

Universidade Federal Fluminense  
Instituto de Física

*Redes Neurais Atratoras em Topologias  
Complexas*

Carlos Eduardo Cardoso Galhardo

Orientador: Dr. Thadeu J. P. Penna

Dissertação apresentada como requisito parcial para obtenção do  
título de Mestre em Física.

Niterói, 7 de maio de 2006

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

“The whole is more than the sum of its parts”

- Aristóteles

À Maria Florencio da Silva

# Índice

<b>1</b>	<b>Redes Neurais</b>	<b>4</b>
1.1	Fundamentos de Neurofisiologia . . . . .	4
1.2	Breve evolução histórica do estudo de redes neurais . . . . .	9
1.2.1	O neurônio de McCulloch e Pitts . . . . .	9
1.2.2	Perceptrons . . . . .	12
1.2.3	Do perceptron a rede neural atratora (ANN) . . . . .	15
1.3	Memória Associativa . . . . .	16
1.4	Aprendizado em Redes Neurais . . . . .	18
1.5	Modelo de Hopfield . . . . .	19
1.5.1	Evolução temporal . . . . .	23
1.5.2	Energia e a matriz sináptica . . . . .	28
1.5.3	Estados espúrios e saturação . . . . .	29
1.6	Conclusões . . . . .	32
<b>2</b>	<b>Redes Complexas</b>	<b>33</b>
2.1	Definições Básicas . . . . .	34
2.2	Grafo de Erdős-Réyl . . . . .	39
2.3	Redes na física . . . . .	41
2.4	Redes no Mundo Real . . . . .	43
2.5	Novos Modelos . . . . .	44
2.6	Sistema Nervoso Real . . . . .	52

2.7	Conclusão . . . . .	53
<b>3</b>	<b>Resultados Numéricos</b>	<b>55</b>
3.1	Descrição do procedimento . . . . .	55
3.2	Padrões sem correlação . . . . .	57
3.3	Bacia de Atração . . . . .	65
3.4	Padrões com correlação . . . . .	71
3.5	Superposição e a correlação . . . . .	75
3.6	Bacia de atração e padrões com correlação . . . . .	79
3.7	Conclusão . . . . .	83
<b>4</b>	<b>Conclusões</b>	<b>84</b>

# Lista de Figuras

1.1	Neurônio biológico típico . . . . .	5
1.2	Sinapse química. . . . .	6
1.3	Esboço do potencial de ação. Despolarização da membrana (à esquerda). Impulso percorrendo o axônio, em direção as sinapses (à direita) . . . . .	9
1.4	Neurônio formal deMcCulloch e Pitt. Círculo grande, que representa o soma. Existem N pontos conectados, logicamente, ao soma por meio de setas na figura. Essas caixas fazem o papel dos axônios pré sinápticos, podendo ativar a sinapse com um pulso, ou não. As sinapses são representadas pelas setas. . . . .	10
1.5	As operações booleanas <b>not</b> (a), <b>and</b> (b) e <b>or</b> (c) representadas através de neurônios de McCulloch e Pitts. . . . .	11
1.6	A operação booleana <b>xor</b> mostra-se mais complexa, com o encadeamento de sete neurônios. . . . .	12
1.7	Esquema do perceptron de Rosenblatt. . . . .	13
1.8	Exemplo de perceptron de várias saídas. . . . .	14
1.9	Exemplo de perceptron de várias camadas. . . . .	15
1.10	Metáfora do vales e colinas. M1 e M2 são padrões armazenados, mínimos globais. X é um atrator espúrio, mínimo local. . . . .	30
2.1	Exemplo de um grafo com 6 vértices e 7 arestas. . . . .	35
2.2	Exemplo de um digrafo . . . . .	36

2.3	A rede possui 8 maneiras de ligar um nó a dois vizinhos distintos, ou seja, 8 conexões triplas. Apenas 1 triângulo. Então pela equação 2.2, $C = 3/8$ . Olhando agora a definição local, temos os coeficientes de clustering $C_i$ , eq. (2.3), da direita para esquerda, valendo: $1/6, 1, 1, 0$ e 0. Dessa forma, utilizando a equação 2.4: $C = 13/30$ . . . . .	38
2.4	Grafo aleatório para dois valores de $p$ . . . . .	40
2.5	Redes $1d$ e $2d$ com $L=7$ . . . . .	43
2.6	rede de small-world . . . . .	45
2.7	Histograma do número de vértices pela distribuição de conexões para redes de small world, com $p = 0.2(+)$ , $p = 0.4(*)$ , $p = 0.7(\square)$ , $p = 1.0(\blacksquare)$ e <i>Erdős-Rényi</i> ( $\times$ ), $K = 100$ em 10000 realizações. . . . .	47
2.8	Rede aleatória e a rede livre de escala. Os nós marcados em cinza atentam para a condensação de arestas. . . . .	48
2.9	Histograma do número de vértices pela distribuição de conexões para $N = 5000$ , $m = n = 50$ em 100000 realizações. . . . .	51
2.10	(a) Histograma do número de vértices pela distribuição de conexões. (b) $P(k)$ . Para ambos os gráficos $N = 100(+)$ , $1000(\times)$ , $10000(*)$ , $100000(\square)$ , $m = n = 10$ em 10000 realizações. . . . .	51
3.1	Superposição vs $\alpha$ . Para $p = 1.0 (\times)$ , e grafo aleatório ( $\square$ ) $N = 5000$ e $K = 50$ (padrões sem correlação). . . . .	58
3.2	Superposição vs $\alpha$ , para rede de small-world, de $p = 0.0$ até $p = 1.0$ acrescido de 0.1, da esquerda para direita. $N = 5000$ , $K = 100$ , média em 100 realizações. Padrões sem correlação . . . . .	59
3.3	Superposição vs $\alpha$ para tamanho finito, $p = 0.4$ . $N = 1000 (+)$ , $N = 2000 (\times)$ , $N = 5000 (*)$ , $N = 10000 (\square)$ e $N = 20000 (\blacksquare)$ . . . . .	60



3.4	Superposição vs $\alpha$ . Redes de small world $p = 0.0$ (+), $p = 0.2$ ( $\times$ ), $p = 0.4$ (*) e $p = 1.0$ (■). Rede livre de escala ( $\square$ ) . $N = 5000$ $K = 100$ para as redes sw. Para as redes livre de escala $N = 5000$ , $m = n = 100$ . Média em 100 realizações. . . . .	61
3.5	$\varphi$ vs $\alpha$ . Rede small world com $p = 0.0$ (+), $p = 0.2$ ( $\times$ ), $p = 0.4$ (*) e $p = 1.0$ (■) com $N = 5000$ e $K = 100$ . Rede livre de escala ( $\square$ ), $N = 5000$ e $m = n = 50$ . Média em 100 realizações para todas as redes.	62
3.6	Superposição vs $\alpha$ para níveis maiores de diluição. $N = 5000$ para todas as simulações. (a) $K = 50$ para as redes sw e $m = n = 50$ para a rede livre de escala. (b) $K = 10$ para sw e $m = n = 10$ para livre de escala. Para ambos os gráficos small world com $p = 0.0$ (+), $p = 0.2$ ( $\times$ ), $p = 0.4$ (*), $p = 1.0$ (■) e livre de escala ( $\square$ ) . . . . .	63
3.7	Superposição vs padrões armazenados $N = 5000$ , $p = 0.4$ . $K = 4$ (+), $K = 6$ ( $\times$ ), $K = 10$ (*), $K = 50$ ( $\square$ ) e $K = 100$ (■) . . . . .	64
3.8	$\varphi$ vs $\alpha$ para níveis maiores de diluição, $N = 5000$ para todas as redes. (a) $K = 50$ para as redes small world, e $m = n = 25$ para a rede livre de escala. (b) $K = 10$ para sw $m = n = 5$ para livre de escala. Em ambos os gráficos: $p = 0.0$ (+), $p = 0.2$ ( $\times$ ), $p = 0.4$ (*), $p = 1.0$ (■) e livre de escala ( $\square$ ) . . . . .	65
3.9	Bacia de atração para sw $p = 1.0$ , $N = 5000$ e $k = 100$ . $P = 5$ (+), $P = 10$ ( $\times$ ), $P = 15 = 5$ (*) e $P = 20$ ( $\square$ ) . . . . .	66
3.10	Bacia de atração, para Hopfield não diluição $N=1000$ . $\alpha = 0.05$ , $\alpha = 0.1$ e $\alpha = 0.13$ da esquerda para a direita. . . . .	67
3.11	Bacia de atração para diversas geometrias, $N = 5000$ , $K = 100$ , $\alpha = 0.05$ e $\alpha = 0.15$ em $10^4$ realizações. . . . .	67
3.12	Bacia de atração para diversas geometrias, com $N = 5000$ , $K = 100$ , $\alpha = 0.05$ e $\alpha = 0.15$ . sw com $K = 100$ , e livre de escala com $m = n = 50$ . Média em $10^4$ realizações. . . . .	68

3.13	(a) $\Phi_{100\%}$ vs $p$ e (b) $\Phi_{97\%}$ . Para ambos os gráficos $N = 5000$ e $K = 200$ . $P = 1 (+)$ , $P = 2 (\times)$ , $P = 5 (*)$ , $P = 10 (\square)$ e $P = 20 (\blacksquare)$ . Média em 1000 realizações. . . . .	70
3.14	(a) $\Phi_{100\%}$ vs $p$ e (b) $\Phi_{97\%}$ . Para ambos os gráficos $N = 5000$ , $K = 100$ . $P = 1 (+)$ , $P = 2 (\times)$ , $P = 5 (*)$ , $P = 10 (\square)$ e $P = 20 (\blacksquare)$ . Média em 1000 realizações. . . . .	71
3.15	Superposição vs carga. Para ambos os gráficos $N = 5000$ e $K = 100$ . Média em 100 realizações, com atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40% (a) $sw$ com $p = 0.2$ (b) $sw$ com $p = 0.4$ . . . . .	72
3.16	Superposição vs carga. para ambos os gráficos $N = 5000$ , $K = 100$ e média em 100 realizações. Atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40% (a) $sw$ com $p = 1.0$ (b) livre de escala . . . . .	73
3.17	Eficiência vs carga. Para ambos os gráficos $N = 5000$ , $K = 100$ e médias em 100 realizações. Atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40%. Da direita para a esquerda. (a) $sw$ com $p = 1.0$ (b) livre de escala . . . . .	74
3.18	Eficiência vs carga. Para ambos os gráficos $N = 5000$ , $K = 100$ e média em 100 realizações. Atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40%. Da direita para a esquerda. (a) $sw$ com $p = 0.4$ (b) $sw$ com $p = 0.2$ . . . . .	74
3.19	Superposição vs carga. $N = 5000$ , $K = 100$ Média em 100 realizações. Atividade de 43%. $p = 0.2 (+)$ , $p = 0.4 (\times)$ , $p = 1.0 (*)$ e rede livre de escala ( $\square$ ) . . . . .	75
3.20	Superposição vs $a$ . $K = 100$ , $P = 10$ para $sw$ com $p = 1.0$ . $N = 1000$ (+), $N = 2000 (x)$ , $N = 5000 (*)$ , $N = 10000 (\square)$ e $N = 20000 (\blacksquare)$ . Média em 100 realizações . . . . .	76

3.21	Superposição vs atividade. $P = 10$ e $N = 5000$ para $sw$ com $p = 1.0$ em diferentes valores de $K$ . $K = 10$ (+), $K = 20$ ( $\times$ ), $K = 30$ (*), $K = 40$ ( $\square$ ), $K = 60$ ( $\blacksquare$ ), $K = 80$ ( $\circ$ ), $K = 100$ ( $\bullet$ ) e $K = 200$ ( $\triangle$ ). Média em 100 realizações . . . . .	77
3.22	Superposição vs $a$ para as geometrias de small world com com $p = 0.2$ (+) , $p = 0.4$ ( $\times$ ), $p = 1.0$ (*) e rede livre de escala ( $\square$ ). Para $\alpha = 0.05$ em (a) e $\alpha = 0.10$ em (b). Médias em 100 realizações. . . . .	78
3.23	Superposição vs $a$ para as geometrias de small world com $p = 0.2$ (+) , $p = 0.4$ ( $\times$ ), $p = 1.0$ (*) e a rede livre de escala ( $\square$ ). Para $\alpha = 0.15$ em (a) e $\alpha = 0.20$ em (b). Médias em 100 realizações. . . . .	79
3.24	Bacia de atração para as geometrias de small world com $p = 0.2$ (a) e $p = 0.4$ (b). 45% (+), 44% ( $\times$ ), 43% (*), 42% ( $\square$ ) e 41% ( $\blacksquare$ ). Média em 1000 realizações. . . . .	80
3.25	Bacia de atração para as geometrias de small world com $p = 1.0$ (a) e a rede livre de escala (b). 45% (+), 44% ( $\times$ ), 43% (*), 42% ( $\square$ ) e 41% ( $\blacksquare$ ). Média em 1000 realizações. . . . .	81
3.26	Bacia de atração. $N = 5000$ e $\alpha = 0.2$ para $sw$ $p = 0.4$ em diferentes valores de de atividade. 50% (+), 45% ( $\times$ ), 43% (*), 41% ( $\square$ ) e 40% ( $\blacksquare$ ). Média em 1000 realizações. . . . .	81
3.27	Eficiência vs $p$ . $N = 5000$ e $P = 5$ em diferentes valores de de atividade. 50% (+), 48% ( $\times$ ), 45% (*), 42% ( $\square$ ) e 40% ( $\blacksquare$ ). Média em 1000 realizações. . . . .	82

# Agradecimentos

Agradeço ao Thadeu por ter me aceito como aluno de mestrado e me orientado na pesquisa em sistemas complexos. Além disso agradeço a ele por todo apoio, em principal nessa última etapa, aonde que, além de orientador, ele se mostrou um grande amigo.

Ao Marcus Moldes, pela grande lição que foi a convivência com ele nesses dois anos. Jamais poderia imaginar que um telefonema de apoio durante o curso de eletromagnetismo iria render uma amizade tão frutífera.

À Lívia Carvalho, pela paciência quase materna na solução dos mais variados problemas.

Ao professor Jurgen Stilck, pela boa vontade e disposição em sanar dúvidas e bom trabalho realizado na coordenação durante esses dois anos de mestrado.

A todo o pessoal da biblioteca, pelo ótimo trabalho e pela boa vontade em me atender sempre, mesmo com todos os meus atrasos.

Aos amigos Bernardo Coutinho e Klauko Motta, pela conversas, ajudas e discussões a respeito de Física, programação e da vida.

Aos meu familiares, pelo apoio e compreensão nesses anos dois anos.

Wellington Gomes Dantas, Gabriel Brandão e Rian Rezende por serem os poucos seres humanos que compram as minhas discussões, sejam elas filosóficas, científicas ou aleatórias.

Aquino Espíndola e à Cinthya Valeska pelo curso instantâneo de LaTeX.

Aos camaradas de apartamento, Marciano Carneiro e André Fernandes pela paciência com a minha rotina de vida nada linear.

Aos amigos da pós, Daniel Dias, Thiago Lobo, Luciana Rios, Minos Martins, Luciano Fonseca da Silva, Eliel Eleutério de Farias, Gabriela Nieva de Oliveira, Claudia Rocha e todos alunos que me acolheram bem e tornaram a UFF um ótimo ambiente de trabalho.

Aos amigos distantes, mas que sempre se preocuparam com o meu progresso e deram apoio: Andre Gonçalves de Lima, Denise Assafrao, Davi Melo de Oliveira, Marcelo Martins, Douglas Draia, Carla Sales, Alessandra Monzo, Amanda Amaro, Cassandro Emer, Karina Marcon, Rodolfo Funa e Tadeu Saldanha.

Aos companheiros de pensionato Bruno, Davi, Luciano, Evandro, Eric, Rodrigo e Felipe que foram as minhas primeiras amizades em Niterói.

A Marcello Sandim, pois sabemos que os verdadeiros amigos aparecem na hora do aperto.

A Mariana Brandão, pelo apoio e incentivo.

E a todas as outras pessoas que eu provavelmente esqueci de tomar nota nesse pedaço de papel, mas que tiveram importância na conclusão desse trabalho.

E, a Sylvester Stallone, pela marcante atuação como Rocky Balboa quando vo-ciferou a frase que se tornou filosofia de trabalho: “Não jogue a toalha!”.

A CAPES e a FAPERJ pelo apoio financeiro.

# Resumo

Neste trabalho estudamos algumas variações no modelo de Hopfield para redes neurais atratoras. Primeiro, com intuito de aproximar o modelo de uma rede de neurônios reais, fizemos uma diluição controlada, de forma a gerar redes de topologias complexas. Essa escolha foi embasada em evidências experimentais [1, 2]. Estudamos então a relação entre a eficiência na recuperação de padrões armazenados e as topologias. Com intuito de melhor averiguar o papel da geometria da rede no modelo, introduzimos padrões correlacionados e estudamos como cada rede responde ao aumento da correlação entre os padrões.

# Abstract

In this work we present some modifications in the Hopfield model for attractor neural networks. First, in order to bring the model closer to a biological network of neurons, we introduce a controlled dilution generating complex topologies. This modification is supported by experimental evidences [1, 2]. Then, study the relation between efficiency and the topologies, in the recovery stored patterns. We study how the network topology play a role in the model, by introducing correlated patterns and studing the performance of each topology as the correlation between patterns is increased.

# Introdução

Alguns dos problemas científicos enfrentados pela humanidade nos últimos 20 anos mostraram-se resistentes a abordagem tradicional. A divisão do conhecimento científico nas tradicionais grandes áreas não foi capaz de trazer esclarecimentos a problemas como: o funcionamento do sistema imunológico, as funções cerebrais, a evolução das espécies, entre outros. Assim, abordagens interdisciplinares fizeram-se necessárias para trazer luz ao entendimento desses sistemas. Alguns sistemas que integram esse conjunto de problemas são chamados de sistemas complexos.

É difícil definir um sistema complexo, mas de forma intuitiva podemos caracterizá-lo por suas propriedades. A propriedade que mais chama atenção em um sistema complexo é a auto-organização. Imagine que um conjunto de elementos (células ou investidores por exemplo) que interagem de forma aleatória e, a partir de um certo momento, sem a existência de uma unidade central controladora, se organizam, apresentando fenômenos claramente coletivos (organismos multicelulares para as células ou *crash* para os investidores). Ou seja, um comportamento emergente surge, diferente do que poderíamos esperar a partir das interações locais do sistema.

Uma das empreitadas mais ousadas da ciência é o entendimento do sistema nervoso central. Desde a antiguidade, as propriedades do intelecto humano já eram alvo de estudo. Aristóteles em “On Memory and Reminiscence” [3] já fazia observações a respeito da memória humana e convencionou leis para suas expressões. No estudo das funções do sistema nervoso central (SNC), a biologia, química, medicina e psicologia fizeram um grande avanço, mas pouco ainda se sabe a respeito de fenômenos



cerebrais, como, por exemplo a inteligência, memória e aprendizado.

Na década de 40, as ciências exatas começaram a dar suas primeiras contribuições no entendimento do cérebro. Com o advento do computador, engenheiros e matemáticos puderam modelar e simular o comportamento neuronal na tentativa de entender como uma rede de neurônios funciona.

No final da década de 70 a teoria de sistemas desordenados, em principal os sistemas magnéticos conhecidos como vidros de spin, estava em seu apogeu. Dada a semelhança entre um spin e o neurônio matemático e as sinapses e as interações magnéticas, J. Hopfield [4] propôs um modelo de vidro de spins para os neurônios, dando início a uma nova área de pesquisa em Física, a das redes neurais. O modelo de Hopfield é uma simplificação forte dos processos celulares e cerebrais e tem como intuito modelar uma das principais propriedades cerebrais, a memória. A analogia consiste em afirmar que cada estado memorizado pela rede é um mínimo de energia da hamiltoniana do modelo de vidro de spin, fazendo assim o problema inverso, isto é, dado um conjunto de mínimos quais serão as sinapses (conexões) que regem esse processo? Para resolver essa questão, Hopfield introduziu uma regra de aprendizado, que consiste em uma forma de montar a matriz sináptica (matriz das conexões). Para que esse modelo seja solúvel analiticamente, Hopfield pagou um preço caro do ponto de vista biológico, pois teve que impor que todos os neurônios são conectados entre si. Além disso é necessário que os padrões armazenados na rede sejam descorrelacionados. Para poder definir uma hamiltoniana sugeriu que as sinapses fossem simétricas, enquanto em um neurônio real a informação tem um sentido bem definido.

O nosso trabalho tem como objetivo estudar duas simplificações desse modelo: a necessidade de padrões não correlacionados e a conectividade. O modelo de Hopfield se mostrou capaz de descrever propriedades importantes do cérebro, mas não aborda uma das propriedades mais surpreendentes da rede de neurônios reais: a plasticidade estrutural. Essa é capacidade cerebral de criar novas conexões, que ocorre com grande intensidade nos estágios iniciais do desenvolvimento cerebral, mas se mantém

ao longo da vida do indivíduo [5]. Com a aproximação de uma rede totalmente conectada o modelo não é capaz de prever os efeitos da plasticidade estrutural no processo de armazenamento-recuperação de uma memória.

Trabalhos averiguando o processo de diluição em redes neurais desse tipo já existem na literatura [6, 7, 8], mas nenhuma leva em conta o papel da correlação entre os padrões na diluição. Estamos interessados em entender como a correlação interage com a topologia da rede, pois é conhecido que o cérebro apresenta áreas especializadas para tarefas distintas. Trabalhos de análise de dados experimentais [1, 2] mostram que o sistema nervoso central, em diferentes escalas, apresenta uma topologia complexa, este trabalho fará uma análise dessas arquiteturas perante a correlação.

No capítulo 1 vamos rapidamente introduzir as propriedades neurofisiológicas para sustentar o modelo de Hopfield. Esse será descrito e suas principais propriedades serão discutidas.

No capítulo 2 vamos fazer um estudo introdutório de teoria dos grafos, para caracterizar as principais topologias complexas e discutiremos as propriedades mais importantes.

No capítulo 3 faremos um estudo do modelo de Hopfield nas arquiteturas complexas, revisando recentes resultados na literatura a fim de validar nossa abordagem, esclarecer incompatibilidades e complementar resultados. Por fim apresentaremos nossos resultados no estudo dessas redes na armazenagem de padrões correlacionados.

Finalmente, apresentamos as conclusões dos nosso trabalho e as futuras perspectivas.

# Capítulo 1

## Redes Neurais

Neste capítulo faremos o embasamento teórico que sustenta nossas simulações. Para abordar o tema de uma forma clara, começaremos com uma introdução bastante simplificada à Neurofisiologia. Em seguida, para contextualizar o trabalho de Hopfield e entender o que ele significou na teoria de redes neurais, um resumo histórico é apresentado. Para finalizar construiremos o modelo de Hopfield de uma forma intuitiva apresentando os resultados clássicos, sem demonstrá-los.

### 1.1 Fundamentos de Neurofisiologia

Podemos dizer que a grande área da ciência que se encontra a teoria de redes neurais é a Neurociência que é, por construção, uma área multidisciplinar. Existem vários enfoques para tratar um problema multidisciplinar e aqui daremos a nossa colaboração como físicos da complexidade, tentando entender o ente real através de modelos simplificados. O objetivo principal desse tipo de abordagem não é exaurir o problema, mas buscar um entendimento global sobre o sistema e seu comportamento. Mas, mesmo assim, é de extrema importância manter um elo entre o modelo e o objeto de estudo, para que nossos resultados possam colaborar com áreas correlatas que estudam o mesmo sistema em sua distinta perspectiva. Como o sistema abordado é uma rede de neurônios, é de fundamental importância o conhecimento

de conceitos básicos da neurofisiologia do sistema nervoso.

O sistema nervoso nos vertebrados é dividido em sistema nervoso central (SNC) e sistema nervoso periférico e, de forma simplificada, podemos reduzir a dois elementos básicos, *neurônios* e as *sinapses*.

O neurônio pode ser dividido em 3 partes, Figura (1.1): os *dendritos* são prolongamentos geralmente muito ramificados responsáveis pela recepção de estímulos. O *soma* é a unidade central responsável pelo processamento de sinais recebidos pelos dendritos e também por manter os processos metabólicos da célula. A transmissão do sinal é feita por um extenso prolongamento que atua como condutor: o *axônio*. Os axônios têm muitas ramificações em suas regiões terminais e cada uma pode formar uma conexão com um dendrito ou corpos celulares.

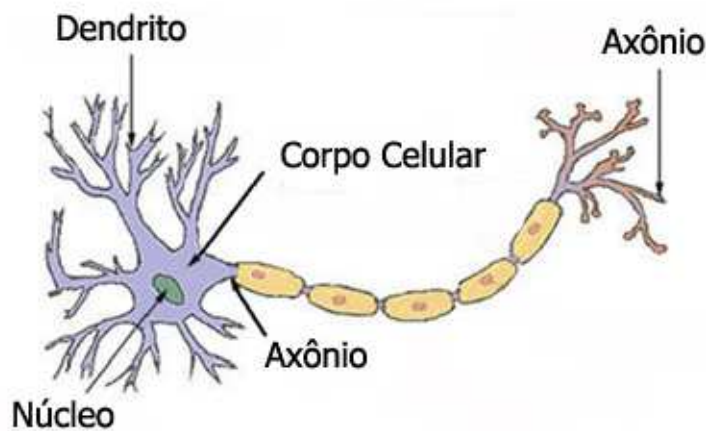


Figura 1.1: Neurônio biológico típico

Existem, nos neurônios, conexões de dois tipos: *sinapses elétricas*, mais simples e evolutivamente mais antigas, que permitem a transferência direta da corrente iônica de uma célula para outra. Esse tipo de sinapse é comum em invertebrados. Em vertebrados aparecem pouco, durante etapas da formação do sistema nervoso. De forma geral a transmissão sináptica no sistema nervoso de um homem adulto é

*química* (Figura (1.2)). As membranas pré e pós-sinápticas são separadas por uma fenda com largura de 20 a 50 nm - a *fenda sináptica*. A passagem do impulso nervoso nessa região é feita, então, por substâncias químicas: os *neuro-hormônios*, também chamados *mediadores químicos* ou *neurotransmissores*, liberados na fenda sináptica. O terminal axonal típico contém dúzias de pequenas vesículas membranosas esféricas que armazenam neurotransmissores - as *vesículas sinápticas*.

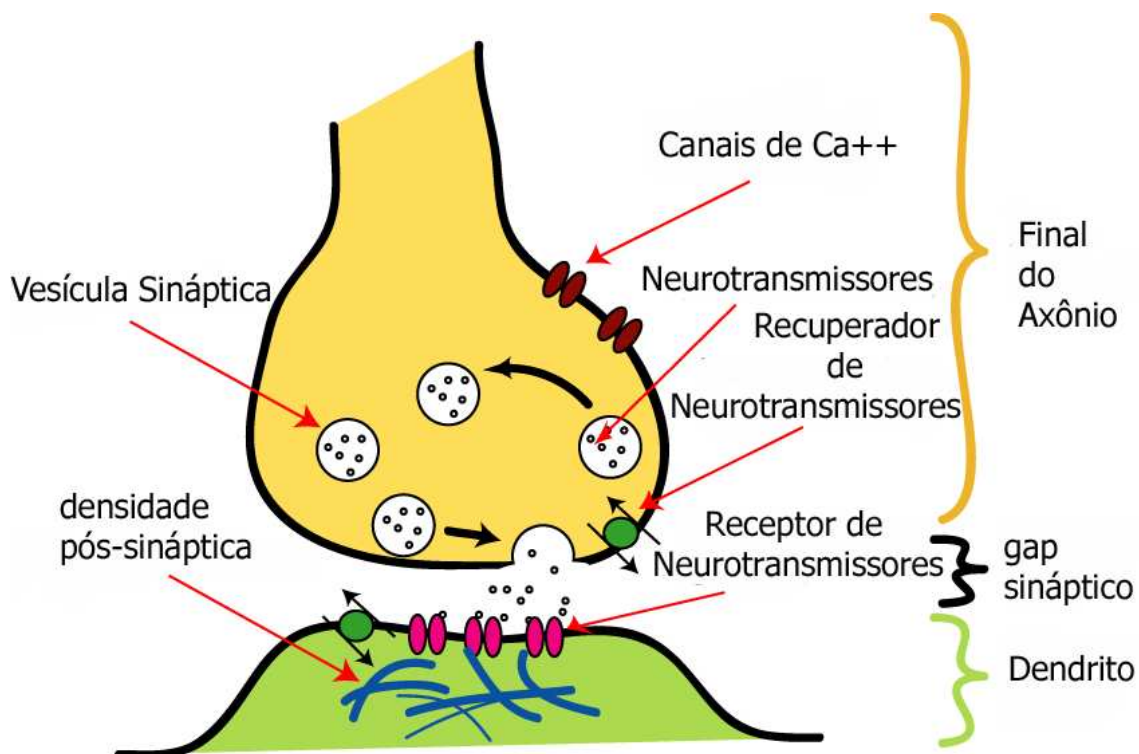


Figura 1.2: Sinapse química.

O axônio que chega na sinapse transmite a informação do que foi computado pelo soma ao dendrito da próxima célula. O neurônio que contribui com o axônio para a sinapse é chamado de *pré-sináptico*, e o outro de *pós-sináptico*. Um fato neurofisiológico importante é que o SNC possui cerca de  $10^{11}$  neurônios e cada neurônio recebe em média  $10^3$  a  $10^4$  sinapses e, em sua maioria, de neurônios distintos. Mas olhando dois pontos do cérebro com uma distância apreciável, nota-se que as sinapses entre esses dois pontos são escassas. Nosso córtex cerebral é um mosaico de

autômatos com grande número de conexões locais, que funcionam como módulos de processamento individuais, conectados esparsamente entre si permitindo assim processamento em paralelo.

Existem vários tipos de neurônios no sistema nervoso, variando em tamanho, estrutura e função. Existe também uma série de classificações, uma delas, importante para nós, é a classificação funcional:

**1. Neurônios receptores ou sensitivos (aférentes):** São os neurônios responsáveis pela entrada de sinais no sistema nervoso. Eles recebem estímulos sensoriais das células receptoras e os propagam para o sistema nervoso central.

**2. Neurônios motores (eferentes):** Esses são os responsáveis pela saída de sinais do sistema nervoso para os outros sistemas. Transmitem, por exemplo, os impulsos motores possibilitando o movimento muscular.

**3. Neurônios associativos ou interneurônios:** estabelecem ligações entre os neurônios receptores e os neurônios motores. Cerca de 90% dos neurônios são neurônios associativos, e estão localizados no sistema nervoso central (SNC).

Nosso interesse será modelar neurônios associativos com sinapses químicas com o intuito de estudar o fenômeno conhecido como memória associativa.

Antes de continuar no nosso objetivo de estudo, é necessário entender a dinâmica bioquímica que ocorre no cérebro. No seu estado não excitado, o interior do neurônio é negativamente carregado proporcionando uma diferença de potencial em relação ao meio externo de, aproximadamente,  $-70$  mV.

De forma simplificada podemos pensar que existem três íons responsáveis por essa polarização: o potássio, o sódio e o cloro. O íon  $\text{Na}^+$  está em maior concentração fora do que dentro da célula e a taxa de entrada de sódio na membrana é baixa, devido a sua pouca permeabilidade ao sódio. O íon  $\text{K}^+$  está em maior concentração dentro do que fora da célula. Como a membrana é permeável ao sódio, ele tende a se distribuir igualmente tanto no meio extracelular quanto no meio intracelular. Como a célula é mais negativa dentro do que fora, existe um potencial que atrai os íons de potássio para dentro da célula. Mas, mesmo assim, a diferença ainda seria

menor do que a observada. Essa polarização é mantida porque a membrana celular mantém um mecanismo de controle conhecido como *bomba de sódio-potássio*. Ela bombeia ativamente, isto é, com dispêndio de energia, íons de sódio para fora da célula e íons de potássio para dentro. Somando esse efeito à impermeabilidade da membrana ao sódio, ocorre um acúmulo de íons positivos no meio extracelular.

Quando o neurônio pré-sináptico estimula a célula pós sináptica, uma mudança na permeabilidade de sódio ocorre. Se a célula fica mais permeável com essa excitação dizemos que a sinapse é *excitatória*. Se a sinapse contribui para a polarização, dizemos que ela é *inibitória*. A essa alteração na polarização causada pela envio de um sinal do neurônio anterior dá-se o nome de Potencial Pós Sináptico (PPS). O efeito conjunto dos PPS são integrados espacialmente e temporalmente pelo soma e se esses atinjam um *limiar*, em torno de  $-30\text{mV}$ , um evento brusco ocorre na célula (*potencial de ação*). Começa uma despolarização em cadeia, começando no soma e se propagando para o axônio, gerando assim um sinal elétrico em direção às sinapses na Figura (1.3). Quando a membrana plasmática na região sináptica é excitada por esse sinal, as vesículas de neurotransmissores são liberadas na fenda sináptica e o neurônio pós sináptico irá receber um PPS, que pode ser polarizador ou despolarizador, dependendo da sinapse.

É importante notar que esse efeito que dispara o potencial de ação é do tipo “tudo ou nada”, se os impulsos não forem suficientes para que o potencial limiar seja atingindo, nenhum efeito notável ocorre no neurônio.

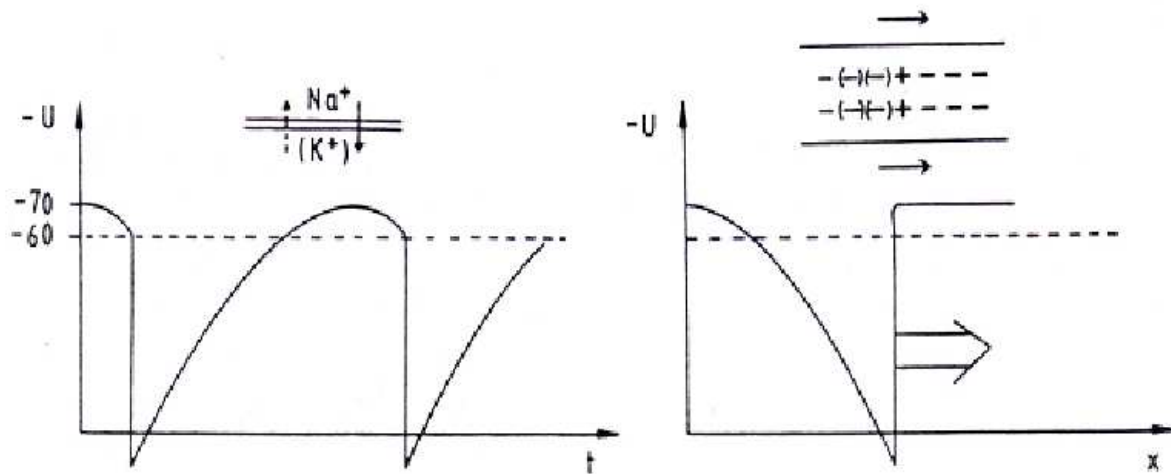


Figura 1.3: Esboço do potencial de ação. Despolarização da membrana (à esquerda). Impulso percorrendo o axônio, em direção as sinapses (à direita)

## 1.2 Breve evolução histórica do estudo de redes neurais

### 1.2.1 O neurônio de McCulloch e Pitts

Em 1943, Warren McCulloch e Walter Pitts publicaram no *Bulletin of Mathematical Biophysics* [9] o artigo que hoje é considerado o marco zero na teoria de redes neurais: “*A Logical Calculus of Ideas Immanent in Nervous Activity*”. A idéia central do artigo é modelar um neurônio que age como dispositivo binário, sua saída pode ser um pulso ou um não pulso. Para isso calcula-se somas ponderadas das entradas com os pesos pré-estabelecidos, podendo ser negativos no caso de uma sinapse inibitória, ou positivos no caso excitatório. Se essa soma for maior que um dado limiar, então o neurônio emite um pulso. Com esse mecanismo simples é possível implementar uma rede de neurônio que realize o cálculo de qualquer função booleana.



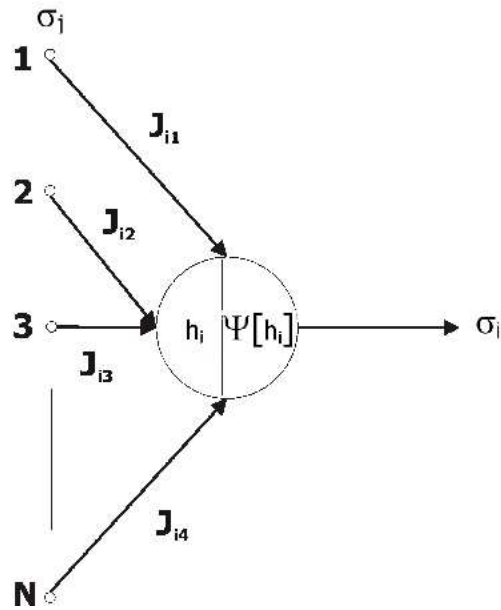


Figura 1.4: Neurônio formal de McCulloch e Pitt. Círculo grande, que representa o soma. Existem  $N$  pontos conectados, logicamente, ao soma por meio de setas na figura. Essas caixas fazem o papel dos axônios pré sinápticos, podendo ativar a sinapse com um pulso, ou não. As sinapses são representadas pelas setas.

A cada sinapse é associado um parâmetro  $J_{ij}$ , onde o índice  $i$  subscrito representa o neurônio pós sináptico, e o índice  $j$  refere-se aos canais de entrada desse neurônio  $i$ . O valor de  $J_{ij}$  é a eficiência da sinapse, ou peso sináptico, que serão adicionados ao soma  $i$  se o canal  $j$  estiver ativo naquele momento 1.4. Cada neurônio opera da seguinte forma:

- Em um dado instante de tempo, alguns canais de entradas são ativados;
- O soma recebe esse PPS que é a soma linear das eficiências  $J_{ij}$  de todos os canais ativados;
- A soma é comparada com o limiar de ativação, se for maior o neurônio dispara.

Seja  $\sigma_j$  o estado do  $j$ -ésimo neurônio, que tem valor 1 se o neurônio dispara, 0

caso contrário  $h_i$  o PSP do neurônio  $i$ , então:

$$h_i = \sum_{j=1}^N J_{ij} \sigma_j \quad (1.1)$$

onde  $N$  é o número de neurônios pré-sinápticos. A resposta do neurônio  $i$ , então pode ser calculada:

$$\sigma'_i = \Psi[h_i > T_i] \quad (1.2)$$

sendo  $\Psi$  uma função verdade, que assume valor 1 caso o argumento seja verdadeiro, e valor zero, caso contrário.

A essência filosófica da proposta de McCulloch e Pitts é colocar a inteligência em equivalência ao cálculo de predicados, esses por sua vez podem ser implementados via funções booleanas. Visto de uma forma bastante simplificada o cérebro é composto por uma rede de neurônios que tem a capacidade básica de resolver essas funções.

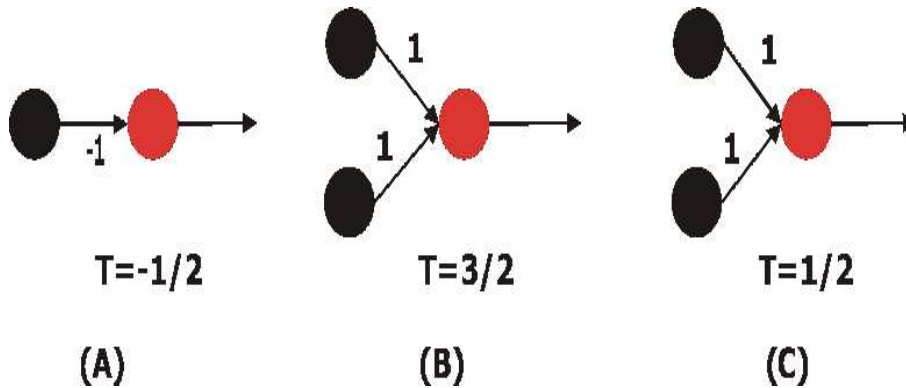


Figura 1.5: As operações booleanas **not** (a), **and** (b) e **or** (c) representadas através de neurônios de McCulloch e Pitts.

O diagrama deve ser lido da seguinte maneira. Olhemos o item (a) na Figura (1.5). O neurônio à esquerda dispara,  $\sigma(t) = 1$ , se uma dada proposição  $A$  é verdadeira. A sinapse tem peso  $-1$ , então, em  $t+1$ , o neurônio da direita que possui limiar  $T = -0,5$  não disparará. Se  $\sigma(t) = 0$ , ou seja, o neurônio da direita não disparou em  $t$ , o neurônio da esquerda disparará, pois  $0 > T$ . Então o disparo do neurônio da direita é a **negação** da proposição  $A$ .

No diagrama (b), suponha que o neurônio superior dispara se a proposição A for verdadeira, e o outro neurônio à esquerda dispara se B for verdadeira. Ambas as sinapses possuem peso 1, e o neurônio pós sináptico possui limiar  $T = 1,5$ .

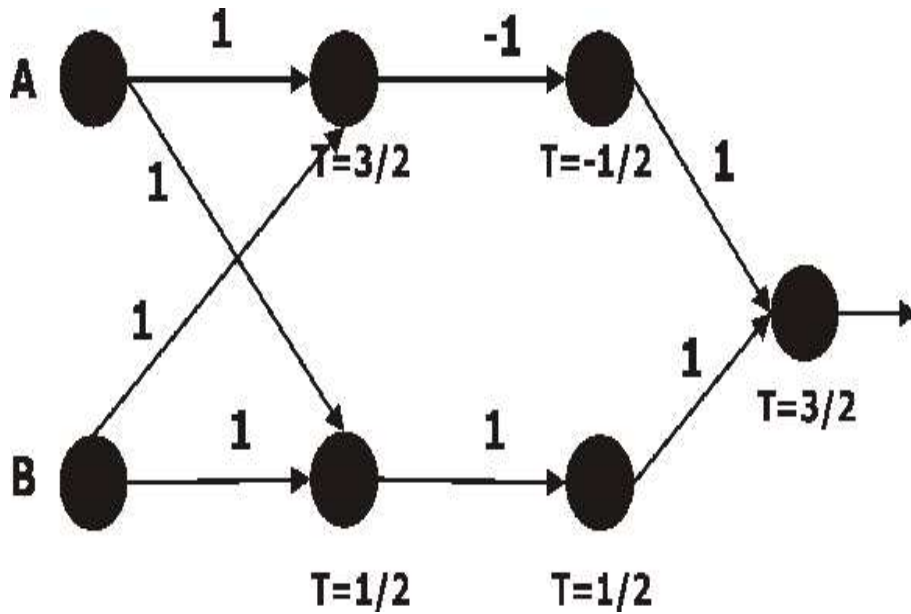


Figura 1.6: A operação booleana **xor** mostra-se mais complexa, com o encadeamento de sete neurônios.

Se no instante  $t$  uma for verdadeira e a outra não, o neurônio à direita não disparará em  $t + 1$ . Mas, se A e B forem verdadeiras ao mesmo tempo, o neurônio pós sináptico irá disparar em  $t + 1$ . Logo, (b) é a operação booleana **and**. Seguindo o mesmo raciocínio, em (c) os pesos são escolhidos de tal maneira que a operação booleana resultante seja o **or**. Para montar a operação **xor** (ou exclusivo) serão necessários mais neurônios, como vemos na 1.6.

### 1.2.2 Perceptrons

No final da década de 50, Rosenblatt deu prosseguimento ao trabalho de McCulloch e Pitts, e montou uma rede de neurônios e a batizou de *perceptron*, Figura 1.7. Como vimos anteriormente, a implementação de funções booleanas simples ocorrem

diretamente através de ajustes dos parâmetros, peso sináptico e potencial limite, porém para a implementação de uma função arbitrária, a escolha dos parâmetros se torna não trivial e, dependendo do número de variáveis, o trabalho pode se tornar inviável. A proposta de Rosenblatt foi desenvolver um método eficaz para determinação desses parâmetros.

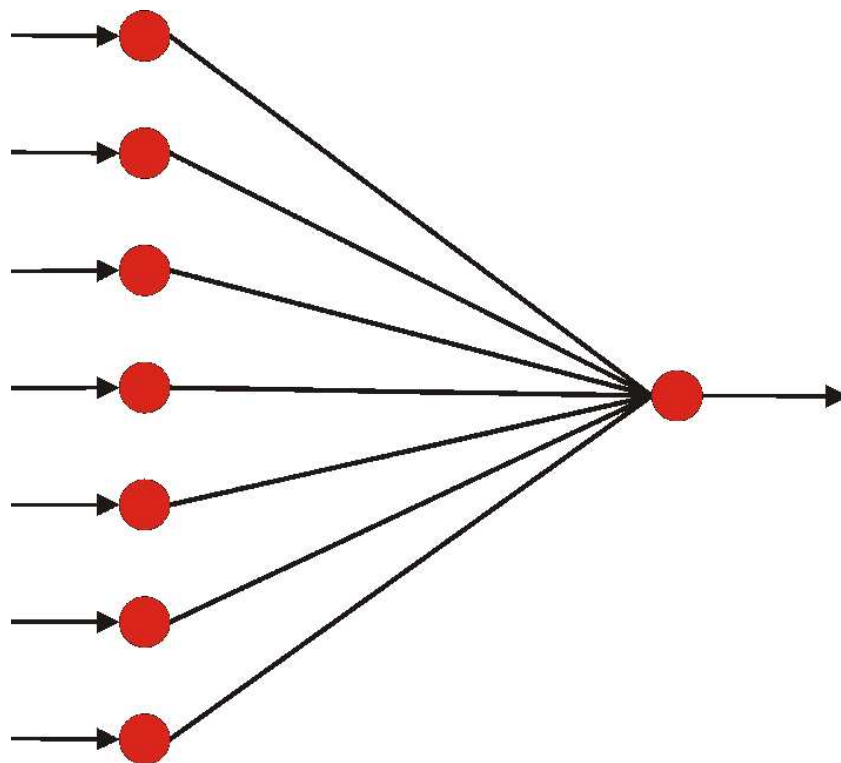


Figura 1.7: Esquema do perceptron de Rosenblatt.

Para resolver o problema dos pesos sinápticos era necessário *ensinar* a rede como responder de forma correta para cada entrada. Uma maneira intuitiva de realizar essa tarefa é através da apresentação de exemplos. Apresentam-se exemplos à rede para o qual conhecemos as saídas e ajustam-se os parâmetros até que todos os exemplos sejam corretamente respondidos pela rede. Esse tipo de aprendizado é conhecido como *aprendizado supervisionado*. A rede pode, em alternativa, aprender por um processo *não supervisionado*. Ela possui uma dinâmica pré estabelecida e calcula para um dado conjunto de entrada uma função, chamada de função custo.

Essa função é minimizada no processo e a rede retorna uma saída. Os padrões que a rede identifica estão diretamente relacionados com a função custo atribuída ao conjunto de dados. Mais a frente, quando definirmos o modelo de Hopfield, veremos uma rede neural que possui aprendizado não supervisionado.

O passo subsequente foi implementar múltiplas saídas, como mostra a Figura (1.8). Essa rede foi a primeira proposta para um modelo associativo. Dada uma entrada de  $N$  componentes a rede retornaria um vetor binário  $Q$ -dimensional. Ou seja, para um dado vetor de entrada  $\sigma_i(t)$  a rede associa um vetor de saída  $\sigma'_i(t)$ , e não apenas uma resposta do tipo função verdade.

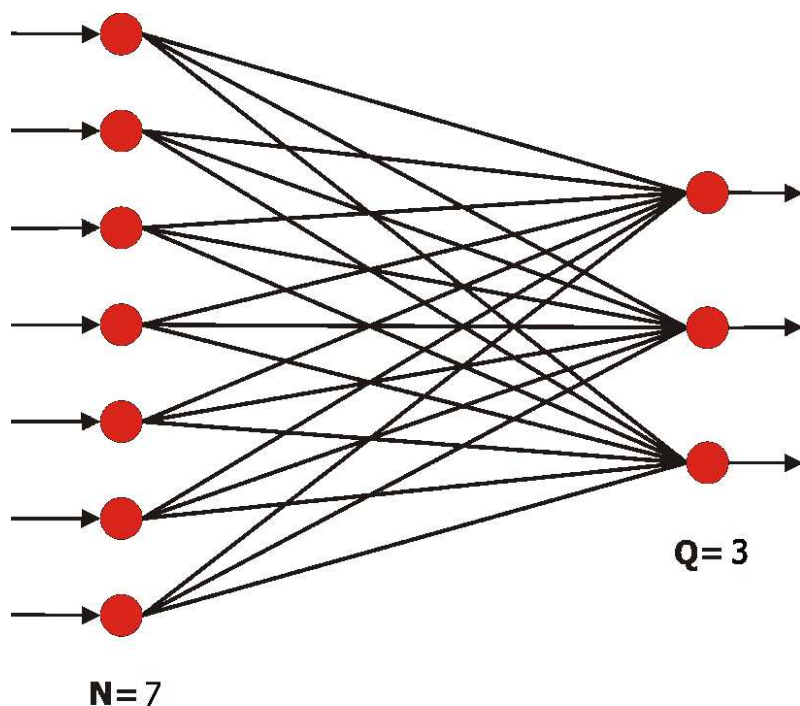


Figura 1.8: Exemplo de perceptron de várias saídas.

Da forma apresentada por Rosenblatt o perceptron possui restrições fortes como, por exemplo, ser incapaz de resolver a função **xor**. Minsky e Papert escreveram “*Perceptrons*” [10] um livro importante na história das redes neurais, abordando o perceptron de forma rigorosa. Nesse monografia é demonstrado que o perceptron é

um separador linear, sendo impossível resolver funções não lineares com o mesmo. Minsky conjecturou que o mesmo seria verdade para redes de mais camadas, mas ele estava errado. O perceptron de várias camadas é capaz de aproximar, qualquer função contínua em um intervalo fechado dos reais, com apenas uma camada oculta, como na Figura (1.9).

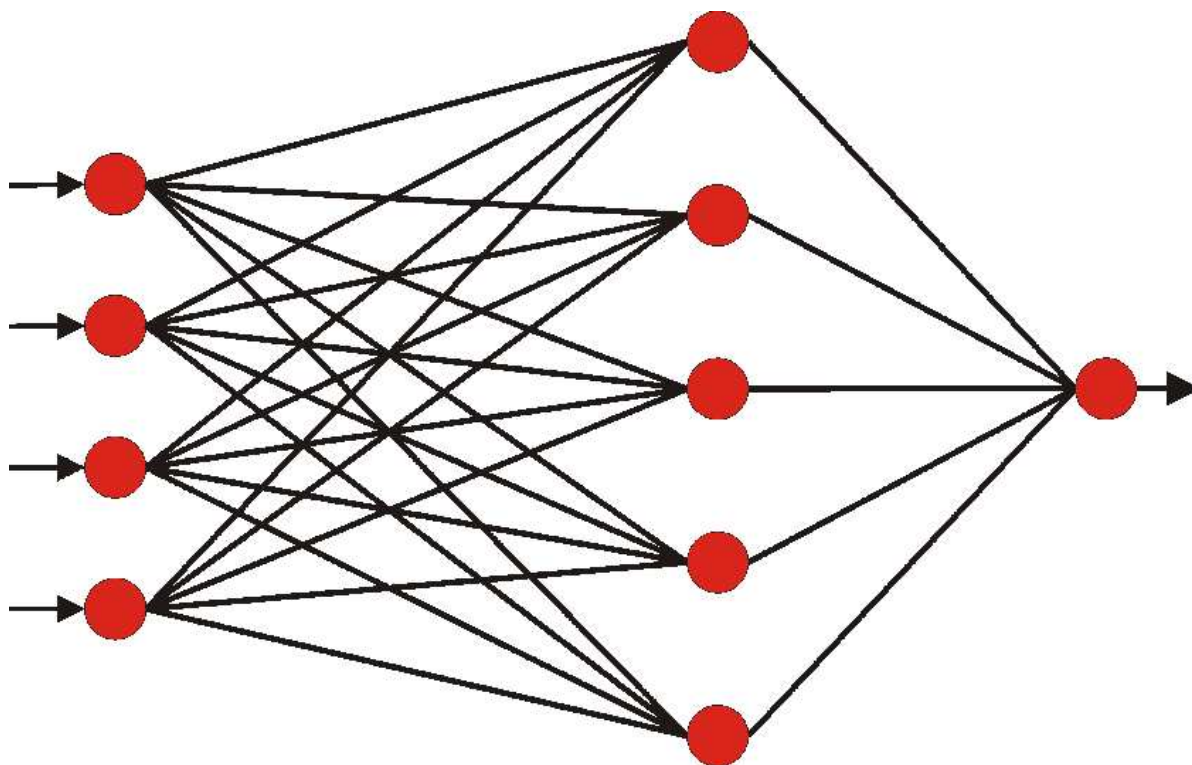


Figura 1.9: Exemplo de perceptron de várias camadas.

### 1.2.3 Do perceptron a rede neural atratora (ANN)

O próximo trabalho significativo na história das redes neurais acontece por volta de 10 anos depois do livro *Perceptrons*. A idéia diferencial, que dá um salto evolutivo na teoria de redes neurais, foi conectar a camada de saída do perceptron à sua camada de entrada, fazendo assim um sistema retro-alimentado. Deixamos de ter um sistema do tipo entrada-saída e passamos para um *sistema dinâmico*.

Seja o vetor  $\sigma(t)$  a representação do estado dos  $N$  neurônios na primeira camada no instante  $t$  e  $\sigma'(t)$  o estado da segunda camada. Damos uma entrada no instante  $t$  para rede, no instante posterior a rede nos entrega uma resposta  $\sigma'(t)$ . Então a rede é realimentada e a saída da camada dois é colocada como entrada para camada um.

$$\sigma(t + 1) = \sigma'(t) \quad (1.3)$$

Para cada neurônio  $i$  temos,

$$\sigma_i(t + 1) = \Psi[h_i(t + 1) - T_i] \quad (1.4)$$

$$h_i(t + 1) = \sum_{j=1}^N J_{ij}\sigma_j(t) \quad (1.5)$$

Deixamos de ter um projetor linear, capaz de responder a estímulos e passamos para um sistema dinâmico que vaga num espaço de  $2^N$  possíveis combinações, com uma complexa auto-alimentação. Um propriedade importante dessas redes é a possibilidade de existirem atratores em sua dinâmica, pontos tais que  $\sigma_i(t + 1) = \sigma_i(t)$  para todo  $i$ .

### 1.3 Memória Associativa

Antes de prosseguirmos, precisamos elucidar o conceito *memória* do ponto de vista cognitivo e matemático. A palavra memória é geralmente associada ao ato de lembrar uma experiência vivida ou a imagem mental de algo, ou alguém. Com um pouco mais de reflexão percebemos que o ato de lembrar ou encontrar pode ocorrer tanto de forma ativa, quando somos inquiridos em uma prova, ou de forma passiva, escapando do nosso consciente, quando, por exemplo, estamos dirigindo ou falando. De uma forma razoável, podemos definir memória como a habilidade de reter e utilizar informação ou conhecimento adquirido.

A teoria cognitiva divide a memória em três diferentes processos: decodificar, armazenar e recuperar. Podemos também classificar as memórias quanto aos tipos:

memória sensorial, memória de curto prazo (*short term memory STM*) e memória de longo prazo (*long term memory LTM*). Para entender o primeiro tipo de memória vamos fazer uma ilustração através de uma experiência cognitiva realizada por Sperling [11]: mostra-se aos participantes 3 vetores de letras, cada um com 4 caracteres, e pede-se para que eles lembrem o maior número de letras possível. Em uma segunda etapa, apresenta-se novos 3 vetores de 4 elementos e pede que cada participante lembre de um dado vetor (dizendo a linha em que ele se encontra). Na primeira etapa os participantes lembram, em média, de 4 letras. Na segunda etapa, grande parte dos participantes se lembra da exata sequência na linha pedida. Isso mostra que, por alguns poucos instantes, o participante manteve em sua mente toda a informação da grade (consegue localizar em uma dada linha a sequência de letras), mas ela decai rapidamente no tempo (não consegue se lembrar de todas as letras). Já a memória de curto prazo é aquela responsável por guardar as palavras de uma frase lidas num texto enquanto você as digita. Se não estiver em exercício essa lembrança decai no tempo, sendo esquecida. Essa é a memória que nos permite manipular informação mentalmente. A memória de longo prazo é a que contém nosso conhecimento a respeito do mundo e as memórias do passado. Aqui é importante fazer uma ressalva, essas definições estão sendo expostas simplificadaamente, pois possuem várias subdivisões e complicações que escapam do escopo desse texto.

Do ponto de vista matemático, uma memória associativa  $\Theta = \{\xi^\mu\}$  é um conjunto de  $P$  vetores, ou padrões, armazenados numa rede neural. Quando um vetor de mesma dimensão é apresentado a rede, ela retorna o vetor  $\xi^\mu$  *mais próximo*. Veja que no momento que falamos “mais próximo”, estamos colocando um critério de distância, logo uma métrica. Como queremos modelar neurônios, que apresentam um comportamento do tipo “tudo ou nada”, podemos optar por dois tipos básicos de representação: 0 e 1 ou -1 e 1. A primeira representação é intuitiva e foi utilizada anteriormente, a segunda é necessária por motivos analíticos. São equivalentes e a transformação de uma para a outra é trivial. Se optarmos por vetores binários, a métrica é a *distância de Hamming* definida pelo número de bits diferentes entre um



vetor e outro. Na representação -1 e 1, a métrica é a usual, definida pelo produto escalar.

É importante notar que a modelagem de memória associativa não está necessariamente atrelada à rede neural atratora. Podemos atacar o problema através uma rede neural atratora desde que exista um método para garantir que os padrões memorizados sejam atratores na dinâmica e que esta dinâmica leve o vetor apresentado ao padrão mais próximo. Ou podemos utilizar uma rede em camada do tipo perceptron aonde dada uma entrada, a saída será o vetor  $\xi^\mu$  mais próximo. Não entraremos no mérito de discutir qual a melhor abordagem, mas algumas ressalvas devem ser feitas:

- Como dito anteriormente, o processo associativo possui várias etapas e para realizar cada etapa o sistema nervoso possui estruturas especializadas;
- A anatomia cerebral possui um indiscutível fluxo, que pode ser determinado por projeções de sinais de uma camada para outra camada. Mas é conhecido também que em  $1\text{ mm}^3$  de córtex existe a mesma ordem de grandeza de sinapses conectadas com o exterior, como sinapses conectadas interiormente, ou seja, anatomicamente o cérebro é distinto de ambas as abordagens teóricas;
- A experiência também traz evidências de estados de atividade cerebral persistente. Isso ocorre desde primitivos circuitos neuronais como os de invertebrados, até no altamente sofisticado cérebro humano.

## 1.4 Aprendizado em Redes Neurais

Não entramos em detalhe a respeito da matriz  $J_{ij}$ , equação 1.9 e, até o momento, ela é apenas um parâmetro a ser ajustado no modelo. Mas isso não é verdade, a matriz das sinapses desempenha um importante papel no aprendizado da rede, em especial nas memórias associativas. Foi proposta por Donald Hebb [12], uma teoria para a função das sinapses no processo de aprendizado:

“Quando um axônio de uma célula A está suficientemente perto para excitar uma célula B e repetidamente ou de forma persistente participa no seu disparo, tem lugar algum processo de crescimento ou câmbio metabólico em alguma ou ambas células, de tal forma que a eficácia de A, como uma das células que causam o disparo de B, é aumentada.”

Mediante este processo, as memórias são armazenadas na rede, distribuídas pelas suas sinapses. Em nosso trabalho iremos nos deter a regra Hebbiana de aprendizado proposta por Hopfield, que é expressa matematicamente da seguinte forma:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^{\mu} \xi_j^{\mu} \quad (1.6)$$

Um fato biologicamente importante é que a regra de Hopfield é local no sentido topológico, dependendo unicamente das atividades dos neurônios pós e pré sinápticos.

## 1.5 Modelo de Hopfield

O modelo de Hopfield propõe uma dinâmica para redes neurais atratoras possuindo várias propriedades interessantes. Mais adiante trataremos com detalhe o modelo, por enquanto preocupar-nos-emos com as fortes hipóteses neurobiológicas tomadas nessa abordagem:

- O neurônio, individualmente, não apresenta memória.

Assumindo que o potencial limite é independente do tempo, e o sinal  $h_i(t + 1)$  depende unicamente do estado da atividade no tempo anterior. A dinâmica neuronal é *markoviana*.

- O axônio do mesmo neurônio pode ter qualquer número de sinapses excitatórias e inibitórias.

Ter ambos os tipos de sinapses é vital para que o sistema apresente um comportamento interessante, possuindo uma variedade de atratores distintos.

- Totalmente conectado

Cada neurônio recebe informação de todos os outros neurônios na rede, e pode mandar informação para qualquer outro. Como vimos em 1.2, o córtex cerebral não é totalmente conectado. Essa imposição não tem fundamento biológico.

O trabalho de Hopfield, o qual temos por referência fundamental, introduz mais duas adicionais simplificações, que por motivo de coesão iremos apresentá-las aqui.

- Conectividade simétrica

Isso significa que a eficiência sináptica de neurônio  $j$  para o neurônio  $i$  é idêntica à eficiência inversa, do neurônio  $i$  para o  $j$ . Essa forte hipótese não tem nenhuma motivação biológica e surge da necessidade de definir uma função energia para o sistema.

- A dinâmica é não sincrônica.

Essa hipótese não tem apelo analítico como a anterior e também não é motivada biologicamente, mas é tomada em via de ser menos restritiva. Para entender essa restrição suponha que a atualização ocorresse conjunta, todos os neurônios tomassem suas decisões para uma mesma configuração. Os neurônios disparam quando acumulam suficiente PSP e atingem o potencial limiar, recebendo sinais com intervalo de tempos diversos. Dessa forma começando sincronicamente uma dinâmica, rapidamente ficariam desincronizados. Além disso, outros fatores podem interferir na diferença de tempo entre a chegada de informação entre um neurônio e outro, como tamanho do axônio e dendritos assim como da liberação de neurotransmissores na fenda sináptica. Esse tipo de desordem tão detalhada é difícil de ser levada

em conta em um modelo. Então preferimos assumir uma hipótese mais forte, a de atualizar os neurônios em série, com uma seqüência aleatória pré-estabelecida. Dessa forma cada neurônio terá seu PPS calculado como a combinação particular de potenciais ativos.

Vamos definir as grandezas matemáticas que compõe o modelo. Como mencionado anteriormente, representamos a atividade neuronal por uma forma binária (0,1) ou (-1,+1). Vemos que a segunda representação tem um apelo físico pois nos remete ao modelo de Ising para o magnetismo. Então, podemos nos valer da simples transformação:

$$S_i = 2\sigma_i - 1 \quad (1.7)$$

onde  $\sigma$  é atividade neuronal na forma binária. Seja  $U_i(t+1)$  o PPS acumulado no neurônio  $i$  no instante  $t+1$ . Se  $J_{ij}$  for a matriz das conexões sináptica, o PPS pode ser escrito como

$$U_i(t+1) = \frac{1}{2} \sum_{j=1; j \neq i}^N J_{ij}(S_j + 1) \quad (1.8)$$

E o *estado* do neurônio  $i$ , na ausência de ruído, será dado por

$$S_i(t+1) = \text{sign}(U_i(t+1) - T_i) \quad (1.9)$$

e função  $\text{sign}(x)$  é definida da seguinte forma:

$$\text{sign}(x) = \begin{cases} +1 & \text{se } x > 0 \\ -1 & \text{se } x < 0 \end{cases} \quad (1.10)$$

Reescrevendo 1.9,

$$S_i = \text{sign}(h_i + h_i^e) \quad (1.11)$$

tal que

$$h_i = \sum_{j=1; j \neq i}^N J'_{ij} S_j \quad (1.12)$$

e

$$h_i^e = \sum_{j=1; j \neq i}^N J'_{ij} - T_i \quad (1.13)$$

com

$$J'_{ij} = \frac{1}{2} J_{ij} \quad (1.14)$$

Escrevendo dessa maneira podemos interpretar os parâmetros  $h_i$  e  $h_i^e$  como análogos de um sistema magnético de spin  $S_i$  conectado pela matriz de interações  $J_{ij}$ . A variável  $h_i$  seria o campo que o  $i$ -ésimo spin sente devido à interação com seus  $j$  vizinhos, enquanto a variável  $h_i^e$  é o campo magnético externo no sítio  $i$ , que não depende da orientação dos outros spins. Para simplificar o modelo podemos assumir que

$$h_i^e = 0 \quad (1.15)$$

Fazer essa hipótese significa que o limiar de cada neurônio está balanceado com atividade média dos neurônios de toda rede. Eliminar o limiar nos permitirá apresentar uma discussão transparente, que enfoque no papel desempenhado pelas eficiências sinápticas no controle do comportamento coletivo da rede neural atratora (ANN).

Em síntese, a dinâmica neuronal é dada por:

$$S_i = \text{sign}(h_i) \quad (1.16)$$

$$h_i = \sum_{j=1, j \neq i} J_{ij} S_j \quad (1.17)$$

Note que não regressamos ao peso sináptico anterior, mas apenas abolimos  $J'_{ij}$  em favor de uma notação enxuta.

As relações descritas anteriormente são totalmente determinísticas. Dada uma condição inicial o sistema segue por uma trajetória bem definida no espaço de configurações. Na prática isso implica que o neurônio não está sujeito a nenhum tipo de flutuação ou imperfeição em todo o complicado processo biológico descrito na seção 1.2, o que é bastante irreal. Evidências experimentais [13] mostram que devido ao processo de transmissão nas sinapses, as vesículas de neurotransmissores variam em tamanho e quantidade, podendo neurotransmissores serem despejados na sinapses mesmo sem a existência de um sinal potencial de ação. Então uma melhor descrição

para o potencial pós sináptico é através de uma variável gaussiana,

$$P(U_i) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left[-\frac{(U_i - \bar{U}_i)}{2\delta^2}\right] \quad (1.18)$$

Dessa forma a probabilidade do neurônio  $i$  disparar,

$$P(S_i = 1) = \int_{T_i}^{\infty} P(U)dU = \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{\bar{U}_i - T_i}{\delta\sqrt{2}}\right) \right] \quad (1.19)$$

e

$$P(S_i = 1) = \frac{1}{2} \left[ 1 - \operatorname{erf}\left(\frac{\bar{U}_i - T_i}{\delta\sqrt{2}}\right) \right] \quad (1.20)$$

onde

$$\operatorname{erf}(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-t^2} dt \quad (1.21)$$

Utilizando (1.19) e (1.21)

$$P(S_i) = \frac{1}{2} \left[ 1 - \operatorname{erf}\left(\frac{h_i S_i}{\delta\sqrt{2}}\right) \right] \quad (1.22)$$

que pode ser aproximada por

$$P(S_i) = \frac{\exp(\beta h_i S_i)}{\exp(\beta h_i) + \exp(-\beta h_i)} = \frac{1}{2} [1 + \tanh(\beta h_i S_i)] \quad (1.23)$$

com

$$\beta^{-1} = 2\delta\sqrt{2} \quad (1.24)$$

A equação (1.23) é exatamente a equação que descreve um spin de Ising em contato com um banho térmico  $T = \beta^{-1}$ . Note que no limite  $\beta \rightarrow \infty$  retomamos a equação (1.16).

### 1.5.1 Evolução temporal

Podemos pensar em uma rede de neurônios, conectados entre si, emitindo e recebendo impulsos de seus vizinhos. Uma vez que cada neurônio pode assumir 2 valores, a evolução temporal dessa rede se dá num espaço de configurações que é um hipercubo N-dimensional, onde N é o número de neurônios.

Para analisarmos a evolução temporal da rede devemos pensar em como é a sua dinâmica. Podemos assumir uma dinâmica sincrônica, ou *paralela*, onde todos disparam juntos em intervalos bem definidos de tempo. Dessa forma, num instante  $t = n$ , o PPS do neurônio será dado pela atividade dos seus neurônios vizinhos no intervalo de tempo  $(n - 1) < t < n$ . A cada unidade de tempo todos os neurônios repolarizam a sua membrana. Então a probabilidade de transição  $W(p, q)$  de um estado  $q \{S_i^q(n - 1)\}$  da rede para um estado  $p \{S_i^p(n)\}$  é dada por:

$$W(p, q) = \prod_{i=1}^N P(S_i^p) = \frac{\exp(\beta \sum_i h_i^q S_i^p)}{\prod_i [\exp(\beta h_i^q) + \exp(-\beta h_i^q)]} \quad (1.25)$$

com

$$h_i^q(n) = \sum_{j, j \neq i} J_{ij} S_j^q(n - 1) \quad (1.26)$$

Os índices superiores são para reforçar que em um passo de tempo todos os neurônios tiveram a chance de mudar de estado.

A outra hipótese é uma dinâmica assincrônica, ou *serial*, onde em um intervalo  $\delta t$  apenas um neurônio pode vir a mudar de estado. Contaremos então uma unidade de tempo quando todos os  $N$  neurônios da rede tiveram oportunidade de trocar de estado. Essa diferente forma de contar tempos é importante para relacionar resultados em simulações feitas com dinâmicas distintas. Dessa forma:

$$\delta t = \frac{1}{N} \quad (1.27)$$

Se, no instante de tempo  $t = n$ , o  $i$ -ésimo neurônio muda de estado no intervalo  $n < t < n + \delta t$  o próximo neurônio terá seu PPS baseado na mesma configuração do neurônio  $i$ , com exceção da troca de estado ocorrida naquele neurônio.

A respeito da probabilidade de transição entre dois estados,  $p$  e  $q$ , podemos afirmar:

- Se dois estados diferem pela mudança de atividade de mais de um neurônio,  $W(n, m) = 0$

Suponha por que a rede esteja numa dinâmica serial, e o tempo seja discreto, a probabilidade de encontrar a rede num instante de tempo  $t = (n + 1)\delta t$  é

$$\rho_p(n + 1) = \sum_q W(p, q)\rho_q(n) \quad (1.28)$$

A soma deve ser feita em todos os  $2^N$  estados, uma vez que a rede, em  $t = n$ , poderia estar em qualquer um desses estados. Escrever a equação (1.28) é equivalente a equação mestra da rede. Se pensarmos na distribuição de probabilidades como um vetor de rede com  $2^N$  coordenadas, a equação (1.26) pode ser escrita como

$$|\rho(n + 1)\rangle = W |\rho(n)\rangle \quad (1.29)$$

onde  $W$  é a matriz estocástica,  $|\rho(n)\rangle$  é o vetor de estado da rede no tempo  $t = n$ . Note que aqui pegamos emprestada a notação da mecânica quântica por uma questão de familiaridade. Seja  $|\rho(0)\rangle$  a condição inicial da rede, então

$$|\rho(n)\rangle = W^n |\rho(0)\rangle \quad (1.30)$$

Estamos interessados no estado assintótico da rede, pois desejamos um sistema tal que dada uma condição inicial (um padrão apresentado) o seu estado estacionário é o atrator mais próximo dele (ele retorne o padrão armazenado com maior semelhança). Então precisamos saber quais propriedades são necessárias à  $W$  para que o sistema possua uma diversidade de estados estacionários, que possuam dependência da condição inicial, em outras palavras, quais características de  $W$  para que o sistema não seja *ergódico*. Para qualquer matriz estocástica valem as seguintes propriedades:

- 1- Como  $W(p, q)$  é uma probabilidade condicional,  $W(p, q) \geq 0$  e  $\sum_p W(p, q) = 1$ ;
- 2- Qualquer autovalor  $\lambda$  de  $W$  satisfaz a condição  $|\lambda| \leq 1$ , isto no plano complexo;
- 3- O autovetor associado ao autovalor  $\lambda = 1$  corresponde ao autovetor de componentes não negativas;



Além desses resultados, precisaremos enunciar uma definição para entendermos o teorema de Perron-Frobenius.

*Definição:* Uma matriz estocástica é dita irredutível se para cada par  $(p, q)$  existe  $l > 0$  tal que  $T^l(p, q) > 0$ .

Uma forma intuitiva de enxergar essa definição é que uma matriz irredutível é uma matriz que não pode ser escrita na forma de uma matriz triangular,

$$\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \quad (1.31)$$

onde  $A$  e  $D$  são matrizes quadradas. Quando uma matriz estocástica pode ser escrita nessa forma, ela é dita redutível.

É importante entender que toda matriz irredutível representa uma rede em que qualquer estado  $q$  é acessível a partir de um estado arbitrário  $p$ . Ou seja, é possível transitar entre todos os estados da rede.

*Teorema de Perron-Frobenius:* O autovalor  $\lambda = 1$  de uma matriz irredutível é não degenerado e seu autovetor correspondente possui todas as componentes estritamente positivas.

Suponha  $W$  tal que valha o teorema e sejam  $|\lambda_i\rangle$  o autovetor a direita, e  $\langle\lambda_i|$  o autovetor a esquerda do autovalor  $\lambda_i$ . Utilizando o teorema espectral,

$$W = \sum_i \lambda_i |\lambda_i\rangle \langle\lambda_i| \quad (1.32)$$

Substituindo (1.32) em (1.30) temos,

$$|\rho(n)\rangle = W_n |\rho(0)\rangle = \sum_i \lambda_i^n |\lambda_i\rangle \langle\lambda_i| |\rho(0)\rangle \quad (1.33)$$

reescrevendo

$$|\rho(n)\rangle = \lambda_1^n |\lambda_1\rangle \langle\lambda_1| |\rho(0)\rangle + \sum_{i \neq 1} \lambda_i^n |\lambda_i\rangle \langle\lambda_i| |\rho(0)\rangle \quad (1.34)$$

No limite  $n \rightarrow \infty$ , o segundo termo desaparece pois todos os autovalores possuem módulo menor que 1. No primeiro termo a direita  $\langle\lambda_1| |\rho(0)\rangle = 1$  pois  $|\rho(0)\rangle$

está normalizado e o teorema garante que todas as componentes de  $\langle \lambda_1 |$  são estritamente positivas. Dessa forma

$$|\rho(n)\rangle = |\lambda_1\rangle \quad (1.35)$$

o que implica que depois de um tempo suficientemente longo a condição inicial é esquecida e qualquer configuração é levada ao mesmo atrator. Esse resultado tão útil em Física, ironicamente, é o contrário do que desejamos para uma rede neural atratora. Dessa forma precisamos encontrar uma maneira de evitar o teorema de Perron-Frobenius. Para invalidar o teorema de maneira simples basta que a matriz  $W$  seja redutível. Vemos que isso é claramente impossível para uma dinâmica paralela com temperatura não nula, pois o sistema tem livre trânsito entre os estados. Mesmo para a dinâmica seqüencial com ruído a matriz continua irredutível. Mas se a dinâmica for a temperatura nula e existir ao menos um atrator, a matriz  $W$  poderá ser escrita facilmente na forma triangular, pois existirá uma coluna composta apenas por 1 na diagonal e 0 em todo resto bastando uma permutação simples para colocá-la na forma triangular.

Para escapar desse problema, à temperatura finita, é necessário que  $W$  seja tal que a diferença entre os autovalores subseqüentes diminuam a medida que o sistema aumenta, então quanto tomarmos o limite termodinâmico existirá uma degenerescência assintótica. Foge do escopo desse texto explicar essa degenerescência. Mas uma coisa é clara, a topologia da rede é crucial para os efeitos emergentes desse tipo. Como as novas arquiteturas abordadas irão influenciar esse comportamento, existirá uma superposição dos novos efeitos emergentes com os efeitos da diluição não aleatória. Com intuito de isolar o efeito de novas topologias e suas influências na capacidade de armazenamento optamos por realizar apenas simulações em  $T = 0$ , e daqui iremos nos referir apenas a rede à temperatura nula.

## 1.5.2 Energia e a matriz sináptica

Seja uma rede neural de  $N$  variáveis  $S_i = \pm 1$ , e a função

$$E = \frac{1}{2} \sum_{i,j,i \neq j} J_{ij} S_i S_j \quad (1.36)$$

com  $J_{ij}$  qualquer. Suponhamos que em algum instante de tempo um dado neurônio  $k$  muda de estado, ou seja,  $S_k \rightarrow -S_k$ . Podemos calcular a variação na função  $E$ , dada por:

$$\Delta E = S_k \sum_{j,j \neq k} J_{kj} S_j + S_k \sum_{j,j \neq k} J_{jk} S_j \quad (1.37)$$

Analisemos essa equação. Pelas equações (1.14) e (1.15) o primeiro termo é negativo, uma vez que  $S_k$  mudou de sinal. Mas nada podemos afirmar a respeito do segundo termo, e não conseguimos inferir nada sobre  $\Delta E$ . Mas se a matriz for simétrica o segundo termo é igual ao primeiro e  $E$  diminui. Assim  $E$  pode ser escolhida como função energia para o problema, que decresce junto com a dinâmica, tendo por pontos mínimos os atratores. Note que se  $J_{kk} \neq 0$  o argumento exposto acima não seria válido.

Para uma dinâmica paralela considere

$$E = - \sum_i \left| \sum_{j,j \neq i} J_{ij} S_j \right| \quad (1.38)$$

Usando a identidade

$$|x| = x \cdot \text{sign}(x) \quad (1.39)$$

temos

$$E = - \sum_i \left( \sum_{j,j \neq i} J_{ij} S_j \right) \text{sign} \left( \sum_{j,j \neq i} J_{ij} S_j \right) \quad (1.40)$$

Mas

$$S_i(t+1) = \text{sign} \left( \sum_{j,j \neq i} J_{ij} S_j \right) \quad (1.41)$$

logo

$$E(t) = - \sum_{i,j,j \neq i} J_{ij} S_i(t+1) S_j(t) \quad (1.42)$$

para  $J_{ij}$  simétrica podemos escrever

$$\Delta E = E(t+1) - E(t) = - \sum_i [S_i(t+2) - S_i(t)] \sum_{i,j \neq i} J_{ij} S_j(t+1) \quad (1.43)$$

Note que se  $S_i(t+2) \neq S_i(t)$ ,  $\Delta E$  é negativo. De fato o estado de  $S_i(t+2)$  é dado por (usando eq (1.41))

$$S_i(t+2) = \text{sign}(h_i(t+1)) = \text{sign}\left(\sum_{j \neq i} J_{ij} S_j(t+1)\right) \quad (1.44)$$

Então  $S_i(t+2)$  e o somatório em  $j$  possuem o mesmo sinal. Por hipótese,  $S_i(t+2) \neq S_i(t)$ . Se  $S_i(t+2)$  é positivo, assim também será o somatório em  $j$ ,  $S_i(t)$  será negativo. Dessa forma, a diferença de energia entre os dois estados será negativa. Se  $S_i(t+2)$  é negativo, o somatório em  $j$  será negativo.  $S_i(t)$  será positivo, e contribuirá negativamente pois está sendo subtraído. Então o sinal do termo entre colchetes é negativo. Levando em conta os dois sinais fora do colchetes,  $\Delta E < 0$ . Então  $E$  define uma função de Lyapunov. Analisemos agora caso em que  $S_i(t) = S_i(t+2)$ . Existe duas possibilidades para esse caso,  $S_i(t)$  é constante, ou seja, chegamos em um estado estacionário, ou temos um ciclo duplo.

Em nossas simulações, por questão de tradição e clareza nos resultados, optamos apenas por dinâmica seqüencial.

### 1.5.3 Estados espúrios e saturação

Conseguimos escapar da ergodicidade e escrever uma função energia para o problema, mas nada garante ainda que os padrões armazenados segundo a equação (1.6) serão os pontos fixos.

Até agora discutimos a respeito da dinâmica, da existência ou não de atratores mas ainda não temos uma forma eficaz de medir o estado da rede. Para isso vamos introduzir uma nova grandeza, a “*superposição*”. A superposição de um dado padrão  $\mu$ , é definida como a distância entre o atual estado da rede e o padrão, e é

representada por  $m_\mu(t)$

$$m_\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i(t) \quad (1.45)$$

Iremos aqui citar alguns resultados importantes a respeito dos atratores para rede à  $T = 0$ .

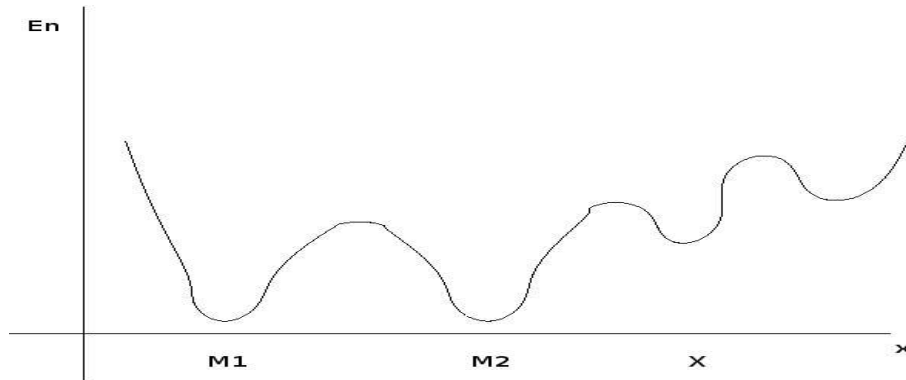


Figura 1.10: Metáfora do vales e colinas. M1 e M2 são padrões armazenados, mínimos globais. X é um atrator espúrio, mínimo local.

Se os  $P$  padrões armazenados são aleatórios, e o número de padrões é bem menor que o número de sinapses, valem as seguintes propriedades:

- Os  $P$  padrões, assim como seus simétricos, são atratores [4];
- Os padrões são mínimos absolutos da função de Lyapunov [13];
- A bacia de atração dos padrões são largas.

Entendemos por bacia de atração de um padrão, toda a região do espaço de configurações que se dado um estímulo dentro dessa, o padrão em questão é recuperado. Então, quando dizemos que a bacia é larga, mesmo que o estado inicial tenha poucos neurônios alinhados com o padrão, ele será recuperado.

- a rede possui uma grande quantidade de *atratores espúrios, ou estados espúrios*;

- Existe um atrator espúrio para cada combinação ímpar simétrica, totalizando  $3^P$  atratores espúrios devido a essas misturas simétricas;
- Atratores espúrios são meta estáveis, isto é, são mínimos locais da função energia, como mostra a Figura 1.10;

Até então as propriedades descritas transformam as redes neurais atratoras em bons candidatos para aplicações em inteligência artificial ou em problemas de computação. Mas quando pensamos em qualquer tipo de aplicação estamos interessados na capacidade do sistema, dessa forma uma questão que segue naturalmente, é o quanto é capaz de armazenar uma rede neural atratora? Antes de respondermos o quanto ela é capaz, é necessário definir a capacidade. Podemos pensar em bits armazenados por sinapses ou neurônios, ou em padrões armazenados por sinapses ou neurônios. Pensando de uma forma aplicada é importante a questão do tamanho da bacia de atração dos padrões, pois uma rede com bacias largas pode recuperar um padrão à custa de uma pequena informação apresentada. Enunciemos agora algumas propriedades da redes neurais atratoras relativas à capacidade de armazenamento.

- Costuma-se definir a carga da rede por:

$$\alpha \equiv \frac{P}{N-1} \simeq \frac{P}{N} \quad (1.46)$$

Note que  $N-1$  é o número de sinapses por neurônio. Como o modelo de Hopfield é totalmente conectado, o número de neurônios serve como aproximação para a número médio de sinapses.

- Existe um número crítico de padrões armazenados,  $P_c$ . Para qualquer número de padrões armazenados abaixo de  $P_c$  a rede recupera o padrão com acerto de 99%. Costuma-se definir:

$$P_c \equiv \alpha_c N \quad (1.47)$$

- Quando  $\alpha$  assume valores maiores que  $\alpha_c$  nenhum padrão é recuperado. Para a rede totalmente conectada, e a regra de aprendizado hebbiana proposta por Hopfield,  $\alpha_c = 0.14$ .

## 1.6 Conclusões

Neste capítulo apresentamos o conceito de redes neurais. Começamos por mostrar um pouco da sua evolução histórica, com o intuito de contextualizar o trabalho de físicos nessa área. Para isso fizemos uma introdução simplificada de neurofisiologia, apresentando as principais características de uma rede de neurônios reais, com intuito de trazer entendimento a respeito do ente modelado. Apresentamos um modelo para a memória associativa, o modelo de Hopfield, e discutimos suas principais propriedades matemáticas e restrições.

# Capítulo 2

## Redes Complexas

A teoria de redes complexas é uma área multidisciplinar e pode ser vista como intercessão entre a mecânica estatística e a teoria dos grafos. Um grande volume de trabalhos vem sendo publicados na literatura abordando esse tema, pois esse tipo de modelagem se mostra cada vez mais útil para o entendimento de vários sistemas como a Internet, WWW (World Wide Web), redes sociais entre indivíduos, rede de negócios entre companhias, redes de interações biológicas, como proteína-proteína, cadeias alimentares, citações acadêmicas e redes neurais. Não temos como objetivos aqui esgotar o assunto, e deixamos indicados para leitores mais assíduos três ótimos artigos de revisão a respeito do tema [14, 15, 16] e dois livros [17, 18].

Na matemática a teoria dos grafos é uma das bases da matemática discreta e está solidamente desenvolvida. Nas ciências sociais desde os anos 30 existem estudos das conexões sociais com objetivo de entender o funcionamento da sociedade, por exemplo, o estudo de pessoas e relações de trabalhos entre elas, ou a amizade entre crianças em uma escola. Mas ambas as teorias abordaram sempre redes de poucos elementos.

Recentemente o foco na pesquisa de redes trasladou, de poucos nós, para redes de larga escala com milhões, até bilhões, de elementos. Com rápida melhora dos computadores nos últimos 20 anos, foi possível recolher e analisar uma quantidade de dados muito maior que as análises anteriores. Questões antes levantadas para



redes pequenas, como: “qual é o vértice mais importante para a conectividade dessa rede?” perderam o sentido quando nos deparamos com redes de milhões de nós. Dessa forma as questões foram reformuladas, e, por exemplo, a questão anterior se transforma em: “qual a percentagem de nós necessária para que a conectividade da rede caia significativamente?” Essa é uma pergunta tipicamente estatística. Com essa mudança de escala tivemos uma mudança na abordagem analítica. Eis então que a colaboração dos físicos na área se torna importante trazendo o ferramental da mecânica estatística para estudar esses sistemas.

Outro problema é que por tratarmos de redes muito grandes, a representação gráfica perde sentido, a não ser em casos particulares. Mesmo com os modernos programas gráficos uma análise no olho de uma rede dessas proporções traz pouca informação, quando traz.

Este capítulo será dividido em 3 seções. A primeira corresponde ao estudo experimental de redes. Esse é um problema que requer muito trato, pois se tratando de redes tão grandes não podemos em vários casos, como a Internet por exemplo, contemplar toda rede mas apenas amostras de seu contexto. Então temos que caracterizar a rede por um conjunto de grandezas. Na primeira seção trataremos dessas grandezas, e de como são medidas nas redes reais, para que essas sejam caracterizadas. Já na seção dois, partiremos para a modelagem. Mostraremos os modelos mais comuns, sem suas generalizações e calcularemos algumas de suas grandezas. Na última seção iremos ver os trabalhos experimentais a respeito de rede neuronal da C. Elegans.

## 2.1 Definições Básicas

A primeira barreira a ser vencida para continuar o nosso estudo é a do jargão. Entre os físicos e matemáticos já temos uma diferença considerável nos termos utilizados para descrever os sistemas. Essa diferença começa quando nós, físicos cha-

mamos o sistema de *rede*, enquanto os matemáticos o chamam de *grafos*. Quando passamos para outras áreas, como ciências sociais ou biologia, a barreira da nomenclatura cresce. Então tentaremos aqui fazer o elo entre a nomenclatura da física e da matemática, uma vez que as outras não serão importantes para o nosso trabalho em questão.

Uma rede, ou grafo, nada mais é que um conjunto de vértices, ou nós, ligados por conexões, ou arestas. Uma rede pode ter diferentes tipos de vértices assim como diferentes tipos de arestas. Pensando no exemplo de redes sociais, um vértice pode representar mulheres ou homens, pessoas de diferentes etnias, idades, ou várias outras coisas. As arestas podem representar amizades, relações profissionais ou proximidade geográfica. Elas podem possuir pesos, representando por exemplo, o quão duas pessoas se conhecem. Formalizando podemos escrever: um grafo é dado por um conjunto não vazio  $V$  de vértices, um conjunto  $E$  de arestas, e uma função  $w : E \rightarrow P(V)$  que associa a cada elemento de  $E$  um subconjunto de elemento de  $V$ , interpretado como os pontos terminais da aresta. Se esse subconjunto de  $V$  é maior que um par, temos uma hiperaresta. Um grafo que possui uma hiperaresta é dito hipergrafo. Hipergrafos não serão de uso para nosso trabalho, mas possuem

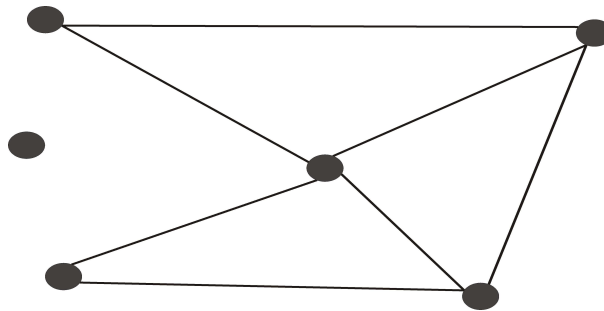


Figura 2.1: Exemplo de um grafo com 6 vértices e 7 arestas.

utilidades para representações sociais, como por exemplo vínculos familiares.

Grafos cujas arestas possuem sentido são chamados de grafos direcionados, ou apenas digrafos. Um grafo desse tipo pode ser utilizado para representar chamadas telefônicas ou e-mails, uma vez que a informação tem uma direção única de fluxo.

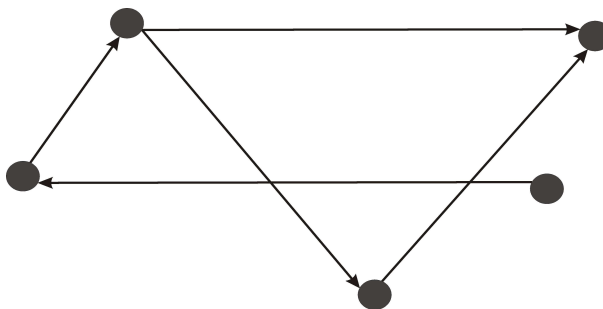


Figura 2.2: Exemplo de um digrafo

Um digrafo consiste em um conjunto  $V$  de vértices, um conjunto  $E$  de arestas e mapas  $s, t : E \rightarrow V$ , onde  $s(e)$  é a fonte e  $t(e)$  é o alvo da aresta  $e$ . Se um digrafo apresenta caminhos fechados dizemos que ele é cíclico, caso contrário acíclico. Cadeias alimentares são exemplos de digrafos quase perfeitamente acíclicos.

O que os matemáticos chamam de vértices, na literatura física encontramos pelo nome de sítio. É o elemento fundamental do sistema, que representa o ente a ser estudado. É necessário definirmos também o grau do vértice. O grau de um vértice é o número de arestas que chega no mesmo. Se estivermos tratando de um digrafo, existirá o grau de entrada e o grau de saída. Mas como não trabalharemos com digrafos, essas diferenciações não serão importantes. Os físicos chamam essa grandeza de conectividade. É importante notar que a conectividade média para um grafo não é o número de arestas dividido pelo número de nós, pois cada aresta liga dois vértices, contribuindo para o grau de dois nós distintos, então a conectividade média, ou o grau médio, é dado pelo dobro de arestas dividido pelo número de vértices.

Vamos discutir agora algumas propriedades mais práticas dos grafos. A primeira delas é o *caminho geodésico*. Antes de definir o caminho geodésico é necessário definirmos o que é um caminho em um grafo.

Um caminho do vértice  $v_i$  ao vértice  $v_j$  é uma seqüência de vértices tal que  $(v_k, v_{k+1})$  com  $i \leq k \leq |j - 1|$  pertence a  $E$ . Ou seja, se existe um caminho entre dois vértices podemos, andando por cima do grafo, sair de um vértice e chegar ao outro. Dizemos que os vértices são conectados quando existe um caminho entre

eles. Quando existe um caminho para quaisquer dois vértices de um grafo ele é dito conexo. O caminho geodésico é o caminho com menor número de vértices, ou seja, o caminho mais curto. Note que estamos utilizando uma idéia de medir distância sem definirmos bem uma métrica. Podemos pensar, por questões de simplicidade, que o número de vértices que compõe um caminho nos dá o seu tamanho, dessa forma, um caminho será maior que outro se possui mais vértices. E será dito menor se possuir menos vértices. Representaremos o tamanho do caminho geodésico por  $d_{ij}$ . Note que essas definições foram feitas para grafos, e não digrafos pois para esses teríamos que pensar em que sentido apontam as arestas. Dessa forma podemos definir para um grafo o caminho geodésico médio:

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \quad (2.1)$$

por definição  $d_{ii} = 0$ . Existem vários algoritmos para medir o caminho em um grafo. Mas a definição 2.1 só tem validade para um grafo conexo. Quando tratamos de grafos não conexos, ou se redefine  $l$  [15], ou se pudermos separá-lo em um conjunto pequeno de subgrafos conexos, podemos calcular  $l$  de acordo com equação 2.1 para cada subgrafo. Uma outra grandeza importante a ser definida é o coeficiente de “clustering”. Dado um nó com dois vizinhos o coeficiente de clustering nos diz qual a probabilidade de que esses estejam conectados. Na literatura encontramos duas formas distintas de calcular essa grandeza. Uma foi importada da literatura de redes sociais, conhecida como “fração de triplas transitivas”, definida por:

$$C = \frac{3 \times \text{número total de triângulos}}{\text{número total de vértices triplamente conectados}} \quad (2.2)$$

aonde triplamente conectado é um vértice com duas arestas terminando em vértices distintos. Essa definição que facilita cálculos teóricos possui sua concorrente. Watts e Strogatz [2] definiram de uma forma distinta o coeficiente de clustering. Propuseram um valor local dado por

$$C_i = \frac{\text{número de triângulos conectados ao vértice } i}{\text{números de triplas centradas em } i} \quad (2.3)$$

Para vértices de grau 0 ou 1, cujo denominador será 0, atribuímos o  $C_i = 0$ . O coeficiente de clustering da rede toda fica

$$C = \frac{1}{n} \sum_i C_i \quad (2.4)$$

Em 2.2 calculamos o número total de triângulos e dividimos pelo total de conexões, ou seja fazemos a fração média. Na definição 2.3 o que fazemos é calcular em cada nó o número de triângulos fechados por conexões triplas, e tiramos a média desse cálculo, ou seja fazemos a média das frações. Para entender melhor a diferença olhemos a Figura 2.1.

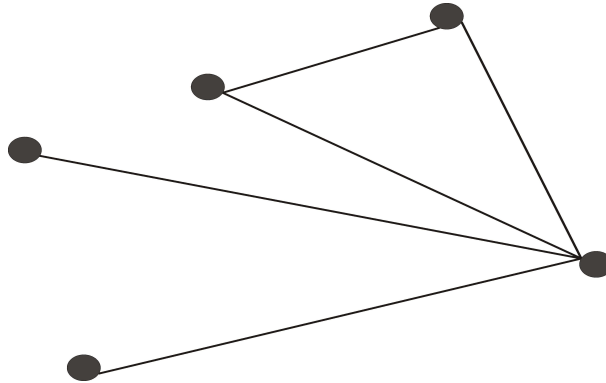


Figura 2.3: A rede possui 8 maneiras de ligar um nó a dois vizinhos distintos, ou seja, 8 conexões triplas. Apenas 1 triângulo. Então pela equação 2.2,  $C = 3/8$ . Olhando agora a definição local, temos os coeficientes de clustering  $C_i$ , eq. (2.3), da direita para esquerda, valendo:  $1/6, 1, 1, 0$  e  $0$ . Dessa forma, utilizando a equação 2.4:  $C = 13/30$ .

Essa segunda forma de calcular o coeficiente de clustering é conhecida na literatura das ciências sociais como “densidade da rede”. Independente do coeficiente de clustering utilizado ele se mostra sempre maior em redes complexas do que em grafos meramente aleatórios. Existem indícios que o valor de  $C$  tende para um valor finito quando  $n \rightarrow \infty$ .

Vemos que, em ambas as definições, o coeficiente de clustering mede a quantidade de caminhos fechados triangulares. Existem generalizações para essa grandeza, cir-

cuitos de 4, 5 ou mais vértices, porém abordar essas grandezas foge do escopo desse texto.

A última importante grandeza a ser definida é a distribuição de graus. Anteriormente definimos que o grau de um dado vértice  $i$  é o número de arestas conectadas a ele. A distribuição de graus  $P(k)$  é a probabilidade de um dado nó escolhido ao acaso, com igual probabilidade, tenha grau  $k$ .

## 2.2 Grafo de Erdős-Rényi

O grafo de Erdős-Rényi [19] é importante para nosso trabalho uma vez que ele remonta às primeiras tentativas de diluição em uma rede neural. Solomonoff e Rapoport [20] e, independentemente, Erdős-Rényi propuseram um modelo para uma rede aleatória que segue uma idéia bem simples. Suponha um conjunto de  $n$  vértices e cada vértice está conectado a um outro com uma dada probabilidade  $p$ . Isso define o que Erdős e Rényi chamaram de  $G_{n,p}$ . Na linguagem física,  $G_{n,p}$  é um *ensemble* em que um grafo com  $m$  arestas possui probabilidade  $p^m(1-p)^{M-m}$  onde  $M$  é o número máximo de arestas possível dado por  $M = \frac{1}{2}n(n-1)$ . Erdős e Rényi definiram outro modelo  $G_{n,m}$  que é o ensemble de todos os grafos com  $n$  vértices e exatamente  $m$  arestas, e cada grafo aparece com mesma probabilidade. Notemos que esses ensembles são análogos ao ensemble canônico e gran-canônico da mecânica estatística. Em analogia é comum nos referirmos a essas redes, como redes em equilíbrio, a distribuição de graus é independente do tempo. Dorogovtsev e Mendes [17] usam essa abordagem de ensembles para gerar redes complexas com distribuições de graus distintas, inclusive de “cauda grossa” (fat-tail).

Pela construção do grafo, um dado vértice  $i$  possui  $k$  arestas com probabilidade  $p^k$ , e não está conectado com  $N-1-k$  vértices com probabilidade  $(1-p)^{N-1-k}$ . Mas existem  $C_{N-1}^k$  maneiras de combinar essas conexões. Dessa forma a probabilidade

de que o  $i$ -ésimo vértice possua  $k$  arestas é:

$$P(k_i = k) = C_{N-1}^k p^k (1-p)^{N-1-k} \quad (2.5)$$

No limite  $N \rightarrow \infty$ ,

$$P(k_i = k) = C_{N-1}^k p^k (1-p)^{N-1-k} \simeq e^{-pN} \frac{(pN)^k}{k!} \quad (2.6)$$

Mas  $\langle k \rangle = pN$ ,

$$P(k_i = k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad (2.7)$$

que é a distribuição de Poisson. Por esse motivo é comum ver esse tipo de rede aleatória ser chamada por grafo aleatório de Poisson.

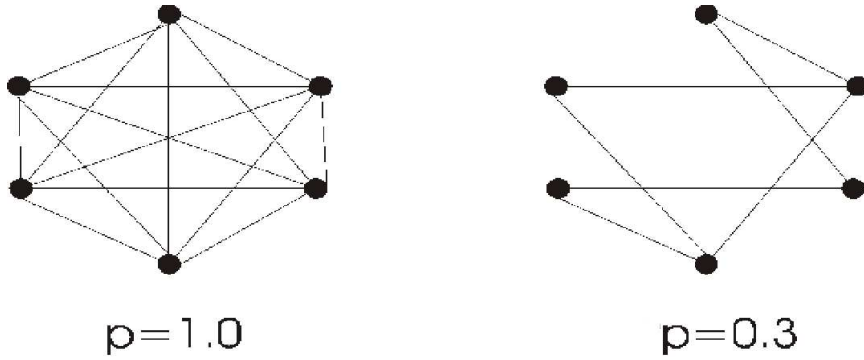


Figura 2.4: Grafo aleatório para dois valores de  $p$

O coeficiente de clustering também é facilmente obtido. Pela definição em 2.3 a probabilidade de haver um triângulo conectado ao vértice  $i$  é  $p^3$ , pois a probabilidade de que cada nó esteja conectado a um outro nó qualquer é  $p$ . Da mesma forma a probabilidade de existir uma ligação tripla centrada em  $i$  é  $p^2$ . Então o coeficiente de clustering para a rede aleatória com distribuição de Poisson é:

$$C = \frac{1}{N} \sum_i C_i = p = \frac{\langle k \rangle}{N} \quad (2.8)$$

O caminho geodésico médio do grafo de Erdős-Rényi é um pouco mais difícil de ser obtido, pois, dependendo da probabilidade  $p$  escolhida, o grafo pode ou não ser

conexo. O estudo do diâmetro desses grafos tem sido estudado por vários autores [21]. O que podemos facilmente é garantir que para um grafo conexo,  $l$  escala com  $\ln(N)$ . Para grafos aleatórios conexos, apresentamos então sem demonstrar,

$$l \sim \frac{\ln(N)}{\ln(\langle k \rangle)} \quad (2.9)$$

## 2.3 Redes na física

Modelos em grafos têm tradição na física, sob a alcunha de redes regulares (ou no inglês, *lattices*). As redes na física apareceram na mecânica estatística, como o modelo de Ising. O entendimento do magnetismo tem uma importância fundamental na história da física, pois é um dos fenômenos quânticos de resultados macroscópicos e o primeiro grande sucesso na descrição de um fenômeno físico por um modelo de rede.

O problema era descrever os estados magnéticos da matéria, o ferromagnetismo e o paramagnetismo. Uma substância ferromagnética apresenta magnetização espontânea, na ausência de campo magnético externo. O ferromagnetismo ocorre em materiais cristalinos como ferro, o níquel e o cobalto. A matéria quando paramagnética não possui magnetização espontânea, mas podem apresentar magnetização quando induzido por um campo magnético externo.

Pela teoria eletromagnética clássica a interação que surge no interior desses materiais seria causada pela força entre dipolos magnéticos, que é proporcional a  $1/r^3$ . O problema é que uma interação desse tipo não poderia ser responsável pelo fenômeno do ferromagnetismo, pois decai muito rapidamente com a distância. Mas o efeito mais intrigante é que os materiais ferromagnéticos quando expostos a uma alta temperatura perdiam as suas propriedades ferromagnéticas, e se mostrando como materiais paramagnéticos. Ou seja, sofriam uma *transição de fase*.

A descrição completa desse fenômeno é dada pela mecânica quântica. Imagine dois elétrons de átomos vizinhos com orbitais  $\phi_1(r_1)$  e  $\phi_2(r_2)$  que se superpõem



localmente. Se eles possuem o mesmo spin, a função de onda será anti-simétrica

$$\psi(r_1, r_2) = \phi_1(r_1)\phi_2(r_2) - \phi_2(r_1)\phi_1(r_2) \quad (2.10)$$

Se  $r_1 = r_2 = r$ , então  $\psi(r, r) = 0$ . Ou seja, a probabilidade de encontrá-los próximos é nula. Essa proibição de dois elétrons interagentes possuírem o mesmo estado quântico é conhecida como Princípio de exclusão de Pauli. Devido à repulsão elétrica entre eles, o distanciamento reduz a energia potencial, tornando essa configuração mais estável. O modelo de Ising [22] propõe uma interação local entre os spins através de uma hamiltoniana baseada na interação anteriormente descrita

$$H = -B \sum_i \sigma_i - J \sum_{\langle ij \rangle} \sigma_i \sigma_j \quad (2.11)$$

onde a variável  $\sigma_i$  representa o estado do  $i$ -ésimo spin, assumindo valores  $\pm 1$ . O primeiro termo representa a interação externa com o campo magnético externo, e a segundo termo a interação entre os spins vizinhos. Quando a constante  $J$  é positiva a energia diminui com os spins se alinhando, ou seja, ferromagnetismo. Quando  $J < 0$  a energia é mínima quando os spins estão antiparalelos, nesse caso o modelo descreve o antiferromagnetismo.

O somatório do segundo termo é feito na vizinhança, que é definida em função da rede. Na física é costume referir a essas redes pela dimensão  $d$ , de  $Z$  (conjunto dos inteiros) que contém seus pontos (Figura (2.5)). Dessa forma, uma rede quadrada plana seria uma rede em  $2d$  e cada sítio pode ser descrito por um ponto em  $Z^2$ . Se a célula original fosse um cubo, teríamos uma rede cúbica em  $3d$ , com cada ponto representado por um trio de coordenadas. Para  $d > 3$ , ela é dita um hipercubo. Dessa forma um vizinho é um sítio da rede que possua mesma coordenada do sítio em questão, com exceção de uma, que pode variar em uma unidade. Assim, na rede quadrada planar, os vizinhos do sítio  $(x, y)$  são os pontos da forma  $(x \pm 1, y)$  e  $(x, y \pm 1)$ , totalizando quatro vizinhos. Para rede quadrada, é comum nos referirmos ao tamanho da aresta  $L$ , em vez do número total de nós.

Em outras áreas da física modelos de rede também são comuns, a exemplo no estado sólido, estruturas cristalinas se organizam nas chamadas redes de Bravais.

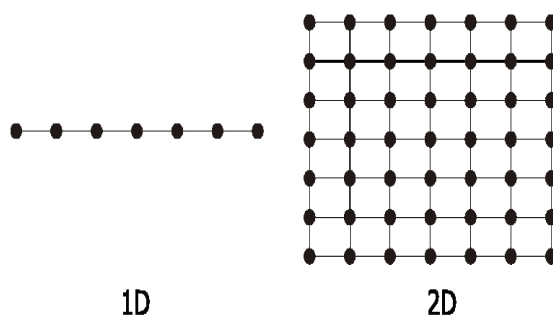


Figura 2.5: Redes  $1d$  e  $2d$  com  $L=7$ .

Essas estruturas são caracterizadas pela sua periodicidade, como se fossem feitas a partir da reprodução de uma célula que apresenta formas geométricas bem definidas, como por exemplo quadrados, triângulos, cubos ou hexágonos.

## 2.4 Redes no Mundo Real

O primeiro passo é identificar as grandes áreas existentes e localizarmos o problema que queremos tratar, para em seguida definir as entidades que nos trarão algum tipo de informação a respeito do sistema estudado. Newman [15] divide as redes complexas em 4 grandes grupos. Seguiremos a divisão proposta por ele em nosso texto.

Redes sociais, envolvem pessoas e as interações entre elas. Pode ser relação matrimonial entre famílias, interações entre empresas no mundo corporativo ou amizade entre pessoas. Na análise desse tipo de rede nasceu o famoso experimento de "small-world" feito por Milgram [23, 24]. O experimento consistiu em enviar uma carta da Costa Oeste do Estados Unidos e essa deveria chegar a um indivíduo alvo da Costa Leste, com a restrição de que a carta só poderia viajar passando de mão em mão. Apenas um quarto das cartas chegaram ao seu destino, mas o impressionante resultado dessa experiência foi que cada carta passou, em média, por apenas 6 pessoas. Esse experimento, apesar de simples e surpreendente, é um dos poucos relatos de estudo de redes sociais em larga escala, normalmente essas são estudadas em amos-

tras pequenas, pois os dados são recolhidos através de questionários de uma forma bastante trabalhosa.

O segundo tipo de rede complexa são as redes de informação. Um exemplo trivial desse tipo de rede são as citações acadêmicas, aonde cada artigo é um vértice e as arestas são diretas, no sentido do artigo citado. Como um artigo feito nos anos 50 não pode citar um artigo dos anos 80, essa rede apresenta uma estrutura acíclica. Outra importante rede de informação é a World Wide Web, aonde as páginas são os vértices e as arestas são os hyperlinks. Ao contrário da rede de citações, a WWW é cíclica, afinal não há ordem natural das páginas, e nenhum vínculo impede que existam caminhos fechados.

O terceiro tipo de redes são as tecnológicas, redes feitas com o intuito de distribuir algum tipo de serviço. A malha rodoviária e ferroviária são exemplos desse tipo de rede. A rede de transporte aéreo, assim como a de distribuição de energia elétrica, aonde a malha é representada pela as arestas e os vértices são as cidades [25]. Outra rede dessa classe bastante estudada é a Internet, a rede de conexões físicas entre computadores.

A última classe é a de redes biológicas. Os caminhos das reações metabólicas, aonde os substratos e produtos são ligados por arestas direcionadas. Outra rede biológica bastante estudada são as cadeias alimentares, aonde um vértice representa uma espécie em um ecossistema e a aresta direcionada representa uma relação predatória da espécie A para com a B. O foco de nosso trabalho, as redes neurais, são redes biológicas e se enquadram nessa classe.

## 2.5 Novos Modelos

Em 1998, Watts e Strogatz propuseram um modelo que apresenta apenas um parâmetro de controle,  $p$ , que é a probabilidade de que uma conexão seja rearranjada. O algoritmo por trás do modelo pode ser descrito em dois passos:

- Começa-se com uma rede unidimensional fechada, como um anel. Cada nó está conectado com  $K/2$  vértices à esquerda e a direita.
- Cada aresta à direita de todo nó na rede terá sua outra extremidade religada com probabilidade  $p$ . Sendo que conexões duplicadas e auto-ligações não são permitidas. Esse processo cria  $pNK/2$  ligações de longo alcance na rede.

Então, varrendo  $p$  no intervalo  $[0,1]$  podemos contemplar a transição entre ordem ( $p=0$ ) e aleatoriedade ( $p=1$ ). Esse modelo advém de uma abordagem de laços sociais, as suas amizades são de pessoas próximas a você, que moram na mesma rua, estudam no mesmo colégio, entretanto alguns possuem amigos distantes, pessoas que se mudaram ou conhecidos de viagem. A localidade é representada no modelo pela rede regular, e as conexões de longa distância seriam esses contatos mais raros.

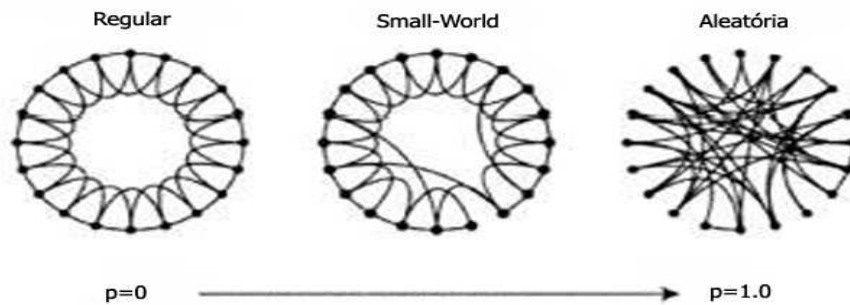


Figura 2.6: rede de small-world

Passemos então para as propriedades da rede de small-world. A solução exata para o caminho geodésico médio ainda é um problema em aberto. Barthélemy e Amaral [26], proporam uma conjectura onde  $l$  possua uma lei de escala. Recentemente, vários autores [27, 28, 29] por meio de argumentos analíticos e simulações numéricas chegaram a uma nova lei de escala que hoje é amplamente aceita,

$$l \sim \frac{N^{1/d}}{K} f(pKN) \quad (2.12)$$

onde  $f(x)$  que obedece aos seguintes valores assintóticos

$$f(x) = \begin{cases} \text{const se } x \ll 1 \\ \ln(u)/u \text{ se } x \gg 1 \end{cases} \quad (2.13)$$

Existe na literatura um esforço para o cálculo de  $l$  de forma exata, mas isso ainda não foi totalmente realizado. Albert e Barabási [14] em seu artigo de revisão mostram, que dada certas circunstâncias, o cálculo de  $l$  pode ser feito exatamente.

Além do pequeno caminho geodésico médio, da ordem de  $\log(N)$ , a rede de small world apresenta alto coeficiente de clustering. Para esse cálculo partiremos da rede regular, com coeficiente de clustering  $C_i(0)$ . Para  $p > 0$  o cluster em torno de  $i$  só continuará inteiro se as três conexões que formam o triângulo não forem religadas, dessa forma,

$$C_i(p) \sim C_i(0)(1 - p)^3. \quad (2.14)$$

Note que essa argumentação é heurística pois certos elementos não são levados em conta como, por exemplo, dependendo de onde a nova aresta for religada, podemos ter a formação de um novo cluster. Mas Barrat e Weigt [30] mostraram que a diferença entre essa expressão e o valor correto do coeficiente de clustering vai a zero para uma rede suficientemente grande.

A última grandeza que falta ser investigada para esse modelo é a distribuição de graus. Para  $p = 0$  a distribuição é uma função delta centrada em  $K$ . Para valores de  $p$  diferentes de zero, a desordem começa a surgir na rede. Pelo processo de re-ligação das arestas, cada nó tem ao menos  $K/2$  arestas no final do processo. Para  $K > 2$  não existirá vértice isolado e a rede é, normalmente, conexa. Evitando alguns cálculos, a distribuição de graus é dada por

$$P(k) = \sum_{n=0}^{f(k,K)} C_{K/2}^n (1-p)^n p^{K/2-n} \frac{(pK/2)^{k-K/2-n}}{(k-K/2-n)!} e^{-pK/2} \quad (2.15)$$

para  $k \geq K/2$ , onde  $f(k, K) = \min(k - K/2, K/2)$ . A forma dessa distribuição é similar a de Poisson, com um pico mais pronunciado em torno do valor médio. Na figura 2.7 plotamos o histograma de conexões para redes de small-world com

4 distintos parâmetros de desordem. Para comparar plotamos a distribuição de conexões de uma rede aleatória. A rede aleatória foi montada da seguinte maneira: sorteia-se um par de nós aleatoriamente, e estes são ligados entre si. Repetimos esse processo até que tenhamos  $KN/2$  arestas na rede, dessa forma cada nó está conectado a um outro nó da rede com probabilidade  $P_{ij} = K/N$ .

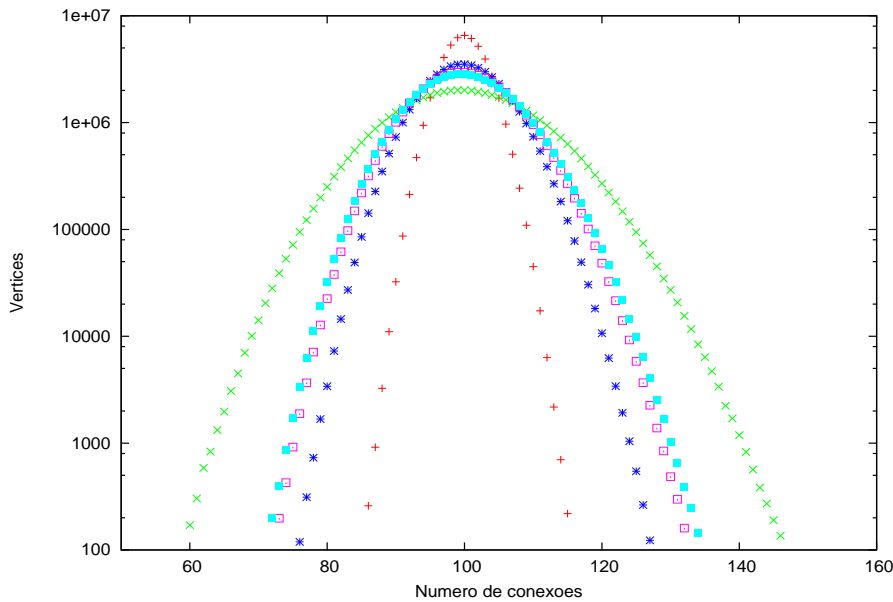


Figura 2.7: Histograma do número de vértices pela distribuição de conexões para redes de small world, com  $p = 0.2(+)$ ,  $p = 0.4(*)$ ,  $p = 0.7(\square)$ ,  $p = 1.0(\blacksquare)$  e *Erdős-Rényi* ( $\times$ ),  $K = 100$  em 10000 realizações.

Para terminar essa seção iremos discutir agora a última topologia estudada em nosso trabalho, as redes livres de escala. No final da década passada e início da década corrente, vários trabalhos na literatura mostraram que redes reais, como a Internet, possuem uma distribuição de graus em forma de uma lei de potência. Alguns outros trabalhos mostraram redes reais com distribuições diferentes, mas também de cauda longa, distinta do grafo de Erdős-Rényi. Então, uma pergunta importante que não possuía resposta é: “qual o mecanismo que faz com que uma rede sem escala ocorra?”. O modelo de Barabási-Albert traz uma resposta para pergunta.

Antes de continuarmos uma nota histórica deve ser feita. O modelo de Barabási-Albert não foi o primeiro modelo a conseguir uma distribuição de graus em forma de uma lei de potência. Derek de Solla Price [31], em 1965, usou o termo *cumulative advantage* para designar o que hoje é amplamente aceito como explicação do surgimento de uma distribuição de lei de potência para os graus [32]. Atualmente utilizamos o termo *prefferential attachment*, criado por Barabási e Albert [33]. As idéias de Price são fruto do trabalho de Herbert Simon [34] que mostrou que as leis de potência surgem quando “os mais ricos ficam mais ricos”.

O modelo de Barabási-Albert pode ser resumido no seguinte algoritmo para determinar o crescimento da rede:

- A rede começa com  $m$  nós, todos conectados entre si.
- A cada passo de tempo  $t$  adiciona-se um novo nó  $j$  a rede, com  $n$  arestas tal que  $n \leq m$ . Essas arestas são ligadas ao sitio  $i < m + t$ , com probabilidade  $p(k_i)$ , dada por:

$$p_{j \rightarrow i}(k_i) = \frac{k_i}{\sum_j k_j} \quad (2.16)$$

Onde  $k_i$  é o grau do  $i$ -ésimo nó. Depois de  $t$  passos a rede possui  $N = t + m$  e  $e = (m - 1) + nt$  arestas. Quando  $t \rightarrow \infty$  a distribuição de conexões é dada por  $P(k_i) \propto k_i^{-3}$ . Na figura na Figura (2.8).

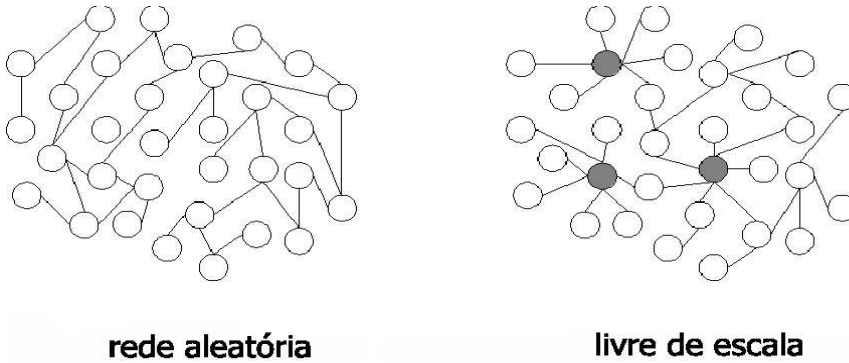


Figura 2.8: Rede aleatória e a rede livre de escala. Os nós marcados em cinza atentam para a condensação de arestas.

Vamos mostrar, por duas abordagens distintas, que o algoritmo de Barabási-Albert realmente nos leva a uma lei de potência. A primeira abordagem é analítica, tratando o grau dos vértices como uma variável contínua, e em seguida vamos mostrar uma abordagem com resultados numéricos de nossas simulações.

A abordagem analítica [14, 17] tem como aproximação tratar o grau como uma variável contínua. Para redes no limite termodinâmico essa abordagem é razoável e produz bons resultados. Se  $k_i$  é uma variável contínua, vale a relação dinâmica

$$\frac{\partial k_i}{\partial t} = np_{j \rightarrow i}(k_i) = n \frac{k_i}{\sum_j k_j} \quad (2.17)$$

Mas

$$\sum_j k_j = 2nt - n \quad (2.18)$$

então

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad (2.19)$$

Levando em consideração que cada nó  $i$  está submetido às condições iniciais  $k_i(t_i) = n$  onde  $t_i$  é o instante que o nó foi adicionado à rede. Dessa forma a solução de (2.17) é

$$k_i(t) = n \left( \frac{t}{t_i} \right)^\beta \quad \text{com } \beta = 1/2 \quad (2.20)$$

Assim, a distribuição de grau evolui para cada nó como uma lei de potência. Então podemos calcular a probabilidade acumulada,  $P(k_i < k)$  que é a probabilidade de um nó possuir grau menor ou igual a  $k$ , para então calcularmos  $P(k)$ .

$$P(k_i < k) = P\left(t_i > n^{1/\beta} \frac{t}{k^{1/\beta}}\right) \quad (2.21)$$

Supondo que adicionamos os nós em intervalos iguais de tempo, a probabilidade de termos  $t = t_i$  é uma constante tal que

$$P(t_i) = \frac{1}{m + t} \quad (2.22)$$



então utilizando 2.22 em 2.21 temos,

$$P\left(t_i > n^{1/\beta} \frac{t}{k^{1/\beta}}\right) = 1 - \frac{n^{1/\beta} t}{k^{1/\beta}(t+m)} \quad (2.23)$$

Dessa forma,  $P(k)$

$$P(k) = \frac{\partial P(k_i < k)}{\partial k} = \frac{2n^{1/\beta} t}{n+t} \frac{1}{k^{1/\beta+1}} \quad (2.24)$$

no limite assintótico ( $t \rightarrow \infty$ )

$$P(k) \sim 2n^{1/\beta} k^{-\gamma} \text{ com } \gamma = 1/\beta + 1 = 3 \quad (2.25)$$

O resultado interessante desse modelo, é que o expoente crítico não depende de  $n$ , o que está de acordo com os resultados numéricos.

Note que alguns efeitos de tamanho finito aparecem em redes livre de escala, como observamos na Figura (2.9). A lei de potência é perdida na cauda da distribuição, e um pequeno “morro” se forma. Esse problema ocorre devido a uma memória da rede em relação às condições iniciais [35]. Conforme aumenta o tamanho da rede o efeito diminui, como podemos notar na Figura (2.10). Stauffer et al [36], mostraram que para uma rede de  $N = 10000$  e  $n = m = 2$  o efeito é imperceptível e o morro desaparece.

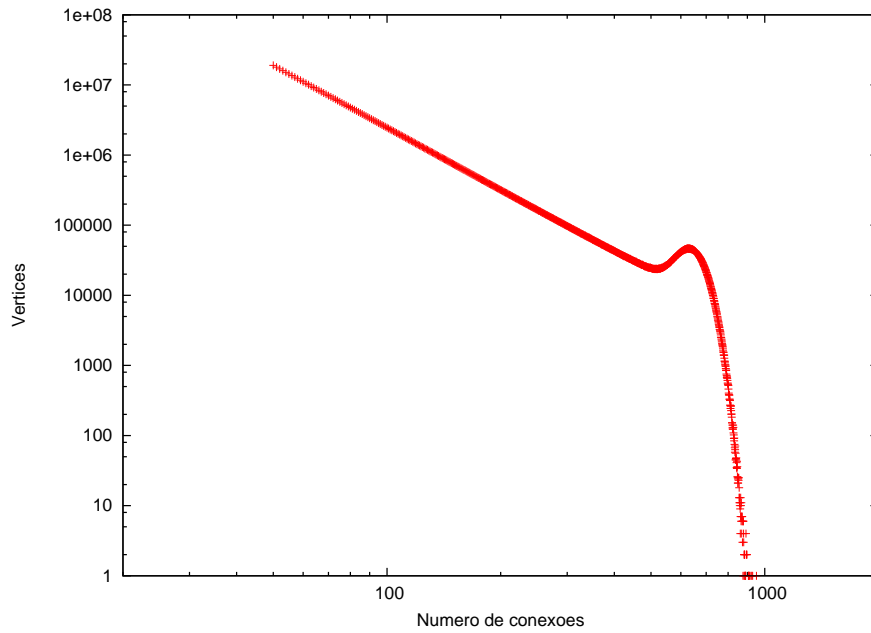


Figura 2.9: Histograma do número de vértices pela distribuição de conexões para  $N = 5000$ ,  $m = n = 50$  em 100000 realizações.

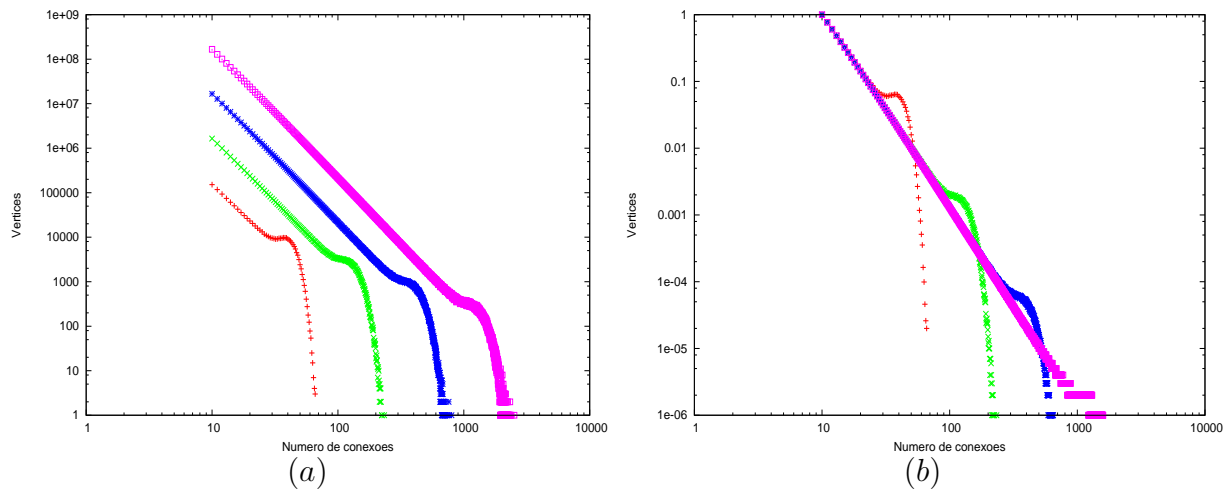


Figura 2.10: (a) Histograma do número de vértices pela distribuição de conexões. (b)  $P(k)$ . Para ambos os gráficos  $N = 100(+)$ ,  $1000(\times)$ ,  $10000(*)$ ,  $100000(\square)$ ,  $m = n = 10$  em 10000 realizações.

Para ilustrar que a cauda gaussiana é efeito de memória do sistema, e conseqüentemente do seu tamanho finito, realizamos simulações para uma rede com a mesma conectividade, mas para diferentes tamanhos. O resultado está disposto na figura 2.10.

O caminho geodésico médio da rede de Barabási-Albert ainda é um problema em aberto. Resultados de simulações numéricas apontam para uma lei de escala com uma função logarítmica [14]. Recentemente Bollobás e Riordan mostraram [37], analiticamente, que uma correção é necessária, e a lei de escala seria  $l \sim \ln(N)/\ln(\ln(N))$ .

Também não existe predição analítica para o coeficiente de clustering da rede livre de escala. O que temos na literatura são simulações numéricas que mostram que o coeficiente de clustering da rede de Barabási-Albert é cerca de 5 vezes maior que a de um grafo aleatório de mesmo tamanho [14].

## 2.6 Sistema Nervoso Real

Como mencionamos anteriormente o número de neurônios no cérebro é da ordem de  $10^{11}$ , o que mesmo para a teoria de redes complexas é um número grande. Se olharmos o cérebro do polvo, o número de neurônios cai para  $10^8$ . Se formos olhar um ser mais primitivo, como o pequeno verme da espécie *Caenorhabditis elegans* (*C. Elegans*) temos apenas 302 neurônios. Por ser um organismo simples, a rede neural da *C. Elegans* é muito bem estudada. Mas por que estudar uma criatura tão primitiva? Apesar de primitiva ela apresenta etapas biológicas com características similares a alguns problemas centrais na biologia humana. A *C. Elegans* se reproduz via uma única célula que passa por estágios complexos, possui divisão embrionária em seguida passa para uma fase de diferenciação celular. Ela apresenta um “cérebro” na forma de um anel em torno da faringe que exibe um comportamento de aprendizado rudimentar. Ela também apresenta envelhecimento. Logo,

essa simples criatura reúne vários problemas da biologia moderna: fase embrionária, diferenciação celular, desenvolvimento, sistema nervoso central com aprendizado e envelhecimento.

Watts e Strogatz [2] construíram um grafo simples da rede neural desse ser, ignorando alguns detalhes da rede como o tipos de sinapses e eliminando alguns neurônios não pertencentes ao nervo central. Eles encontraram uma rede com coeficiente de clustering 5 vezes maior do que um grafo aleatório equivalente, e um caminho geodésico médio de mesma ordem. Cinco vezes maior não é uma diferença de uma ordem de grandeza, mas devemos lembrar que a própria rede é muito pequena. O importante desse resultado é que ele dá indícios de que para redes maiores, a clusterização seja algo importante para o desempenho das funções do sistema nervoso central. O sistema nervoso da *C.Elegans* foi modelado por outros autores, como Satoru Morita et al [38], que propuseram um outra modelagem baseada num modelo geométrico multidimensional.

## 2.7 Conclusão

No capítulo anterior discutimos o modelo de Hopfield, e vimos que uma de suas fortes simplificações é a total conectividade da rede. Com o intuito de prover uma topologia mais próxima do real para o modelo, dedicamos o atual capítulo a uma discussão de grafos e redes complexas.

Definimos os conceitos básicos da teoria de grafos, avançamos para os exemplos clássicos em física e matemática. Mostramos em contra partida que as redes no mundo real são bem diferentes desses exemplos, e novos modelos são necessário para essas.

Apresentamos e discutimos as propriedades do modelo de Barabási-Albert para redes livre de escala, e o modelo de Small-World de Watts-Strogatz.

Finalizamos com um exemplo do sistema nervoso da *C. Elegans*, aonde podemos

conjecturar que essas novas topologias apontam a direção para a verdadeira topologia de um sistema nervoso central.

# Capítulo 3

## Resultados Numéricos

Neste capítulo iremos apresentar os principais resultados desses dois anos de trabalho. Nosso objetivo consiste em estudar o comportamento do modelo de Hopfield diluído em topologias complexas. Para isso efetuamos uma busca rigorosa, varrendo uma grande extensão dos parâmetros de controle.

### 3.1 Descrição do procedimento

Como dito anteriormente, uma propriedade de grande interesse no estudo de redes neurais é a sua capacidade de armazenamento. Aqui estaremos interessados em duas grandezas: o número total de padrões armazenados  $P$ , e o número total de padrões armazenados por conectividade média  $K$ , que definimos por carga,

$$\alpha = \frac{P}{K} \tag{3.1}$$

Para calcular a superposição  $m_\mu$  em função da carga utilizamos o seguinte algoritmo:

- 1- Geramos a rede que servirá como topologia para modelo de Hopfield, definimos os elementos da matriz simétrica das conexões,  $c_{ij} = c_{ji}$ . Onde  $c_{ij} = 1$  se  $i$  e  $j$  estão conectados, e  $c_{ij} = 0$  se não;
- 2- Gera-se aleatoriamente os padrões que serão aprendidos pela rede,  $\xi^\mu$  com  $\mu =$

1, ..., P, descorrelacionados na média;

3- Montamos a matriz das sinapses, etapa de aprendizado da rede, dada por,

$$J_{ij} = \frac{c_{ij}}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \quad (3.2)$$

4- É dado como estímulo à rede um de seus padrões;

5- Fazemos uma atualização seqüencial segundo as equações (1.16) e (1.17) numa ordem pré-escolhida aleatoriamente. Esse passo é repetido até que a rede atinja um atrator, isto é, nenhum de seus neurônios mude de estado;

6- Calculamos a superposição com o padrão inicial, segundo a equação (1.45);

7- Repetimos 1-6 até termos uma boa estatística da superposição;

8- Aumentamos o número de padrões em uma unidade.

9- Repetimos 1-7 até atingirmos um dado valor de carga suficientemente alto para observarmos os efeitos de transição.

Com o intuito de economizar memória e otimizar o algoritmo, não armazenamos a matriz sináptica em um *array* de  $N \times N$  elementos, como em redes totalmente conectadas. Ao invés disto utilizamos uma estrutura de dados composta por três vetores: *PesoSin*, *sítio-i*, *sítio-j*. As variáveis *sítio-i* e *sítio-j* são responsáveis por guardar os valores dos índices  $i$  e  $j$  da matriz das conexões, quando essa for igual a unidade. A variável *PesoSin* armazenará o valor da sinapse entre os neurônios *sítio-i* e *sítio-j*. Para entender melhor essa estrutura, imaginemos o seguinte exemplo: O sítio  $i = 3$  está conectado aos sítios  $i = 2, 5, 17$  com os respectivos pesos sinápticos  $J_{32} = 1, J_{35} = -2, J_{37} = 2$ . Dessa forma teríamos:

Tabela 3.1: Estrutura de dados para matriz esparsa

	Ligações					
PesoSin (valor do elemento da matriz)	1	-2	2	1	-2	2
sítio-i (coordenada $i$ do elemento)	3	3	3	2	5	17
sítio-j (coordenada $j$ do elemento)	2	5	17	3	3	3

Além da economia em memória, esse tipo de armazenamento garante mais uma vantagem. Se os vetores *sítio-i* e *sítio-j* estiverem em ordem, durante a dinâmica não precisaremos realizar  $N^2$  operações, e apenas  $KN$ . Para ordenar o vetor utilizamos o algoritmo quicksort, [39] que possui complexidade  $N\log_2(N)$ . Dessa forma reduzimos a complexidade do algoritmo de  $N^2$  para  $N\log_2(N)$ . Note que esse procedimento é necessário durante a fase de aprendizado dos padrões.

Para entender como o modelo de Hopfield se porta perante essas novas topologias utilizaremos a superposição no equilíbrio  $m_\mu$ , definida em (1.45). Será necessário também definir uma nova grandeza, a eficiência  $\varphi_\mu$ . A eficiência é a fração de vezes que o padrão  $\mu$  foi recuperado de forma correta. Para padrões sem correlação, um padrão é recuperado corretamente se [13]

$$\varphi = |m_\mu| + 0.03 \geq 1 \quad (3.3)$$

## 3.2 Padrões sem correlação

Primeiramente, fizemos uma comparação entre as redes aleatórias, Erdős-Rényi com o a rede small-world (*sw*) com  $p = 1$ , Figura (3.1). Sabemos que ambas as redes possuem as conexões distribuídas aleatoriamente, mas com distribuição de graus um pouco distinta e a rede *sw* é conexa, enquanto o grafo de Erdős-Rényi pode não ser.

Na Figura (2.7) observamos uma diferença apreciável na distribuição de conectividade das redes, mas na Figura (3.1) vemos que essa diferença não é relevante para o modelo, e sim na aleatoriedade das conexões se mostra o diferencial. A aleatoriedade nas conexões está intimamente ligada com o coeficiente de clustering e a



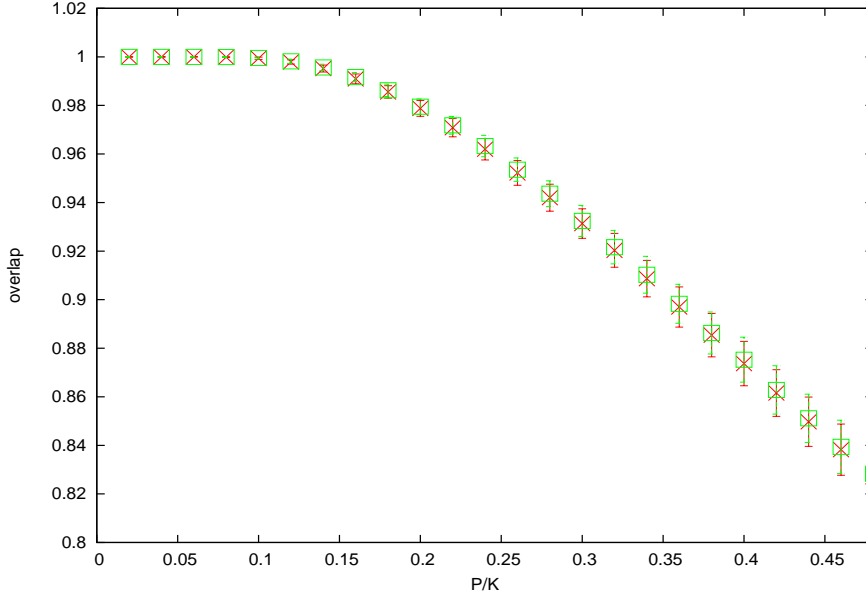


Figura 3.1: Superposição vs  $\alpha$ . Para  $p = 1.0$  ( $\times$ ), e grafo aleatório ( $\square$ )  $N = 5000$  e  $K = 50$  (padrões sem correlação).

caminho geodésico médio.

No gráfico (3.2) podemos ver o comportamento da rede de small-world para diferentes valores de desorganização. Esse resultado já foi obtido por McGraw et al [7]. Percebemos que para  $p > 0.4$  a diferença de desempenho é muito pequena. Dessa forma trataremos as redes de small-world em 4 estágios: Rede com baixa organização  $p = 0.2$ , a rede em torno da transição  $p = 0.4$ , rede aleatória  $p = 1.0$  e rede regular  $p = 0$ . A rede livre de escala foi simulada segundo a algoritmo descrito na seção 2.5. Nossa primeira preocupação foi verificar o efeito de tamanho finito da rede, ou seja analisar o limite  $N \rightarrow \infty$ . O resultado é mostrado na Figura (3.3). Vemos que para  $N \geq 5000$  o comportamento se torna indiferente do tamanho da rede. Como o comportamento de uma rede neural é *self-averaging*, o aumento do tamanho diminui a intensidade da flutuação. Dessa forma consideramos que uma rede de 5000 neurônios é um tamanho bom para prosseguirmos nas investigações. Esse comportamento apresentado para *sw* com  $p = 0.4$  e  $K = 50$  se repete para todas as outras redes em distintos valores de conectividade. Com o intuito de não carregar

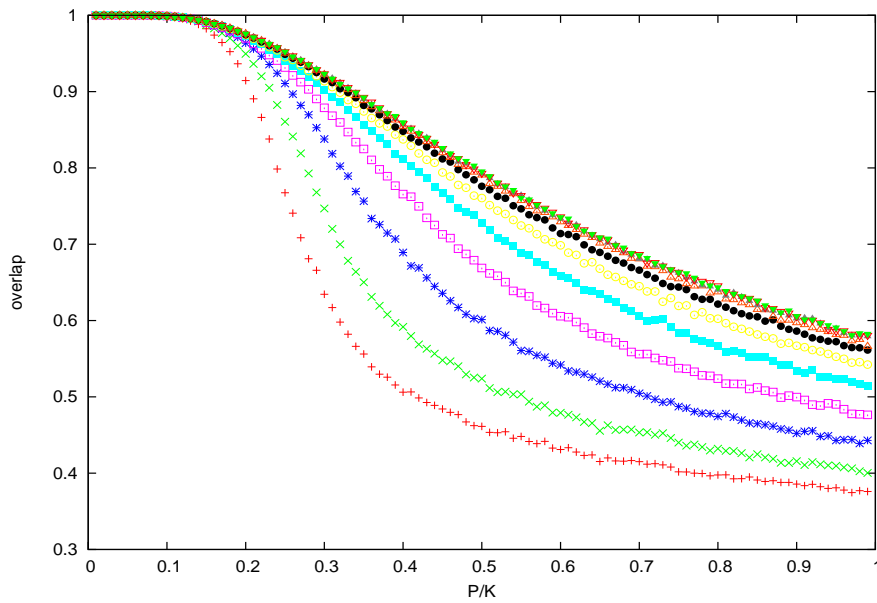


Figura 3.2: Superposição vs  $\alpha$ , para rede de small-world, de  $p = 0.0$  até  $p = 1.0$  acrescido de 0.1, da esquerda para direita.  $N = 5000$ ,  $K = 100$ , média em 100 realizações. Padrões sem correlação

o texto com muitos gráficos de pouca informação, omitiremos esses resultados no presente texto.

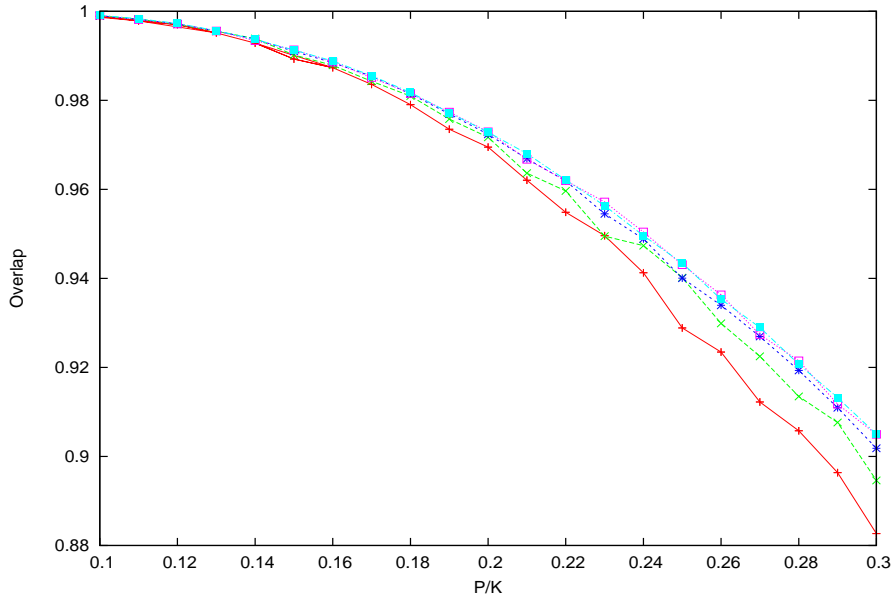


Figura 3.3: Superposição vs  $\alpha$  para tamanho finito,  $p = 0.4$ .  $N = 1000$  (+),  $N = 2000$  ( $\times$ ),  $N = 5000$  (\*),  $N = 10000$  ( $\square$ ) e  $N = 20000$  ( $\blacksquare$ )

Então refizemos os resultados apresentados em [7] com o objetivo de validar nossos algoritimos. O resultado pode ser apreciado na Figura (3.4). Para  $\alpha < 0.4$  o desempenho da rede livre de escala é inferior a  $sw$  com  $p = 0.4$ . Mas como a livre de escala ( $SF$ ) possui uma atenuação mais lenta, conforme agregamos mais padrões o seu desempenho relativo às outras aumenta.

Apesar dessa melhora relativa da rede SF em relação às outras, uma questão importante deve ser respondida. Até que carga armazenada a superposição corresponde à fração de padrões recuperados totalmente ou é fruto de uma recuperação de estados espúrios?

Sabemos que o modelo de Hopfield, no limite termodinâmico, seja ele todo conectado ou diluído aleatoriamente, apresenta uma transição descontínua da recuperação total dos padrões para a não recuperação. Para apreciar a transição e entendermos melhor o resultado anterior verificamos o comportamento da eficiência como pode ser visto na Figura (3.5). Podemos ver que para essas novas geometrias a transição continua descontínua entre as duas fases.

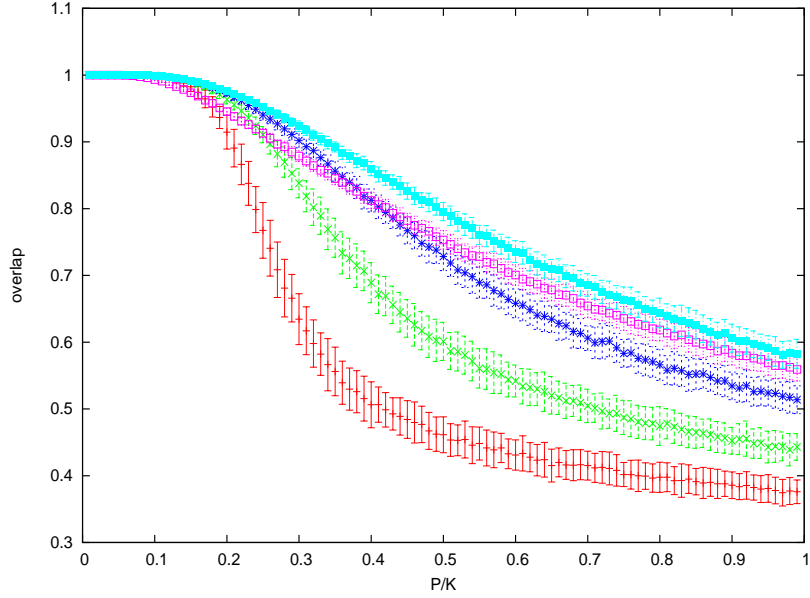


Figura 3.4: Superposição vs  $\alpha$ . Redes de small world  $p = 0.0$  (+),  $p = 0.2$  ( $\times$ ),  $p = 0.4$  (\*) e  $p = 1.0$  (■). Rede livre de escala ( $\square$ ).  $N = 5000$   $K = 100$  para as redes sw. Para as redes livre de escala  $N = 5000$ ,  $m = n = 100$ . Média em 100 realizações.

Note que nossos valores de transição não são boas estimativas para encontrar  $\alpha_c$  mas, do ponto de vista prático, nos garante uma boa informação a respeito do que está ocorrendo. Em uma rede neural o interesse é na recuperação completa de padrões, e não em um estado espúrio que possui apenas um pedaço da informação original. Podemos ver que a rede aleatória possui a melhor efetividade, recuperando 100% das vezes no intervalo de  $0.0 < \alpha \leq 0.20$ , ou seja 10 padrões recuperados corretamente, e ainda apresentando uma recuperação parcial na faixa de  $0.20 < \alpha \leq 0.24$ . Isso é cerca de 50% em relação à rede SF. Um fato intrigante é que a rede sem escala apresentou uma eficiência inferior a rede regular. Para mostrar que esses resultados não são frutos de particularidades dos parâmetros escolhidos, realizamos simulações para outros valores de  $K$  e  $m$ .

Na Figura (3.6) é mostrado o efeito do aumento da diluição na superposição das redes. Para uma conectividade de  $K = 50$ , Figura (3.6) (a) vemos que o compor-

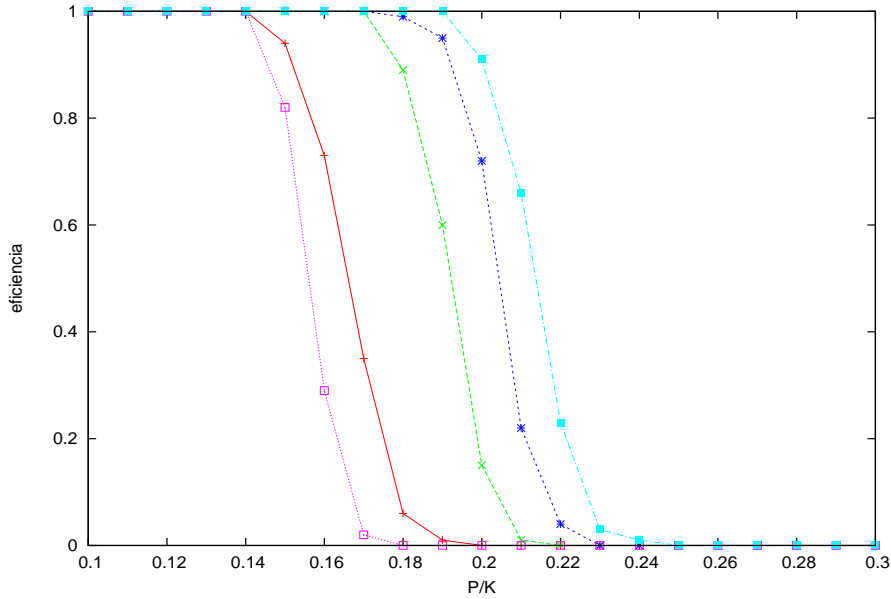


Figura 3.5:  $\varphi$  vs  $\alpha$ . Rede small world com  $p = 0.0$  (+),  $p = 0.2$  ( $\times$ ),  $p = 0.4$  (\*) e  $p = 1.0$  ( $\blacksquare$ ) com  $N = 5000$  e  $K = 100$ . Rede livre de escala ( $\square$ ),  $N = 5000$  e  $m = n = 50$ . Média em 100 realizações para todas as redes.

tamento é semelhante ao apresentado na Figura (3.4) e mostra-se em total acordo com o obtido em [7]. Já para uma diluição mais severa, obtemos uma mudança no comportamento. A rede  $SF$  não consegue superar a rede  $sw$  com  $p = 0.4$ .

Assim como previsto na teoria, o aumento da diluição aumenta a superposição, quando medimos esse em função da carga na rede [13]. Mas notemos que quanto mais diluída a rede, menos padrões ela é capaz de recuperar. Para mostrar com clareza a diferença na recuperação dos padrões, mostramos na Figura (3.7) a superposição versus número de padrões armazenados para  $sw$  com  $p = 0.4$  com diferentes conexões. Note que para a  $K = 100$ , a rede recupera por volta de 22 padrões com  $m \geq 0.97$ . Esse número cai para 10 padrões na rede com  $K = 50$ , vai a 3 na rede  $K = 10$ , e é apenas 2 para  $K = 6$  e na rede de  $K = 4$  a recuperação é a trivial, de apenas 1 parâmetro recuperado totalmente. Esses resultados estão expostos na Figura (3.8).

Nela podemos perceber que a recuperação total de padrões, se torna quase indis-

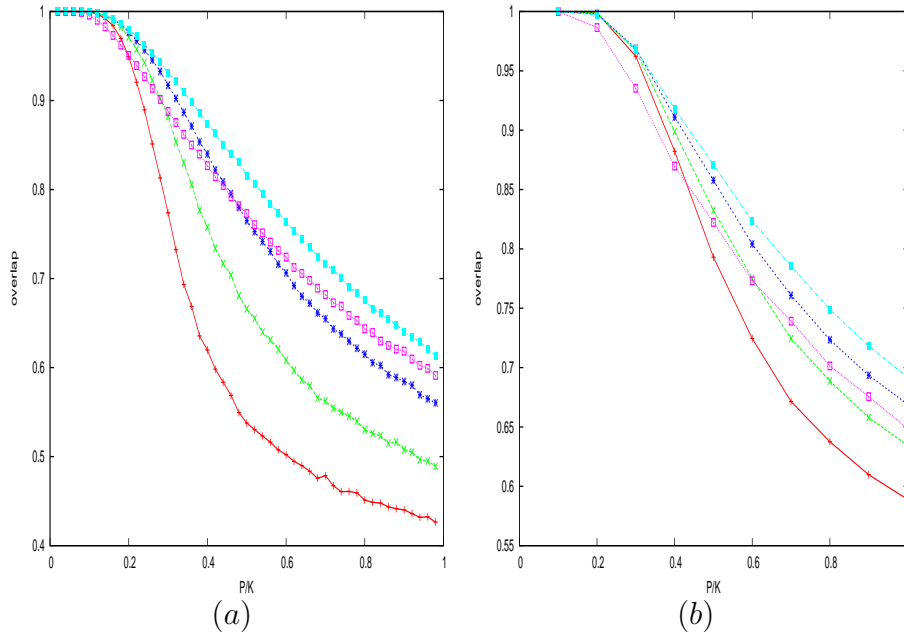


Figura 3.6: Superposição vs  $\alpha$  para níveis maiores de diluição.  $N = 5000$  para todas as simulações. (a)  $K = 50$  para as redes  $sw$  e  $m = n = 50$  para a rede livre de escala. (b)  $K = 10$  para  $sw$  e  $m = n = 10$  para livre de escala. Para ambos os gráficos small world com  $p = 0.0$  (+),  $p = 0.2$  ( $\times$ ),  $p = 0.4$  (\*),  $p = 1.0$  (■) e livre de escala (□)

tinguível para uma conectividade muito pequena,  $K \leq 10$ , mostrando uma pequena vantagem da rede  $sw$  com  $p = 0.4$  em relação às outras topologias. Sabendo que eficiência é uma variável booleana, com variância finita, não podemos afirmar com clareza que a essa rede possui uma melhora a baixa conectividade. É mais razoável entender que com baixa conectividade a transição acontece com poucos padrões, e o acréscimo de apenas um padrão na rede já causa uma grande diferença em sua carga. Em outras palavras, olhar uma rede com conectividade  $K < 10$  é ter uma baixa resolução a respeito das diferenças das topologias estudadas. Além disso podemos apreciar a variação da efetividade pela carga para uma conectividade distinta.

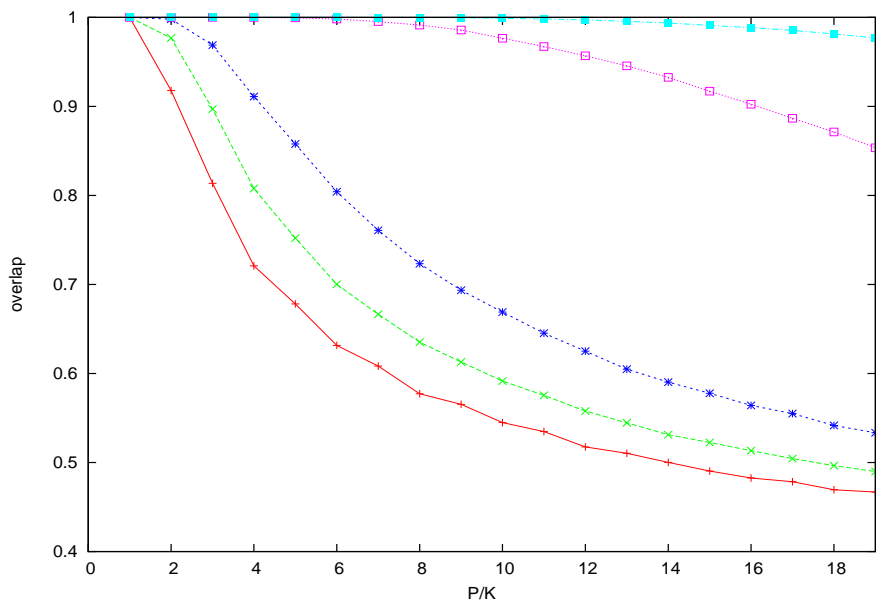


Figura 3.7: Superposição vs padrões armazenados  $N = 5000$ ,  $p = 0.4$ .  $K = 4$  (+),  $K = 6$  (x),  $K = 10$  (\*),  $K = 50$  (□) e  $K = 100$  (■)

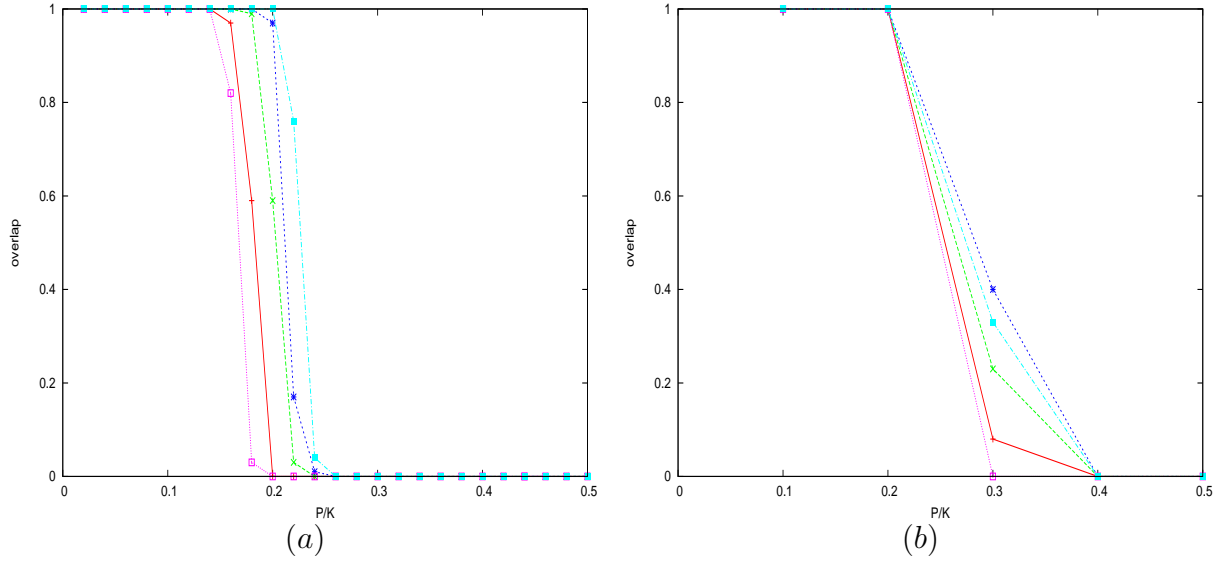


Figura 3.8:  $\varphi$  vs  $\alpha$  para níveis maiores de diluição,  $N = 5000$  para todas as redes. (a)  $K = 50$  para as redes small world, e  $m = n = 25$  para a rede livre de escala. (b)  $K = 10$  para  $sw$   $m = n = 5$  para livre de escala. Em ambos os gráficos:  $p = 0.0$  (+),  $p = 0.2$  (x),  $p = 0.4$  (\*),  $p = 1.0$  (■) e livre de escala (□)

### 3.3 Bacia de Atração

Além de medirmos a capacidade de recuperação de informação temos interesse em mais uma característica das redes neurais, o tamanho médio das bacias de atração. Quanto maior a bacia de atração, maior é a capacidade da rede de recuperar um padrão, mesmo que o estímulo dado só tenha uma fração da informação daquele padrão. Para medir a bacia, o estímulo dado a rede é um padrão  $\mu$  com um ruído, ou seja com a superposição inicial menor do que um.

Na Figura 3.9 apresentamos a medida da bacia de atração para a rede aleatória a distintos níveis de armazenamento. Note que a bacia diminui com o aumento de  $\alpha$ . Para  $\alpha = 0.05$  e  $\alpha = 0.1$  a recuperação dos padrões é perfeita até uma distância de 95% do padrão original. A essa distância ocorre uma transição súbita para a não recuperação. Já para  $\alpha = 0.15$  a recuperação do padrão inicial se mantém em 99% até por volta de uma superposição inicial de 0.08 quando cai subitamente. Em



$\alpha = 0.2$  vemos que a rede está próxima ao limite de carga, mas mesmo assim se mantém constante a superposição final, em torno de 97%, até uma súbita transição no valor de 0.08 para a superposição inicial.

Isso mostra que as bacias para a rede diluída são muito maiores que para o modelo totalmente conectado, que é apresentado na Figura (3.10).

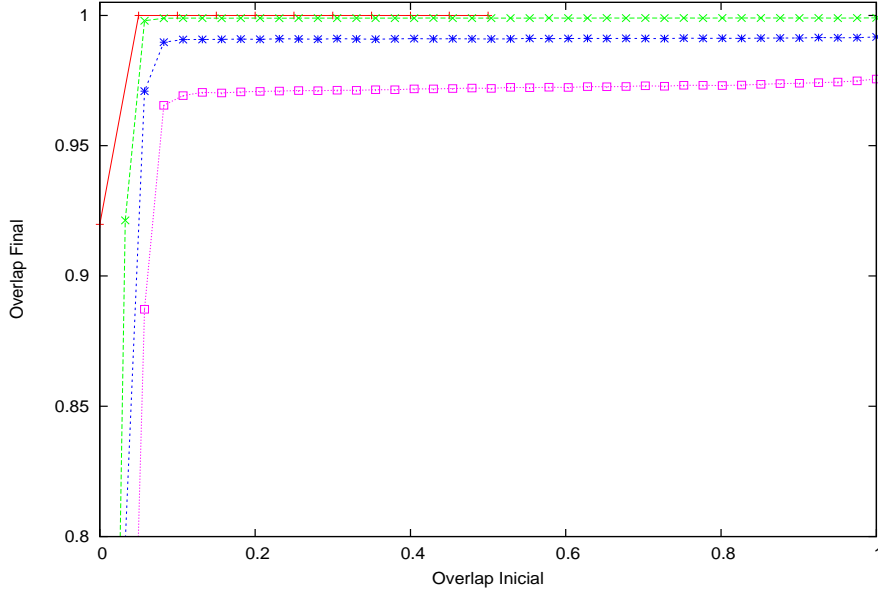


Figura 3.9: Bacia de atração para  $sw p = 1.0$ ,  $N = 5000$  e  $k = 100$ .  $P = 5$  (+),  $P = 10$  (x),  $P = 15 = 5$  (\*) e  $P = 20$  (□)

No modelo não diluído, para uma carga pequena, sabemos da existência do grande número de atratores espúrios, oriundos da combinação de padrões armazenados. Dessa forma, quanto mais distante dos padrões ficamos, maior a possibilidade de que a dinâmica comece na bacia de um desses atratores. Já no modelo diluído, o número de atratores espúrios é menor, e o afastamento do estímulo inicial dado tem menor chance de cair na bacia de atração de um atrator espúrio. Assim ou ele já está suficientemente longe para cair na bacia de outro atrator, ou ainda cai na bacia do mesmo. Vemos que conforme aumentamos a carga, mais a recuperação fica comprometida, e atratores espúrios começam a aparecer em maior número. Nas Figuras (3.11) e (3.12) podemos contemplar a bacia de atração para as outras geometrias.

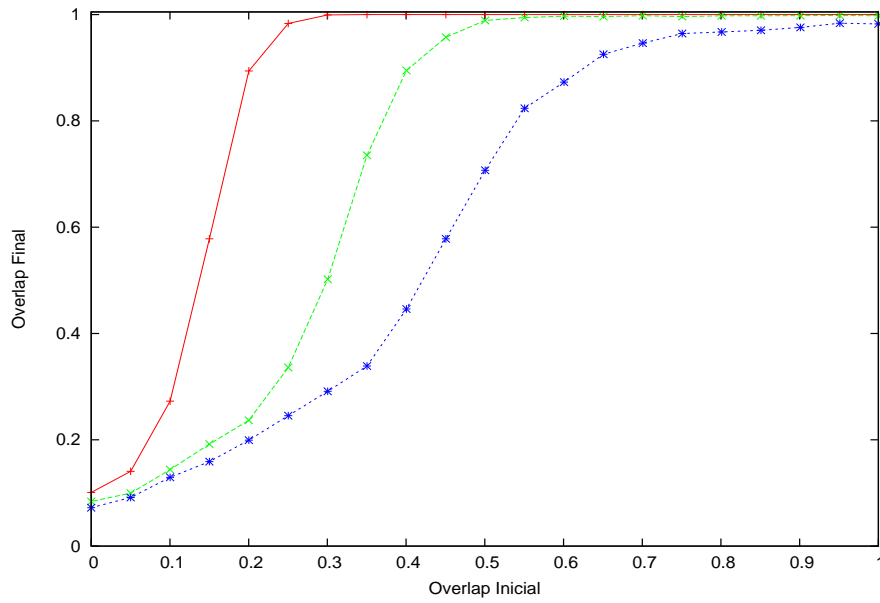


Figura 3.10: Bacia de atração, para Hopfield não diluição  $N=1000$ .  $\alpha = 0.05$ ,  $\alpha = 0.1$  e  $\alpha = 0.13$  da esquerda para a direita.

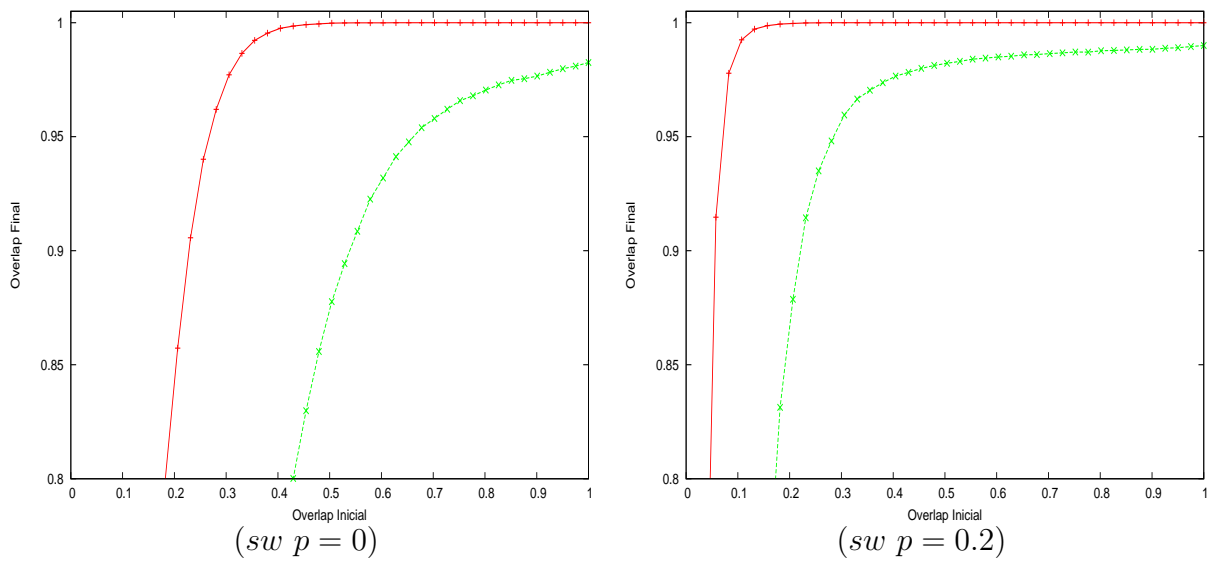


Figura 3.11: Bacia de atração para diversas geometrias,  $N = 5000$ ,  $K = 100$ ,  $\alpha = 0.05$  e  $\alpha = 0.15$  em  $10^4$  realizações.

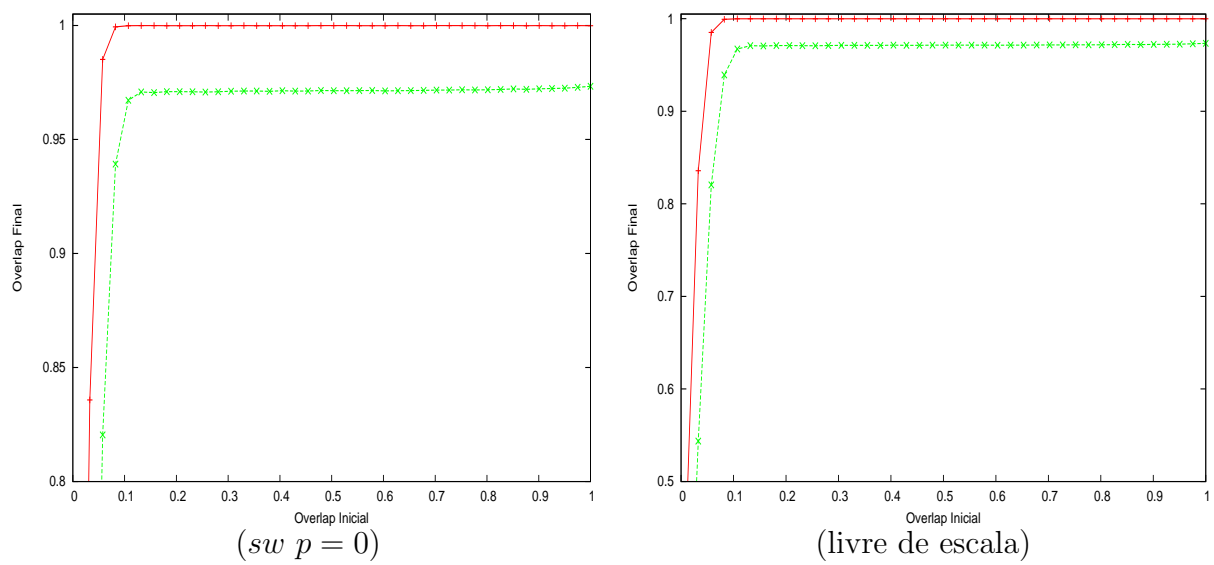


Figura 3.12: Bacia de atração para diversas geometrias, com  $N = 5000$ ,  $K = 100$ ,  $\alpha = 0.05$  e  $\alpha = 0.15$ . sw com  $K = 100$ , e livre de escala com  $m = n = 50$ . Média em  $10^4$  realizações.

A Figura (3.11) não apresenta grandes novidades. Como sabemos a rede altamente organizada apresenta uma recuperação ruim, como mostrado em [40]. A rede  $sw$  com  $p = 0.2$  ainda possui poucas conexões de longa distância, e isso se reflete na diminuição considerável do tamanho da bacia com o aumento da carga na rede.

Para  $sw$  com  $p = 0.4$  tem um desempenho pouco inferior a rede  $sw$  com  $p = 1.0$ , pois sua superposição para  $\alpha = 0.15$  se estabiliza por volta de 98%. Já a rede SF que mostrou um rendimento inferior as redes  $sw$ , em contrapartida apresentou uma bacia de atração tão larga quanto a rede  $sw$  com  $p = 1.0$ .

Para entender melhor esse comportamento da bacia de atração, uma simulação interessante foi proposta por Morelli et al [6] para as redes de  $sw$ . Apresenta-se a rede um estímulo gerado aleatoriamente, deixa-se a rede evoluir até um estado estacionário, e verifica-se a superposição da rede com cada um dos padrões armazenados. Defini-se então a eficiência total,  $\Phi_{100\%}$ , que tem valor igual a unidade se  $|m_\mu| = 1$ ,  $\mu \in \{P\}$ , e é 0 caso contrário. Refizemos essas simulações para duas redes, com os resultados expostos nas Figuras (3.13) (a) e (3.14) (a) em perfeito acordo com a literatura, garantindo bom desempenho de nossos algoritmos.

Alguns fatos chamaram atenção nessa simulação. Como comentado por Morelli et al [6], a rede com valores de  $p$  em torno de 0.4 apresentam melhor eficiência do que a rede aleatória. Vemos um estranho fato na Figura (3.13) (a). A rede, para  $p < 0.3$  e  $p > 0.5$  com 10 padrões armazenados se mostra superior a para 5 padrões, na mesma região de desordem. Mas o mais curioso é que esse resultado não se reproduziu quando diluímos mais a rede, em (3.14) (a).

Fizemos uma modificação na proposta inicial, diminuimos a rigidez da eficiência total, e a redefinimos para:  $\Phi_{97\%} = 1$  se  $|m_\mu| + 0.03 \geq \mu \in \{P\}$  e  $\Phi_{97\%} = 0$  caso contrário. Essa definição menos rígida é a extensão da equação (3.3). Utilizando essa nova definição refizemos as simulações, apresentadas nas Figuras (b) (3.13) e (b) (3.14).

Podemos notar, fora a distinta manifestação para  $P = 10$  em (3.13) que a melhor eficiência em torno de  $p = 0.4$  não depende da conectividade, quando essa varia em

curta escala, mas sim da carga na rede.

A diferença visível de ambas as figuras quando comparamos  $\Phi_{100\%}$  e  $\Phi_{97\%}$  é o ganho de eficiência para  $\alpha = 0.2$ . Isso mostra que em uma região próxima a crítica, os mínimos da bacia são encontrados em uma região de superposição  $m_\mu > 0.97$  e não no ponto  $m_\mu = 1$ .

Sabemos que número de elementos no espaço amostral dos estímulos da rede é  $2^N$ . Dessa forma, é necessário uma grande estatística para obtermos  $\Phi_{100\%}$  e  $\Phi_{97\%}$  com clareza. A nossa estatística nos permite perceber o comportamento geral, mas, infelizmente, está abaixo dos nossos objetivos. Morelli et al realizaram  $10^4$  simulações por ponto e analisaram tamanhos até  $5 \times 10^4$ , e nós realizamos  $10^3$  simulações por ponto.

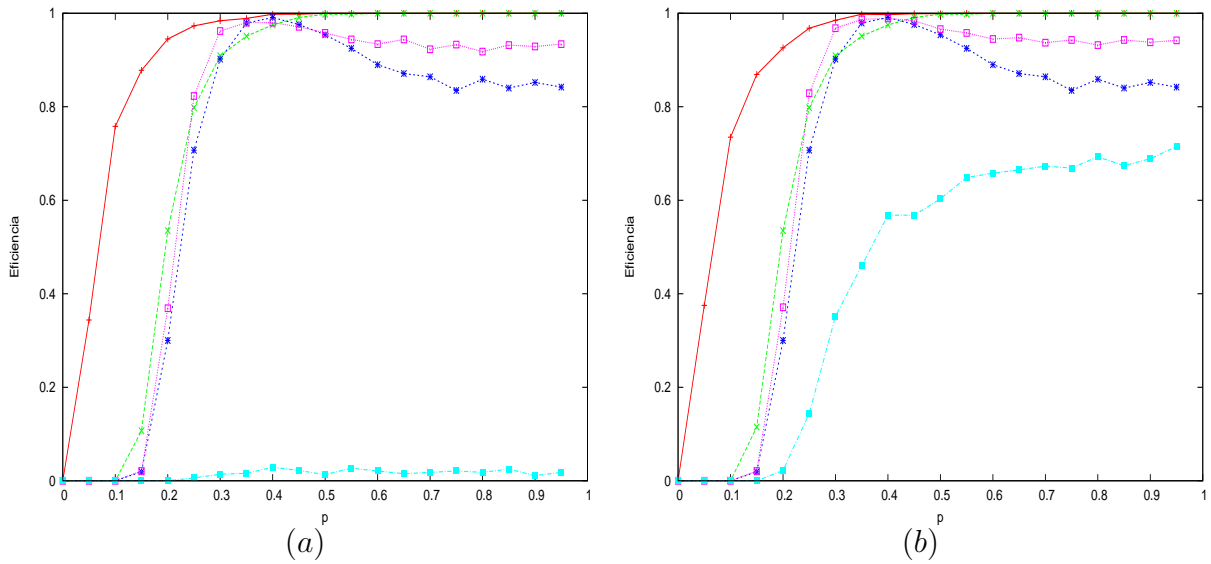


Figura 3.13: (a)  $\Phi_{100\%}$  vs  $p$  e (b)  $\Phi_{97\%}$ . Para ambos os gráficos  $N = 5000$  e  $K = 200$ .  $P = 1$  (+),  $P = 2$  (x),  $P = 5$  (\*),  $P = 10$  ( $\square$ ) e  $P = 20$  ( $\blacksquare$ ). Média em 1000 realizações.

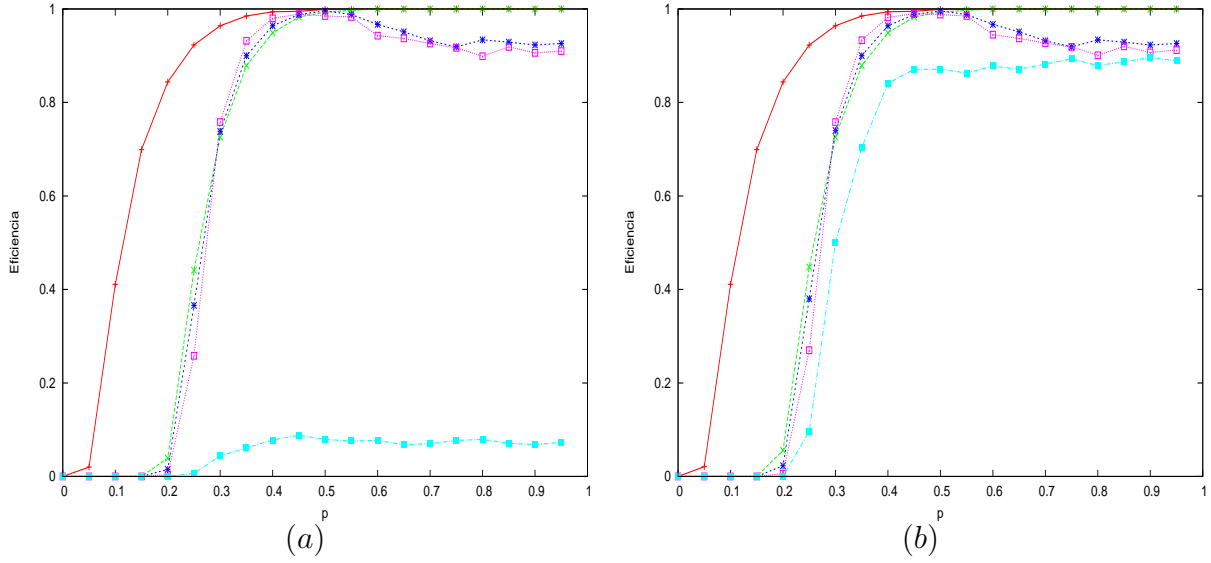


Figura 3.14: (a)  $\Phi_{100\%}$  vs  $p$  e (b)  $\Phi_{97\%}$ . Para ambos os gráficos  $N = 5000$ ,  $K = 100$ .  $P = 1$  (+),  $P = 2$  (×),  $P = 5$  (\*),  $P = 10$  (□) e  $P = 20$  (■). Média em 1000 realizações.

### 3.4 Padrões com correlação

Resolvemos introduzir nos padrões uma leve correlação e analisar como a regra Hebbiana proposta por Hopfield se comporta em função do aumento da correlação. A primeira vista essa abordagem pode parecer ingênua e equivocada, pois já existem meios para abordar padrões com correlação como proposto em [41] ou pelo método da pseudoinversa [13]. Primeiramente desejamos manter o apelo biológico da proposta original, o que descarta pseudoinversa devido a sua não localidade. A regra proposta por Amit é interessante, mas mesmo assim ainda possui uma não localidade, ao supor um *bias* igual para todos os padrões. Além disso desejamos fazer desse estudo uma introdução para uma regra de aprendizado baseada no crescimento da rede.

Então o primeiro problema atacado foi o da variação da superposição com a atividade da rede. Para controlar a correlação dos padrões utilizamos a atividade média, que é definida como a quantidade média de neurônios ativos em cada padrão

armazenado. Nas Figuras (3.15) e (3.16) mostramos, para as 4 redes de interesse, a superposição versus carga para diversos valores de atividade.

Podemos observar para todas as geometrias que pequenos valores de carga, atividade por volta de 45%, o desempenho da rede é pouco afetado. Mas com um pequeno aumento da atividade ocorre uma deterioração súbita da superposição. Para valores inferiores a 40% a correlação já compromete severamente a regra de aprendizado.

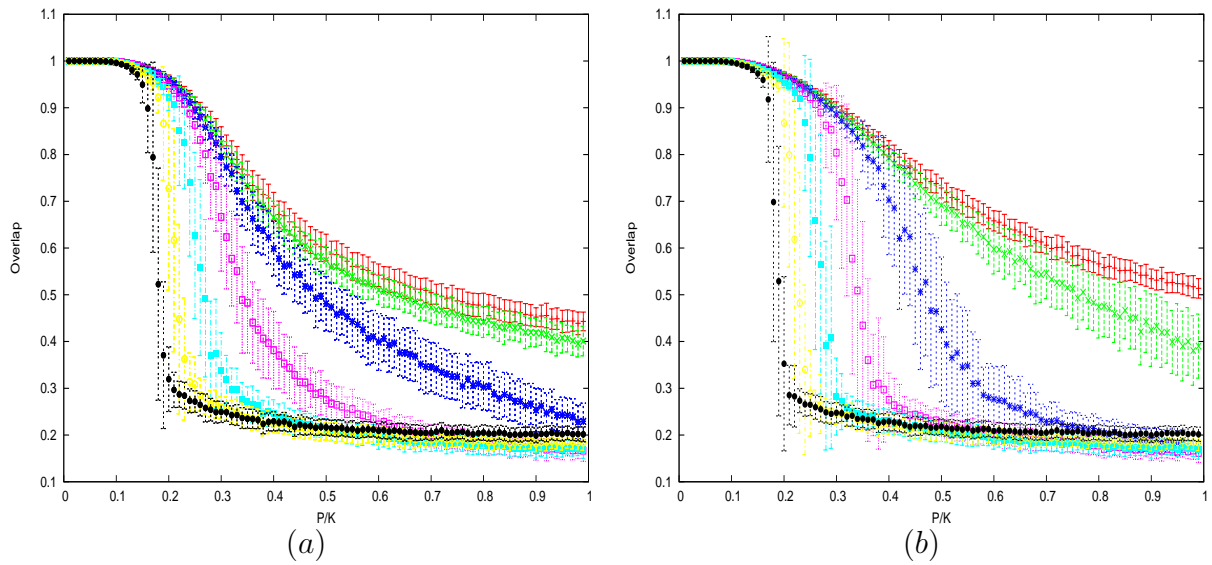


Figura 3.15: Superposição vs carga. Para ambos os gráficos  $N = 5000$  e  $K = 100$ . Média em 100 realizações, com atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40%  
 (a)  $sw$  com  $p = 0.2$  (b)  $sw$  com  $p = 0.4$

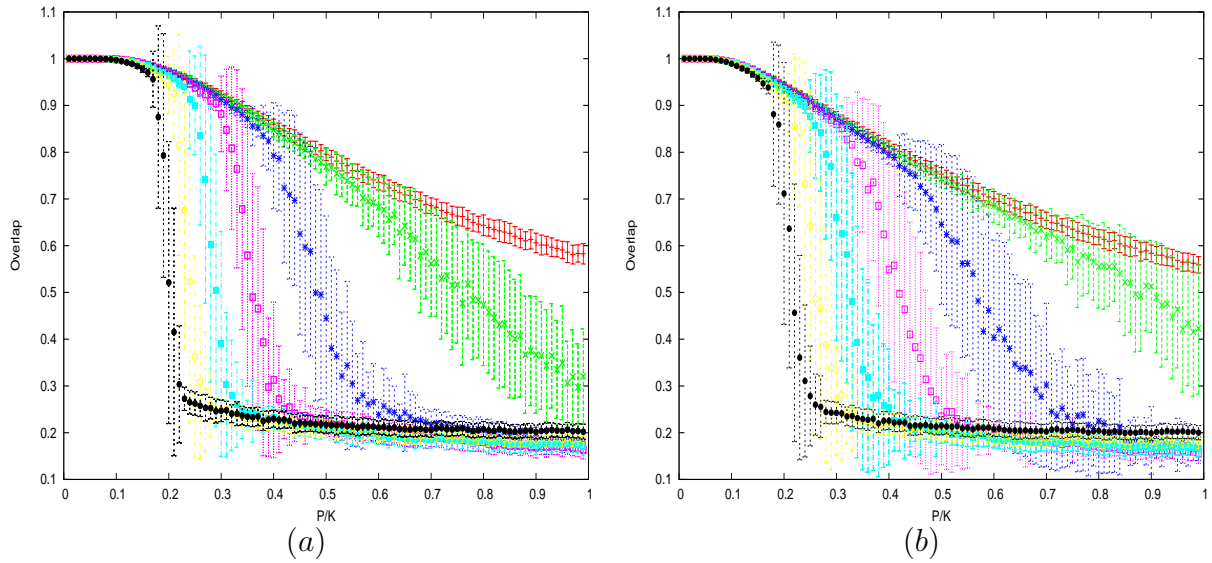


Figura 3.16: Superposição vs carga. para ambos os gráficos  $N = 5000$ ,  $K = 100$  e média em 100 realizações. Atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40% (a)  $sw$  com  $p = 1.0$  (b) livre de escala

Analisamos também a eficiência mediante o aumento de correlação e observamos a mesma questão, para atividade maior que 45% a mudança no comportamento da rede é pouca, mas para atividades abaixo desse patamar, o desempenho da rede piora rapidamente.

Nas Figuras 3.17 e 3.18 podemos ver uma notável diferença da eficiência nas redes  $sw$  e na rede SF. Para atividade maior que 43% a rede de SF não apresenta queda significativa, enquanto as redes de  $sw$  perderam 8%. Conforme diminuimos a atividade, mais significativa a perda se torna para essas redes. Com atividade de 40% a perda nas redes  $sw$  é de 25% enquanto a rede SF tem apenas uma perda de 10%.

Aqui começamos a evidenciar as primeiras pistas que procurávamos: para arquiteturas distintas a correlação possui efeitos diferentes.



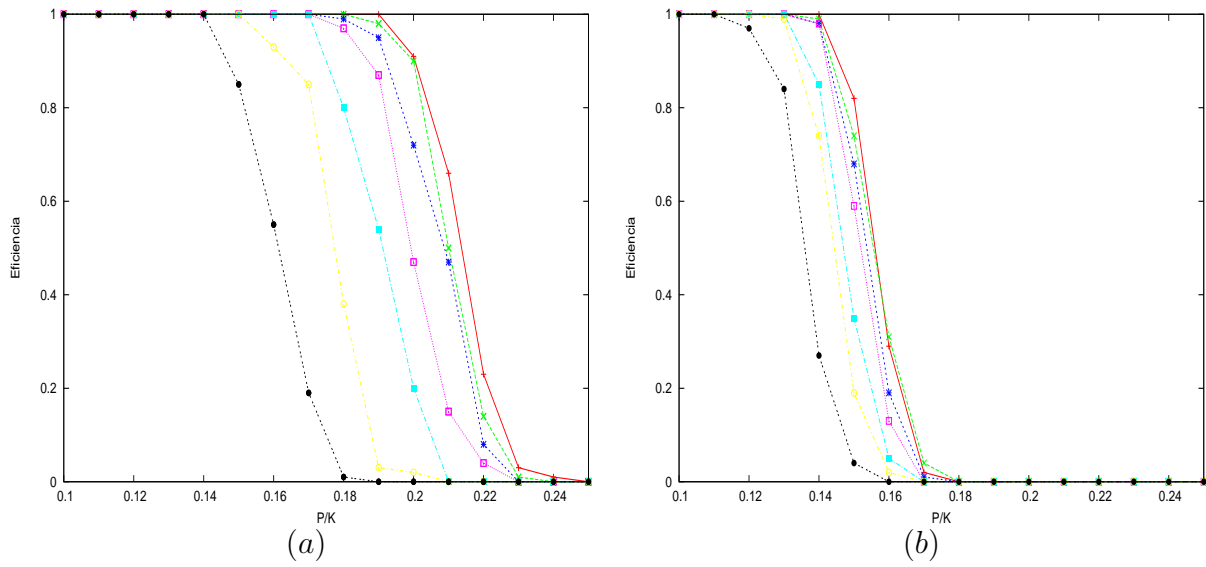


Figura 3.17: Eficiência vs carga. Para ambos os gráficos  $N = 5000$ ,  $K = 100$  e médias em 100 realizações. Atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40%. Da direita para a esquerda. (a)  $sw$  com  $p = 1.0$  (b) livre de escala

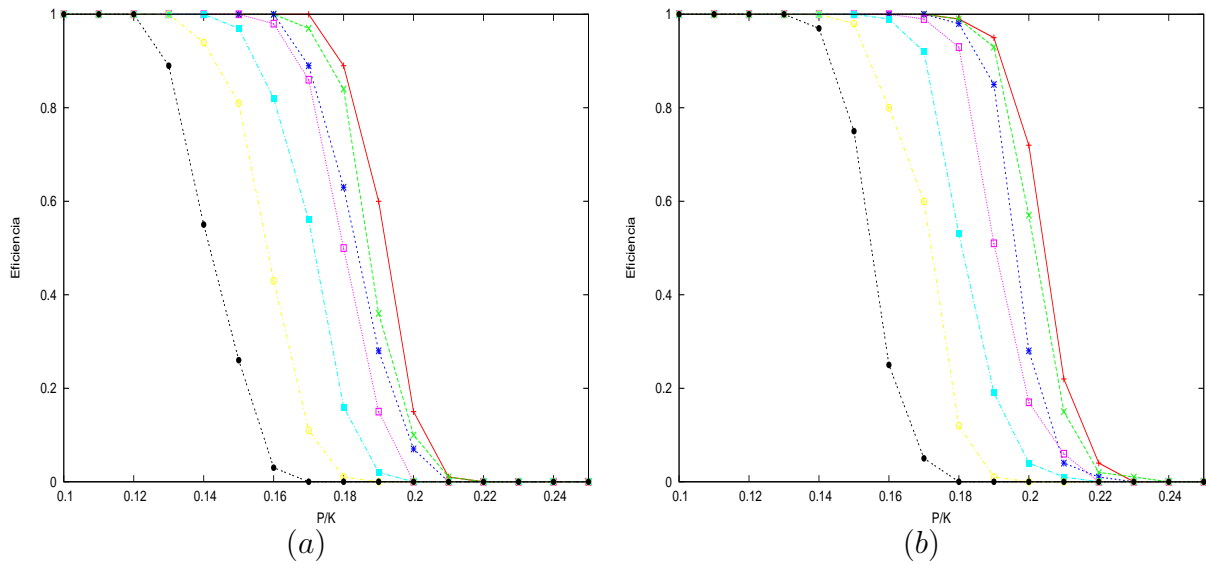


Figura 3.18: Eficiência vs carga. Para ambos os gráficos  $N = 5000$ ,  $K = 100$  e média em 100 realizações. Atividade de 50%, 45%, 44%, 43%, 42%, 41% e 40%. Da direita para a esquerda. (a)  $sw$  com  $p = 0.4$  (b)  $sw$  com  $p = 0.2$

Essa melhora da rede de SF frente às outras pode ser evidenciada na Figura , onde comparamos o desempenho de todas as redes para uma atividade de 43%.

A topologia livre de escala apresenta, para  $0.35 < \alpha < 0.5$  um desempenho melhor que as outras redes. Apesar de ser uma região de carga alta, e com isso a eficiência já é nula, podemos perceber que o comportamento global, de possuir um superposição que cai mais lentamente com aumento da carga, se manteve perante a correlação, indicando novamente que a rede sem escala possui maior capacidade de tratar padrões correlacionados. Já as outras redes, o aumento da correlação acentua a transição de recuperação para não recuperação em um sistema finito. A rede *sw* com  $p = 0.2$  mostra uma recuperação elevada na região  $0.5 < \alpha < 1.0$  .

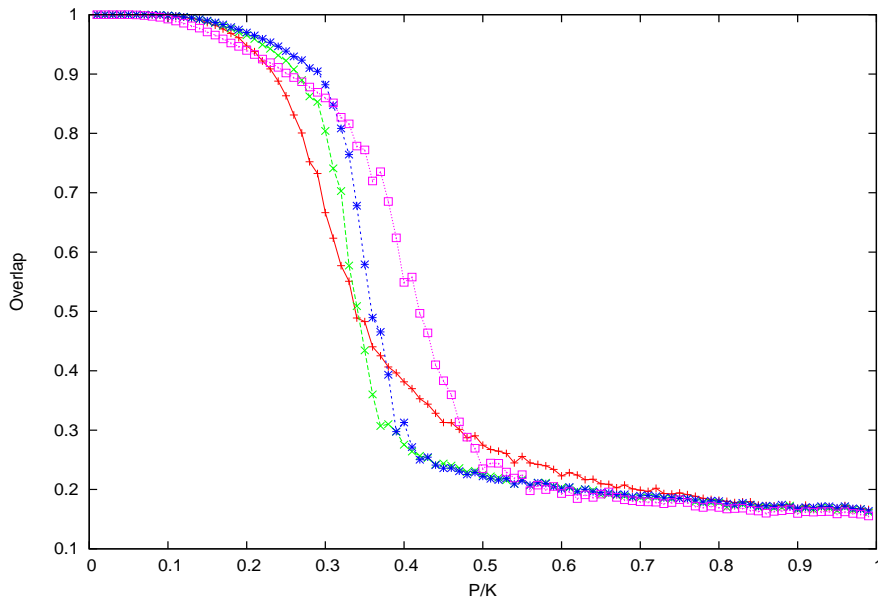


Figura 3.19: Superposição vs carga.  $N = 5000$ ,  $K = 100$  Média em 100 realizações. Atividade de 43%.  $p = 0.2$  (+),  $p = 0.4$  (x),  $p = 1.0$  (\*) e rede livre de escala (□)

### 3.5 Superposição e a correlação

Os efeitos vistos anteriormente ainda não deixam claro o que ocorre na rede neural quando a correlação é aumentada. Para isso, propomos um novo experi-

mento. Dada uma carga, arquitetura e conectividade fixa, investigamos o comportamento da superposição em função da atividade. Por simplicidade nos referiremos à *atividade*/100 por  $a$ .

Nossa primeira preocupação foi descobrir um tamanho do sistema que possibilitasse uma análise satisfatória. O resultado é mostrado para a rede  $sw$  com  $p = 1.0$  como exemplo na Figura (3.20). As outras redes não fogem a esse comportamento. Optamos por  $N=5000$  para seguir nas simulações, pois sugere estar suficientemente próximo ao estado estacionário, e possui um custo computacional aceitável para cobrir a quantidade de simulações que foram necessárias para varrer o espaço de parâmetros.

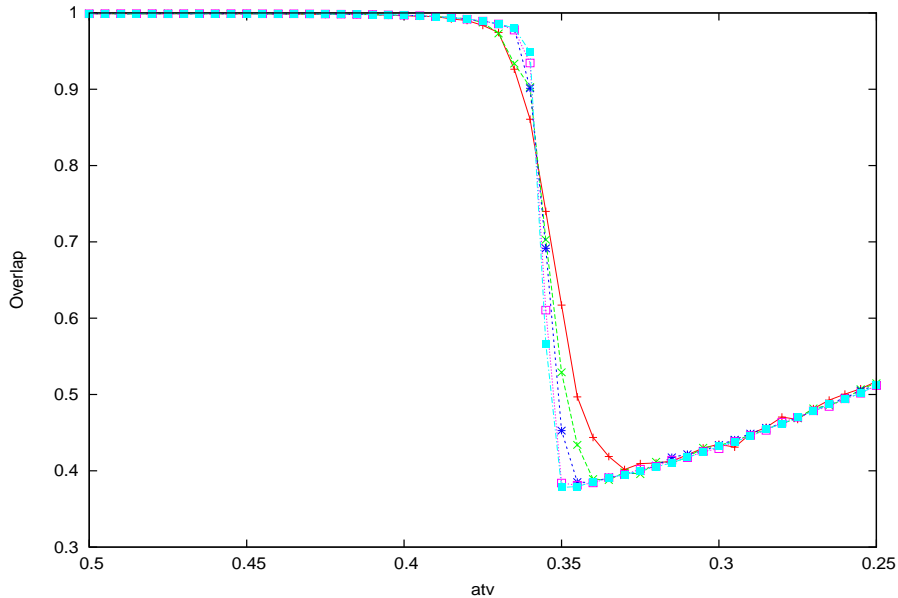


Figura 3.20: Superposição vs  $a$ .  $K = 100$ ,  $P = 10$  para  $sw$  com  $p = 1.0$ .  $N = 1000$  (+),  $N = 2000$  (x),  $N = 5000$  (\*),  $N = 10000$  (□) e  $N = 20000$  (■). Média em 100 realizações

Vemos na Figura (3.20) que uma acentuada transição ocorre entre as duas fases, recuperação e não recuperação, por volta de  $a = 0.35$ . Além disso, para  $a > 0.35$  o gráfico começa um crescimento, aparentemente, linear. Isso se dá devido ao crescimento da correlação, e essa passa a ser dominante no processo. O próximo

passo dado é ver o efeito da diluição nesse sistema e nos certificarmos que não estamos observando nenhum resultado peculiar.

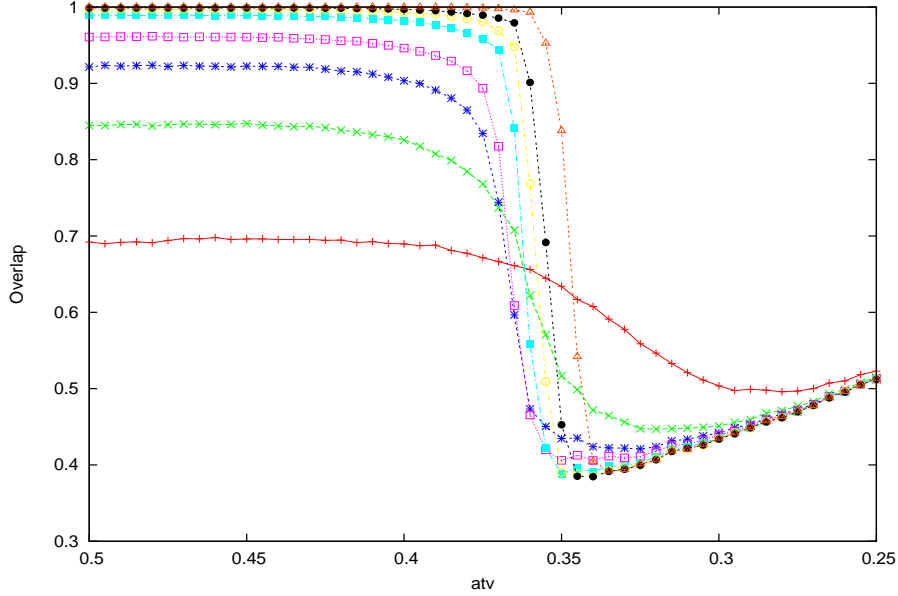


Figura 3.21: Superposição vs atividade.  $P = 10$  e  $N = 5000$  para  $sw$  com  $p = 1.0$  em diferentes valores de  $K$ .  $K = 10$  (+),  $K = 20$  (×),  $K = 30$  (\*),  $K = 40$  (□),  $K = 60$  (■),  $K = 80$  (o),  $K = 100$  (●) e  $K = 200$  (△). Média em 100 realizações

Nas Figuras (3.22) e (3.23) vemos o efeito para as 4 geometrias para 4 valores de  $\alpha$ . Em (3.22) (a) temos a rede armazenando poucos padrões, observamos que a transição das redes ocorrem em dois pares. As geometrias  $sw$  com  $p = 0.2$  e  $p = 0.4$  mostram-se cerca de 1% inferior as redes com  $p = 1.0$  e livre de escala. Essa diferença não é suficiente para nenhuma afirmação definitiva. O mesmo ocorre na Figura (3.22) (b), com a distinção que a transição não é mais por volta de  $a = 0.3$  mas  $a = 0.35$ . Isso nos mostra que o aumento da carga na rede influencia essa transição. Na Figura (3.23) (a) vemos que a transição recua no eixo  $x$ , para o valor de  $a = 37$ . Já a uma carga maior, em (b) da mesma figura vemos que a SF leva vantagem perante as outras redes. Mas, vale lembrar, que a esse nível de carga, todas as redes de small-world ainda possuem eficiência maior do que zero, o que não é verdade para a rede livre de escala. Dessa forma o que apreciamos em (3.23)

(b) é fruto daquele mesmo efeito de superposição lentamente atenuado, apreciado anteriormente. Esse efeito de atenuação da superposição já foi mostrado em [42], mas utilizando um distinto algoritmo para conseguir uma rede com distribuição de conectividade dada por uma lei de potência. Além disso os resultados obtidos não são para redes com distintos  $\gamma$ , expoente da distribuição de graus da rede livre de escala, e distintas conectividades, dificultando uma comparação direta com os nossos resultados.

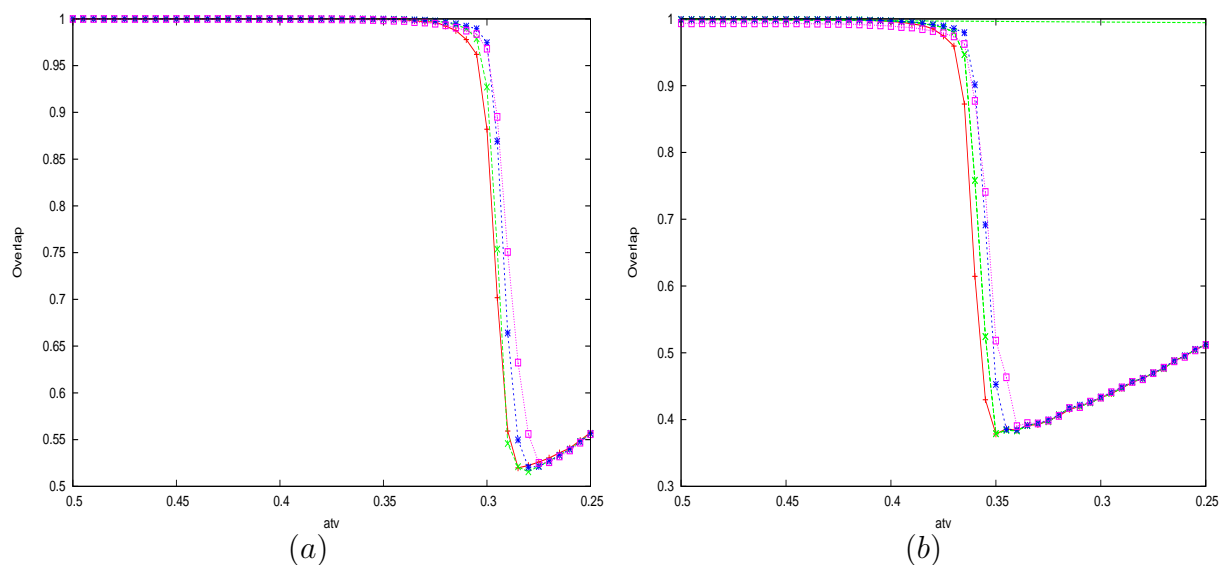


Figura 3.22: Superposição vs  $a$  para as geometrias de small world com com  $p = 0.2$  (+) ,  $p = 0.4$  (x),  $p = 1.0$  (\*) e rede livre de escala ( $\square$ ). Para  $\alpha = 0.05$  em (a) e  $\alpha = 0.10$  em (b). Médias em 100 realizações.

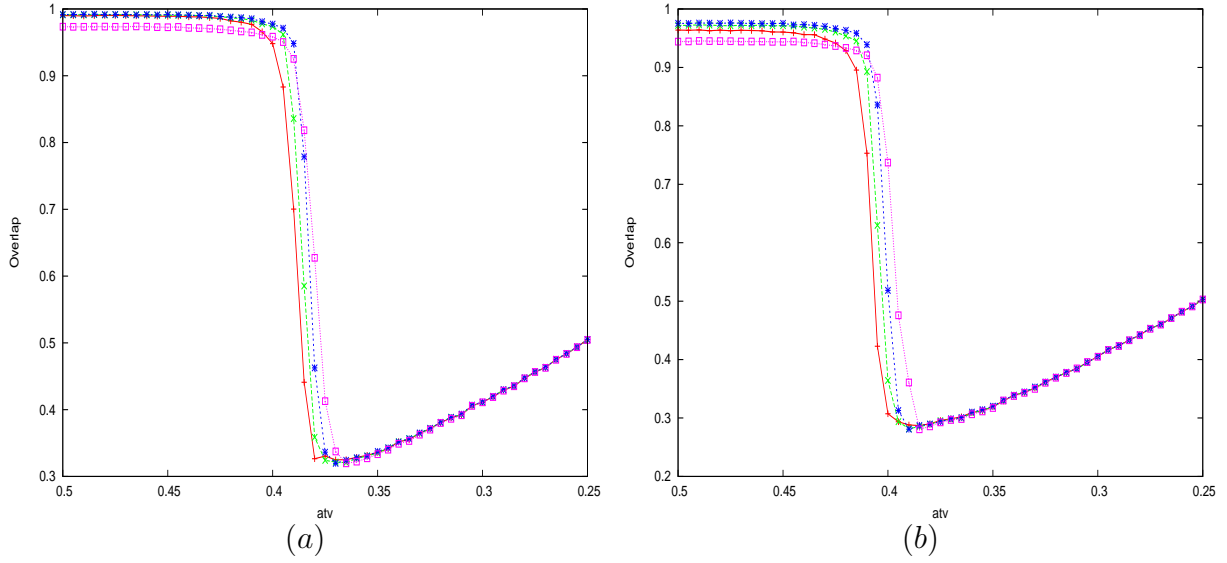


Figura 3.23: Superposição vs  $a$  para as geometrias de small world com  $p = 0.2$  (+),  $p = 0.4$  ( $\times$ ),  $p = 1.0$  ( $*$ ) e a rede livre de escala ( $\square$ ). Para  $\alpha = 0.15$  em (a) e  $\alpha = 0.20$  em (b). Médias em 100 realizações.

### 3.6 Bacia de atração e padrões com correlação

Realizamos o estudo da bacia de atração quando a rede armazena padrões correlacionados. O resultado comparativo é o mesmo que sem atividade, a bacia cresce conforme a rede aumenta a aleatoriedade nas redes de  $sw$ , e a rede livre de escala apresenta uma bacia tão boa quanto a  $sw$  com  $p = 1.0$ .

Nas Figuras (3.24) e (3.25) olhamos, para como se comporta a bacia de atração frente a uma carga de  $\alpha = 0.15$ . Vemos que ela cai com a aumento da correlação, mesmo que a correlação esteja menor que a correlação de transição, Figura 3.23.

Se aumentarmos a carga para  $\alpha = 0.20$ , vemos que a bacia caiu bruscamente para  $a = 0.41$  como apreciamos na Figura (3.26). Para esse mesmo valor de carga a rede passa a não recuperar corretamente padrões para atividade em torno de 41%. Ou seja, quando passamos da atividade crítica a rede perde a capacidade de recuperar os padrões e a bacia diminui rapidamente.

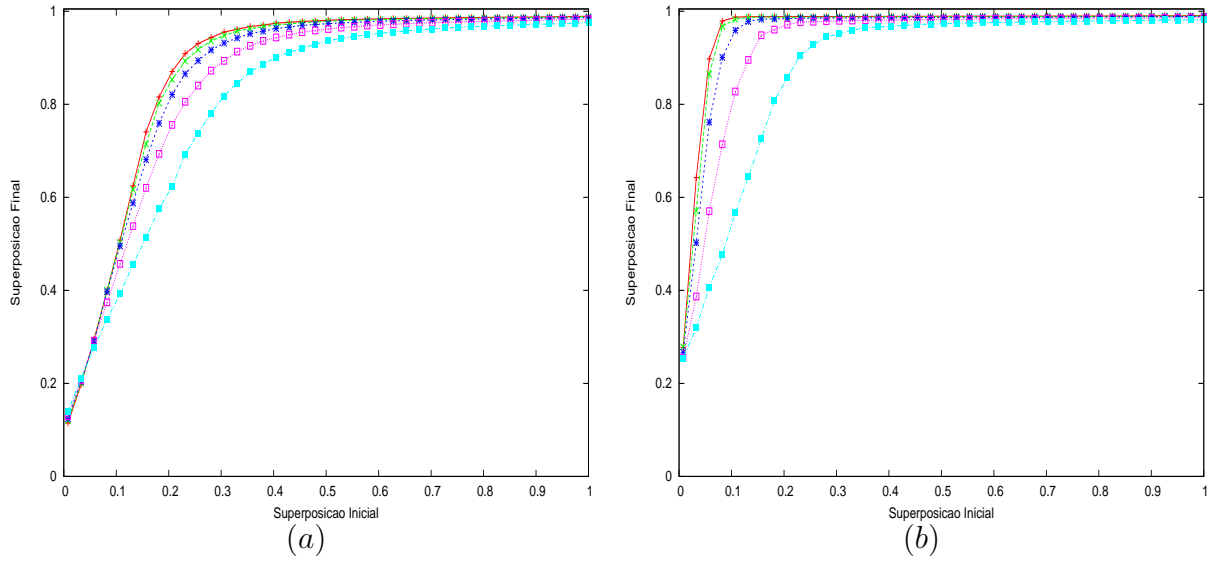


Figura 3.24: Bacia de atração para as geometrias de small world com  $p = 0.2$  (a) e  $p = 0.4$  (b). 45% (+), 44% (×), 43% (\*), 42% (□) e 41% (■). Média em 1000 realizações.

Realizamos também a simulação proposta por Morelli [6], variando a atividade e percebemos que para um estímulo dado aleatoriamente para rede, a eficiência cai mas as curvas mantem o mesmo estilo de formato, com a distinção que o pico translada de  $p = 0.4$ , em  $a = 0.5$ , e vai para  $p = 0.3$  em  $a = 0.4$ .

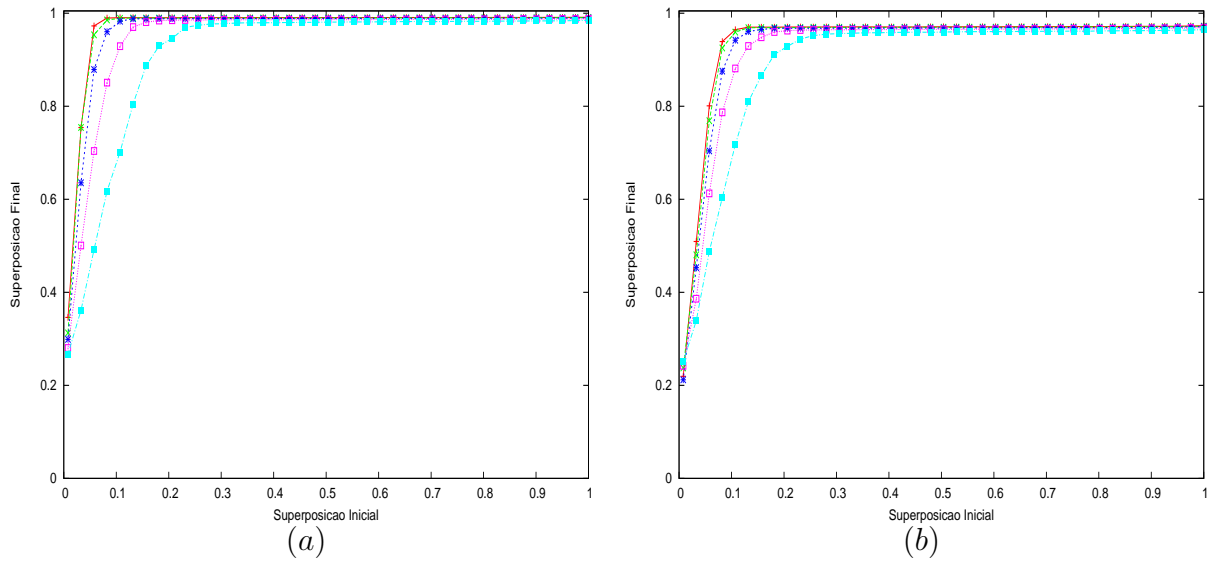


Figura 3.25: Bacia de atração para as geometrias de small world com  $p = 1.0$  (a) e a rede livre de escala (b). 45% (+), 44% (×), 43% (\*), 42% (□) e 41% (■). Média em 1000 realizações.

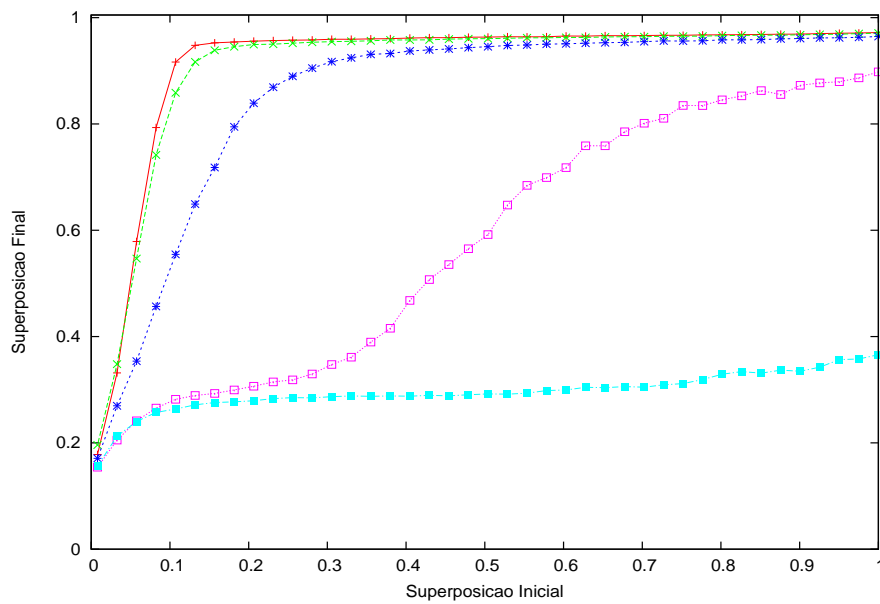


Figura 3.26: Bacia de atração.  $N = 5000$  e  $\alpha = 0.2$  para  $sw$   $p = 0.4$  em diferentes valores de de atividade. 50% (+), 45% (×), 43% (\*), 41% (□) e 40% (■). Média em 1000 realizações.



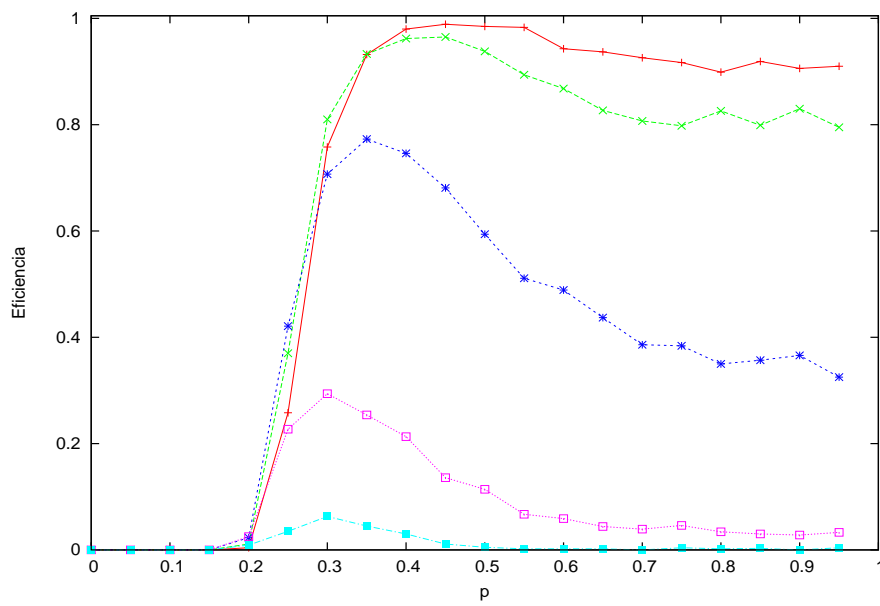


Figura 3.27: Eficiência vs  $p$ .  $N = 5000$  e  $P = 5$  em diferentes valores de de atividade. 50% (+), 48% (x), 45% (\*), 42% (□) e 40% (■). Média em 1000 realizações.

## 3.7 Conclusão

Nesse capítulo fizemos um estudo dos principais trabalhos que discutem redes neurais em redes complexas. Refizemos os principais resultados [7, 6, 36] e contribuímos com o cálculo das bacias de atração para as novas topologias. E mostramos que a grandeza efetividade, se levemente redefinida, pode trazer mais informações a respeito das topologias. Mostramos também que para uma rede de poucas conexões ( $K \leq 10$ ), as diferenças topológicas não ficam evidenciadas.

Analisamos a superposição e a efetividade quando os padrões são correlacionados. Mostramos que a rede livre de escala apresenta, percentualmente, a menor perda com o aumento da correlação. Mostramos que o aumento da atividade não deteriora lentamente a regra de aprendizado, mas temos uma transição aguda para uma dada atividade, entre dois estados da rede: recuperando os padrões e não recuperando. O valor dessa “atividade crítica” depende da topologia e da carga armazenada, sendo o efeito da carga mais forte. Analisamos a bacia de atração em função da atividade e vimos que essa decresce lentamente com o aumento da atividade, se  $a$  estiver antes da transição.

# Capítulo 4

## Conclusões

Em nosso trabalho investigamos o comportamento de redes neurais atratoras, via modelo de Hopfield [4]. A modelagem de redes neurais possui várias simplificações sem nenhum apelo biológico. Nesse trabalho atacamos duas delas, a conectividade do modelo e a necessidade de armazenar padrões descorrelacionados. Primeiramente fizemos uma abordagem diferente das tradicionais [13] para diluição, e nos guiamos por uma recente corrente que surgiu nos últimos anos na física, o estudo de redes complexas. Por redes complexas entendemos uma área multidisciplinar que uniu a mecânica estatística e a teoria dos grafos com intuito de estudar redes reais de milhões de componentes, fazendo uso de ferramentas matemáticas já pre estabelecidas pela teoria dos grafos mas efetuando uma abordagem estatística com o ferramental oriundo da física. O sucesso dessa abordagem em diversas áreas do conhecimento rapidamente se voltou para o maior e mais importante sistema que pode ser modelado por uma rede, o cérebro. Estudos numéricos foram feitos [2] em cima de redes neurais reais e deram indícios de que essas redes possuem características de modelos de redes complexas. Então tivemos na literatura vários artigos [6, 7, 36, 42] com o objetivo de estudar modelos de redes neurais em redes complexas.

Nossa contribuição para a literatura foi investigar o desempenho e a bacia de atração em função da correlação dos padrões armazenados. Podemos concluir com o nosso estudo que:

- Para  $\alpha < \alpha_c$ , rede com diluição aleatória possui melhor desempenho na recuperação de padrões armazenados, fato já conhecido na literatura;
- A bacia de atração mostrou-se proporcional à aleatoriedade da rede, e provavelmente está associada com o caminho geodésico médio;
- Analisamos a recuperação de padrões das geometrias estudadas a um dado  $\alpha$  em função do aumento da correlação entre os padrões, e notamos uma acentuada transição entre duas fases, uma de recuperação e uma de não recuperação;
- Estudamos para as redes small world a recuperação de um estado inicial aleatório em função da desorganização da rede, e como Morelli et al [6] encontramos um pico de recuperação em torno de  $p = 0.4$ . Mostramos que quando armazena padrões correlatos o pico se translada continuamente para algo em torno de  $p = 0.3$ , e cai rapidamente com o aumento da correlação;

Como perspectivas temos a intenção de investigar o papel do "morro" na distribuição de graus para as redes livre de escala utilizando medidas propostas por Dorogovtsev et al [35] para contornar o problema de tamanho finito. Além disso uma análise do comportamento em função de  $\gamma$ , uma vez que o único trabalho feito nessa direção [42] não deixa totalmente claro o papel do expoente no desempenho do modelo de Hopfield diluído em uma rede complexa.

A importante continuidade desse trabalho é a possibilidade de criarmos uma nova regra de aprendizado baseada na diluição da rede. Kürten [8] efetuou um estudo desse tipo e mostrou que é possível armazenar padrões correlatos e apresenta capacidade de carga superior ao modelo de Hopfield tradicional. Essa abordagem tem um grande apelo biológico, pois existem evidências experimentais que o desenvolvimento cerebral é feito com o rearranjar das sinapses [43]. Então, em continuidade a esse trabalho direcionaremos nossos esforços no intuito de desenvolver uma regra para rearranjar as sinapses e investigar a topologia resultante.

# Referências Bibliográficas

- [1] O.Sporns, G.Tononi, G.M.Eldelman. *Cerebral Cortex* **10**, 127–141 (2000).
- [2] D.J. Watts, S. H. Strogatz. *Nature* **393(6684)**, 440–442 (1998).
- [3] Aristóteles. *On Memory and Reminiscence*. e-book (350 AC).
- [4] J.J.Hopfield. *Proc.Natl.Acad.Sci.USA* **79**, 2554 (1982).
- [5] W. H. G. J. P. , Elly Nedivi1. *PLoS Biology* **4 (2)**, 271–280 (2006).
- [6] L.G. Morelli, G Abramson, M.N. Kuperman. *European Physical Journal B* **38**, 495–500 (2004).
- [7] P.M. McGraw, M. Menzinger. *Physical Review E* **68**, 047102 (2003).
- [8] K.E.Kürten. *J.Phys. I France* **2**, 615–624 (1992).
- [9] W. McCulloch, W. Pitts. *Bulletin of Mathematical Biophysics* **5**, 115–133 (1943).
- [10] M.L. Minsky, S.A. Papert. *Perceptrons– Expanded Edition*. MIT Press (1988).
- [11] G.Sperling. *Psychological Monographs* **74**, 1–29 (1960).
- [12] D.Hebb. *The organization of Behaviours*. Jonh Wiley (1949).
- [13] D. Amit. *Modeling Brain Function*. Cambridge University Press (1989).
- [14] A.L Barabási, R. Albert. *Rev. Mod. Phys.* **74**, 47 (2002).

- [15] M. Newman. *SIAM Review* **45(2)**, 167–256 (2003).
- [16] S.N. Dorogovtsev, J.F.F. Mendes. *Advances in Physics* **51**, 1079–1187 (2002).
- [17] S. J.F.F.Mendes. *Evolution of Networks - From Biological Nets to the Internet and WWW*. Oxford University Press (2003).
- [18] B. Bollobás. *Random Graphs*. Academic Press (1985).
- [19] P.Erdős, A.Rényi. *Publ. Math. Debrecen* **6**, 290–297 (1959).
- [20] R.Solomonoff, A.Rapoport. *Bull. Math. Biophys.* **13**, 107–117 (1951).
- [21] F. Chung, L. Lu. *Adv. Appl. Math.* **26**, 257 (2001).
- [22] E.Ising. *Zeitschr. f. Physik* **31**, 253–258 (1925).
- [23] S.Milgram. *Psychology Today* **2**, 60–67 (1967).
- [24] J.Travers, S.Milgram. *Sociometry* **32**, 425–443 (1969).
- [25] M. Argollo de Menezes ,A.L. Barabási. *Phys Rev Lett* **93(6)**, 068701 (2004).
- [26] M.Barthélémy,L.A.N.Amaral. *Phys. Rev. Lett.* **82**, 3180–3183 (1999).
- [27] A.Barrat, M.Weigt. *em <http://arxiv.org/abs/cond-mat/> página 9903323* (1999).
- [28] Argollo de Menezes M., C.F.Moukarzel, T.J.P.Penna. *Europhys. Lett.* **50**, 574 (2000).
- [29] M. Barthélémy,L.A.N.Amaral. *Phys. Rev. Lett.* **82**, 5180(E) (1999).
- [30] A.Barrat, M.Weigt . *Eur. Phys. J. B* **13**, 547 (2000).
- [31] D. S.Price. *Science* **149**, 510–515 (1965).
- [32] D. S.Price. *J.Amer.Soc.Inform.Sci.* **27**, 292–306 (1976).
- [33] A.L Barabási, R. Albert. *Science* **286**, 509 (1999).

- [34] H.A.Simon. *Biometrika* **42**, 425–440 (1955).
- [35] S NN Dorogovtsev, J F F Mendes, A N Samukhin. *Phys. Rev. E* **63**, 062101 (2001).
- [36] D Stauffer, A Aharony, L da Fontoura Costa, J Adler. *Eur. Phys. J. B* **32**, 385–399 (2003).
- [37] B.Bollobas , O.Riordan . *Combin.* **24**, 5–34 (2004).
- [38] Satoru Morita , Ken ichi Oshio , Yuko Osana , Yasuhiro Funabashi , Kotaro Oka , Kiyoshi Kawamura . *Physica A* **298**, 553–561 (2001).
- [39] *Numerical Recipes*. <http://library.lanl.gov/numerical/bookcpdf.html>.
- [40] J.W. Bohland, A.A. Minai. *Neurocomputing* **38-40**, 489 (2001).
- [41] I.Kanter, H.Sompolinsky. *Phys. Rev.* **A35**, 380 (1986).
- [42] J.J. Torres, M.A. Muñoz, J. Marro, P.L. Garrido. *Neurocomputing* **58-60**, 229–234 (2004).
- [43] J.P. Changeux, A. Danchin. *Nature* **264**, 705 (1976).

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)



[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)