|--|----|
| v | Variável de estado no modelo LTM | 85 |
| N_i | Conjunto de vizinhos do vértice i | 86 |
| θ | Parâmetro do modelo LTM | 86 |
| D | Densidade de sítios ativados no modelo LTM | 86 |
| A | Parâmetro de ordem do modelo LTM | 86 |

SUMÁRIO

| | | |
|----------|--|----|
| 1 | INTRODUÇÃO | 25 |
| 2 | TEORIA DOS GRAFOS | 27 |
| 2.1 | REPRESENTAÇÕES DE GRAFO | 28 |
| 2.2 | GRAU | 30 |
| 2.3 | COMPRIMENTO, DISTÂNCIA E DIÂMETRO | 31 |
| 2.4 | COMPONENTES E ÁRVORES | 31 |
| 2.5 | DISTRIBUIÇÃO DE GRAUS | 32 |
| 3 | REDES | 35 |
| 3.1 | REDES ALEATÓRIAS | 35 |
| 3.1.1 | Distribuição de graus das redes aleatórias | 37 |
| 3.1.2 | Diametro das redes aleatórias | 38 |
| 3.2 | REDES LIVRE DE ESCALA | 38 |
| 3.2.1 | Distribuição de graus das redes livre de escala | 38 |
| 3.2.2 | Diâmetro das rede livre de escala | 39 |
| 4 | CONECTOMA | 41 |
| 4.1 | MEDIDAS DA REDE DO CONECTOMA | 44 |
| 5 | TRANSIÇÃO DE FASE E EXPOENTES CRÍTICOS 51 | |
| 5.1 | SISTEMAS FINITOS | 53 |
| 6 | PERCOLAÇÃO | 55 |
| 6.1 | FUNÇÃO GERADORA | 57 |
| 6.2 | NORMALIZAÇÃO E MOMENTOS | 58 |
| 6.3 | POTÊNCIAS DA FUNÇÃO GERADORA | 58 |
| 6.4 | DISTRIBUIÇÃO DE EXCESSO DE GRAU | 59 |
| 6.5 | FUNÇÃO GERADORA DO EXCESSO DE GRAUS | 61 |
| 6.6 | FUNÇÃO GERADORA PARA PEQUENOS <i>CLUSTERS</i> E MÉTODO DA CAVITAÇÃO | 62 |
| 6.7 | <i>CLUSTER</i> GIGANTE | 65 |
| 6.8 | TAMANHO MÉDIO DO <i>CLUSTER</i> O QUAL UM VÉRTICE ESCOLHIDO ALEATORIAMENTE FAZ PARTE..... | 66 |
| 6.9 | TAMANHO MÉDIO DE UM PEQUENO <i>CLUSTER</i> | 68 |
| 6.10 | EQUAÇÃO DO <i>CLUSTER</i> GIGANTE | 69 |
| 6.11 | CALCULO NUMÉRICO | 71 |
| 6.12 | FSS NO CONECTOMA | 72 |
| 6.13 | SIMULAÇÃO NO CONECTOMA | 74 |
| 7 | MODELO DE LIMIAR LINEAR | 85 |
| 8 | CONCLUSÃO E PERSPECTIVAS FUTURAS .. | 91 |
| | REFERÊNCIAS | 93 |

1 INTRODUÇÃO

Neste trabalho estudei o conectoma (HAGMANN et al, 2008), uma rede gerada através de técnicas não invasivas de ressonância magnética convencional e imagem de difusão por ressonância magnética, DSI (do inglês *diffusion spectrum imaging*) recriando a estrutura do córtex humano discretizado em pequenas regiões macroscópicas com área de aproximadamente $1,5\text{cm}^2$, chamadas de regiões de interesse (ROIs, do inglês, *regions of interest*), juntamente com o trato nervoso. O conectoma por tratar do cérebro de uma maneira mais simples na forma de grafo pode ser estudando usando a teoria dos grafos, campo da matemática que estuda as redes ou grafos. Dois estudos foram feitos usando métodos das transições de fase, no primeiro retiram-se ROIs ou as as conexões entre as ROIs a fim de medir o limite em que a rede fica desconectada em vários pequenos grupos desconectados de ROIs. Este processo é conhecido na física estatística como percolação e é um modelo simples para estudar doenças neurodegenerativas como Alzheimer (BERNDT; RUPPIN; REGGIA, 1996, p. 404), doença que causa a perda de neurônios e sinapses no córtex cerebral e em algumas regiões subcorticais.

Fiz algumas comparações para descobrir a natureza topológica do conectoma com as redes aleatórias, redes livres de escala e as redes de 'mundo pequeno' (WATTS e STROGATZ, 1998), as quais são amplamente utilizadas como modelos de redes complexas. Uma das comparações foram os expoentes críticos obtidos na percolação, o qual demonstra que o conectoma está mais próximo de uma rede livre de escala. Com a simulação de ataque (ALBERT; BARABÁSI; JEONG, 2000), o qual são retirados primeiramente as ROIs mais conectadas, pude concluir que o conectoma também pode ser descrito como uma rede de 'mundo pequeno' (ACHARD et al, 2006).

O segundo estudo foi feito usando o modelo de limiar linear (LTM, do inglês *Linear Threshold Model*) (GRANOVETTER, 1978), o qual simula uma difusão no conectoma (MIŠIĆ et al, 2015). O modelo inicia com uma perturbação em uma das ROIs e a evolução é dependente da proporção dos estados dos vizinhos. O modelo LTM também apresentou uma transição de fase de segunda ordem e uma possível pesquisa a ser realizada é verificar a correlação do ponto crítico θ_c e dos expoentes críticos da transição de fase do modelo LTM em redes semelhantes ao conectoma de grupos saudáveis e doentes, podendo ser possível caracterizar certos tipos de doença no cérebro.

Para configurar uma transição de fase de segunda ordem com os expoentes críticos em um sistema finito é necessário o uso de redes com tamanhos diferentes e como o grafo experimental empregado aqui tem um tamanho fixo, utilizei sub-redes derivadas do grafo original através de cortes em um sistema de coordenadas cartesiano. O próximo passo é utilizar ferramentas da teoria de grupos de renormalização para redimensionar a rede de forma rigorosa (GARCIA-PÉREZ; BOGUÑA; SERRANO, 2018) e obter os pontos críticos e os expoentes críticos para os dois estudos.

2 TEORIA DOS GRAFOS

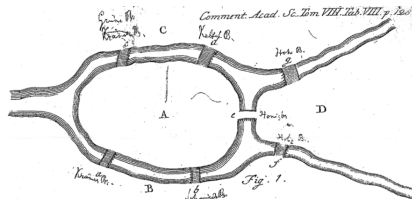
A teoria dos grafos é um ramo da matemática o qual lida com os grafos ou redes e é aplicada em várias áreas da Física, da Biologia (MASHAGHI; RAMAZANPOUR; KARIMPOUR V., 2004), da Sociologia (FELD, 1991), Química (BALABAN, 1985) e Neurociência (HAGMANN et al, 2008). Este capítulo inicia-se com uma breve história da teoria dos grafos e expõe alguns termos usados ao longo do trabalho.

O início da teoria é comumente atribuído ao artigo do matemático Leonhard Euler de 1736 (NEWMAN; WATTS; BARABÁSI, 2011, p. 2) (HARRY, 2018, p. 1). No trabalho, Euler fala sobre a cidade de Königsberg ¹, antiga Prússia, o qual há uma ilha que faz o rio Prególia se ramificar em dois braços. Nesta região haviam sete pontes e o problema, chamado de Sete pontes de Königsberg, consistia em obter uma rota, o qual um viajante devesse passar apenas uma vez por todas as pontes. Euler substituiu o mapa da cidade por um diagrama (fig.1), representando as porções de terra por letras maiúsculas e as pontes por letras minúscula (BIGGS; LLOYD; WILSON, 1976, p. 1). Na teoria dos grafos o que seria as porções de terra é chamado de **vértices**, as pontes de **arestas** e a rota do viajante de **caminho Euleriano**. O problema consiste em dizer se existe um caminho Euleriano no grafo. O matemático provou que para existir é necessário que, uma vez que o caminho precisa entrar e sair do vértice, exceto o início e o fim, o grafo pode no máximo conter dois vértices com grau ímpar. O **grau** de um vértice é um número de arestas ligado a ele, no caso do artigo a quantidade de pontes ligando as porções de terra. Como todos os quatro vértices no grafo de Königsberg possuem grau ímpar não existe um caminho (NEWMAN; WATTS; BARABÁSI, 2011, p. 2).

Um grafo é também conhecido como rede e neste trabalho os termos serão usados indiscriminadamente. Na sua forma mais simples, uma rede ou grafo, é nada mais que um conjunto de elementos discretos, os vértices, e um conjunto de conexões, as arestas, que conectam os elementos. Normalmente uma aresta conecta dois vértices distintos, mas pode ocorrer da aresta ligar o vértice a si mesmo, neste caso a aresta é chamada de **laço** ou então mais de uma aresta conecta os mesmos dois vértices e as arestas coletivamente ligadas chamam-se **arestas paralelas** (NEWMAN, 2010, p. 110). Um grafo também pode ser **direcionado** ou **não direcionado**, no caso direcionado as arestas possuem um único sentido ligando um vértice a outro em uma certa

¹A cidade hoje se chama Kaliningrado e fica na Rússia

Figura 1 – Diagrama das pontes de Königsberg feito por Euler no artigo original de 1736.



Fonte: Artigo de Leonhard Euler, "Solutio problematis ad geometriam situs pertinentis" publicado 1941 no *Commentarii academiae scientiarum Petropolitanae* 8. Disponível em: <<http://eulerarchive.maa.org>>.

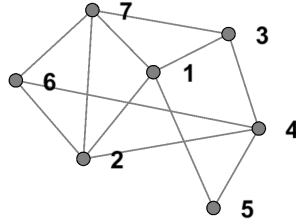
ordem e no não direcionado a aresta apenas liga os vértices sem direção preferencial como no caso das pontes de Königsberg.

2.1 REPRESENTAÇÕES DE GRAFO

As redes ou grafos não direcionados são desenhados colocando cada vértice em um ponto e representando as arestas por curvas que ligam os pontos (fig. 2). Para redes direcionadas utilizam-se vetores para representar uma aresta direcionada, porém esse tipo de rede não é usado aqui. Os vértices e arestas podem ser respectivamente pessoas e amigos, computadores e linhas de comunicação, substâncias químicas e reagentes, artigos científicos e citações, regiões do cérebro e conexões neuronais, e por abstrair os detalhes do problemas, a teoria dos grafos é capaz de descrever importantes características topológicas que seriam impossíveis se todos os detalhes fossem mantidos (NEWMAN; WATTS; BARABÁSI, 2011, p. 3).

Embora representar as redes como imagens seja interessante à medida que há muitos vértices ou arestas pode ser muito complicado para o olho compreender e sua utilidade fica limitada. Por isso é preciso representar matematicamente uma rede, uma delas é a **lista de adjacência**. Considere uma rede não direcionada com n vértices e para cada um rotula-se um número inteiro de 1 a n (fig. 2). Não importa qual vértice tem qual rótulo apenas que ele deve ser único, assim pode-se usar os rótulos para se referir a qualquer vértice sem ambiguidade. Denotando as arestas ligando os vértices i e j por (i, j) , toda a rede pode ser especificada pelo dado valor de n e a lista de todas as arestas. Para a figura 2, $n = 7$ vértices e as arestas são:

Figura 2 – Representação de uma rede ou grafo não direcionado



Cada ponto representa um vértice e cada curva uma aresta. Os números são os rótulos de cada vértice.

Fonte: Criado pelo autor com o software Gephi.

$(1,2), (1,3), (1,5), (1,7), (2,4), (2,6), (2,7), (3,4), (3,7), (4,5), (4,6)$ e $(6,7)$.

Uma segunda representação é a chamada **matriz de adjacência**. A matriz de adjacência \mathbb{A} de uma rede simples, isto é, sem laços ou arestas paralelas, possui os elementos A_{ij} tal que:

$$A_{ij} = \begin{cases} 1 & \text{se existe uma aresta entre os vértice } i \text{ e } j \\ 0 & \text{outro caso} \end{cases} \quad (2.1)$$

Novamente pode-se representar a figura 2 agora pela matriz de adjacência:

$$\mathbb{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Vale notar que esta matriz possui a diagonal principal nula, pois não há laços e também que é simétrica porque o grafo é não direcionado portanto os elementos A_{ij} e A_{ji} são iguais. É possível representar laços e ou arestas paralelas com matriz de adjacência. Arestas paralelas são representadas pelo elemento de matriz correspondentes A_{ij} igual a multiplicidade da aresta, por exemplo uma aresta dupla entre os vértices i e j é representado por $A_{ij} = A_{ji} = 2$. Um laço no vértice i é representado pelo elemento de matriz A_{ii} na diagonal da matriz igual a 2, não 1, porque o laço começa no vértice i e termina no vértice i , ou seja, ambas as extremidades estão conectadas ao vértice i . Muitos

dos resultados matemáticos relativos à matriz de adjacência funcionam igualmente bem para redes com ou sem laços, mas apenas fazendo com que os elementos da diagonal da matriz sejam iguais a 2 em vez de 1 (ou caso existam laços paralelos, múltiplos de 2) (NEWMAN, 2010, p. 112).

Uma rede ou grafo também pode ser **ponderado**, tendo suas arestas uma força ou peso, isto é, um valor associado, geralmente um número real. O valor associado as arestas pode ser por exemplo na rede da Internet, a quantidade de dados que flui ao longo dela, numa rede de companhia área pode ser a distância entre as regiões ou numa rede do cérebro pode ser a densidade de conexões. Uma rede ponderada também é representada pela matriz de adjacência tendo seus elementos de matriz o peso associado a aresta, $A_{ij} = w_{ij}$ sendo w_{ij} o peso da aresta que liga os vértice i e j (NEWMAN, 2010, p. 113).

2.2 GRAU

Como já mencionado o grau de um dado vértice é quantidade de arestas ligadas a ele. Neste trabalho denota-se o grau de um vértice i por k_i . Para um grafo não direcionado com n vértices o grau pode ser escrito em termos da matriz de adjacência como:

$$k_i = \sum_{j=1}^n A_{ij} \quad . \quad (2.2)$$

A expressão vale para laços, porém somente se usar corretamente o elemento de matriz igual a 2. Cada aresta em um grafo não direcionado possui duas extremidades e se há m arestas no total então há $2m$ extremidades nas arestas. O numero de extremidades é igual a soma dos graus de todos os vértices, portanto:

$$2m = \sum_{i=1}^n k_i \quad , \quad (2.3)$$

ou

$$m = \frac{1}{2} \sum_{i=1}^n k_i = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} \quad . \quad (2.4)$$

O grau médio de um vértice numa rede não direcionada é:

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^n k_i = \frac{2m}{n} \quad , \quad (2.5)$$

o qual foi usada a expressão (2.3) (NEWMAN, 2010, p. 134).

2.3 COMPRIMENTO, DISTÂNCIA E DIÂMETRO

Caminho em uma rede é qualquer sequência de vértices tal que todos os consecutivos pares de vértices na sequência estejam conectados por uma aresta, ou seja, a rota através da rede que passa de vértice em vértice por meio de uma aresta. O comprimento de um caminho é o número de arestas que atravessam o caminho. As arestas podem ser atravessadas mais de uma vez e devem ser contadas separadas sempre que são atravessadas. Define-se também o comprimento médio de uma rede como a média de todos os comprimentos de um caminho da rede (NEWMAN, 2010, p. 136).

Distância Geodésica, também chamada somente de distância, é o menor comprimento entre dois vértices em um grafo. É possível que não exista uma distância geodésica entre dois vértices se eles não estão conectados na rede, isto é, eles estão em diferentes **componentes** (veja na seção 2.4). Neste caso diz-se que a distância geodésica entre os vértices é infinita, porém é apenas uma convenção e não significa muito além do fato dos vértices não estarem conectados. As distâncias não são necessariamente únicas e é possível ter duas ou mais distâncias com o mesmo tamanho ligando um dado par de vértices (NEWMAN, 2010, p. 139).

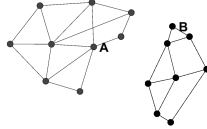
O diâmetro de um grafo ou rede, representado pela letra d , é a maior distância geodésica entre todos os pares de vértices na rede, dado que existe um caminho, isto é, estão no mesmo componente (NEWMAN, 2010, p. 140).

2.4 COMPONENTES E ÁRVORES

Como mencionado na seção anterior, é possível que uma par de vértices não possua um caminho ou uma distância geodésica. A rede da figura 3 é dividida em dois subgrupos de vértices desconectados, portanto não há uma distância geodésica definida para vértices de subgrupo da

esquerda com o subgrupo da direita. Cada subgrupo de vértices é chamado de componente. No caso da figura são duas componentes, uma com 9 vértices e outra com 8 vértices.

Figura 3 – Componentes de uma rede

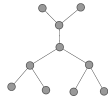


Uma rede dividida em 2 componentes. Não há uma distância geodésica definida para os vértices A e B de diferentes componentes.

Fonte: Criado pelo autor com o software Gephi.

Árvores na teoria dos grafos são redes não direcionadas, a qual seus componentes não contém *loops* (veja figura 4). Uma peculiaridade das árvores é que, já que não há *loops*, existe apenas um caminho para qualquer par de vértices e o número de arestas é igual ao número de vértices menos um, quer dizer, árvores com n vértices possuem $n - 1$ arestas (NEWMAN, 2010, p. 128).

Figura 4 – Representação de uma árvore



Note que se for adiciona mais uma aresta necessariamente haverá um *loop*, portanto em árvores com n vértices há sempre $n - 1$ arestas

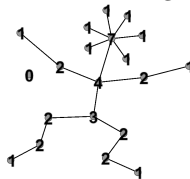
Fonte: Criado pelo autor com o software Gephi.

2.5 DISTRIBUIÇÃO DE GRAUS

Como definido na seção 2.2, o grau de um vértice é número de arestas ligados a ele. Na figura 5 temos uma rede com $n = 20$ vértices, o qual 1 tem grau 0, 10 tem grau 1, 6 tem grau 2, 1 tem grau 3, 1 tem grau 4 e 1 tem grau 7. Definindo p_k como a fração de vértices de uma rede com grau k , temos para a rede da figura: $p_0 = \frac{1}{20}$, $p_1 = \frac{10}{20}$, $p_2 = \frac{6}{20}$, $p_3 = \frac{1}{20}$, $p_4 = \frac{1}{20}$, $p_5 = 0$, $p_6 = 0$, $p_7 = \frac{1}{20}$ e $p_k = 0$ para $k > 7$. As quantidade p_k representam a **distribuição de graus** de uma rede. O valor p_k é também a probabilidade de um vértice escolhido

aleatoriamente ter grau k . O conhecimento de uma dada distribuição de graus é muito relevante e assumiu um importante papel na descoberta das redes livre de escala (veja seção 3.2), sabendo a forma de p_k pode-se determinar muitos fenômenos de rede, como a robustez e a propagação de vírus (BARABÁSI, 2016), porém somente a distribuição de graus não dá todas informações da estrutura da rede (NEWMAN, 2010, p. 245).

Figura 5 – Grafo mostrando o grau de cada vértice



Os números representam o grau de cada vértice.
 Fonte: Criado pelo autor com o software Gephi.

3 REDES

Neste capítulo são discutidos dois modelos principais de redes, as redes aleatórias e as redes livre de escala, ambas utilizadas no estudo de redes complexas.

3.1 REDES ALEATÓRIAS

As redes aleatórias foram e são a fonte de muitas das ideias nos estudos de redes. São grafos que têm suas arestas distribuídas aleatoriamente. O uso de redes aleatórias é repetidamente usado nos estudos de redes complexas porque grafos com uma topologia complexa e com princípios organizacionais desconhecidos em geral parecem aleatórios. O início dos estudos da rede é dado para Paul Erdős e Alfréd Rényi (ERDŐS e RÉNYI, 1959), após descobrirem métodos probabilísticos frequentemente úteis para lidar com problemas na teoria dos grafos (ALBERT e BARABÁSI, 2002).

Há duas formas de gerar uma rede complexa, a primeira é fixar uma quantidade n de vértices e uma quantidade m de arestas e distribuir aleatoriamente, geralmente não permitindo que existam arestas paralelas ou laços, modelo conhecido na matemática por $G(n, m)$. A segunda maneira é fixar uma quantidade n de vértices e uma probabilidade b dos vértices se ligarem, conhecido por $G(n, b)$. Neste caso as arestas não são fixas e é possível, mas improvável, que não haja arestas ou mesmo que todos vértices estejam ligados. O modelo $G(n, b)$ é o mais utilizado, o mais simples de obter várias propriedades e é o discutido neste trabalho. Ele foi primeiramente estudado por Solomonoff e Rapoport (SOLOMONOFF e RAPOPOR, 1951), porém é mais associado aos trabalhos de Erdős e Alfréd Rényi e pode ser chamado de modelo de Erdős-Rényi ou rede aleatória de Erdős-Rényi.

Na realidade a definição de grafos aleatórios não é em termos de um único grafo, mas em termos de um ensemble, ou seja, uma distribuição de probabilidades sob todas as possíveis redes. $G(n, b)$ é um ensemble de redes com n vértices o qual cada grafo G aparece com probabilidade:

$$P(G) = b^m (1 - b)^{\binom{n}{2} - m} \quad , \quad (3.1)$$

onde m é o número de arestas e $\binom{n}{2}$ é o coeficiente binomial. Neste caso estamos considerando apenas grafos simples e a probabilidade de

grafos não simples é nula. (NEWMAN, 2010, p. 400).

Para $G(n, b)$ o número de arestas não é fixo e o número de grafos com n vértices e exatamente m arestas é igual ao número de maneiras de arranjar as posições de arestas de $\binom{n}{2}$ pares distintos de vértices. Cada grafo aparece com a mesma probabilidade $P(G)$ dado pela equação (3.1), logo a probabilidade total de obter um grafo com m arestas do ensemble é:

$$P(m) = \binom{\binom{n}{2}}{m} b^m (1-b)^{\binom{n}{2}-m} \quad , \quad (3.2)$$

que é a distribuição binomial. O número médio de arestas é:

$$\begin{aligned} \langle m \rangle &= \sum_{m=0}^{\binom{n}{2}} m P(m) = \sum_{m=0}^{\binom{n}{2}} m \frac{\binom{n}{2}!}{\left(\binom{n}{2} - m\right)! m!} b^m (1-b)^{\binom{n}{2}-m} \\ &= \binom{n}{2} b \sum_{m=1}^{\binom{n}{2}} \frac{\left(\binom{n}{2} - 1\right)!}{\left(\left(\binom{n}{2} - 1\right) - (m-1)\right)! (m-1)!} b^{m-1} (1-b)^{\left(\binom{n}{2}-1\right)-(m-1)} \\ &= \binom{n}{2} b \sum_{m=1}^{\binom{n}{2}} \binom{\binom{n}{2}-1}{m-1} b^{m-1} (1-b)^{\left(\binom{n}{2}-1\right)-(m-1)} \\ &= \binom{n}{2} b \sum_{l=0}^{\binom{n}{2}-1} \binom{\binom{n}{2}-1}{l} b^l (1-b)^{\left(\binom{n}{2}-1\right)-l} \quad \text{onde } l = m-1 \\ &= \binom{n}{2} b \sum_{l=0}^h \binom{h}{l} b^l (1-b)^{h-l} \quad \text{onde } h = \binom{n}{2} - 1 \\ &= \binom{n}{2} b (b + (1-b))^h = \binom{n}{2} b \quad . \end{aligned} \quad (3.3)$$

A equação (3.3) nos mostra que número médio de arestas na rede é igual a probabilidade b entre qualquer par de vértices, multiplicado pelo número de pares. Usando este resultando e sabendo que o grau médio num grafo é dado por $\langle k \rangle = \frac{2m}{n}$ (eq. 2.5), o grau médio em $G(n, b)$ é:

$$\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n} \binom{n}{2} b = (n-1)b \quad , \quad (3.4)$$

um resultado esperado, pois um vértice num grafo aleatório tem pro-

bilidade b independente de se ligar a qualquer um dos $n - 1$ outros vértices (NEWMAN, 2010, p. 401).

3.1.1 Distribuição de graus das redes aleatórias

A probabilidade de um vértice se ligar com k outros vértices e nenhum outro é $b^k(1 - b)^{(n-1)-k}$. Existe $\binom{n-1}{k}$ maneiras de escolher esses k outros vértices, portanto a probabilidade total é:

$$p_k = \binom{n-1}{k} b^k (1 - b)^{(n-1)-k} \quad , \quad (3.5)$$

o qual é a distribuição binomial novamente. Ou seja, $G(n, b)$ tem uma distribuição binomial.

Muitas vezes faz-se uma aproximação para redes com um número muito grande de vértices. Pode-se escrever usando a equação (3.4):

$$\ln[(1 - b)^{(n-1)-k}] = (n - 1 - k) \ln\left(1 - \frac{\langle k \rangle}{n-1}\right) \quad , \quad (3.6)$$

no limite $n \rightarrow \infty$

$$(n - 1 - k) \ln\left(1 - \frac{\langle k \rangle}{n-1}\right) \approx (n - 1 - k) \left(-\frac{\langle k \rangle}{n-1}\right) \approx -\langle k \rangle \quad , \quad (3.7)$$

onde usa-se a expansão de Taylor do logaritmo. Usando a exponencial nos dois lados, tem-se $(1 - b)^{n-1-k} = e^{-\langle k \rangle}$. Também usa-se:

$$\binom{n-1}{k} = \frac{(n-1)^k}{k!} \quad . \quad (3.8)$$

Novamente usando equação (3.4), a equação (3.5) no limite de muitos vértices toma a forma:

$$p_k \approx \frac{(n-1)^k}{k!} b^k e^{-\langle k \rangle} = \frac{(n-1)^k}{k!} \left(\frac{\langle k \rangle}{n-1}\right)^k e^{-\langle k \rangle} = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad , \quad (3.9)$$

esta é uma distribuição de Poisson, portanto no limite para n muito grande $G(n, b)$ tem uma distribuição de Poisson e alguns autores utilizam o nome rede aleatória de Poisson (NEWMAN, 2010, p. 402).

3.1.2 Diametro das redes aleatórias

No artigo *The Diameters of Sparse Random Graphs*, CHUNG e LU (2001) estimaram o diâmetro das redes aleatórias para vários valores de b , utilizando o tamanho das vizinhanças de um dado conjunto de vértices. Os dados foram sintetizado por ALBERT e BARABÁSI (2002), o qual informa que para redes com o mesmo número de n e b os valores de diâmetro variam muito pouco, geralmente concentrados em:

$$d = \frac{\ln(n)}{\ln(bn)} = \frac{\ln(n)}{\ln(\langle k \rangle)} \quad , \quad (3.10)$$

porém, para valores de $bn = \langle k \rangle < 1$ o grafo é composto por várias árvores isoladas e o diâmetro é igual ao diâmetro das árvores.

3.2 REDES LIVRE DE ESCALA

Uma rede livre de escala é quando sua distribuição de graus tem a forma de lei de potência. Muitas redes reais possuem este tipo de distribuição, dois exemplos são a rede da Internet (ALBERT; JEONG; BARABÁSI, 1999) e as rede de interação proteína-proteína (MASHAGHI; RAMAZANPOUR; KARIMPOUR V., 2004).

3.2.1 Distribuição de graus das redes livre de escala

Uma rede livre de escala tem por definição a distribuição de graus na forma de lei de potência, isto é:

$$p_k = Y k^{-\lambda} \quad , \quad (3.11)$$

sendo Y uma constante de normalização e λ o expoente da lei de potência. A distribuição de graus deve ser normalizada da forma:

$$\sum_{k=0}^{\infty} p_k = 1 \quad , \quad (3.12)$$

portanto para as redes livres de escala:

$$Y \sum_{k=0}^{\infty} k^{-\lambda} = 1 \quad , \quad (3.13)$$

$$Y = \frac{1}{\sum_{k=0}^{\infty} k^{-\lambda}} = \frac{1}{\zeta(\lambda)} \quad , \quad (3.14)$$

$\zeta(\lambda)$ é a função zeta de Riemann e a distribuição de grau fica:

$$p_k = \frac{k^{-\lambda}}{\zeta(\lambda)} \quad . \quad (3.15)$$

Na prática, poucos são os fenômenos empíricos que obedecem uma lei de potência para todos os valores e geralmente à lei de potência se aplica apenas à cauda da distribuição de graus (CLAUSET; SHALIZI; NEWMAN, 2009). Para distribuições onde somente a cauda é uma lei de potência, é necessário declarar um valor mínimo k_{min} onde a lei vale, isto é, o valor inicial da cauda. Então a distribuição é escrita na forma:

$$p_k = \frac{k^{-\lambda}}{\sum_{k=k_{min}}^{\infty} k^{-\lambda}} = \frac{k^{-\lambda}}{\zeta(\lambda, k_{min})} \quad , \quad (3.16)$$

$\zeta(\lambda, k_{min})$ é a função zeta de Riemann generalizada. Pode-se aproximar a soma da função zeta de Riemann generalizada por uma integral:

$$Y = \frac{1}{\sum_{k=k_{min}}^{\infty} k^{-\lambda}} \approx \frac{1}{\int_{k_{min}}^{\infty} k^{-\lambda} dk} = (\lambda - 1)k_{min}^{\lambda-1} \quad , \quad (3.17)$$

então a distribuição de graus para a cauda com lei de potência pode ser aproximada por:

$$p_k \approx \frac{(\lambda - 1)}{k_{min}} \left(\frac{k}{k_{min}} \right)^{-\lambda} \quad (3.18)$$

(NEWMAN, 2010, p. 257).

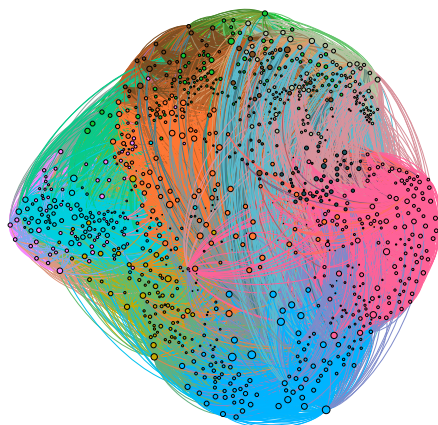
3.2.2 Diâmetro das rede livre de escala

COHEN e HAVLIN (2003) mostraram que para redes livres de escala com expoentes entre $2 < \lambda < 3$ o diâmetro da rede é da ordem de $d \approx \ln[\ln(n)]$ e para expoente $\lambda > 3$ o diâmetro é da ordem de $d \approx \ln(n)$. Se $\lambda = 3$ BARABÁSI (2016) afirma que $d \approx \frac{\ln(n)}{\ln[\ln(n)]}$ e para $\lambda = 2$, d é uma constante independente do número de vértices n .

4 CONECTOMA

Conectoma é uma representação estrutural do cérebro humano na forma de grafo, o qual possibilita uma análise quantitativa da conectividade do cérebro com o uso de ferramentas matemáticas. O termo conectoma (originalmente do inglês *connectome*), fazendo analogia com a palavra genoma, foi cunhado por SPORNS; TONONI e KÖTTER (2005) para designar uma rede de elementos e conexões do cérebro. HAGMANN et al (2008) publicaram um artigo sobre o mapeamento de regiões do córtex cerebral humano utilizando técnicas não invasivas de imagem para a construção de uma matriz de conexão estrutural de alta resolução, o conectoma utilizado neste trabalho representado na figura 6.

Figura 6 – Representação do conectoma



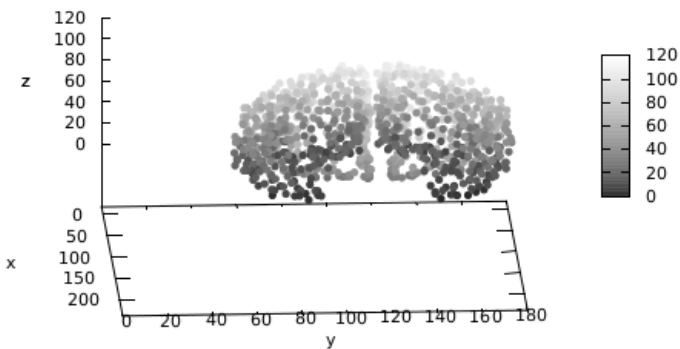
Representação do conectoma.

Fonte: Criado pelo autor com o software Gephi.

Estudando 5 participantes saudáveis de idade entre 24 e 32 anos,

foram necessárias algumas etapas para a construção da rede. A primeira delas, para a criação dos vértices da matriz, foi através da ressonância magnética convencional, utilizado o tempo de relaxação longitudinal, chamado T1, para gerar o contraste na imagem. A ressonância magnética utiliza um campo magnético intenso que faz com que os spins se alinhem com o campo. Ao aplicar um pulso transversal de radio-frequência, a magnetização resultante longitudinal dos spins se torna nula, T1 é o tempo para que a magnetização resultante longitudinal volte 63% de seu valor inicial (POOLEY, 2005, p. 1091) e como cada tecido possui um valor de T1 diferente é possível diferencia-los. Com o uso de campos magnéticos auxiliares forma-se um gradiente de intensidade único em cada ponto podendo os sinais serem traduzidos em uma imagem em 3 dimensões. A imagem dos participantes foi então dividida em massa branca, córtex e núcleos cerebelares profundos e o córtex foi particionado, através de um software, em 66 regiões de interesse (ROIs do inglês *regions of interest*) com localizações topográficas idênticas, apesar da anatomia individual de cada participante. Essas mesmas regiões foram subdivididas em pequenas e compactas regiões de aproximadamente $1,5\text{cm}^2$ resultando 998 ROIs cobrindo todo o córtex (fig. 7) (HAGMANN et al, 2008).

Figura 7 – Localização das 998 ROIs do conectoma recobrando o córtex cerebral



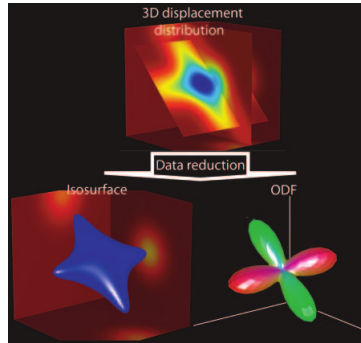
Cada ponto representa uma região do cérebro de aproximadamente $1,5\text{cm}^2$, são essas regiões que dão origem aos vértices da rede. A escala em cinza representa a altura no eixo z para melhor visualização.

Fonte: Imagem criada pelo autor com o software gnuplot.

Para criar as arestas da rede foi necessário o uso da imagem de difusão por ressonância magnética, DSI, técnica que utiliza a difusão da água nos tecidos, baseado nos experimentos de HAHN (1954), o qual demonstrou uma diminuição do sinal dos spins em um líquido imerso em campo magnético heterogêneo, e nas equações desenvolvidas por TORREY (1956) que descrevem o fenômeno.

O tecido neuronal consiste em axônios firmemente compactados e coerentemente alinhados, circundados por células da glia e, muitas vezes, organizados em feixes, conseqüentemente as moléculas de água são impedidas de se movimentar na direção perpendicular a orientação axonal resultando em uma anisotropia da densidade de probabilidade da difusão de água (HAGMANN et al, 2006, p. 209). Para se obter a imagem é necessário que cada volume de resolução medido do cérebro, os voxels (unidade semelhante ao pixel, porém em 3 dimensões) tenha uma densidade de probabilidade de difusão atribuída também em 3 dimensões, tendo portanto uma imagem em 6 dimensões. Para representar a imagem em 6 dimensões a primeira abordagem é substituir a função de densidade de probabilidade de difusão por uma isosuperfície, uma superfície que passa por todos os pontos de igual valor, neste caso, a probabilidade de difusão da água. Porém uma técnica mais comumente usada que é menos sensível a ruídos envolve a computação de uma função de distribuição de orientação (ODF do inglês *orientation distribution function*) que pode ser considerada como uma esfera deformada cujo raio em um dada direção é proporcional à soma dos valores das funções de densidade de probabilidade de difusão nessa direção. Para facilitar a visualização, codifica-se a superfície da ODF de acordo com a direção de difusão ($[x, y, z] = [r, b, g]$, onde r =vermelho, b =azul e g =verde)(Figura 8) (HAGMANN et al, 2006, p. 210). Com cada voxel preenchido com uma ODF se faz a tractografia, processo para a construção de curvas em 3D com coerência máxima de difusão, chamada de fibras. Cada voxel começa com o mesmo número de fibras para cada direção, isto é, se um voxel possui uma ODF com duas direções, é iniciado 30 fibras para cada direção, num total de 60 fibras. O ponto de partida é escolhido aleatoriamente no espaço do voxel. Para cada ponto de partida a trajetória da fibra é formada da mesma forma que o fluxo de um campo vetorial com restrições adicionais em algumas localizações, ocorrendo às vezes múltiplas orientações. Para contornar o problema utiliza-se o crescimento das fibras para cada direção oposta com um passo fixo de 1mm. Quando entra num novo voxel a fibra continua a crescer na direção do ODF do novo voxel, cuja a orientação deve ser próxima a fibra que entrou. Se isto resultar em uma mudança

Figura 8 – Redução de dados de um voxel em 6 dimensões obtido por DSI



Na primeira imagem a densidade de probabilidade da difusão de água obtida no DSI e abaixo a isosuperfície e a função de distribuição de orientação com a codificação de cores.

Fonte: Artigo Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond. Disponível em: <<https://pubs.rsna.org/doi/pdf/10.1148/rg.26si065510>>.

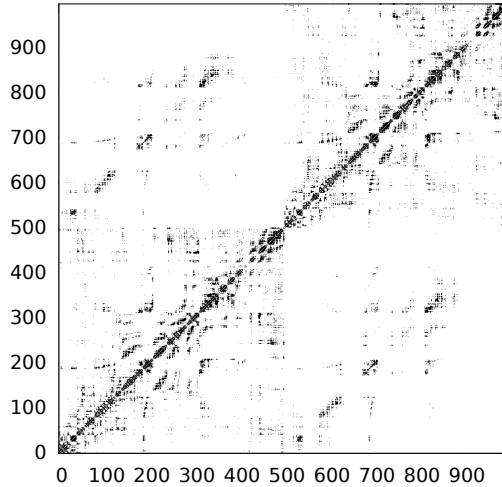
de direção maior que $30^\circ/\text{mm}$ a fibra pára de crescer. O processo de crescimento de uma fibra só é válido se o começo e o fim saem da região delimitada pela superfície da massa branca do cérebro. Para produzir o conectoma deste trabalho foram usados 3 milhões de pontos de partidas, os quais entre metade e dois terços conectaram áreas corticais e, portanto, foram gravados (HAGMANN et al, 2008).

Unindo as ROIs criadas na primeira etapa com a tractografia média dos 5 participantes obteve-se um grafo com a conectividade estrutural do cérebro, sendo cada ROI um vértice, num total de 998 e as arestas sendo as fibras que ligam duas ROIs totalizando 17865 (fig. 10). O peso de uma aresta é definido como a densidade de fibras (o número de conexões por unidade de área) entre os vértices. A rede foi então representada em uma matriz de adjacência mostrada na figura 9, sendo cada ponto no elemento de matriz A_{ij} uma aresta que liga os vértices i e j .

4.1 MEDIDAS DA REDE DO CONECTOMA

A principal medida de uma rede é a sua distribuição de graus. A figura 11 mostra a distribuição de graus do conectoma juntamente com as distribuições binomial e de Poisson, distribuições de uma rede

Figura 9 – Matriz de adjacência binária do conectoma



Os eixos são os vértices rotulados de 1 a 998 e os pontos representam a existência de uma conexão, isto é, uma aresta entre os vértices.

Fonte: Criado pelo autor com o software gnuplot.

aleatória (seção 3.1.1), o qual usei a equação 3.4 para obter o valor b da distribuição binomial e a distribuição de Poisson possui média igual à do conectoma, $\langle k \rangle = 35.8$. Pode-se ver apenas pela distribuição que o conectoma não pode ser modelado por uma rede aleatória.

A melhor forma de saber se há uma lei de potência é discretizar de forma logarítmica a distribuição de graus e representar no gráfico log-log (CLAUSET; SHALIZI; NEWMAN, 2009). Na figura 12 a distribuição de graus está desta forma e observa-se que ela não possui uma lei de potência, o que deveria acontecer é um decrescimento uniforme das barras, como uma escada com degraus regulares. Pode-se aproximar uma lei de potência para cauda (veja seção 3.2.1) escolhendo $k_{min} = 40$ e o melhor ajuste ¹ é um expoente $\lambda = 6.30 \pm 10.93$, um decaimento tão rápido quanto a de uma exponencial.

Segundo o artigo de AMARAL et al (2000), modelando um custo para se adicionar conexões nas redes livres de escalas, parece difícil que as redes estruturais do cérebro possam exibir uma distribuição de lei de potência através de uma ampla extensão de graus. Como os vértices do cérebro, independentemente de como são definidos, estão espacialmente

¹melhor ajuste feito pelo software gnuplot

Figura 10 – Passos para a criação do conectoma

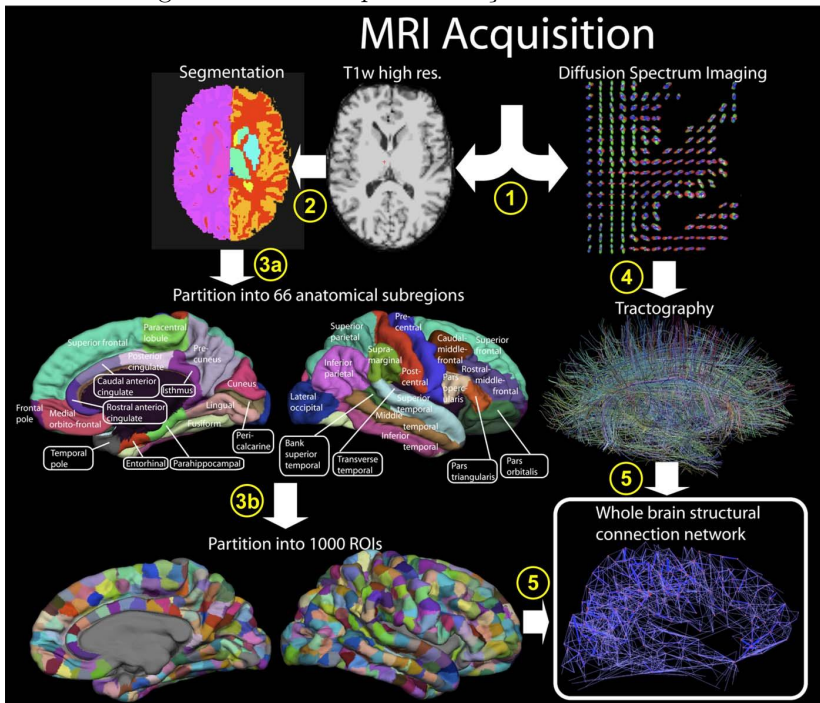


Diagrama mostrando os passos para a criação do conectoma a partir das imagens de ressonância magnética.

Fonte: Artigo Mapping the Structural Core of Human Cerebral Cortex.

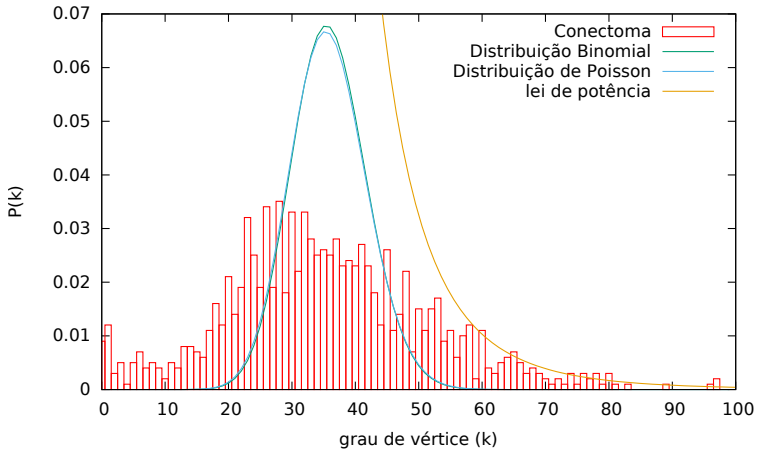
Disponível em:

<https://commons.wikimedia.org/wiki/File:Connectome_extraction_procedure.jpg>.

fixos, deve haver um limite máximo na densidade de conexões que pode ser mantida devido às restrições espaciais e metabólicas (SPORNS, 2011, p. 105). Embora haja evidências de uma organização não-aleatória e não-regular das redes cerebrais em larga escala, o quanto as redes em nível celular são aleatórias ou regulares é muito menos compreendido. Uma das principais razões é a relativa falta de dados da conectividade celular adquiridos em um formato que permita a análise de grafos. Os conjuntos de dados presentes consistem em números muito pequenos de neurônios e não abrangem redes inteiras ou relatam probabilidades de conexão entre células e tipos de células em vez de reais padrões de conexão (SPORNS, 2011, p. 105) .

Uma medida interessante feita no conectoma é o coeficiente de

Figura 11 – Distribuição de graus do conectoma junto com as distribuições binomial, de Poisson e uma lei de potência para cauda



Em vermelho, em formas de barras, a distribuição de graus do conectoma, em verde a distribuição binomial com $b = 0.036$ da equação 3.4 e $n = 998$ e em azul a distribuição Poisson com $n = 998$ e média $\langle k \rangle = 35.8$ (mesma do conectoma), em amarelo uma lei de potência para a cauda com $k_{min} = 40$ e $\lambda = 6.30$

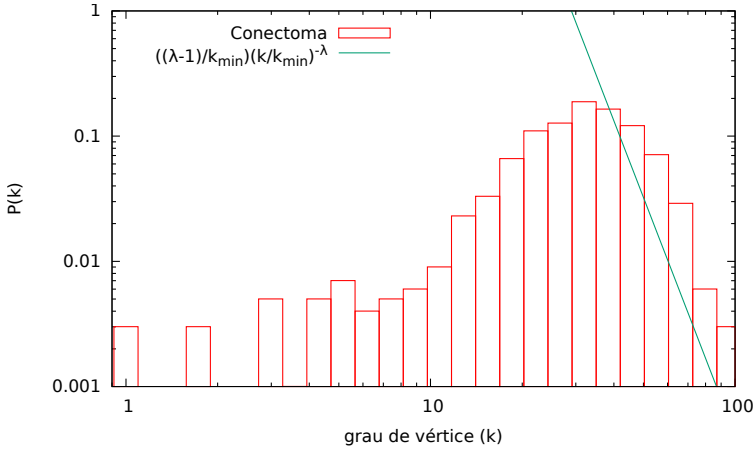
Fonte: Imagem criada pelo autor com o software gnuplot.

agrupamento. Este coeficiente mede a probabilidade de dois vizinhos de um vértice serem também vizinhos (NEWMAN, 2010, p. 262). É uma propriedade comum das rede sociais e representa os círculos de amigos ou conhecidos medindo, de 0 a 1, o quanto os membros se conhecem entre si. (ALBERT e BARABÁSI, 2002). O coeficiente de agrupamento local para um vértice i é definido como:

$$C_i = \frac{\text{número de pares de vizinhos de } i \text{ que estão conectados}}{\text{número de pares de vizinhos de } i}. \quad (4.1)$$

Para o calculo de C_i é preciso passar por todos os pares distintos de vértices que são vizinhos de i , contar esse número de pares que estão conectados um com outro e dividir pelo número total de pares, cujo valor é $\frac{1}{2}k_i(k_i - 1)$ (NEWMAN, 2010, p. 202). O coeficiente de agrupamento global da rede é a média dos coeficientes de agrupamentos locais (ALBERT e BARABÁSI, 2002), representada pela letra C . No conectoma $C = 0.46$ com desvio padrão $\sigma_C = 0.13$, isso demonstra o quanto o conectoma está conectado, como exemplo de comparação temos as redes aleatória e as redes livre de escala. Para as redes aleatórias a probabilidade que

Figura 12 – Distribuição de graus do conectoma em log-log e melhor ajuste de lei de potência



Distribuição de graus do conectoma com discretização da distribuição em forma logarítmica. Na melhor das hipóteses escolhendo $k_{min} = 40$ temos um ajuste próximo de uma lei de potência com expoente $\lambda = 6.30 \pm 10.93$

Fonte: Criado pelo autor com o software gnuplot.

qualquer dois vértices sejam vizinhos é a mesma igual a $b = \frac{\langle k \rangle}{n-1}$ e portanto $C_{aleatoria} = \frac{\langle k \rangle}{n-1}$ (NEWMAN, 2010, p. 403). Usando a mesma média de graus e o número de vértices do conectoma para comparar com uma rede aleatória o que se obtém é $C_{aleatoria} = 0.036$, este é um valor com uma ordem de grandeza diferente do conectoma. Para as redes livres de escala o coeficiente de agrupamento é aproximadamente uma lei de potência do tipo $C_{livre-de-escala} \approx n^{-0.75}$ (ALBERT e BARABÁSI, 2002), usando novamente o número de vértices do conectoma $C_{livre-de-escala} = 0.0056$, novamente um valor com ordens de grandeza diferente do conectoma.

O comprimento médio do conectoma, descrito na seção 2.3, possui valor $l = 0.21$. Para rede aleatória, escolhendo um vértice aleatório e calculando o número de vizinhos que estão a m passos de distância deste vértice demonstra-se que $l_{aleatoria} = \frac{\ln(n)}{\ln(\langle k \rangle)}$ (NEWMAN; STROGATZ; WATTS, 2001). Usando $n = 998$ e $\langle k \rangle = 35.8$ pode-se comparar o conectoma com a rede aleatória sendo $l_{aleatoria} = 1.93$.

A medida do diâmetro do conectoma foi $d = 6$, muito próximo do diâmetro das redes livres de escala com $\lambda > 3$ e $n = 998$, o qual possui $d_{livre-de-escala} \approx 6.9$ (veja seção 3.2.2), diferentemente das redes

aleatórias com o mesmo número de vértices e mesma média de graus do conectoma, o qual tem diâmetro igual $d_{aleatoria} = 1.9$ (veja seção 3.1.2).

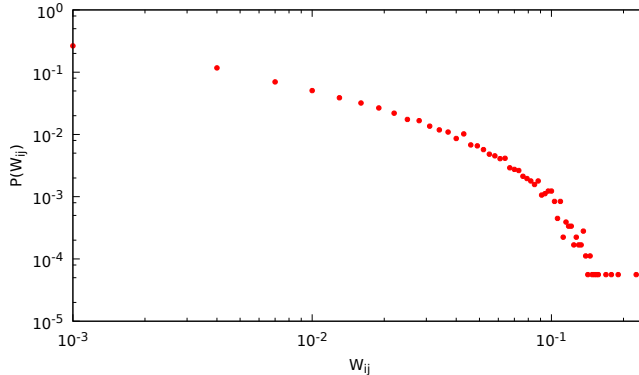
Um outro tipo de rede complexa é a de 'mundo pequeno', que como o nome sugere possui a propriedade de ligar qualquer dois vértices apenas por poucas arestas (WATTS e STROGATZ, 1998). Para dizer se uma rede é de mundo pequeno é preciso que:

$$\sigma_{SW} = \frac{C/C_{aleatoria}}{l/l_{aleatoria}} > 1 \quad , \quad (4.2)$$

(HUMPHRIES e GURNEY, 2008) no caso do conectoma $\sigma_{SW} = 117$ muito maior que um, provando que é uma rede de mundo pequeno..

A distribuição de pesos de arestas, $P(w_{ij})$, é exposta no gráfico log-log da figura 13. Como já mencionado, o peso é a densidade de fibras entre os vértices.

Figura 13 – Distribuição de pesos das arestas do conectoma

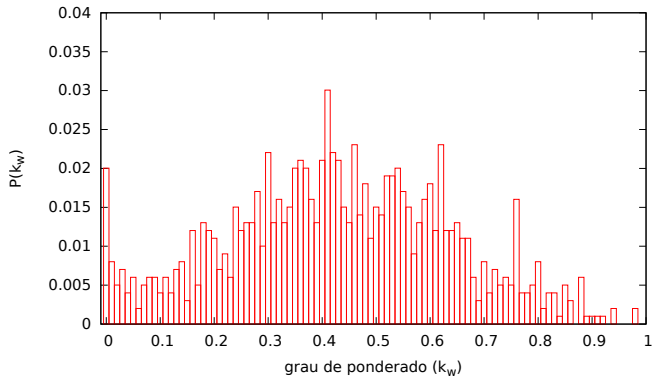


Distribuição de pesos das arestas do conectoma em um gráfico log-log. O peso da aresta é definido como a densidade de fibras entre os vértices.

Fonte: Criado pelo autor com o software gnuplot.

Na figura 14 é exposto um gráfico com o grau ponderado k_w do conectoma, isto é, a soma de todos pesos das arestas que ligam um vértice, utilizado para o modelo de limiar linear descrito no capítulo 7.

Figura 14 – Distribuição de graus ponderados do conectoma



Distribuição de graus ponderado do conectoma, definido como a soma de todos pesos das arestas que ligam o vértice.

Fonte: Criado pelo autor com o software gnuplot.

5 TRANSIÇÃO DE FASE E EXPOENTES CRÍTICOS

É costume fazer analogia com água para introduzir o estudo de transições de fase. A água está na fase líquida a temperatura ambiente e pressão atmosférica, mas se resfriada a temperatura abaixo de 273,15K fica na fase sólida e se aquecida acima 373,15K se encontra na fase de vapor. Em cada uma dessas temperaturas, o material sofre uma mudança de propriedades, uma **transição de fase** (CALLEN, 1985, p. 215).

Para cada transição é definido um **parâmetro de ordem**, uma quantidade física usada para descrever a transição, o qual deve ser nulo em uma região e não nulo em outra região delimitada pelo ponto de transição de fase. No caso do fenômeno da transição de fase líquido-gás, o parâmetro de ordem é a diferença de densidade entre líquido e gás.

As transições de fase podem ser de primeira ou segunda ordem. Em uma transição de primeira ordem, regiões macroscópicas completamente diferentes aparecem na temperatura de transição, por exemplo, gelo e líquido no caso do derretimento da água. De uma região para outra a configuração microscópica muda de maneira abrupta. No nível macroscópico, isso se traduz na descontinuidade em certas primeiras derivadas dos potenciais termodinâmicos e na existência de um calor latente. Nas transições de fase de segunda ordem as primeiras derivadas dos potenciais termodinâmicos são contínuas, enquanto que certas derivadas de segunda ordem divergem. Nenhuma descontinuidade é observada no nível microscópico, apenas uma divergência no tamanho das flutuações características (LESNE e LAGUÁS, 2012, p. 6).

O comprimento de correlação é definido como a medida do alcance da função de correlação que por sua vez mede como as flutuações das variáveis microscópicas estão relacionadas (STANLEY, 1971, p. 46). Transições de fase de segunda ordem tem a peculiaridade de ter o comprimento de correlação ξ divergindo de acordo com uma lei de potência próximo a temperatura crítica. Define-se a temperatura reduzida t como a medida da distância da temperatura crítica T_c :

$$t = \frac{T - T_c}{T_c} \quad , \quad (5.1)$$

e portanto ξ diverge da forma:

$$\xi \sim |t|^{-\nu} \quad . \quad (5.2)$$

Isso significa que qualquer dois componentes do sistema estão correlacionados, e as flutuações se propagam por todo o sistema (NEWMAN; STROGATZ; WATTS, 2001, p. 77).

O problema básico na teoria de transição de fase é o comportamento de um dado sistema na vizinhança do ponto crítico. O comportamento é marcado pelo fato de várias quantidades físicas pertencentes ao sistema apresentarem singularidades no ponto crítico devido à lei de potência divergente no comprimento de correlação. É habitual expressar essas singularidades em termos de leis de potência caracterizado por um conjunto de **expoentes críticos** que determina a natureza qualitativa do comportamento crítico de um dado sistema (PATHRIA e BEALE, 2011, p. 435). Na equação (5.3) tem-se um exemplo de como um parâmetro de ordem ρ de um sistema se comporta próximo e abaixo da temperatura crítica T_c :

$$\rho \sim |t|^\beta \quad . \quad (5.3)$$

Outra medida do comportamento do sistema, proporcional à segunda derivada de um potencial termodinâmico, é a compressibilidade isotérmica κ_T para os fluídos ou então a susceptibilidade magnética χ para sistemas magnéticos, a qual diverge da forma:

$$\kappa_T \sim |t|^{-\gamma} \quad , \quad (5.4)$$

ν , β e γ são expoentes críticos e são mais interessantes que a própria temperatura crítica, pois T_c depende dos detalhes microscópicos do sistema, enquanto os expoentes críticos são universais dependendo apenas de alguns parâmetros fundamentais como a dimensionalidade do sistema, a dimensionalidade do parâmetro de ordem, simetria do parâmetro de ordem e alcance das interações. Um dado conjunto de valores para expoentes críticos define uma **classe de universalidade** (YEOMANS, 1992, p. 27). Sabendo a classe de universalidade de um sistema é possível trabalhar com um modelo mais simples da mesma classe e obter propriedades que valem para ambos. Outro aspecto relevante que originou vários métodos chamados de grupos de renormalização é a invariância de escala na física do estado crítico (LESNE e LAGUËS, 2012, p. 30).

Porém, os comportamentos das equações (5.3), (5.4) e a divergência de ξ só acontecem no limite termodinâmico, isto é, quando a quantidade de elementos do sistema tende a infinito. Na próxima seção é descrito o método para estimar expoentes críticos em sistemas finitos.

5.1 SISTEMAS FINITOS

Para sistemas finitos precisa-se usar o método de escala de tamanho finito (FSS do inglês *finite size scaling*), que consiste numa forma de extrair os expoentes observando como as medidas variam ao mudar o tamanho de um sistema. Eliminado $|t|$ das equações (5.2) e (5.4) e usando χ em vez κ_T a fim de minimizar a notação, pode-se escrever que na vizinhança da transição de fase:

$$\chi \sim \xi^{\gamma/\nu} \quad , \quad (5.5)$$

o qual não diverge para sistemas finitos, o que ocorre é um truncamento do comprimento de correlação para um sistema de tamanho linear L , isto é, $\xi > L$, mas para $\xi \ll L$ o valor de χ deve ser o mesmo que em sistemas infinitos. Expressa-se isso da forma:

$$\chi = \xi^{\gamma/\nu} \chi_0(L/\xi) \quad , \quad (5.6)$$

onde χ_0 é uma função adimensional de uma única variável e com propriedades:

$$\begin{cases} \chi_0(x) = \text{const.} & \text{para } x \gg 1 \\ \chi_0(x) \sim x^{\gamma/\nu} & \text{quando } x \rightarrow 0 \end{cases} \quad . \quad (5.7)$$

A equação (5.6) contém ξ que é desconhecido, por isso define-se uma nova função adimensional $\tilde{\chi}$ tal que:

$$\tilde{\chi}(x) = x^{-\gamma} \chi_0(x^\nu) \quad . \quad (5.8)$$

Fazendo uso da equação (5.2) pode-se retirar a dependência de ξ e escrever:

$$\chi = L^{\gamma/\nu} \tilde{\chi}(L^{1/\nu}t) \quad . \quad (5.9)$$

A equação (5.9) mostra todas as dependências de L em χ para o comportamento finito da susceptibilidade próximos a temperatura crítica. A função $\tilde{\chi}(x)$ é chamada de função de escala para a susceptibilidade e mesmo que esta função seja desconhecida sabe-se que:

$$\tilde{\chi}(x) \rightarrow x^{-\gamma} (x^\nu)^{\gamma/\nu} = \text{const.} & \text{para } x \rightarrow 0 \quad . \quad (5.10)$$

Isso significa que $\tilde{\chi}$ é finito na origem e portanto próximo da temperatura crítica. Para cada tamanho de sistema L temos um valor de $\chi_N(t)$

a uma temperatura t . Reescrevendo a equação (5.9) na forma:

$$\tilde{\chi}(L^{1/\nu}t) = L^{-\gamma/\nu}\chi_L(t) \quad . \quad (5.11)$$

Pode-se ter uma estimativa da função de escala $\tilde{\chi}$ para diferentes valores da variável de escala $x = L^{1/\nu}t$ para cada tamanho de sistema. Como a função de escala deve ser a mesma para todos tamanhos, todos os dados devem ficar sobrepostos no mesmo gráfico. O mesmo processo é feito para o parâmetro de ordem ρ :

$$\rho \sim L^{-\beta/\nu}\tilde{\rho}(L^{1/\nu}t) \quad (5.12)$$

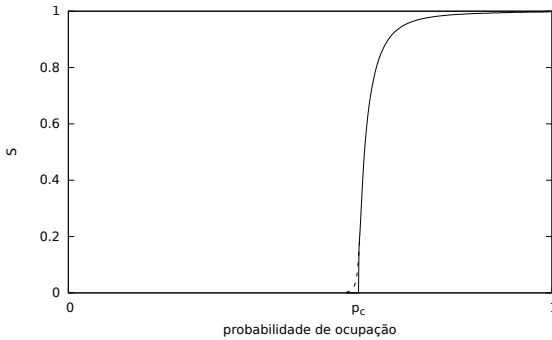
(NEWMAN; STROGATZ; WATTS, 2001).

6 PERCOLAÇÃO

Remover um único vértice ou uma única aresta tem um impacto limitado na integridade de uma rede, porém remover diversos pode ocasionar sua ruptura em vários componentes isolados, quanto mais vértices ou arestas são retirados, maiores são as chances de danificar a rede. A pergunta a ser feita é quantos vértices ou arestas são necessários para que a rede se quebre em vários componentes isolados? Para responder isso utiliza-se a teoria de percolação (BARABÁSI, 2016, p. 273).

No processo de percolação os vértices ou arestas podem estar em um de dois estados, o estado **ocupado** ou vazio. Inicialmente a rede se encontra vazia e na **percolação de sítios** ocupam-se os vértices e na **percolação de ligação** ocupam-se as arestas. Tradicionalmente utiliza-se a variável p para parametrizar os processos, ela é a fração de vértices ou arestas que estão ocupados e é chamada de **probabilidade de ocupação**. Em função de p contam-se o tamanho de **clusters**, definido como o grupo de vértices ocupados conectados por seus vizinhos (STAUFFER e AHARONY, 1991, p. 4). O tamanho de um *cluster* é definido como a soma dos vértices ocupados do *cluster* dividido pelo total de vértices ocupados. Os *clusters* inicialmente estão isolados, mas a medida que as arestas ou vértices são ocupados eles se juntam formando um grande *cluster* do tamanho proporcional a toda a rede ocupada, este grande *cluster* chama-se de **cluster gigante**. É neste ponto que forma o *cluster* gigante que acontece uma transição de fase de segunda ordem, discutido no capítulo 5. O parâmetro de ordem na percolação é o tamanho do *cluster* gigante, até o valor crítico p_c não existe um *cluster* gigante e portanto o parâmetro de ordem é zero, depois vai crescendo continuamente. Isto ocorre para redes infinitas, com redes finitas acontece um efeito de tamanho finito, de forma que a parâmetro de ordem cresce um pouco antes do ponto crítico (veja fig. 15) (NEWMAN, 2010, p. 607).

Figura 15 – Exemplo de uma transição de fase de segunda ordem na percolação



Variando a probabilidade de ocupação até o um valor crítico p_c não há um *cluster* gigante e portanto o seu tamanho $S = 0$ (linha sólida), depois do ponto crítico tem-se uma variação contínua do parâmetro de ordem S , caracterizando uma transição de fase de segunda ordem. Para redes infinitas a transição de fase é representada pela linha sólida, para redes finitas ocorre um efeito de tamanho finito de forma a "arredondar" a transição mostrado pela linha pontilhada.

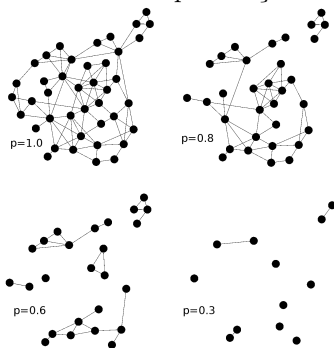
Fonte: Criado pelo autor com o software gnuplot.

A percolação pode ser entendida de outra forma, em vez de se ocupar os vértices ou arestas eles são retirados. Quando $p = 1$ todos os vértices ou arestas estão ocupados ou o mesmo que os vértices ou arestas não terem sido removidos e quando $p = 0$ não há vértices ou arestas ocupados, isto é, todos foram removidos (NEWMAN, 2010, p. 594), desta forma a percolação é utilizada para modelar alguns fenômenos do mundo real, como por exemplo falhas nos roteadores da internet (ALBERT; BARABÁSI; JEONG, 2000), possíveis ataques físicos a redes de comunicação (BARAN, 1964) e na doença de Alzheimer (BERNDT; RUPIN; REGGIA, 1996, p. 404).

A forma mais simples de se ocupar (ou se retirar) os vértices ou arestas na percolação é a ocupação (remoção) aleatória. Outra forma é simular um ataque aos componentes mais conectados, isto é, retirar primeiramente os vértices com maiores graus para a percolação de sítios (ALBERT; BARABÁSI; JEONG, 2000) ou então na percolação de ligação, retiram-se primeiro as arestas de maior peso.

Na figura 16 eu demonstro a percolação de sítios na interpretação da retirada de vértices. Na primeira figura $p = 1.0$, nenhum vértice foi retirado e há um *cluster* gigante de tamanho 1. Quando p é variando, o tamanho do *cluster* gigante vai diminuindo e no ponto crítico não há mais um *cluster* gigante, apenas pequenos *clusters* isolados.

Figura 16 – Processo de percolação em uma rede



Exemplo de percolação de sítios em um grafo para vários valores de probabilidade de ocupação p . (a) $p = 1.0$, existe um *cluster* gigante de tamanho 1. (b) $p = 0.8$, ainda existe um *cluster* gigante com tamanho 0.875 (c) $p = 0.6$, não há mais um *cluster* gigante, apenas pequenos *clusters* (d) $p = 0.3$, há apenas pequenos *clusters*.

Fonte: Criado pelo autor com o software Gephi.

6.1 FUNÇÃO GERADORA

Antes de descrever a equação que rege a transição de fase na percolação é relevante trazer alguns fundamentos da matemática. O primeiro é a função geradora, uma maneira de codificar uma sequência infinita de números a_n , tratando-os como os coeficientes de uma série de potências. Sendo uma distribuição de graus p_k , a função geradora desta distribuição $g_0(z)$ ¹ é definida como:

$$g_0(z) = p_0 + p_1z + p_2z^2 + p_3z^3 + \dots = \sum_{k=0}^{\infty} p_k z^k \quad (6.1)$$

(NEWMAN, 2010, p. 429). Aparentemente não se vê o benefício de criar uma função geradora, mas ao longo deste capítulo o leitor deve estar convencido de que é uma poderosa ferramenta.

¹O subscripto 0 é porque há mais funções geradoras no trabalho e convém já fazer a distinção.

6.2 NORMALIZAÇÃO E MOMENTOS

Se a distribuição de probabilidade é normalizada

$$g_0(1) = \sum_{k=0}^{\infty} p_k = 1 \quad . \quad (6.2)$$

Pode-se obter o primeiro momento da uma distribuição de graus derivando a sua função geradora de modo que:

$$g'_0(z) = \sum_{k=0}^{\infty} k p_k z^{k-1} \quad , \quad (6.3)$$

$$g'_0(1) = \sum_{k=0}^{\infty} k p_k = \langle k \rangle \quad . \quad (6.4)$$

O segundo momento:

$$z \frac{d}{dz} \left(z \frac{d g_0(z)}{dz} \right) = \sum_{k=0}^{\infty} k^2 p_k z^k \quad , \quad (6.5)$$

$$\left(z \frac{d}{dz} \right)^2 g_0(z) \Big|_{z=1} = \langle k^2 \rangle \quad , \quad (6.6)$$

e de modo generalizado o l -ésimo momento:

$$\left(z \frac{d}{dz} \right)^l g_0(z) \Big|_{z=1} = \langle k^l \rangle \quad (6.7)$$

(NEWMAN, 2010, p. 433).

6.3 POTÊNCIAS DA FUNÇÃO GERADORA

Suponha m números inteiros $k_i, i = 1, 2, \dots, m$, o quais são números aleatórios independentes retirados da distribuição de graus p_k com função geradora $g_0(z)$. A probabilidade que eles tomem um conjunto particular de valores $\{k_i\}$ é $\prod_i^m p_{k_i}$ e a probabilidade π_s que os valores sorteados somem exatamente s é a soma dessas probabilidades sobre

todos os conjuntos $\{k_i\}$ que somam s :

$$\pi_s = \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} \delta(s, \sum_{i=0}^k k_i) \prod_{i=0}^m p_{k_i} \quad , \quad (6.8)$$

onde $\delta(a, b)$ é o delta de Kronecker.

A função geradora $h(z)$ da distribuição π_s é:

$$\begin{aligned} h(z) &= \sum_{s=0}^{\infty} \pi_s z^s \\ h(z) &= \sum_{s=0}^{\infty} z^s \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} \delta(s, \sum_{i=0}^k k_i) \prod_{i=0}^m p_{k_i} \\ &= \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} z^{\sum_i k_i} \prod_{i=0}^m p_{k_i} \\ &= \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} z^{k_1} z^{k_2} \dots z^{k_m} \prod_{i=0}^m p_{k_i} \\ &= \sum_{k_1=0}^{\infty} z^{k_1} \dots \sum_{k_m=0}^{\infty} z^{k_m} \prod_{i=0}^m p_{k_i} = \left[\sum_{k=0}^{\infty} z^k p_k \right]^m \\ &= [g_0(z)]^m \quad . \end{aligned} \quad (6.9)$$

Isso quer dizer que pode-se calcular a distribuição de probabilidade da soma dos graus de m vértices escolhidos aleatoriamente de uma rede, dado uma função geradora da distribuição de graus (NEWMAN, 2010, p. 434). Esta é uma propriedade relevante para uso no estudo das redes e que será explorada na seção 6.6.

6.4 DISTRIBUIÇÃO DE EXCESSO DE GRAU

Se um vértice é escolhido aleatoriamente e seguindo por uma de suas arestas (assumindo que tenha pelo menos uma, claro) até um novo vértice, qual a probabilidade deste novo vértice ter grau k ? Já foi discutido na seção 2.2 na página 30 que um grafo não direcionado com m arestas possui duas extremidades para cada aresta, excluindo a primeira escolhida, isto é $2m - 1$, a aresta tem probabilidade $k/(2m - 1)$ de terminar em algum vértice de grau k . No limite de grafos de tamanho grande, o qual m se torna grande (assumindo que a distribuição de

grau e o grau médio se mantenham constantes) pode-se ignorar o -1 e escrever apenas $k/2m$.

Dado que p_k é a fração total de vértices no grafo com grau k , o número total de destes vértices é np_k e a probabilidade da aresta terminar em qualquer vértice com grau k é:

$$\frac{k}{2m} \times np_k = \frac{n}{2m} kp_k = \frac{kp_k}{\langle k \rangle} \quad , \quad (6.10)$$

o qual foi usada a equação (2.5).

A média do grau de um vizinho de um vértice é:

$$\sum_k k \frac{kp_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} \quad . \quad (6.11)$$

Pode-se demonstrar que a média do grau do vizinho é maior que a média do grau do grafo $\langle k \rangle$ fazendo uma conta simples:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{1}{\langle k \rangle} (\langle k^2 \rangle - \langle k \rangle^2) = \frac{\sigma_k^2}{\langle k \rangle} \quad , \quad (6.12)$$

onde σ_k^2 é a variância da distribuição de grau que deve ser positiva ou nula somente se todos os vértices possuem o mesmo grau. Portanto:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle > 0 \quad , \quad (6.13)$$

ou ainda:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle \quad (6.14)$$

(NEWMAN, 2010, p. 447).

Isto demonstra o porque a maioria das pessoas têm menos amigos em média do que os seus amigos têm, o chamado paradoxo das amizades (FELD, 1991).

O número de arestas fixado em um vértice, excluído aquele do qual se chega é chamado **excesso de graus** de um vértice e é apenas uma menos do total de graus. É possível calcular a distribuição de probabilidade do excesso de grau utilizando a equação (6.11). A probabilidade q_k de ter um excesso de grau k é a probabilidade de ter grau

total $k + 1$ e substituindo $k \rightarrow k + 1$ na equação (6.11):

$$q_k = \frac{(k + 1)p_{k+1}}{\langle k \rangle} . \quad (6.15)$$

Esta distribuição é chamada de **distribuição de excesso de graus** e é a distribuição de probabilidade do número de arestas fixadas a um vértice quando este é alcançado seguindo uma de suas arestas (NEWMAN, 2010, p. 449).

6.5 FUNÇÃO GERADORA DO EXCESSO DE GRAUS

A função geradora do excesso de graus é simplesmente:

$$g_1(z) = \sum_{k=0}^{\infty} q_k z^k , \quad (6.16)$$

basta substituir q_k da equação (6.15) em $g_1(z)$. Rescrevendo:

$$g_1(z) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k + 1)p_{k+1} z^k , \quad (6.17)$$

e fazendo a substituição $k + 1 \rightarrow k'$ e notando que a soma pode começar de zero sem mudar nada

$$g_1(z) = \frac{1}{\langle k \rangle} \sum_{k'=0}^{\infty} k' p_{k'} z^{k'-1} = \frac{1}{\langle k \rangle} g_0'(z) , \quad (6.18)$$

onde $g_0'(z)$ é a derivada da função geradora da distribuição de graus. Como $g_0'(1) = \langle k \rangle$ (eq. 6.3), ainda pode-se escrever:

$$g_1(z) = \frac{g_0'(z)}{g_0'(1)} , \quad (6.19)$$

o que demonstra que tendo $g_0(z)$ pode-se obter $g_1(z)$ diretamente (NEWMAN, 2010, p. 450).

6.6 FUNÇÃO GERADORA PARA PEQUENOS *CLUSTERS* E MÉTODO DA CAVITAÇÃO

Suponha que π_s seja a probabilidade de um vértice escolhido aleatoriamente fazer parte de um *cluster* não gigante de tamanho s . A função geradora de π_s é:

$$h_0(z) = \sum_{s=1}^{\infty} \pi_s z^s \quad , \quad (6.20)$$

onde o mínimo de s é um, pois um vértice sempre faz parte de um *cluster* de tamanho um.

Para obter a expressão de π_s é preciso antes demonstrar que os *clusters* não gigantes de um grafo são árvores (veja seção 2.4). Uma árvore com s vértices possui $s - 1$ arestas. O número total de lugares onde seria possível adicionar uma aresta extra ao componente é dado pelo número de pares de vértices distintos menos o número de pares já conectados:

$$\binom{s}{2} - (s - 1) = \frac{1}{2} \frac{s(s - 1)(s - 2)!}{(s - 2)!} - (s - 1) = \frac{1}{2} [(s - 1)(s - 2)] . \quad (6.21)$$

A probabilidade de dois vértices i e j estarem ligados, assumindo que o grau de ambos é não nulo, é dada por:

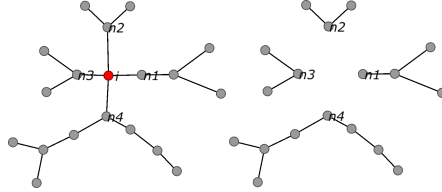
$$p_{ij} = \frac{k_i k_j}{(2m - 1)} = \frac{k_i k_j}{(n \langle k \rangle - 1)} \quad . \quad (6.22)$$

Supondo que os vértices i e j estão no mesmo *cluster*, a probabilidade de uma aresta extra é $(1/2)[(s - 1)(s - 2)]p_{ij}$. Assumindo que s cresce mais devagar que \sqrt{n} como é provado na seção 6.9, e que mantendo-se a distribuição p_k constante no limite de $n \rightarrow \infty$ a probabilidade dos *clusters* não gigantes não serem árvores tende a zero (NEWMAN, 2010, p. 457).

No processo de remover um vértice i de uma árvore juntamente com todas as suas arestas (fig. 17) o que se obtém são vários *clusters* não conectados entre si. O tamanho do *cluster* original, o qual i faz parte, é o tamanho dos *clusters* separados mais um, que é o vértice i . Os vizinhos n_1 , n_2 , n_3 e n_4 estão conectados por uma das arestas do vértice i , porém como a média de um grau vizinho é maior que a média do grafo (veja seção 6.4), os *clusters* não estão distribuídos de acordo com π_s , mas sim por uma distribuição de probabilidade ρ_s , que é a

probabilidade de que um vértice no extremo de uma aresta pertença a um *cluster* pequeno de tamanho s depois que essa aresta é removida.

Figura 17 – Retirada de um vértice numa árvore formando *clusters* isolados



Um pequeno *cluster*, que no limite de redes grandes é uma árvore. Retirando o vértice i é obtido vários *clusters* isolados, o qual segue uma distribuição de probabilidade ρ_s .

Fonte: Imagem criada pelo autor com o software Gephi.

Admitindo que o vértice i do grafo original tenha grau k e após este vértice ter sido retirado a probabilidade que seus k vizinhos pertençam a pequenos *clusters* de tamanho s_i é $P(s|k)$, cuja soma total dos pequenos *clusters* é igual a s . A probabilidade de i pertencer a um pequeno *cluster* de tamanho s dado que ele possui grau k é denotado por $P(s-1|k)$. A probabilidade π_s que i pertença a um pequeno *cluster* de tamanho s é:

$$\pi_s = \sum_{k=0}^{\infty} p_k P(s-1|k) \quad . \quad (6.23)$$

Substituindo a equação (6.23) na equação (6.20):

$$\begin{aligned} h_0(z) &= \sum_{s=1}^{\infty} \sum_{k=0}^{\infty} p_k P(s-1|k) z^s \\ &= z \sum_{s=1}^{\infty} \sum_{k=0}^{\infty} p_k P(s-1|k) z^{s-1} \quad . \end{aligned} \quad (6.24)$$

Fazendo uma substituição $s' = s - 1$

$$h_0(z) = z \sum_{k=0}^{\infty} p_k \sum_{s'=0}^{\infty} P(s'|k) z^{s'} \quad . \quad (6.25)$$

A expressão anterior pode ser reescrita em termos da probabilidade ρ_s utilizando o método da seção 6.3. O tamanho de pequenos clusters $s_i, i = 1, \dots, k$ são números aleatórios e independentes extraídos

da distribuição ρ_s . Definindo a função geradora para a distribuição ρ_s como:

$$h_1(z) = \sum_{s=1}^{\infty} \rho_s z^s \quad . \quad (6.26)$$

A expressão 6.25 fica:

$$h_0(z) = z \sum_{k=0}^{\infty} p_k \left[\sum_{s=0}^{\infty} \rho_s z^s \right]^k \quad (6.27)$$

$$h_0(z) = z \sum_{k=0}^{\infty} p_k [h_1(z)]^k \quad .$$

A função geradora $h_1(z)$ ainda é desconhecida. Considerando novamente a retirada de um vértice i do grafo, qual é a probabilidade ρ_s que um vizinho de i pertença a um pequeno cluster de tamanho s ? Usando o método da cavitação que consiste em afirmar que no limite de grafos de tamanho grande a retirada de um único vértice não afeta significativamente a distribuição de grau, portanto o grafo possui a mesma distribuição que antes da retirada, o que significa que se um vizinho possui grau k a probabilidade i pertencer a um pequeno clusters de tamanho s é $P(s-1|k)$ como anteriormente. Porém o vizinho não segue a distribuição p_k e sim a distribuição do excesso de grau q_k definida na equação (6.15). Logo:

$$\rho_s = \sum_{k=0}^{\infty} q_k P(s-1|k) \quad , \quad (6.28)$$

substituindo na equação (6.26)

$$h_1(z) = \sum_{s=1}^{\infty} \sum_{k=0}^{\infty} q_k P(s-1|k) z^s = z \sum_{s=0}^{\infty} q_k \sum_{k=0}^{\infty} P(s|k) z^s \quad . \quad (6.29)$$

Mais uma vez usando as potências da função geradora pode-se escrever a expressão em termos da probabilidade ρ_s ou $h_1(z)$. Portanto:

$$h_1(z) = z \sum_{s=0}^{\infty} q_k \left[\sum_{s=0}^{\infty} \rho_s z^s \right]^k \quad (6.30)$$

$$h_1(z) = z g_1(h_1(z)) \quad .$$

Resolvendo o sistema abaixo, sabe-se π_s :

$$h_0(z) = zg_0(h_1(z)) \quad , \quad (6.31)$$

$$h_1(z) = zg_1(h_1(z)) \quad (6.32)$$

(NEWMAN, 2010, p. 460).

6.7 CLUSTER GIGANTE

Dado que $h_0(z) = \sum_{s=1}^{\infty} \pi_s z^s$, onde π_s é a probabilidade de um vértice escolhido aleatoriamente pertencer a um pequeno *cluster* de tamanho s , $h_0(1) = \sum_{s=1}^{\infty} \pi_s$ é a probabilidade total de um vértice escolhido aleatoriamente pertencer a um pequeno *cluster*. Vale notar que diferentemente de $g_0(1) = 1$ e $g_1(1) = 1$, $h_0(1)$ não é necessariamente um, pois pode haver um *cluster* gigante no grafo e alguns vértices não pertencerão a nenhum pequeno *cluster*, portanto $\sum_{s=1}^{\infty} \pi_s$ será menor que um. De fato $h_0(1) = \sum_{s=1}^{\infty} \pi_s$ é a fração de vértices que pertencem a pequenos *cluster*. Denotando S como a fração de vértices de um *cluster* gigante temos:

$$S = 1 - \sum_{s=1}^{\infty} \pi_s = 1 - h_0(1) \quad . \quad (6.33)$$

Utilizando a equação (6.31):

$$S = 1 - g_0(h_1(1)) \quad , \quad (6.34)$$

o valor de $h_1(1)$ pode ser obtido da equação (6.32):

$$h_1(1) = g_1(h_1(1)) \quad . \quad (6.35)$$

Utilizando uma notação mais sucinta, escrevemos $h_1(1) = u$ e as equações (6.34) e (6.35) ficam:

$$S = 1 - g_0(u) \quad , \quad (6.36)$$

$$u = g_1(u) \quad . \quad (6.37)$$

Neste caso u é um ponto fixo da função $g_1(z)$, isto é, um ponto onde a função é igual ao seu próprio argumento. A função $g_1(z)$ sempre possui o ponto fixo em $u = 1$, pois $g_1(1) = 1$, mas a solução $S - g_0(1) = 0$ mostra que não há um *cluster* gigante. Porém se há um *cluster* gigante então deve haver um outro ponto fixo além de $u = 1$. A quantidade

$u = h_1(1) = \sum_{s=1}^{\infty} \rho_s$ é a probabilidade total de um vértice alcançado por uma aresta pertencer a um pequeno *cluster* de qualquer tamanho ou equivalente a probabilidade de não pertencer ao *cluster* gigante (NEWMAN, 2010, p. 462).

6.8 TAMANHO MÉDIO DO *CLUSTER* O QUAL UM VÉRTICE ESCOLHIDO ALEATORIAMENTE FAZ PARTE

Nesta seção é calculado $\langle r \rangle$, definido como o tamanho médio do *cluster* o qual um vértice escolhido aleatoriamente faz parte. Como os *clusters* gigantes têm mais vértices, as chances de escolher vértices que fazem parte dele é maior, portanto $\langle r \rangle$ diverge na transição de fase e é uma estimativa tendenciosa do tamanho médio real de um pequeno *cluster*. O cálculo de $\langle r \rangle$ é efetuado da seguinte forma:

$$\langle r \rangle = \frac{\sum_{r=1}^{\infty} r \pi_r}{\sum_{r=1}^{\infty} \pi_r} = \frac{h'_0(1)}{1-S} = \frac{h'_0(1)}{g_0(u)} \quad , \quad (6.38)$$

onde foi usada a equação (6.36) na igualdade final. Diferenciando a equação (6.31):

$$\begin{aligned} h'_0(z) &= g_0(h_1(z)) + z g'_0(h_1(z)) h'_1(z) \\ &= g_0(h_1(z)) + z g'_0(1) g_1(h_1(z)) h'_1(z) \\ &= \frac{h_0(z)}{z} + g'_0(1) h_1(z) h'_1(z) \quad , \end{aligned} \quad (6.39)$$

o qual usa-se a equação (6.19) na segunda igualdade e as equações (6.31) e (6.35) na terceira igualdade. Estabelecendo $z = 1$ e utilizando a expressão (6.33) junto com a notação $h_1(1) = u$.

$$\begin{aligned} h'_0(1) &= h_0(1) + g'_0(1) h_1(1) h'_1(1) \\ &= 1 - S + g'_0(1) u h'_1(1) \quad . \end{aligned} \quad (6.40)$$

Diferenciando a equação (6.32):

$$\begin{aligned} h'_1(z) &= g_1(h_1(z)) + z g'_1(h_1(z)) h'_1(z) \\ &= \frac{h_1(z)}{z} + z g'_1(h_1(z)) h'_1(z) \quad , \end{aligned} \quad (6.41)$$

rearranjando os termos de $h_1'(z)$:

$$\begin{aligned} h_1'(z)(1 - zg_1'(h_1(z))) &= \frac{h_1(z)}{z} \\ h_1'(z) &= \frac{h_1(z)/z}{(1 - zg_1'(h_1(z)))} \end{aligned} \quad (6.42)$$

Novamente fixando $z = 1$ e empregando $h_1(1) = u$:

$$h_1'(1) = \frac{u}{1 - g_1'(u)} \quad , \quad (6.43)$$

combinando as equações (6.40) e (6.43):

$$h_0'(1) = 1 - S + g_0'(1) \frac{u^2}{1 - g_1'(u)} \quad . \quad (6.44)$$

A equação para $\langle r \rangle$ é:

$$\begin{aligned} \langle r \rangle &= \frac{1 - S + \frac{g_0'(1)u^2}{1 - g_1'(u)}}{1 - S} \\ &= 1 + \frac{g_0'(1)u^2}{(1 - g_1'(u))(1 - S)} \\ \langle r \rangle &= 1 + \frac{g_0'(1)u^2}{(1 - g_1'(u))g_0(u)} \quad , \end{aligned} \quad (6.45)$$

onde usa-se a equação (6.36). Um caso simples para a equação (6.45) é quando não há um *cluster* gigante, ou seja, $S = 0$ e por definição $u = 1$, portanto a equação fica:

$$\langle r \rangle = 1 + \frac{g_0'(1)}{1 - g_1'(1)} \quad , \quad (6.46)$$

o qual usa-se que $g_0(1) = 1$. A equação (6.46) mede apenas o tamanho médio de pequenos *clusters* para um vértice aleatório, porém ela diverge para $g_1'(1) = 1$, exatamente onde há um *cluster* gigante, ou seja há uma transição de fase neste ponto.

6.9 TAMANHO MÉDIO DE UM PEQUENO *CLUSTER*

Seja n_s o número de *clusters* de tamanho s , o número de vértices que pertencem a *clusters* de tamanho s é sn_s e a probabilidade de escolher um vértice que pertence a um *cluster* deste tamanho é:

$$\pi_s = \frac{sn_s}{n} \quad , \quad (6.47)$$

onde n é o número total de vértices. Denotando $\langle s \rangle$ como tamanho médio de um pequeno *cluster* e usando a equação (6.33):

$$\langle s \rangle = \frac{\sum_{s=1}^{\infty} sn_s}{\sum_{s=1}^{\infty} n_s} = \frac{n \sum_{s=1}^{\infty} \pi_s}{n \sum_{s=1}^{\infty} \pi_s / s} = \frac{1 - S}{\sum_{s=1}^{\infty} \pi_s / s} \quad , \quad (6.48)$$

e para a soma que restou usa-se o seguinte artifício:

$$\int_0^1 \frac{h_0(z)}{z} dz = \sum_{s=1}^{\infty} \pi_s \int_0^1 z^{s-1} dz = \sum_{s=1}^{\infty} \frac{\pi_s}{s} \quad . \quad (6.49)$$

Pode-se obter $h_0(z)/z$ da equação (6.39):

$$\begin{aligned} \sum_{s=1}^{\infty} \frac{\pi_s}{s} &= \int_0^1 \frac{dh_0}{dz} dz - g'_0(1) \int_0^1 h_1(z) \frac{dh_1}{dz} dz \\ &= \int_0^{1-S} dh_0 - \langle k \rangle \int_0^u h_1 dh_1 \\ &= 1 - S - \frac{1}{2} \langle k \rangle u^2 \quad , \end{aligned} \quad (6.50)$$

onde é utilizado a equações (6.4) e (6.33). Portanto:

$$\langle s \rangle = \frac{1 - S}{1 - S - \frac{1}{2} \langle k \rangle u^2} = \frac{2}{2 - \langle k \rangle u^2 / (1 - S)} \quad . \quad (6.51)$$

Agora verifica-se a afirmação para a equação (6.21) de considerar que os pequenos *clusters* crescem mais devagar que \sqrt{n} é válida e ao contrário da equação (6.45) que diverge no ponto crítico em $S = 0$ e $u = 1$, a equação (6.51) não diverge, exceto para o caso particular $\langle k \rangle = 2$.

6.10 EQUAÇÃO DO *CLUSTER* GIGANTE

Se um vértice que não foi removido de uma rede fizer parte do *cluster* gigante então ele precisa estar ligado por algum de seus vizinhos e de forma equivalente, se ele não faz parte do *cluster* gigante, então não pode estar ligado ao *cluster* gigante por nenhum de seus vizinhos. Como definido anteriormente u é a probabilidade de um vértice não pertencer ao *cluster* gigante ligado por um vizinho particular, se ele possui grau k então a probabilidade total de não pertencer ao *cluster* gigante é u^k . É possível então calcular a média sobre a distribuição de probabilidade de graus p_k e achar a probabilidade média de não pertencer ao *cluster* gigante, isto é:

$$\sum_{k=0}^{\infty} p_k u^k \quad , \quad (6.52)$$

porém esta é a função geradora da distribuição de graus $g_0(u)$ e a probabilidade média do grafo de pertencer ao *cluster* gigante é simplesmente $1 - g_0(u)$. Mas esta é probabilidade média para vértices que não foram removidos, no caso da percolação de sítios, obviamente os que foram removidos não fazem parte do *cluster* gigante. Então a fração de vértices S no *cluster* gigante é igual a fração p de vértices não removidos vezes a probabilidade $1 - g_0(u)$:

$$S = p(1 - g_0(u)) \quad , \quad (6.53)$$

para o caso da percolação de ligação não é necessário levar em conta a fração p de vértices, pois altera-se apenas as arestas e a fração de vértice B para percolação de ligação é:

$$B = 1 - g_0(u) \quad . \quad (6.54)$$

Para o calculo de u é preciso analisar duas possibilidades de não estar ligado ao *cluster* gigante. Supondo um vértice A , no primeiro caso este vértice A ou a aresta em que chegamos a ele foi removido, ou seja, a probabilidade de remoção de um vértice ou aresta é $1 - p$. No segundo caso A está presente, isto é, um vértice ou aresta tem probabilidade p de estar presente, porém não faz parte do *cluster* gigante por nenhum de seus vizinhos, considerando que A tenha k vizinhos, a probabilidade de não estar ligado ao *cluster* gigante por nenhum dos vizinhos é u^k . Somando as duas probabilidade, a probabilidade total de que um vértice

não esteja ligado ao *cluster* gigante por A é $1 - p + pu^k$. Como A é alcançado através de uma aresta, k neste caso é distribuído através de excesso de grau definido na equação (6.15). Fazendo a média nesta distribuição multiplicado pela probabilidade de que ele não esteja ligado ao *cluster* gigante por um vértice A , tem-se a probabilidade média de u :

$$u = \sum_{k=0}^{\infty} q_k(1 - p + pu^k) = 1 - p + p \sum_{k=0}^{\infty} q_k u^k \quad (6.55)$$

$$u = 1 - p + pg_1(u) \quad ,$$

onde $g_1(u)$ é a função geradora para o excesso de grau. As equações (6.53) ou (6.54), dependendo do tipo de percolação, e (6.55) dão a solução completa para o tamanho do *cluster* gigante em um grafo.

Graficamente é possível achar a solução do ponto crítico da transição de fase utilizando a equação (6.55). Para isso utiliza-se a função $g_1(u)$, multiplicado por p e somado a $1 - p$. Isto é equivalente a comprimir o gráfico de $g_1(u)$ até que tenha altura p e deslocar para cima a uma distância $1 - p$. Como mencionado, $u = 1$ sempre será um ponto fixo para o qual não há *cluster* gigante. Porém há outro ponto fixo não trivial tal que $u < 1$. O ponto ou os pontos em que a linha $y = u$ cruza no gráfico resultante é a solução da equação (6.55). Quando $p = p_c$ ocorre que a equação $y(u) = 1 - p_c + p_c g_1(u)$ tangencia a reta $y = u$ e isso é o mesmo que escrever matematicamente:

$$\left. \frac{d}{du} \left(1 - p_c + p_c g_1(u) \right) \right|_{u=1} = 1 \quad . \quad (6.56)$$

Efetuando-se a derivada e isolando p_c o que se tem é:

$$p_c = \frac{1}{g_1'(1)} \quad . \quad (6.57)$$

Usando a definição da função geradora $g_1(z)$ dada pela equação (6.16)

junto com a equação (6.15):

$$\begin{aligned}
 g'_1(1) &= \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k+1)p_{k+1} = \frac{1}{\langle k \rangle} \sum_{k'=0}^{\infty} (k'-1)k'p_{k'} \\
 &= \frac{1}{\langle k \rangle} \left(\sum_{k=0}^{\infty} k^2 p_k - \sum_{k=0}^{\infty} k p_k \right) = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \quad ,
 \end{aligned}
 \tag{6.58}$$

portanto o ponto crítico na percolação é:

$$p_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad .
 \tag{6.59}$$

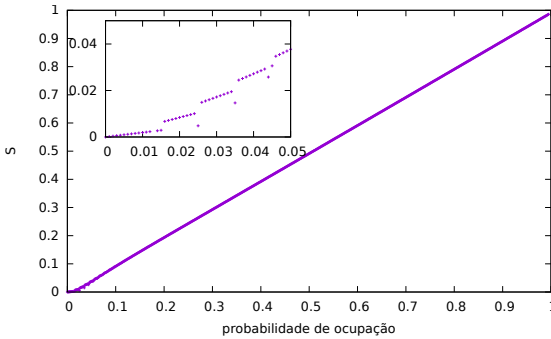
Este resultado vale para qualquer rede conectada aleatoriamente, independentemente da forma da distribuição de conectividade (COHEN et al, 2000).

6.11 CALCULO NUMÉRICO

O parâmetro de ordem usado para a transição de percolação é definido como o número de vértices pertencentes ao maior *cluster* da rede dividido pelo total de vértices no grafo. Aqui usei a letra S para a percolação de sítios e B para a percolação de ligação.

Primeiramente usei o cálculo numérico da equação (6.55) e a função geradora da distribuição de excesso de graus (eq. 6.17) do conectoma para encontrar os pontos fixos e então substituí na equação (6.53) para a percolação de sítios (fig. 18) e na equação (6.54) para a percolação de ligação (fig. 19) para obter onde o valor do *cluster* gigante é nulo, isto é, o ponto crítico. Em ambas as percolações o *cluster* gigante só é nulo na origem, porém devido ao efeito de tamanho finito o ponto crítico é deslocado do ponto crítico do limite termodinâmico (veja fig. 15), ou seja, é provável que o ponto crítico seja próximo da origem, mas não exatamente zero.

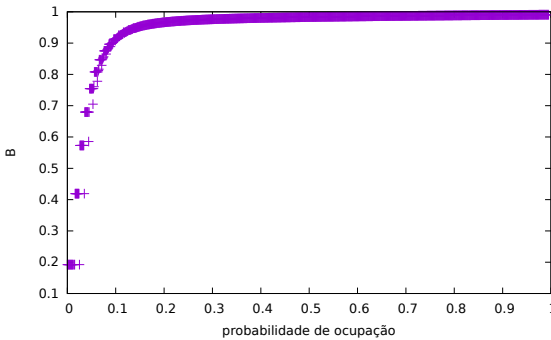
Figura 18 – Cálculo numérico da percolação de sítios para o conectoma



Cálculo numérico do sistema de equações 6.55 e 6.53 com a função geradora da distribuição de excesso de graus (eq. 6.17) do conectoma. Nota-se que o tamanho do *cluster* gigante só é nulo para a probabilidade de ocupação nula (inserção).

Fonte: Imagem criada pelo autor com o software gnuplot.

Figura 19 – Cálculo numérico da percolação de ligação para o conectoma



Cálculo numérico do sistema de equações 6.55 e 6.54 com a função geradora da distribuição de excesso de graus (eq. 6.17) do conectoma. Nota-se que o tamanho do *cluster* gigante só é nulo para a probabilidade de ocupação nula.

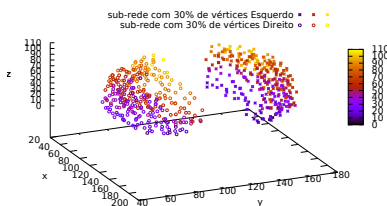
Fonte: Imagem criada pelo autor com o software gnuplot.

6.12 FSS NO CONECTOMA

Na seção 5.1 foi discutido o método de FSS, o qual utiliza um mesmo sistema com tamanhos finitos diferentes a fim de extrair os expoentes críticos de uma transição de fase de segunda ordem. Neste

trabalho algo semelhante foi feito, porém não é possível reproduzir uma rede experimental para tamanhos diferentes, o que pode ser feito é utilizar um grupo de renormalização geométrica, utilizando alguma invariância da rede sob uma mudança de escala para obter um grafo de tamanho menor (GARCIA-PÉREZ; BOGUÑA; SERRANO, 2018). Porém neste trabalho usei uma abordagem mais simples do que a teoria de grupos de renormalização. Para obter uma rede com tamanho menor utilizei recortes do conectoma nos eixos cartesianos da figura 7, obtendo subgrafos com 80%, 50% e 30% do número de vértices originais. Os cortes foram feitos nos eixos x , y e z . Em cada eixo eu encontro as extremidades projetadas do conectoma e então calculo o centro das projeções. A partir destes centros em cada eixo, vou variando o corte para um dos lados até obter a quantidade de vértices que gostaria. Por exemplo, 30% do número de vértices originais do conectoma é 299, na figura 20 o corte foi feito no eixo y e variando a partir do centro este corte para a esquerda até obter exatamente 299 vértices eu criei uma sub-rede representada pelos círculos, novamente a partir do centro variando o corte para a direita eu criei uma sub-rede com o mesmo número de vértices representada pelos asteriscos. O mesmo é feito nos outros dois eixos, criando 6 sub-redes com o mesmo tamanho, totalizando 18 sub-redes, 6 com 299 vértices, 6 com 499 vértices e 6 com 799 vértices. Porém, este processo não permite uma sub-rede com o mesmo número de arestas.

Figura 20 – Corte no eixo y do conectoma obtendo duas sub-redes com 30% de vértices do original



A partir do centro da projeção do eixo y , eu variei este valor para a esquerda até obter uma rede com exatamente 299 vértices (30% do original) mostrado pelos círculos. Depois a partir do centro novamente variei o valor para a direita até o obter uma rede com o mesmo número de vértices, mostrado pelos asteriscos. Este é um exemplo de duas sub-redes com 30% dos vértices do original obtida fazendo um corte no eixo y , o mesmo processo é feito nos outros eixos, obtendo 6 sub-redes com o mesmo número de vértices. As cores são para ajudar a visualizar o valor da componente z dos vértices.

Fonte: Imagem criada pelo autor com o software gnuplot.

6.13 SIMULAÇÃO NO CONECTOMA

Para as simular a retirada de vértices ou arestas na percolação utilizei um processo inverso, o qual é mais eficiente computacionalmente (NEWMAN; STROGATZ; WATTS, 2001). No início o que temos é uma rede vazia sem vértices ou arestas ocupadas, então são adicionados aleatoriamente os vértices ou arestas (dependendo da percolação) e monitora-se a evolução do maior *cluster* da rede. As percolações de sítio e de ligação foram medidas de formas diferentes. Para a percolação de sítios o maior *cluster* da rede foi medido como $Z(p)$ em função da probabilidade de ocupação $p = i/n$ onde i indica o número de vértices que são ocupados na evolução da simulação. Foram feitos 1000 processos independentes para cada rede, totalizando em $Q = 6000$ simulações para cada tamanho e estimado o valor de $S(p)$ como:

$$S(p) = \frac{1}{nQ} \sum_{q=1}^Q Z_q(p) \quad , \quad (6.60)$$

e a susceptibilidade da percolação de sítios $\chi_S(p)$:

$$\chi_S(p) = n \frac{(1/n^2Q) \sum_{q=1}^Q Z_q(p)Z_q(p) - [S(p)]^2}{S(p)} \quad , \quad (6.61)$$

$Z_q(p)$ é o maior *cluster* observado durante a q -ésima simulação quando a probabilidade de ocupação é p .

Para a percolação de ligação o maior *cluster* da rede, Z_r , foi medido não em função de p , mas da quantidade de arestas r ocupadas. Isto porque as sub-redes possuem diferentes quantidade de arestas como já mencionado e para se obter a média foi necessário que todas as simulações estivessem no mesmo intervalo, pois se utilizasse $p = i/m$, sendo i o número de arestas que são ocupados na evolução da simulação, as sub-redes com mais arestas possuiriam mais pontos, dificultando obter a média. Como cada aresta é ocupada com probabilidade p , a probabilidade que exatamente r arestas estão ocupadas é dado por uma distribuição binomial (NEWMAN, 2010, p. 617). Foram feitas $Q = 50$ processo independentes para cada sub-rede e $Q > 200$ para o conectoma original. O número de simulações na percolação de ligação é menor que o da percolação de sítio por demandar mais tempo de computação, isto é, na percolação de ligação é preciso adicionar em cada simulação da rede original 17865 arestas, comparado com a percolação de sítios

que é necessário adicionar apenas 998 vértices em cada simulação da rede original. A média do maior *cluster* para cada rede com r arestas ocupadas, \bar{Z}_r , é:

$$\bar{Z}_r = \frac{1}{Q} \sum_{q=1}^Q Z_r^q \quad , \quad (6.62)$$

Z_r^q indica o maior tamanho de *cluster* observado na rede ou sub-rede durante a q -ésima simulação quando r arestas estão ocupadas. $Z(p)$ é então:

$$Z(p) = \sum_{r=0}^m p_r \bar{Z}_r = \sum_{r=0}^m \binom{m}{r} p^r (1-p)^{m-r} \bar{Z}_r \quad , \quad (6.63)$$

por fim a média das 6 sub-redes (para o conectoma original não foi necessário a soma, apenas a divisão pelo número de vértices):

$$B(p) = \frac{1}{n6} \sum_{i=1}^6 Z_i(p) \quad , \quad (6.64)$$

$Z_i(p)$ indica o maior tamanho de *cluster* observado na média das sub-redes durante a q -ésima simulação quando a probabilidade de ocupação é p . A susceptibilidade da percolação de ligação χ_B foi medida utilizando as etapas:

$$Z^{(2)}(p) = \sum_{r=0}^m p_r \bar{Z}_r \bar{Z}_r = \sum_{r=0}^m \binom{m}{r} p^r (1-p)^{m-r} \bar{Z}_r \bar{Z}_r \quad , \quad (6.65)$$

e para as sub-redes (para a rede original não foi feita a soma abaixo):

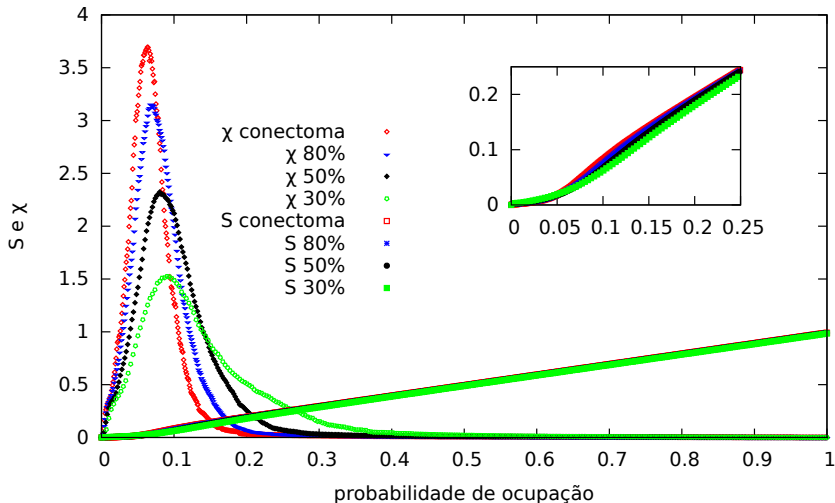
$$\chi_B(p) = \frac{1}{6} \sum_{q=1}^6 \frac{(1/n^2) Z_i^{(2)}(p) - [Z_i(p)/n]^2}{Z_i(p)} \quad . \quad (6.66)$$

A simulação também monitora o tamanho médio dos *clusters* isolados, isto é, a média de todos os *clusters* exceto o maior deles, denotado por $\langle s \rangle$ e $\langle b \rangle$ para as percolações de sítio e de ligação respectivamente.

Na figura 21 e na figura 22 é mostrada a rede original junto com as médias de todas as simulações para as redes parciais do parâmetro de ordem S e B e a susceptibilidade em função da probabilidade de percolação p para a percolação de sítios e para percolação de ligação respectivamente. É possível observar a tendência das flutuações, traduzidas na susceptibilidade χ , de crescer no ponto crítico, ou seja, no

limite termodinâmico ela deve divergir com uma lei de potência. Nos gráficos também é possível ver que o parâmetro de ordem aproxima-se de zero a medida que a rede cresce, basta observar a tendência das curvas de se inclinar mais no ponto de transição e no limite termodinâmico tem-se $S(p \leq p_c) = 0$, $S(p > p_c) > 0$, $B(p \leq p_c) = 0$ e $B(p > p_c) > 0$.

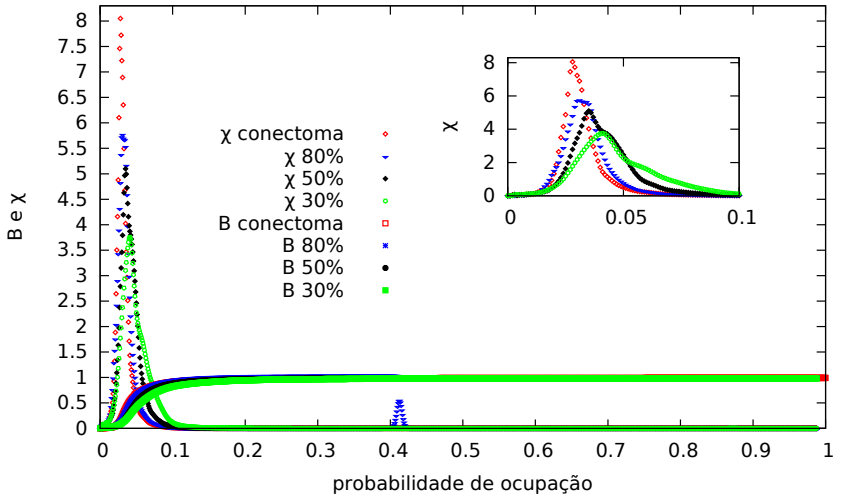
Figura 21 – Parâmetro de ordem χ e susceptibilidade em função de p para percolação de sítios



No gráfico as médias da susceptibilidades e do parâmetro de ordem da percolação em função de p para as sub-redes e o conectoma. Na inserção é possível notar que as curvas de $S(p)$ tendem a se inclinar mais a medida que a rede é maior, tendo no limite termodinâmico $S(p \leq p_c) = 0$ e $S(p > p_c) > 0$. Também observa-se a tendência das flutuações mostradas pela variável $\chi(p)$ de crescer no ponto crítico para as redes maiores, tendo no limite termodinâmico $\chi(p = p_c) \rightarrow \infty$.

Fonte: Imagem criada pelo autor com o software gnuplot.

Figura 22 – Parâmetro de ordem e susceptibilidade em função de p para percolação de ligação



No gráfico as médias da susceptibilidades e do parâmetro de ordem de percolação de ligação em função de p para o conectoma e as sub-redes. Nota-se que as curvas de $B(p)$ tendem a se inclinar mais a media que a rede é maior, tendo no limite termodinâmico $B(p \leq p_c) = 0$ e $B(p > p_c) > 0$. Também observa-se a tendência de $\chi(p)$ crescer para as redes maiores, tendo no limite termodinâmico $\chi(p = p_c) \rightarrow \infty$.

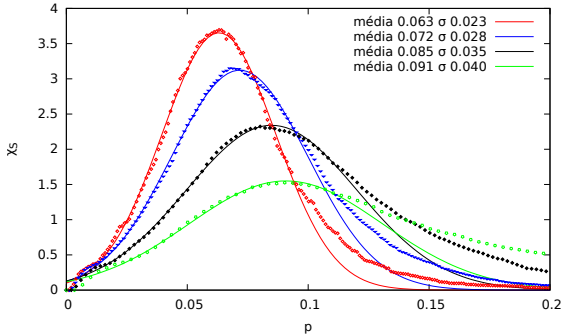
Fonte: Imagem criada pelo autor com o software gnuplot.

Para se obter uma estimativa do expoente crítico do comprimento de correlação usa-se o ponto de χ máximo para cada rede e com os pontos constrói-se em um gráfico log-log o pseudo ponto crítico $p_c(N)$ para cada tamanho de rede menos o ponto crítico p_c no limite termodinâmico versus o tamanho de rede N (RADICCHI; CASTELLANO, 2015).

Usando um ajuste da função gaussiana no ponto máximo de χ para cada tamanho de rede (figs. 23 e 24) e utilizando a média da gaussiana como o pseudo ponto crítico associado com desvio padrão (barras de erro nas figuras 25 e 26), sendo o ponto crítico no limite termodinâmico como zero dado pelo calculo numérico das equações (6.55), (6.53) e (6.54), mesmo que a equação (6.59) apresente outro valor, o segundo momento de uma rede finita não diverge embora seja um valor grande (NEWMAN, 2010, p. 259), aqui o valor do segundo momento do conectoma $\langle k^2 \rangle = 1536.3$ é tratado como infinito e portanto pela equação (6.59) $p_c = 0$, porém p_c no limite termodinâmico é desconhe-

cido e o valor $p_c = 0$ é apenas uma aproximação.

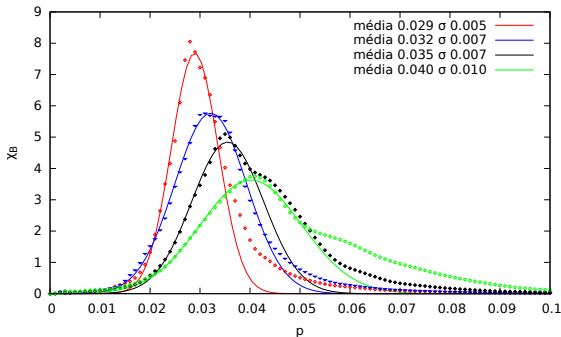
Figura 23 – Ajuste de funções gaussianas para as susceptibilidades de percolação de sítios



Função gaussiana ajustada para ter o pseudo ponto crítico $p_c(N)$ como a média da função ajustada e o erro associado como o desvio padrão σ (barras de erro na figura 26). As cores são as mesmas da figura 21.

Fonte: Imagem criada pelo autor com o software gnuplot.

Figura 24 – Ajuste de funções gaussianas para as susceptibilidades de percolação de ligação



Função gaussiana ajustada para ter o pseudo ponto crítico $p_c(N)$ como a média da função ajustada e o erro associado como o desvio padrão σ (barras de erro na figura 26). As cores são as mesmas da figura 22.

Fonte: Imagem criada pelo autor com o software gnuplot

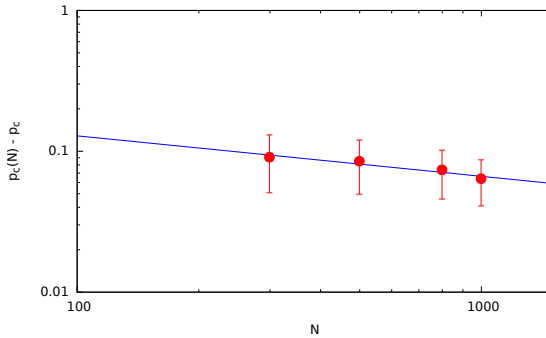
Para a percolação de sítios o expoente crítico do comprimento de correlação encontrado foi $1/\nu = 0.28 \pm 0.06$ (fig. 25) e para a percolação de ligação $1/\nu = 0.27 \pm 0.03$ (fig. 26), ou seja um valor próximo

a $\nu \sim 3$, expoente crítico do comprimento de correlação para uma rede livre de escala com $\lambda > 4$ (COHEN; BEN-AVRAHAM; HAVLIN, 2002). Os valores são apenas aproximações dos reais valores, sendo p_c no limite termodinâmico, utilizado para a estimativa do expoente crítico do comprimento de correlação, desconhecido.

Para descobrir o expoente crítico do parâmetro de ordem é desenhado em um gráfico log-log o valor do parâmetro no pseudo ponto crítico versus o tamanho da rede N (RADICCHI; CASTELLANO, 2015), no caso $S(p_c(N))$ e $B(p_c(N))$ (figuras 27 e 28).

Para a percolação de sítios $\beta/\nu = 0.46 \pm 0.09$ e para a percolação de ligação $\beta/\nu = 0.49 \pm 0.22$. Isso mostra que o decaimento pode ser $\beta \sim 1$, expoente crítico de percolação de campo médio, que pode ser obtido exatamente numa rede de Bethe (STAUFFER e AHARONY, 1991, p. 31) ou uma rede livre de escala com $\lambda > 4$ (COHEN; BEN-AVRAHAM; HAVLIN, 2002), embora seja difícil de confirmar devido ao erros serem muito grandes.

Figura 25 – Gráfico log-log $p_c(N) - p_c$ versus tamanho da rede N para percolação de sítios



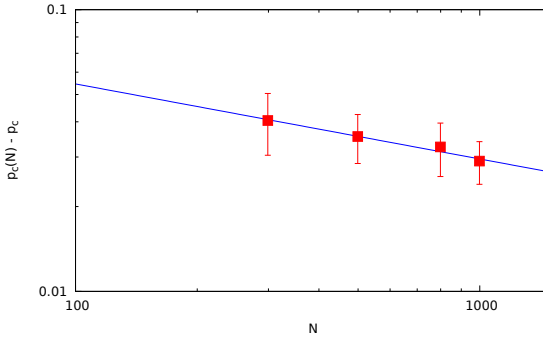
Utilizando o ponto crítico do limite termodinâmico $p_c = 0$, uma aproximação dada pelo cálculo numérico da equação do *cluster* gigante e $p_c(N)$ como a média de uma função gaussiana ajustada no ponto máximo de χ para cada média de tamanho de rede e o erro como o desvio padrão da mesma função gaussiana (fig. 23) estimei o expoente crítico do comprimento de correlação para a percolação de sítios igual a $1/\nu = 0.28 \pm 0.06$.

Fonte: Imagem criada pelo autor com o software gnuplot.

Podemos multiplicar a equação 6.54 por p e recuperamos a equação 6.53, isto é, os parâmetro de ordem B e S estão relacionados por:

$$S = pB \quad , \quad (6.67)$$

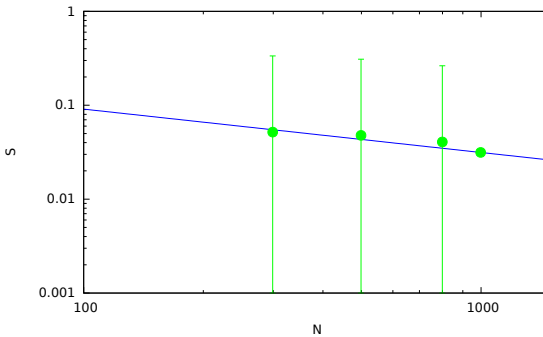
Figura 26 – Gráfico log-log $p_c(N) - p_c$ versus tamanho da rede N para percolação de ligação



Utilizando o ponto crítico do limite termodinâmico $p_c = 0$, uma aproximação dada pelo cálculo numérico da equação do *cluster* gigante e $p_c(N)$ como a média de uma função gaussiana ajustada no ponto máximo de χ para cada média de tamanho de rede e o erro como o desvio padrão da mesma função gaussiana (fig. 24) estimei o expoente crítico do comprimento de correlação para a percolação de ligação igual a $1/\nu = 0.27 \pm 0.03$.

Fonte: Imagem criada pelo autor com o software gnuplot.

Figura 27 – Gráfico log-log do parâmetro de ordem no pseudo-ponto crítico $S(p_c(N))$ versus o tamanho da rede N

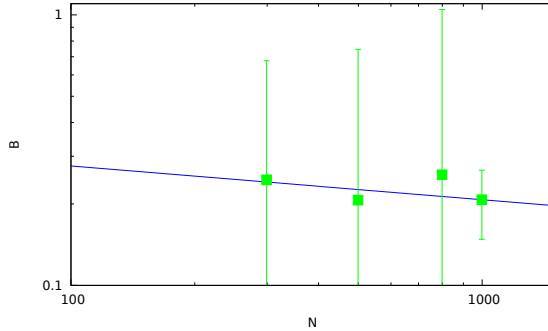


Usando o valor de $S(p_c(N))$ para cada tamanho de rede no pseudo ponto crítico $p_c(N)$ o expoente crítico do parâmetro de ordem $\beta/\nu = 0.46 \pm 0.09$ para a percolação de sítios.

Fonte: Imagem criada pelo autor com o software gnuplot.

o qual demonstra que a percolação de sítios e de ligação são linearmente proporcionais, porém esta equação é válida apenas para o limite em

Figura 28 – Gráfico log-log do parâmetro de ordem no pseudo-ponto crítico $B(p_c(N))$ versus o tamanho da rede N

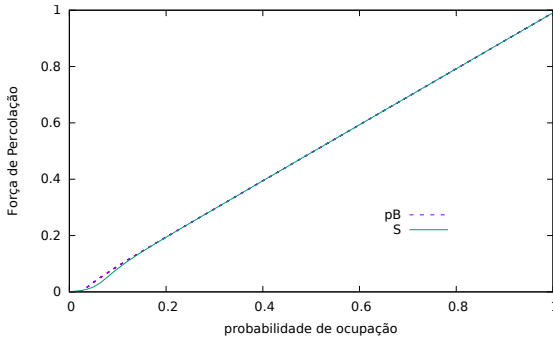


Usando o valor de $B(p_c(N))$ para cada tamanho de rede no pseudo ponto crítico $p_c(N)$ o expoente crítico do parâmetro de ordem para a percolação de ligação $\beta/\nu = 0.49 \pm 0.22$, muito próximo da percolação de sítios.
 Fonte: Imagem criada pelo autor com o software gnuplot.

que que os *clusters* não gigantes de um grafo são árvores (RADICCHI; CASTELLANO, 2015). Como no ponto crítico os parâmetros de ordem de percolação se comportam da forma $S \sim (p - p_c)^{\beta_s}$ para a percolação de sítio e $B \sim (p - p_c)^{\beta_b}$ para a percolação de ligação, a partir da equação (6.67) podemos escrever $(p - p_c)^{\beta_s} \sim p(p - p_c)^{\beta_b}$ e se o valor crítico é estritamente não nulo, os expoentes críticos da percolação de sítios e da percolação de ligação são iguais. Porém para o caso de $p_c = 0$, $p^{\beta_s} \sim p^{\beta_b+1}$, isto é, $\beta_s = \beta_b + 1$. Na figura 29 fiz um gráfico do parâmetro de ordem S e do parâmetro de ordem B multiplicada por p , verificando que a equação (6.67) se sustenta e como os expoentes críticos encontrados nos dois tipos de percolação são praticamente os mesmos, pode-se dizer que o ponto crítico é próximo de zero, mas não nulo.

Futuramente utilizando técnicas mais sofisticadas pretendo encontrar o real valor do ponto crítico no limite termodinâmico e por sua vez estimar os reais valores dos expoentes críticos de comprimento de correlação e do parâmetro de ordem.

Figura 29 – Relação entre as parâmetro de ordem da percolação de sítio e de ligação



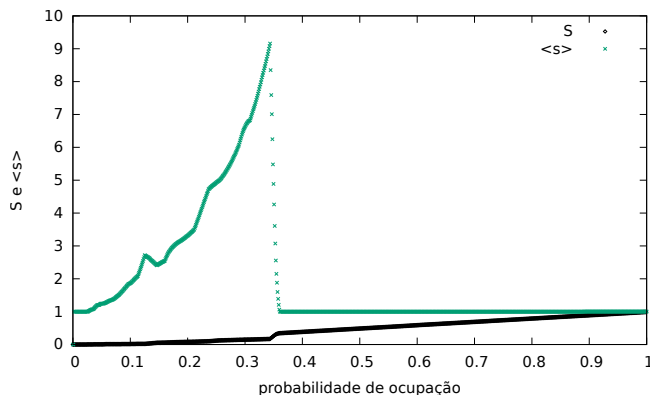
No gráfico a percolação de sítios (linha sólida) e a percolação de ligação multiplicada pela probabilidade de ocupação p (linha pontilhada).

Fonte: Imagem criada pelo autor com o software gnuplot.

Além da percolação com a retirada aleatória dos vértices e arestas, outro tipo percolação simulada foi o ataque, o qual os vértices com maior grau são primeiramente retirados, se mais vértices possuem o mesmo grau então escolhe-se aleatoriamente um deles para se retirar primeiro. O processo utilizado para a simulação foi a mesma da percolação de sítios, tendo no início uma rede vazia, porém desta vez os vértices não são adicionados aleatoriamente, mas numa ordem crescente de graus. Na figura 30 pode-se observar que o ataque é mais efetivo em quebrar a rede comparando com a retirada aleatória dos vértices. Na percolação com remoção aleatória é preciso que se retire quase 97% (sendo o ponto crítico do conectoma $p_c = 0.03$) dos vértices para quebrar a rede original em *clusters* isolados, enquanto que no ataque é necessário retirar 65% ($p_c = 0.35$) dos vértices com maior grau da rede original, mas este resultado demonstra que o conectoma é ainda muito robusto ao ataque devido a rede ser de 'mundo pequeno' se comparado as redes livres de escala (ACHARD et al, 2006). Optou-se por usar o tamanho médio dos *clusters* isolados, definido como a média de todos os *clusters* exceto o maior deles (ALBERT; BARABÁSI; JEONG, 2000), para demonstrar a transição de fase pelo menor nível de flutuação dos valores médios.

E a última simulação de percolação foi um tipo de ataque usando o mesmo procedimento da percolação de ligação no entanto retira-se primeiro as arestas com maior peso. Esta simulação mostra que o ponto crítico decresce, enquanto que na percolação de ligação com remoção

Figura 30 – Ataque no conectoma

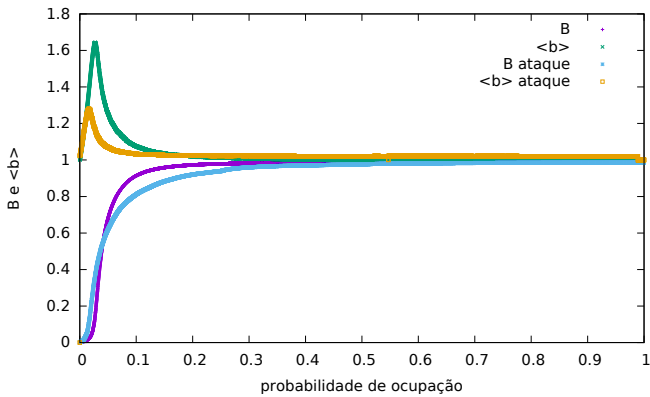


No gráfico o parâmetro de ordem S e o tamanho médio dos *clusters* isolados $\langle s \rangle$, definido como a média de todos os *clusters* exceto o maior deles em função de p , porém os vértices de maior grau são retirados primeiramente. Em comparação com a percolação com remoção aleatória, p_c do ataque é maior, sendo que na primeira é preciso que se retire quase 97% ($p_c = 0.03$) dos vértices do conectoma e no ataque é necessário retirar 65% dos vértices com maior grau.

Fonte: Imagem criada pelo autor com o software gnuplot.

aleatória $p_c \approx 0.030$ para a rede original, no ataque as arestas $p_c = 0.016$ (fig. 31). Novamente optou-se por usar o tamanho médio dos *clusters* isolados pelo mesmo motivo do ataque na percolação de sítios.

Figura 31 – Ataque de arestas no conectoma



No gráfico as parâmetro de ordem e o tamanho médio dos *clusters* isolados (b), definido como a média de todos os *clusters* exceto o maior deles em função de p .

No ataque as arestas de maior peso são retirados primeiramente. Opondo a percolação com remoção aleatória de arestas, o qual $p_c \approx 0.030$, no ataque $p_c = 0.016$.

Fonte: Imagem criada pelo autor com o software gnuplot.

7 MODELO DE LIMIAR LINEAR

Neste capítulo apresento outra parte da pesquisa, descrevendo um modelo simples para simular uma difusão na rede. A proposta de usar o modelo é baseado na ideia de Alan Turing (TURING, 1950). Em 1950 o cientista da computação escreveu a possibilidade de "injetar" uma ideia em uma máquina com inteligência e ela responder de alguma maneira e retornar a quiescência. Ele associa o evento a uma pilha de blocos de material radioativo, o qual a ideia injetada seria um nêutron atingindo a pilha. O nêutron causa uma perturbação que eventualmente se desfaz e se o tamanho do aglomerado é aumentado até uma massa crítica, a perturbação causada pelo nêutron vai causar uma reação em cadeia até destruir a pilha. Correspondendo a esta analogia da pilha de blocos de material radioativo, a maioria das mentes humanas parecem ser "subcríticas" e uma ideia apresentada para estas mentes vai, em média, dar origem a menos de uma ideia como resposta. Mentes supercríticas por outro lado, com uma ideia apresentada, podem dar origem a uma teoria inteira consistindo de muitas ideias.

O modelo de limiar linear (LTM, do inglês *Linear Threshold Model*) foi inicialmente proposto pelo sociólogo Mark Granovetter, o qual um agente adota um estado dependendo da proporção dos estados de seus vizinhos, podendo ser aplicado para difusão de inovações, disseminação de rumores e doenças, adoção de greves, decisão de voto, migração entre outros (GRANOVETTER, 1978). Aqui, como no trabalho de MIŠIĆ et al (2015), o LTM foi utilizado para descrever uma difusão no cérebro, o qual uma perturbação se espalha dependendo da proporção de vizinhos que também foram perturbados. O que é feito é ativar o sistema em um único vértice e ver se a rede se ativa, como a ideia injetada descrita por Turing. A ativação feita na matriz do conectoma é caracterizada pela variável de estado v que pode estar ativo (1) ou inativo (0) para cada vértice em um tempo discreto t , isto é, $v_i(t) = \{0, 1\}$. Uma vez que o vértice adota um estado ativo ele não volta para estado inativo. Para $t = 0$ toda a rede se encontra inativa, exceto por um vértice escolhido aleatoriamente que inicia com o estado ativo, chamado de semente. A rede possui atualização síncrona, isto é, todos os estados são atualizados simultaneamente de acordo com as equações abaixo:

$$v_i(t+1) = \begin{cases} 1 & \text{se } \theta k_{w_i} < \sum_{j \in N_i} v_j(t) \\ 0 & \text{outro caso} \end{cases}, \quad (7.1)$$

onde k_{w_i} representa o grau ponderado do vértice i , isto é, a soma de todos pesos das arestas que ligam o vértice i , N_i é o conjunto de vizinhos de i e θ é um parâmetro do modelo que controla o limiar necessário de vizinhos ativos para que o vértice i seja ativado. Diferentemente do trabalho de MISIĆ et al (2015), o qual o parâmetro θ é mantido fixo para um valor em que qualquer semente sempre ative toda a rede, pois o que se procura é o tempo de adoção para um vértice se ativar, aqui eu vario o θ . Uma única semente pode ativar o sistema inteiro ou então limitar-se a apenas perturbar os primeiros vizinhos, o que de alguma forma parece ser uma transição de primeira ordem, porém, tomando a média de várias simulações independentes, com as sementes escolhidas aleatoriamente, e variando o θ é possível ver uma transição de fase de segunda ordem, o qual média das flutuações é aumentada no ponto crítico. Para cada semente é medida a densidade de sítios D ativados após um tempo máximo ($Tmax$) para que a difusão possa se propagar por toda a rede:

$$D = \frac{1}{n} \sum_{i=0}^n v_i(Tmax) \quad , \quad (7.2)$$

onde $v_i(Tmax)$ é o estado do vértice i no tempo máximo da duração da simulação e n é o número total de vértices. O parâmetro de ordem A neste caso é a média da densidade de sítios das Q simulações com sementes escolhidas aleatoriamente:

$$A = \frac{1}{Q} \sum_{i=0}^Q D_i \quad , \quad (7.3)$$

onde D_i é a densidade de sítios ativados na i -ésima simulação. As flutuações são medidas como a susceptibilidade dada por:

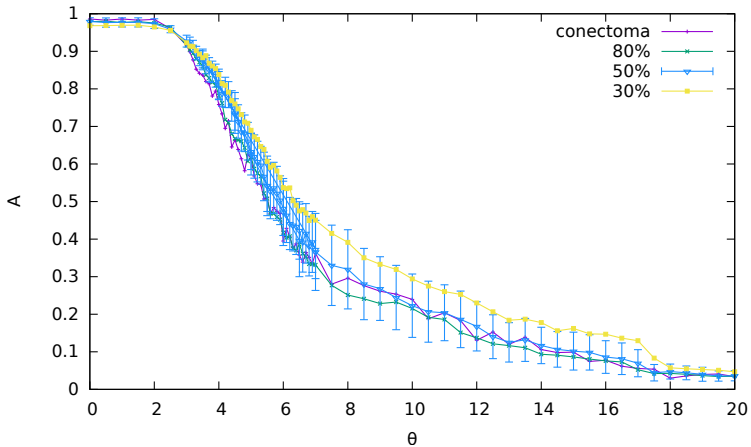
$$\chi = n \frac{(1/Q) \sum_{i=0}^Q A_i A_i - \langle A \rangle^2}{\langle A \rangle} \quad , \quad (7.4)$$

o qual sempre diverge no ponto crítico para processos de contato (FERREIRA; CASTELLANO; PASTOR-SATORRAS, 2012).

Como descrito no capítulo 5 é preciso utilizar mais de uma rede com tamanhos diferentes para se detectar a transição de fase de segunda

ordem. Usando as redes descritas na seção 6.12 e tomando a média do parâmetro de ordem e da susceptibilidade das sub-redes com o mesmo número de vértices é possível caracterizar a transição. Neste estudo, utilizei $Q=2000$ sementes escolhidas aleatoriamente. Na figura 32 vemos que o parâmetro de ordem vai assintoticamente a zero próximo ao ponto crítico a medida que a rede é maior, melhor visualizado na figura 33 como sendo $\theta_c \approx 5$. Já a susceptibilidade, que mede as flutuações

Figura 32 – Parâmetro de ordem do modelo LTM em função de θ



No gráfico o parâmetro de ordem do modelo LTM para o conectoma e as sub-redes descritas na seção 6.12. As linhas são apenas um guia para a melhor visualização.

Pode-se observar o parâmetro indo assintoticamente a zero no ponto crítico a medida que a rede cresce. O erro associado é dado pelo desvio padrão, mostrado apenas para a sub-rede com 50% dos vértices originais.

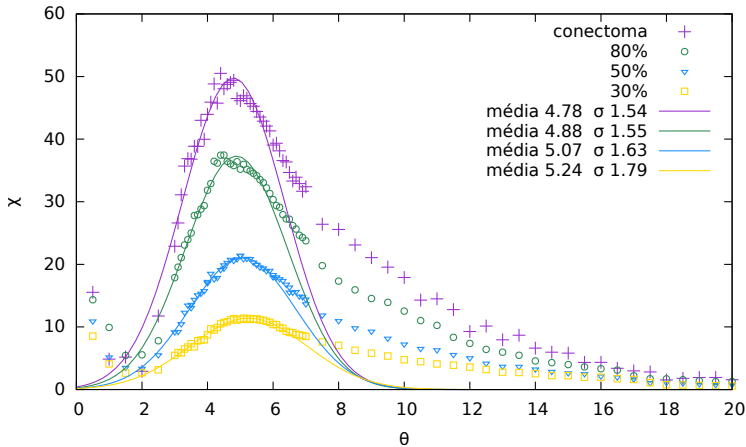
Fonte: Imagem criada pelo autor com o software gnuplot.

do sistema, mostradas na figura 33, cresce no ponto crítico $\theta_c \approx 5$, o qual deve divergir no limite termodinâmico.

Utilizando o pseudo ponto crítico $\theta_c(N)$ como sendo a média de um ajuste da função gaussiana (figura 33) e o ponto crítico no limite termodinâmico θ_c como o pseudo ponto crítico da maior rede, neste caso o conectoma original. Usando o pseudo ponto cr obtive uma estimativo de expoente crítico do comprimento de correlação $1/\nu = 1.55 \pm 0.48$ exibido na figura 34, isto é ν está entre o expoente de correlação espacial da percolação direcionada com dimensão 2 ($1/\nu = 1.3636(33)$) e com dimensão 3 ($1/\nu = 1.71(0)$) (HENKEL; HIRINCHSEN; LÜBECK, 2008, p. 159).

Utilizando o valor de A no pseudo ponto crítico e fazendo um

Figura 33 – Susceptibilidade do modelo LTM em função de θ e ajuste de uma função gaussiana



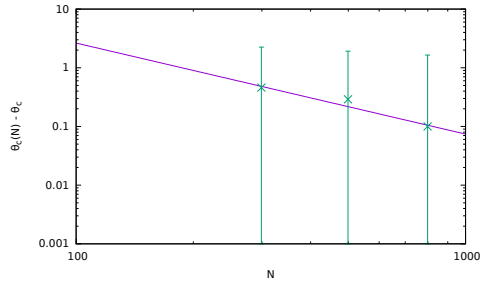
No gráfico a susceptibilidade do parâmetro de ordem do modelo LTM para o conectoma e as sub-redes descritas na seção 6.12. A susceptibilidade é a variável que mede as flutuações que divergem assintoticamente no ponto crítico a medida que a rede cresce.

Fonte: Imagem criada pelo autor com o software gnuplot.

gráfico log-log em função do tamanho da rede obteve o expoente crítico do parâmetro de ordem, mostrado na figura 35, sendo $\beta = 0.059$, muito longe do expoente crítico da percolação direcionada para qualquer dimensão ($\beta = 0.276486(8)$ para dimensão 1, $\beta = 5834(30)$ para dimensão 2 e $\beta = 0.813(9)$ para dimensão 3).

O modelo apresentou uma transição da fase de segunda ordem tendo um ponto crítico $\theta_c \approx 5$. Uma pergunta que poderia ser feita é se o valor deste ponto crítico ou os expoentes críticos da transição de fase, mudam para diferentes cérebros, podendo classificá-los em subcrítico ou supercrítico ou então se é diferente para cérebros saudáveis ou doentes, podendo ser parâmetros para diagnosticar alguma doença? Para responder esta pergunta é necessário ter várias redes obtidas da mesma forma do conectoma, com pacientes doentes e grupos de controle para então simular o LTM e verificar a hipótese.

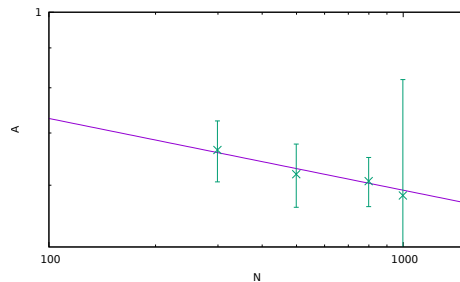
Figura 34 – Gráfico log-log $\theta_c(N) - \theta_c$ versus tamanho da rede N para o modelo LTM



No gráfico temos uma estimativa do expoente de comprimento de correlação $1/\nu = 1.55 \pm 0.48$, sendo $\theta_c(N)$ o pseudo ponto crítico no modelo LTM como a média de uma função gaussiana ajustada no ponto máximo da susceptibilidade do parâmetro de ordem (figura 33) e o ponto crítico θ_c no limite termodinâmico como o pseudo ponto crítico para a maior rede, neste caso o conectoma original.

Fonte: Imagem criada pelo autor com o software gnuplot.

Figura 35 – Gráfico log-log do parâmetro de ordem no pseudo-ponto crítico $A(\theta_c(N))$ versus o tamanho da rede N



Usando o valor de $A(\theta_c(N))$ para cada tamanho de rede no pseudo ponto crítico $\theta_c(N)$ o expoente crítico do parâmetro de ordem para a percolação de ligação

$$\beta/\nu = 0.092 \pm 0.018.$$

Fonte: Imagem criada pelo autor com o software gnuplot.

8 CONCLUSÃO E PERSPECTIVAS FUTURAS

Este trabalho contém noções da teoria dos grafos e visava apresentar simulações em uma rede estrutural do cérebro saudável originada experimentalmente através de técnicas não evasivas de imagem. Nesta dissertação mostrei que o conectoma não pode ser modelado por uma rede aleatória e que está mais próximo de ser modelado por uma rede livre de escala. Através das simulações das percolações e da teoria de transição de fase abordadas, e embora não usei a teoria de grupos de renormalização para utilizar a técnica do FSS com rigor, os resultados convergem para expoentes críticos de uma rede livre de escala com $\lambda > 4$, pois como apresentado no capítulo 6 o que se encontrou para os dois tipos de percolação foi um valor de expoente crítico do comprimento de correlação $\nu \sim 3$ e expoente crítico do parâmetro de ordem $\beta \sim 1$ (COHEN; BEN-AVRAHAM; HAVLIN, 2002). Como os expoentes são os mesmos para as duas percolações, caso o ponto crítico no limite termodinâmico fosse estritamente nulo haveria uma quebra da universalidade da percolação de ligação e de sítios (RADICCHI; CASTELLANO, 2015) e também as redes livres de escala com $\lambda > 4$ possuem ponto crítico não nulo, o qual chega-se a conclusão que o ponto crítico do conectoma no limite termodinâmico deve ser próximo, mas não exatamente zero. Como foi usando $p_c = 0$ para o valor do ponto crítico no limite termodinâmico a fim de obter os expoentes críticos de comprimento de correlação, é necessário realizar novos cálculos para obter os expoentes de forma mais precisa além do real valor do ponto crítico no limite termodinâmico. A simulação do ataque de vértices também revelou que o conectoma é robusto não só para a retirada aleatória dos vértices e arestas, tendo que retirar mais de 90% dos elementos da rede para quebrá-la, mas também, se comparado com as redes livre de escala, para a retirada dos vértices de maior grau, sendo necessário a retirada de 65% desses vértices para que o grafo sinta os efeitos, resultado devido a rede ser de 'mundo pequeno' (ACHARD et al, 2006). Um resultado interessante foi o ataque a arestas de maior peso, diminuindo o ponto crítico da probabilidade de ocupação, sendo que na remoção de arestas aleatórias $p_c \approx 0.030$ quando que no ataque $p_c = 0.016$.

Na outra parte do trabalho, usando o modelo de limiar linear foi possível verificar uma transição também de segunda ordem, com o parâmetro de ordem indo assintoticamente a zero a medida que a rede cresce e as flutuações do sistema divergindo no ponto crítico $\theta_c \approx 5$ e com expoente crítico de comprimento de correlação $\nu \approx 0.645$, que está

entre os expoentes de correlação espacial da percolação direcionada com dimensão 2 e com dimensão 3 (HENKEL; HIRINCHSEN; LÜBECK, 2008, p. 159), porém o expoente crítico do parâmetro de ordem obtido foi $\beta \approx 0.059$, muito longe do expoente crítico da percolação direcionada para qualquer dimensão. Uma pesquisa que pode ser realizada é construir redes similares ao conectoma de indivíduos saudáveis e doentes e verificar como o valor do ponto crítico e dos expoentes críticos na transição de fase variam, tendo a possibilidade de diagnosticar uma doença através do parâmetro θ_c ou dos expoentes críticos da transição de fase de segunda ordem.

O que pretendo para trabalhos posteriores é obter por meio de um cálculo direto o valor dos expoentes críticos de comprimento de correlação ν das percolações além do real valor do ponto crítico p_c no limite termodinâmico. E usando um grupo de renormalização geométrico (GARCIA-PÉREZ; BOGUÑA; SERRANO, 2018) para reduzir a rede o FSS será feito de forma mais precisa, tendo estimativas dos expoentes críticos para os dois estudos de forma rigorosa.

REFERÊNCIAS

ACHARD et al. **A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs**. *The Journal of Neuroscience*, v. 26, p. 63-72, 2006.

ALBERT; BARABÁSI; JEONG. **Error and attack tolerance of complex networks**. *Nature*, v. 406, p. 378-381, 2000.

ALBERT e BARABÁSI. **Statistical mechanics of complex networks**. *Review of Modern Physics*, v. 74, p. 47-98, 2002.

ALBERT; JEONG; BARABÁSI. **Diameter of the World-Wide Web**. *Nature*, v. 401, p. 130-131, 1999.

AMARAL et al. **Classes of small-world networks**. *PNAS*, v. 97, p. 11149-11152, 2000.

BALABAN. **Applications of Graph Theory in Chemistry**. *J. Chem. Inf. Comput. Sci.*, v. 25, p. 334-343, 1985.

BARABÁSI. **Network Science**. Cambridge University Press, Reino Unido, 2016. 456 p. Disponível em: <http://networksciencebook.com/>. Acesso em: 13 jul 2018.

BARAN. **Introduction to Distributed Communications Networks**. Randon Corporation Memorandum, RM-3420-PR, p. 1-37, 1964.

BERNDT; RUPPIN; REGGIA. **Neural Modeling of Brain and Cognitive Disorders**. World Scientific Publishing, Singapura, 1996. 496 p.

BIGGS; LLOYD; WILSON. **Graph Theory, 1736-1936**. Oxford University Press, EUA, 1976. 240 p.

CALLEN. **Thermodynamics and an Introduction to Thermostatistics**. John Wiley e Sons, 2^a ed., Singapura, 1985. 495 p.

CHUNG e LU. **The Diameter of Sparse Random Graphs**. *Advances in Applied Mathematics*, v. 26, p. 257-279, 2001.

CLAUSET; SHALIZI; NEWMAN. **Power-Law Distributions in Empirical Data**. Society for Industrial and Applied Mathematics, v. 51, p. 661-703, 2009.

COHEN; BEN-AVRAHAM; HAVLIN. **Percolation critical exponents in scale-free networks**. Physical Review E, v. 66, p. 036113-1-4, 2002.

COHEN e HAVLIN. **Scale-Free Networks Are Ultrasmall**. Physical Review Letters, v. 90, p. 058701-1-4, 2003.

COHEN et al. **Resilience of the Internet to Random Breakdowns**. Physical Review Letters, 2000. 4626-4628 p.

ERDŐS e RÉNYI. **On random graphs**. Publicationes Mathematicae Debrecen, v. 6, p. 290-297, 1959.

FELD. **Why Your Friends Have More Friends than You Do?** American Journal of Sociology, v. 96, p. 1464-1477, 1991.

FERREIRA; CASTELLANO; PASTOR-SATORRAS. **Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results**. Physical Review E, v. 86, 2012. 041125 p.

GARCIA-PÉREZ; BOGUÑA; SERRANO. **Multiscale unfolding of real networks by geometric renormalization**. Nature Physics, v. 14, p. 583-589, 2018.

GRANOVETTER. **Threshold Model of Collective Behavior**. The American Journal of Sociology, v. 83, 1978. 1420-1443 p.

HAGMANN et al. **Understanding Diffusion MR Imaging Techniques From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond**. RadioGraphics, v. 26, p. 205-223, 2006.

HAGMANN et al. **Mapping the Structural Core of Human Cerebral Cortex: Fundamental physics of mr imaging**. PLoS Biology, v. 6, p. 1479-1493, 2008.

HAHN. **Spin Echoes**. Physical Review, v. 80, p. 580-594, 1954.

HARARY. **Graph Theory**. 1º ed. 1969, CRC Press, EUA, 2018. 274 p.

HENKEL; HIRINCHSEN; LÜBECK. **Non-equilibrium phase transitions**: Volume 1 - absorbing phase transitions. Springer, Países Baixos, 2008. 387 p.

HUMPHRIES e GURNEY. **Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence**. PLoS ONE, v. 3, p. 2051-1-10, 2008.

LESNE e LAGUÄS. **Scale Invariance**: From phase transitions to turbulence. Springer, Alemanha, 2012. 397 p.

MASHAGHI; RAMAZANPOUR; KARIMPOUR V. **Investigation of a protein complex network**. The European Physical Journal B, v. 41, p. 113-121, 2004.

MIŠIĆ et al. **Cooperative and Competitive Spreading Dynamics on the Human Connectome**. Neuron, v. 86, 2015. 1518-1529 p.

NEWMAN. **Networks**: An introduction. Oxford University Press, EUA, 2010. 772 p.

NEWMAN; STROGATZ; WATTS. **Random graphs with arbitrary degree distributions and their applications**. Physical Review E, v. 64, p. 026118-1-17, 2001.

NEWMAN; WATTS; BARABÁSI. **Structure and Dynamics of Networks Princeton Studies in Complexity**. Princeton University Press, EUA, 2011. 582 p.

PATHRIA e BEALE. **Statistical Mechanics**. Elsevier, 3^a ed, EUA, 2011. 718 p.

POOLEY. **AAPM/RSNA Physics Tutorial for Residents**: Fundamental physics of mr imaging. RadioGraphics, v. 25, p. 1087-1099, 2005.

RADICCHI; CASTELLANO. **Breaking of the site-bond percolation universality in networks**. Nature Communications, v. 6, p. 1-7, 2015.

SOLOMONOFF e RAPOPOR. **Connectivity of Random Nets**. Bulletin of Mathematical Biophysics, v. 13, p. 107-117, 1951.

SPORNS. **Networks of the Brain**. Massachusetts Institute of Technology Press, EUA, 2011. 413 p.

SPORNS; TONONI e KÖTTER. **The Human Connectome: A Structural Description of the Human Brain.** PLoS Computational Biology, v. 1, p. 245-251, 2005.

STANLEY. **Introduction to Phase Transition and Critical Phenomena.** Oxford University Press, Países Baixos, 1971. 308 p.

STAUFFER e AHARONY. **Introduction to Percolation Theory.** Taylor & Frances, 2^a ed., Inglaterra, 1991. 181 p.

TORREY. **Bloch Equations with Diffusion Terms.** Physical Review, v. 104, p. 563-565, 1956.

TURING. **Computing machines and intelligence.** Mind, v. 59, 1950. 433-460 p.

WATTS e STROGATZ. **Collective dynamics of 'small-world' networks.** Nature, v. 393, p. 440-442, 1998.

YEOMANS. **Statistical Mechanics of Phase Transitions.** Clarendon Press, Inglaterra, 1992. 153 p.