



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

WAGNER RODRIGUES DE SENA

TESTE DE VALIDADE DE MÉTODOS DE MAXIMIZAÇÃO DE
ENTROPIA PARA CONSTRUÇÃO DE MODELOS COM CORRELAÇÃO
PAR-A-PAR

FORTALEZA

2017

WAGNER RODRIGUES DE SENA

TESTE DE VALIDADE DE MÉTODOS DE MAXIMIZAÇÃO DE ENTROPIA PARA
CONSTRUÇÃO DE MODELOS COM CORRELAÇÃO PAR-A-PAR

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Física. Área de Concentração: Física da Matéria Condensada.

Orientador: Prof. Dr. André Auto Moreira.

FORTALEZA
2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S477t Sena, Wagner Rodrigues de.
Teste de validade de métodos de maximização de entropia para construção de modelos com correlação par-a-par / Wagner Rodrigues de Sena. – 2017.
61 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Física, Fortaleza, 2017.
Orientação: Prof. Dr. André Auto Moreira.
1. Teoria da Informação. 2. Máxima Entropia. 3. Rede Neural. 4. Problema Inverso de Ising. 5. Máquina de Boltzmann. I. Título.

CDD 530

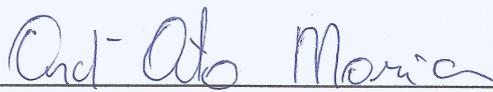
WAGNER RODRIGUES DE SENA

TESTE DE VALIDADE DE MÉTODOS DE MAXIMIZAÇÃO DE ENTROPIA PARA
CONSTRUÇÃO DE MODELOS COM CORRELAÇÃO PAR-A-PAR

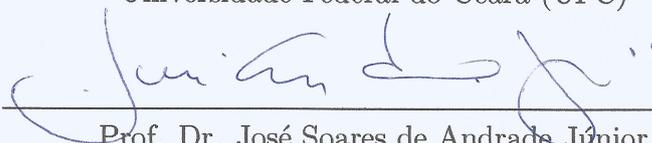
Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Física. Área de Concentração: Física da Matéria Condensada.

Aprovada em 20/02/2017.

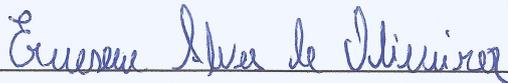
BANCA EXAMINADORA



Prof. Dr. André Auto Moreira (Orientador)
Universidade Federal do Ceará (UFC)



Prof. Dr. José Soares de Andrade Júnior
Universidade Federal do Ceará (UFC)



Dr. Erneson Alves de Oliveira
Universidade de Fortaleza (UNIFOR)

AGRADECIMENTOS

Agradeço principalmente aos meus pais por sempre me incentivarem a continuar estudando e a minha irmã que sempre esteve presente quando precisei.

Ao meu orientador, Prof. Dr. André Auto Moreira, pela paciência e competência em me orientar desde a graduação até este trabalho.

Aos professores do Departamento de Física da Universidade Federal do Ceará, por proporcionarem todo o aprendizado adquirido ao longo da minha formação.

Aos amigos que estão comigo desde a graduação, Victor Nocrato, Jonathan Sales, Daniel Linhares, Matheus Falcão, Pedro Henrique, Laura Barth, Ícaro Lavor e Rafael Farias, por sempre estarem presentes em todas as dificuldades e alegrias encontradas ao longo do caminho.

Agradeço aos meus amigos do laboratório de sistema complexos, Samuel Moraes, Emanuel Fontelles, Felipe Operti, Rilder Pires, Nailson Vasconcelos, Thiago Bento, Calebe Alves e Fernando Almeida pelas conversas e trocas de ideias.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (**CNPq**) pelo apoio financeiro.

RESUMO

No século XXI a humanidade produziu mais novos dados (informações) do que em toda sua história. Entender a natureza dos diversos sistemas que geram essa abundância de dados se tornou um dos grandes desafios desse século. Uma forma de analisar formalmente esses grandes bancos de dados é empregando a teoria da informação desenvolvida por Claude Shannon. Essa teoria permite, usando o princípio da máxima entropia, encontrar as distribuições de probabilidades que melhor descrevem os comportamentos coletivos desses sistemas. Nessa dissertação, discutimos a possibilidade de usar modelos tipo Ising para descrever observações de sistemas reais. Devido a suas limitações, empregar o modelo de Ising implica em supor que os elementos que constituem o sistema real só podem estar em dois estados, por exemplo ativo ou inativo. Além disso, o modelo de Ising da conta apenas de interações entre pares de elementos e desconsidera a possibilidade de interações entre grupos maiores de elementos. Como discutiremos, mesmo com essas limitações tal modelo pode descrever bem resultados observados em alguns sistemas naturais, como por exemplo redes de neurônios. Especificamente, discutiremos resultados de trabalhos anteriores que mostram que usando apenas as médias de atividade de cada neurônio e a correlação entre os mesmo, usando a teoria de Shannon, observa-se que os estados visitados pela rede seguem à distribuição de Ising. Para testar a aplicabilidade desse método em diversos sistemas geramos dados sintéticos, obtidos de modelos tipo Ising em três situações: ferromagnético, anti-ferro e vidro de spins (*spin glass*). Nós chamamos o sistema que gera os dados sintéticos de sistema subjacente. Usamos métodos de maximização de entropia para tentar construir sistemas modelos que consigam reproduzir as média e correlações observadas nos dados sintéticos. Dessa forma, verificamos em que situações nossos métodos conseguem de fato gerar um sistema modelo que reproduza o sistema subjacente que gerou os dados. Esses resultados podem estabelecer um limite de aplicabilidade para a técnica discutida.

Palavras-chave: Teoria da Informação. Máxima Entropia. Rede Neural. Problema Inverso de Ising. Máquina de Boltzmann.

ABSTRACT

In the 21st century humanity has produced more new data (information) than in all its history. Understanding the nature of the various systems that generate this abundance of data has become the great challenge of this century. One way to formally analyze these large databases is to use the information theory developed by Claude Shannon. This theory allows us, using the principle of maximum entropy, to find the distributions of probabilities that best describes the collective behavior of these systems. In this dissertation we discuss the possibility of using Ising models to describe observation of real systems. Due to its limitations, employing the Ising model implies that the elements that constitute the real system can only be in two states, for example active or inactive. In addition, the Ising model counts only interactions between pairs of elements and disregards the possibility of interactions between larger groups of elements. As we will discuss, even with these limitations such a model can well describe results observed in some natural systems, such as networks of neurons. Specifically, we discuss results from earlier work that show that using only the activity averages of each neuron and the correlation between them, using Shannon's theory, we observe that the states visited by the network follow the Ising distribution. In order to test the applicability of this method in several systems we generate synthetic data, obtained from Ising model in three systems: ferromagnetic, antiferromagnetic and spin glass. We call the system that generate the synthetic data as underlying system. We use methods of maximization of entropy to try to construct model systems that can reproduce the mean and correlations observed in the synthetic data. We thus verify in which situations our methods can actually generate a model system that reproduces the underlying system that generated the data. These results may establish a limit of applicability for the technique discussed.

Keywords: Information Theory. Principle of Maximum Entropy. Inverse Ising Problem. Boltzmann Machine.

LISTA DE FIGURAS

Figura 1 – Relação da teoria da informação com outras áreas.	11
Figura 2 – Estrutura do neurônio.	19
Figura 3 – Atividade neural	21
Figura 4 – Exemplos de redes aleatórias.	32
Figura 5 – Acoplamentos e campos.	32
Figura 6 – Exemplo de um estado.	33
Figura 7 – Exemplo de uma passo de Monte Carlo.	34
Figura 8 – Correlações e magnetizações.	36

LISTA DE GRÁFICOS

Gráfico 1 – Entropia de um processo de Bernoulli	15
Gráfico 2 – Potencial de ação.	20
Gráfico 3 – Probabilidade $P(K)$	23
Gráfico 4 – Teste para os valores de $\langle \sigma_i \sigma_j \sigma_k \rangle_{(c)}$	24
Gráfico 5 – Calor específico versus β	36
Gráfico 6 – Ferromagnético para $T > T^*$	41
Gráfico 7 – Ferromagnético para $T \approx T^*$	42
Gráfico 8 – Ferromagnético para $T < T^*$	43
Gráfico 9 – Antiferromagnético para $T > T^*$	45
Gráfico 10 – Antiferromagnético para $T \approx T^*$	46
Gráfico 11 – Antiferromagnético para $T < T^*$	47
Gráfico 12 – <i>Spin Glass</i> para $T > T^*$	50
Gráfico 13 – <i>Spin Glass</i> para $T \approx T^*$	51
Gráfico 14 – <i>Spin Glass</i> para $T < T^*$	52
Gráfico 15 – Calor específico versus temperatura T	53

SUMÁRIO

1	INTRODUÇÃO	10
2	INTRODUÇÃO À TEORIA DA INFORMAÇÃO	11
2.1	Entropia	12
2.2	Princípio da Máxima Entropia	15
2.3	Distância Kullback-Leibler	18
3	REDE NEURAL E MÁXIMA ENTROPIA	19
3.1	Modelando redes de neurônios	20
3.2	Problema inverso de Ising	24
3.3	Métodos aproximativos	26
3.3.1	Naive Mean Field (nMF)	26
3.3.2	Equações de TAP	28
4	RESULTADOS	31
4.1	Dados Sintéticos	31
4.2	Teste de validade	37
4.2.1	Ferromagnético	39
4.2.2	Antiferromagnético	44
4.2.3	<i>Spin Glass</i>	48
5	CONCLUSÃO E PERSPECTIVA	54
	REFERÊNCIAS	56

1 INTRODUÇÃO

No século XXI a humanidade produziu mais novos dados (informação) do que em toda sua história [1]. Entender a natureza dos diversos sistemas que geram essa abundância de dados se tornou um dos grandes desafios desse século [2]. O volume, a variedade, o valor, a veracidade e a velocidade com que essas informações são criadas, constituem os 5Vs que caracterizam o que é chamado de *Big Data* [3].

As fontes que geram esse gigantesco volume de dados estão presentes em diversas áreas: as imagens astronômicas geradas por telescópios [4], os bancos genéticos na biologia [5], os preços de ações sendo atualizados a cada dia, hora e até mesmo a cada minuto na economia [6,7], os dados de colisões de partículas realizadas por aceleradores de partículas [8], como o *Large Hadron Collider (LHC)*, os dados de atividade cerebral [9,10], as mais de 300 horas de vídeos que são enviadas para o YouTube a cada minuto e os mais de 500 milhões de tweets gerados por dia [11]. Pesquisas científicas já estão sendo realizadas para tentar entender a natureza dos diversos sistemas que geram essa abundância de dados. Como por exemplo, as pesquisas nas áreas de economia [12–14], genética [15,16], no estudo de interação proteína-proteína [17,18] e neurociência [19–22].

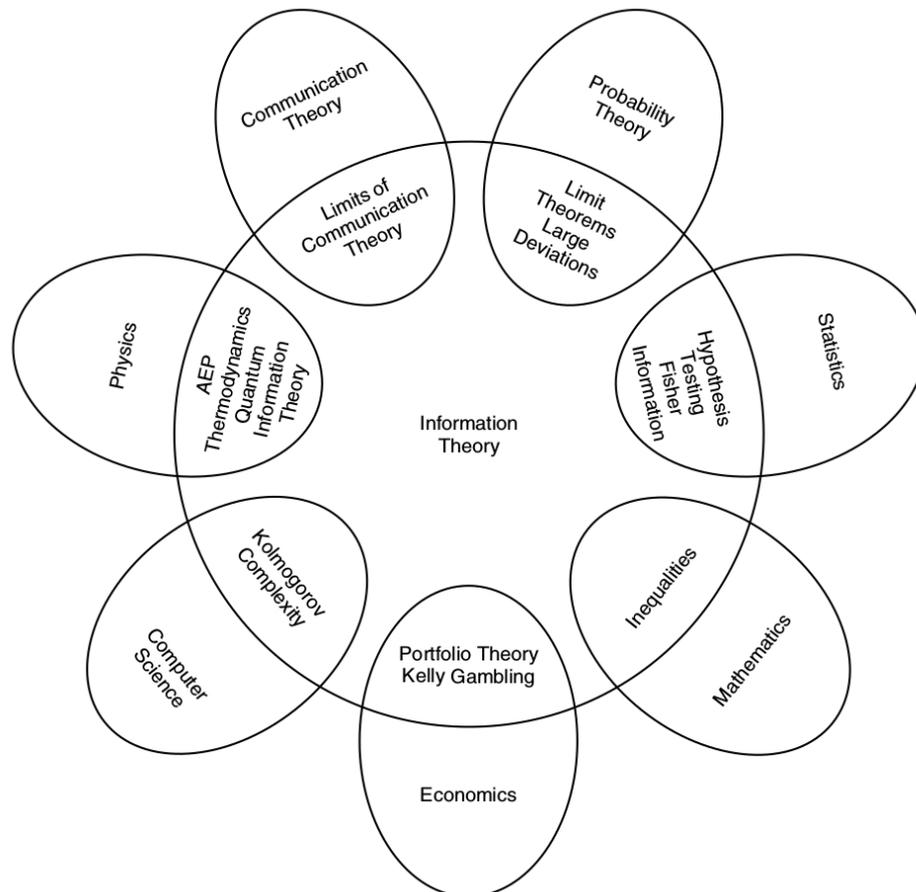
Nessa dissertação, analisaremos a eficiência de um método que captura o comportamento coletivo de um sistema que gera um banco de dados. Para tanto, no Capítulo 2 faremos uma rápida introdução sobre a teoria da informação para então enunciar o princípio da máxima entropia. No Capítulo seguinte, usando o princípio da máxima entropia, encontraremos uma distribuição de probabilidades que descreva o comportamento coletivo de uma rede de neurônios. Nesse mesmo Capítulo, desenvolveremos um método aproximativo para encontrar os parâmetros da distribuição encontrada. Para testar a eficiência desse método, explicaremos no Capítulo 4 como obter dados sintéticos a partir de um sistema subjacente. No Capítulo 5, testaremos a eficiência desse método em diferentes regimes. Por fim, no último Capítulo, apresentaremos as conclusões obtidas e algumas perspectivas.

Todos os códigos necessários para as simulações numéricas foram escritos pelo autor utilizando a linguagem de programação C++. Os gráficos e figuras, elaboradas pelo autor, foram geradas usando os programas gnuplot, gephi e inkscape.

2 INTRODUÇÃO À TEORIA DA INFORMAÇÃO

Para entender a natureza dos diversos sistemas que geram uma abundância de dados, foi desenvolvida uma teoria geral da informação. Claude Shannon é celebrado como o fundador dessa teoria em 1948, ao procurar a melhor forma de codificar uma informação que um emissor queira transmitir para um receptor [23]. Vale ressaltar que aspectos da teoria da informação já estavam presentes na teoria da mecânica estatística, desenvolvida por Boltzmann e Gibbs [24], na virada do século XX. A grande visão de Shannon foi tal que sua teoria não era apenas um ramo da teoria da comunicação, mas sim uma teoria mais geral, com implicações em muitas outras áreas, como: física estatística (termodinâmica) [25, 26], ciência da computação (Complexidade de Kolmogorov) [27], inferência estatística (navalha de Occam) [28] e economia (teoria do portfólio) [29]. A Figura 1 mostra a conexão entre a teoria da informação e outras áreas.

Figura 1 – Relação da teoria da informação com outras áreas.



Fonte: Figura retirada de Cover e Thomas: *Elements of information theory* [30]. Como mostra a figura, a teoria da informação faz interseção com diversos outros campos.

2.1 Entropia

Devemos iniciar nossa discussão dando um sentido matemático ao que é informação. Suponha que tenhamos um experimento que envolva a observação de uma variável aleatória discreta x que pode assumir um número finito de valores, $x \in \{x_1, \dots, x_n\}$, com probabilidades p_1, p_2, \dots, p_n , onde devemos admitir que todas as probabilidades p_i são maiores que zero e normalizadas, isto é, $p_i > 0$ e $\sum_{i=1}^n p_i = 1$. Tendo em vista esse experimento, vamos construir duas funções, h e H . A função h será definida no intervalo $(0, 1]$ e pode ser interpretada como a quantidade de informação associada a um evento de probabilidade p . Então, se o evento $\{x = x_i\}$ com probabilidade p_i é observado, a quantidade $h(p_i)$ é a informação ganha devido a esse evento. Já a função $H(p_1, p_2, \dots, p_n)$ será a informação média associada a todos os eventos possíveis, assim:

$$H(p_1, p_2, \dots, p_n) = \sum_{i=1}^n p_i h(p_i). \quad (2.1)$$

Antes da variável x ser revelada, o experimento possui uma incerteza sobre qual será o valor de x , logo podemos pensar que $h(p_i)$ é a incerteza removida do sistema ao se revelar o evento $x = x_i$ com probabilidade p_i e $H(p_1, p_2, \dots, p_n)$ é a incerteza média removida. A partir de agora vamos nos referir a $H(p_1, p_2, \dots, p_n) \equiv H(x)$ como sendo a incerteza de x .

Com essas definições iniciais, vamos impor quatro axiomas sobre a função incerteza, $H(x)$. Primeiro suponha que todos os valores de x são igualmente prováveis. Chamaremos de $f(n)$ a incerteza associada a n resultados igualmente prováveis, isto é,

$$f(n) = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = \sum_{i=1}^n p_i h(p_i) = h\left(\frac{1}{n}\right). \quad (2.2)$$

Como exemplo, $f(2)$ poderia ser a incerteza de se obter cara ou coroa ao jogar um moeda e $f(10^6)$ poderia ser a incerteza de escolher uma pessoa aleatoriamente em uma cidade com um milhão de habitantes. Esperamos que a incerteza de escolher uma pessoa aleatoriamente em uma cidade seja maior que obter cara ou coroa ao jogar uma moeda. Teremos assim o primeiro axioma:

Axioma 1 $f(n) = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$ deve ser uma função monotonicamente crescente de n . Matematicamente, teremos:

$$f(n) < f(m) \quad \text{para } n < m \quad (n, m = 1, 2, 3, \dots). \quad (2.3)$$

Considere agora um experimento com duas variáveis aleatórias independentes, x e y , x pode assumir os seguintes valores $\{x_1, x_2, \dots, x_n\}$ com iguais probabilidades e

y pode assumir os seguintes valores $\{y_1, y_2, \dots, y_l\}$ também com iguais probabilidades. O experimento de obter x e y possui no total nl resultados igualmente prováveis, cuja incerteza é $f(nl)$. Se o valor de x é revelado, a incerteza sobre y não deve ser alterada, pois consideramos as variáveis como sendo independentes. Então, a incerteza removida do sistema por revelar x deve ser apenas a incerteza de x , $f(n)$. Analogamente para a variável y . Com isso, o processo de se obter x primeiro remove $f(n)$ de incerteza e em seguida obter y remove mais $f(l)$ de incerteza, removendo no total $f(n) + f(l)$ de incerteza. Já o processo de revelar as duas variáveis simultaneamente remove $f(nl)$ de incerteza. A ordem de extração de incerteza não importa, primeiro uma e depois a outra ou as duas simultaneamente, ambas devem remover a mesma quantidade de incerteza do sistema. O segundo axioma diz que:

Axioma 2 *A função f deve ser aditiva:*

$$f(nl) = f(n) + f(l). \quad (2.4)$$

Agora vamos considerar um caso mais geral. Vamos remover a restrição de probabilidades igualmente prováveis e dividir a variável aleatória x em dois grupos, A e B , onde $\{x_1, x_2, \dots, x_r\} \in A$ e $\{x_{r+1}, x_{r+2}, \dots, x_n\} \in B$. Para sortear a variável x vamos realizar o seguinte experimento: primeiro selecionamos um dos dois grupos e depois sorteamos x dentro desse grupo. As probabilidades de selecionar o grupo A ou o grupo B são dadas por:

$$p(A) = p_1 + p_2 + \dots + p_r, \quad (2.5)$$

$$p(B) = p_{r+1} + p_{r+2} + \dots + p_n. \quad (2.6)$$

Já a probabilidade de selecionar x , dado que o grupo A ou B foi escolhido é:

$$p(x_i|A) = \frac{p_i}{p_1 + p_2 + \dots + p_r} = \frac{p_i}{p(A)}, \quad (i = 1, 2, \dots, r), \quad (2.7)$$

ou

$$p(x_i|B) = \frac{p_i}{p_{r+1} + p_{r+2} + \dots + p_n} = \frac{p_i}{p(B)}, \quad (i = r + 1, r + 2, \dots, n). \quad (2.8)$$

Antes de realizar o experimento, a incerteza associada a variável x é $H(p_1, \dots, p_n)$.

Quando realizamos a primeira parte do experimento, sorteando o grupo A ou B , a incerteza removida é $H(p(A), p(B))$. Se o grupo A é escolhido, a incerteza restante com relação a variável x é

$$H\left(\frac{p_1}{p(A)}, \frac{p_2}{p(A)}, \dots, \frac{p_r}{p(A)}\right), \quad (2.9)$$

se o grupo B é escolhido, a incerteza restante é:

$$H\left(\frac{p_{r+1}}{p(B)}, \frac{p_{r+2}}{p(B)}, \dots, \frac{p_n}{p(B)}\right). \quad (2.10)$$

A incerteza média restante depois que o grupo é escolhido é

$$p(A)H\left(\frac{p_1}{p(A)}, \frac{p_2}{p(A)}, \dots, \frac{p_r}{p(A)}\right) + p(B)H\left(\frac{p_{r+1}}{p(B)}, \frac{p_{r+2}}{p(B)}, \dots, \frac{p_n}{p(B)}\right). \quad (2.11)$$

Com isso, escreveremos o terceiro axioma:

Axioma 3 *A incerteza da variável x será a incerteza na escolha de qual grupo será sorteado x mais a incerteza média de sortear x dentro de cada grupo, ou seja:*

$$\begin{aligned} H(p_1, \dots, p_n) &= H(p(A), p(B)) + p(A)H\left(\frac{p_1}{p(A)}, \frac{p_2}{p(A)}, \dots, \frac{p_r}{p(A)}\right) + \\ &+ p(B)H\left(\frac{p_{r+1}}{p(B)}, \frac{p_{r+2}}{p(B)}, \dots, \frac{p_n}{p(B)}\right). \end{aligned} \quad (2.12)$$

E por último, teremos o quarto axioma:

Axioma 4 *$H(p, 1-p)$ é uma função contínua em p .*

Com esses quatro axiomas pode-se provar o seguinte teorema [23, 31]:

Teorema 1 *A única função que satisfaz aos quatro axiomas dados é:*

$$H(p_1, \dots, p_n) = -c \sum_{i=1}^n p_i \ln p_i \quad (2.13)$$

onde c é uma constante arbitrária positiva e a base do logaritmo é qualquer uma maior do que 1.

Quando Shannon chegou a Equação 2.13 ele percebeu que sua equação possuía a mesma forma da entropia de Boltzmann:

$$S = -k_B \sum_i p_i \ln p_i \quad (2.14)$$

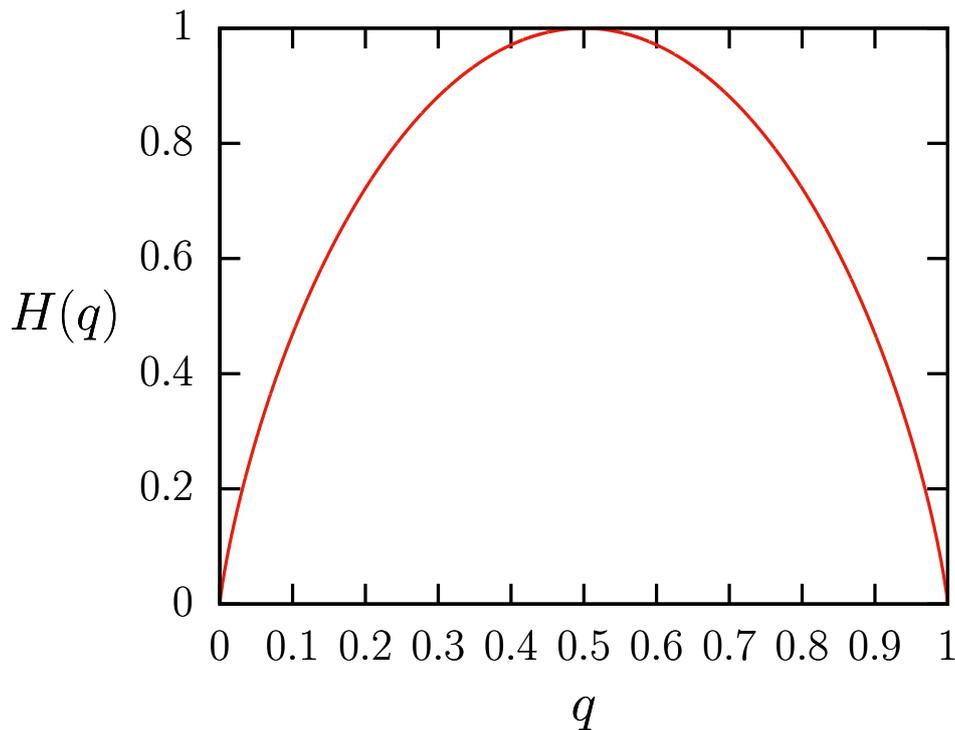
e, por isso, à chamou de entropia. Atualmente é conhecida como entropia de Shannon e, a partir de agora, vamos nos referir a ela por esse nome.

Para o caso em que $c = 1$ e a base do logaritmo é 2, a entropia de Shannon é expressa em *bits*, uma abreviação para *binary digits*. Como exemplo, considere um processo de Bernoulli [32], onde uma variável aleatória x pode assumir apenas dois valores, $x \in \{0, 1\}$, com probabilidades $p(0) = q$ e $p(1) = 1 - q$. A entropia associada a esse evento é:

$$H(q, 1 - q) = -q \log_2 q - (1 - q) \log_2(1 - q). \quad (2.15)$$

O Gráfico 1 mostra o comportamento da entropia, Equação 2.15, em função de q desse processo. Observe que quando $q = 0$ a incerteza é nula, isto é, não há nenhuma informação para se retirar do experimento, já que a probabilidade $p(1) = 1$ e assim sabemos qual será o resultado da variável x . O mesmo acontece quando $q = 1$. Mas quando $q = 0.5$, a entropia é máxima, mostrando a nossa total incerteza sobre o valor de x .

Gráfico 1 – Entropia de um processo de Bernoulli



Fonte: Elaborado pelo autor. Comportamento da entropia de uma processo de Bernoulli em função do parâmetro q .

2.2 Princípio da Máxima Entropia

Suponha, novamente, a variável aleatória x que pode assumir os valores $\{x_1, x_2, \dots, x_n\}$ com probabilidades $\{p_1, p_2, \dots, p_n\}$, mas dessa vez não conhecemos tais probabilidades. Apenas conhecemos o valor médio de alguma função $f(x)$:

$$\langle f(x) \rangle = \sum_{i=1}^n p_i f(x_i) \quad (2.16)$$

e com base nessa informação, queremos encontrar a distribuição de probabilidades, $\{p_i\}$. A princípio, o problema aparenta ser insolúvel, pois precisaríamos de mais $(n-2)$ equações, fora a Equação 2.16 e a condição de normalização

$$\sum_i^n p_i = 1, \quad (2.17)$$

para determinar as probabilidades $\{p_i\}$.

O que podemos fazer é estimar a distribuição de probabilidades, $\{p_i\}$, que descreve a variável x . Para isso devemos usar o Princípio da Máxima Entropia [25], que diz que quando estimamos a distribuição de probabilidades, devemos selecionar aquela distribuição que nos deixa com a maior incerteza restante, máxima entropia e esteja de acordo com qualquer conhecimento prévio. Dessa forma, não precisamos impor nenhuma condição sobre $\{p_i\}$.

Para maximizar a entropia 2.13 sujeita as restrições 2.16 e 2.17, podemos fazer uso dos multiplicadores de Lagrange $\{\lambda_0, \lambda_1\}$, de tal modo que a lagrangiana fica:

$$L = - \sum_{i=1}^n p_i \ln p_i + \lambda_0 \left(\sum_i^n p_i - 1 \right) + \lambda_1 \left(\sum_{i=1}^n p_i f(x_i) - \langle f(x) \rangle \right), \quad (2.18)$$

onde fizemos $c = 1$. A equação de Lagrange nos fornece:

$$\frac{\partial L}{\partial p_i} = 0 \Rightarrow \quad (2.19)$$

$$-(\ln p_i + 1) + \lambda_0 + \lambda_1 f(x_i) = 0 \Rightarrow \quad (2.20)$$

$$p_i = e^{-1+\lambda_0+\lambda_1 f(x_i)} = p_0 e^{\lambda_1 f(x_i)}, \quad (2.21)$$

onde p_0 é determinado pela restrição 2.17:

$$\sum_i p_0 e^{\lambda_1 f(x_i)} = 1 \Rightarrow \quad (2.22)$$

$$p_0 = \frac{1}{Z(\lambda_1)}, \quad (2.23)$$

onde

$$Z(\lambda_1) = \sum_i e^{\lambda_1 f(x_i)} \quad (2.24)$$

é chamada de função de partição. Já λ_1 é determinado mediante a restrição 2.16, que, em termos da função de partição $Z(\lambda_1)$, pode ser reescrita como:

$$\langle f(x) \rangle = \frac{\partial}{\partial \lambda_1} \ln Z(\lambda_1). \quad (2.25)$$

Podemos generalizar para qualquer número m ($m < n$) de funções $f(x)$. Se é conhecida a média

$$\langle f_j(x) \rangle = \sum_i p_i f_j(x_i), \quad (j = 1, 2, \dots, m), \quad (2.26)$$

então a distribuição de probabilidades que maximiza a entropia é dada por

$$p_i = \frac{1}{Z(\lambda_1, \lambda_2, \dots, \lambda_m)} e^{\lambda_1 f_1(x_i) + \lambda_2 f_2(x_i) + \dots + \lambda_m f_m(x_i)}, \quad (2.27)$$

com a função de partição da seguinte forma:

$$Z(\lambda_1, \lambda_2, \dots, \lambda_m) = \sum_i e^{\lambda_1 f_1(x_i) + \lambda_2 f_2(x_i) + \dots + \lambda_m f_m(x_i)}. \quad (2.28)$$

As constantes $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ são determinadas a partir de

$$\langle f_j(x) \rangle = \frac{\partial}{\partial \lambda_j} \ln Z(\lambda_1, \lambda_2, \dots, \lambda_m), \quad (j = 1, 2, \dots, m). \quad (2.29)$$

Usando a distribuição da Equação 2.27, a entropia, Equação 2.13, maximizada fica:

$$H_{max}(x) = \ln Z - \lambda_1 \langle f_1(x) \rangle - \lambda_2 \langle f_2(x) \rangle - \dots - \lambda_m \langle f_m(x) \rangle. \quad (2.30)$$

Para mostrar a relação da teoria da informação com a mecânica estatística, considere que tenhamos uma sistema físico, onde conhecemos apenas a sua energia interna (energia média de cada estado i)

$$U = \sum_i p_i E_i. \quad (2.31)$$

A distribuição que maximiza a entropia é:

$$p_i = \frac{1}{Z(\lambda_1)} e^{\lambda_1 E_i}, \quad (2.32)$$

onde

$$Z(\lambda_1) = \sum_i e^{\lambda_1 E_i}. \quad (2.33)$$

A constante λ_1 pode ser determinada mediante a seguinte relação da termodinâmica

$$\frac{1}{k_B T} = \frac{\partial H}{\partial U}, \quad (2.34)$$

onde k_B é a constante de Boltzmann. Assim, identificamos que

$$\lambda_1 = -\frac{1}{k_B T}, \quad (2.35)$$

e concluimos que a distribuição é

$$p_i = \frac{1}{Z(T)} e^{-\frac{E_i}{k_B T}} \quad (2.36)$$

$$Z(T) = \sum_i e^{-\frac{E_i}{k_B T}}. \quad (2.37)$$

Na mecânica estatística a Equação 2.37 é chamada função partição e é associada a energia livre, $F = -k_B T \ln Z$, enquanto que a Equação 2.36 é a distribuição do ensemble canônico.

2.3 Distância Kullback-Leibler

Outro conceito que surgiu devido à entropia de Shannon e será importante para o nosso objetivo, é saber quão distante uma distribuição de probabilidades está de outra distribuição. Isso pode ser medido usando a Distância Kullback-Leibler [33].

Para duas distribuições $p(x)$ e $q(x)$ sobre o mesmo espaço $x \in X$, a Distância Kullback-Leibler é definida como [30, 34]:

$$D_{KL}(q||p) = \sum_{x \in X} q(x) \ln \frac{q(x)}{p(x)}. \quad (2.38)$$

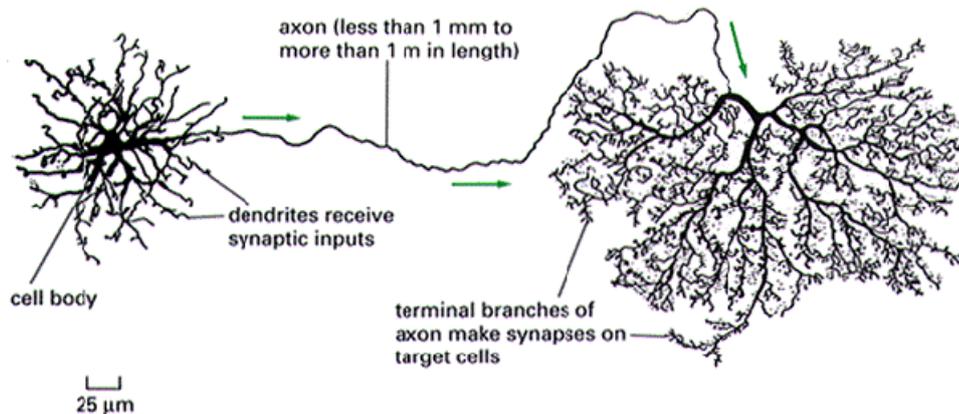
Apesar do nome “distância”, $D_{KL}(q||p)$ não é realmente uma distância métrica entre duas distribuições, pois, pela definição da Equação 2.38, vemos que $D_{KL}(q||p)$ não é simétrica, $D_{KL}(q||p) \neq D_{KL}(p||q)$, e não satisfaz a desigualdade triangular. Vemos que $D_{KL}(q||p) \geq 0$, onde há igualdade se, e somente se, $p(x) = q(x)$. No lugar de pensar em uma distância, podemos pensar em $D_{KL}(q||p)$ como a quantidade de informação que perdemos quando usamos $p(x)$ como aproximação de $q(x)$ [35].

3 REDE NEURAL E MÁXIMA ENTROPIA

O cérebro é um dos objetos mais complexo conhecido e que ainda permanece um mistério para humanidade. Entender o seu funcionamento é um dos principais desafios científicos do século XXI.

O cérebro humano é composto por aproximadamente 10^{11} células individuais [36,37], chamadas neurônios, e que são conectados uns aos outros. Um neurônio é por si só um objeto complexo, apesar que seu comportamento possa ser descrito de uma forma simples. Cada neurônio é formado por três partes distintas: o corpo celular, os dendritos e os axônios, como mostra a Figura 2. Os dendritos e os axônios são conectados ao corpo celular e são responsáveis por receber e enviar sinais elétricos para outros neurônios, respectivamente.

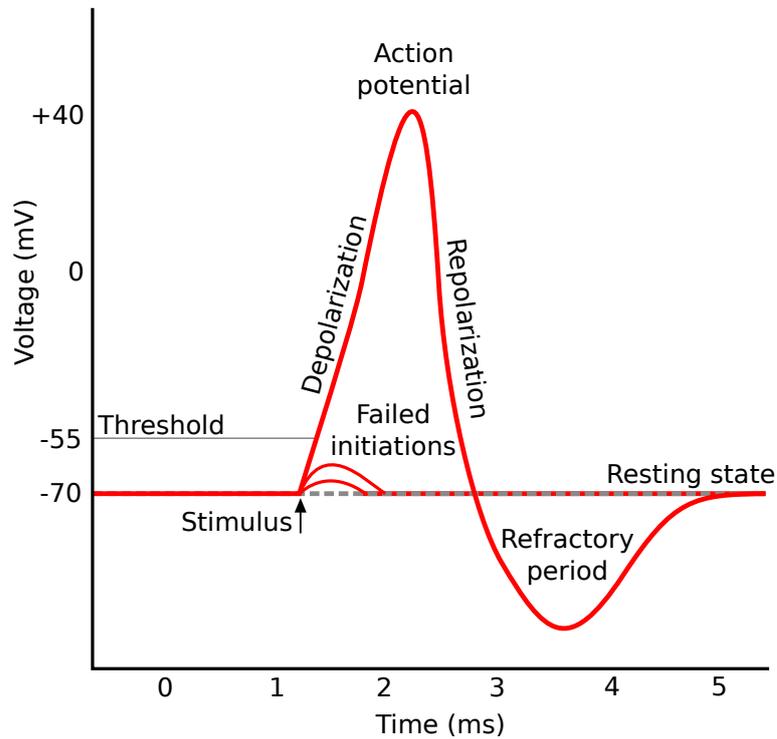
Figura 2 – Estrutura do neurônio.



Fonte: Figura retirada de [38]. Os neurônios são compostos pelo corpo celular, pelos dendritos e pelos axônios. O comprimento do axônio pode atingir, em alguns casos, tamanhos muito superiores a qualquer outra parte do neurônio.

Quando o neurônio não está recebendo sinal de outros neurônios a diferença de potencial na membrana é em torno de -70 mV . Ao receber um sinal que atinge um certo potencial limite (aproximadamente -55 mV) os canais de íons da membrana abrem e a diferença de potencial sobe para $+40\text{ mV}$ por um pequeno período de tempo, da ordem de 1 ms , e logo em seguida cai rapidamente de volta ao estado repouso, como mostra o Gráfico 2. Esse processo é chamado *firing* ou *spiking* e o pico gerado é chamado de potencial de ação.

Gráfico 2 – Potencial de ação.



Fonte: Figura retirada de [39]. Diferença de potencial na membrana de um neurônio ao receber um sinal.

3.1 Modelando redes de neurônios

Discutiremos os resultados obtidos por Tkacik *et al.* em [20] ao estudar um conjunto de neurônios do córtex visual de uma salamandra.

Devido ao comportamento de cada neurônio descrito acima, pode-se dizer que cada neurônio i , em um pequeno intervalo de tempo, está em atividade ($\sigma_i = +1$) ou não ($\sigma_i = -1$) [40]. Assim, o estado da rede neural pode ser descrito pelo conjunto de variáveis binárias, $\sigma = \{\sigma_i\}$. Realizando observações em diferentes momentos, teremos um conjunto de estados $\{\sigma\}$.

$$\sigma^{(1)} = \{+1, -1, -1, \dots, +1, +1, -1\}, \quad (3.1)$$

$$\sigma^{(2)} = \{-1, -1, +1, \dots, +1, -1, -1\}, \quad (3.2)$$

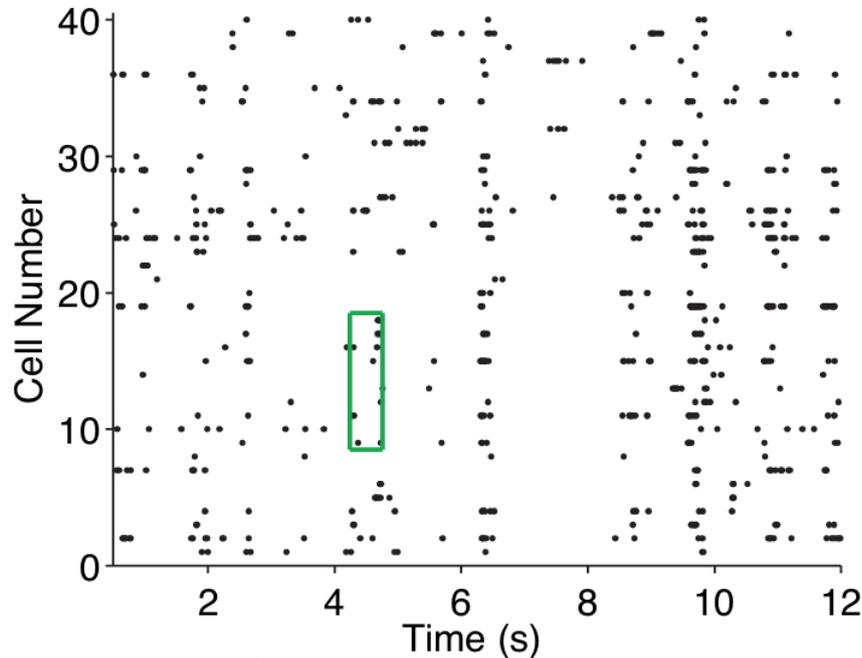
$$\vdots$$

$$\sigma^{(M)} = \{-1, +1, -1, \dots, +1, +1, +1\}. \quad (3.3)$$

A Figura 3 mostra a atividade de 40 neurônios ao longo do tempo, onde os pontos representam neurônios em atividade.

Nosso objetivo é encontrar a distribuição de probabilidades $P_{exp}(\sigma)$ que repro-

Figura 3 – Atividade neural



Fonte: Figura retirada de [19]. Atividade de 40 neurônios obtidas experimentalmente em diferentes momentos.

duz os estados observados experimentalmente. Com as ferramentas obtidas no capítulo anterior podemos estimar uma distribuição que esteja de acordo com as informações que possuímos e que seja o mais simples possível. Do conjunto de M estados obtidos experimentalmente, podemos calcular a atividade média de cada neurônio, $\langle \sigma_i \rangle$, e a correlação entre os mesmos, $C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$, onde

$$\langle \sigma_i \rangle = \sum_{\{\sigma\}} \sigma_i P_{exp}(\sigma) = \frac{1}{M} \sum_{k=1}^M \sigma_i^{(k)} \quad (3.4)$$

e

$$\langle \sigma_i \sigma_j \rangle = \sum_{\{\sigma\}} \sigma_i \sigma_j P_{exp}(\sigma) = \frac{1}{M} \sum_{k=1}^M \sigma_i^{(k)} \sigma_j^{(k)}. \quad (3.5)$$

Observe que C_{ij} é simétrico, $C_{ij} = C_{ji}$, pois $\langle \sigma_i \sigma_j \rangle = \langle \sigma_j \sigma_i \rangle$. Para um sistema de tamanho N , temos no total N médias $\langle \sigma_i \rangle$ e $N(N-1)/2$ correlações C_{ij} distintas.

Usando apenas os dados das atividade 3.4 e correlações 3.5, o princípio da máxima entropia implica que a distribuição de estados é da forma:

$$P(\sigma) = \frac{1}{Z} e^{\sum_{i < j} \alpha_{ij} \sigma_i \sigma_j + \sum_i \lambda_i \sigma_i}, \quad (3.6)$$

$$Z = \sum_{\{\sigma\}} e^{\sum_{i < j} \alpha_{ij} \sigma_i \sigma_j + \sum_i \lambda_i \sigma_i}, \quad (3.7)$$

onde o conjunto $\{\alpha_{ij}, \lambda_i\}$ são os multiplicadores de Lagrange. Vemos que a distribuição obtida é igual à do modelo de Ising [41–43] para spins que interagem par-a-par sob a influência de um campo magnético com energia:

$$H_{Ising}(\sigma) = - \sum_{i < j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (3.8)$$

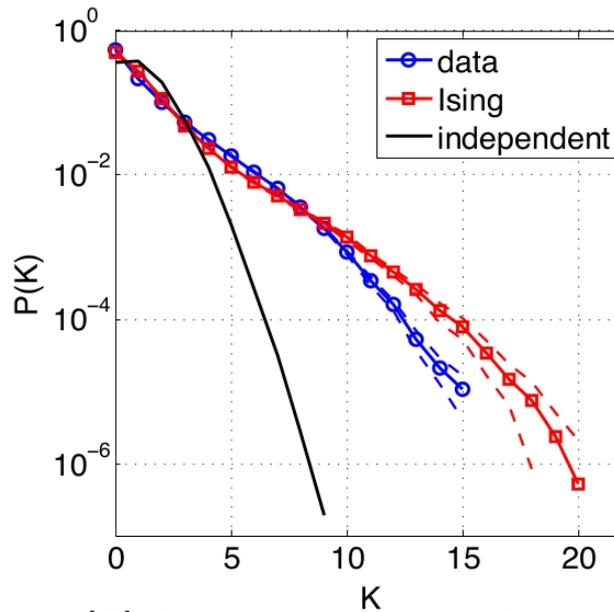
onde h_i é o campo magnético atuando no spin σ_i e J_{ij} é a constante de acoplamento entre os spins σ_i e σ_j . No equilíbrio, a distribuição de Ising em uma temperatura T é:

$$P_{Ising}(\sigma) = \frac{1}{Z_{Ising}} e^{\beta \left(\sum_{i < j} J_{ij} \sigma_i \sigma_j + \sum_{i=1}^N h_i \sigma_i \right)}, \quad (3.9)$$

$$Z_{Ising} = \sum_{\{\sigma\}} e^{\beta \left(\sum_{i < j} J_{ij} \sigma_i \sigma_j + \sum_{i=1}^N h_i \sigma_i \right)}, \quad (3.10)$$

onde β é o inverso da temperatura, $\beta = (k_B T)^{-1}$. Então, a distribuição encontrada, Equação 3.6, é o modelo de Ising para $\beta = 1$. Essa não é a primeira vez que o modelo de Ising é usado para modelar uma rede neural [44,45], mas em tais situações a distribuição de Ising era usada apenas por hipótese sobre o comportamento da rede. Já pelo princípio da máxima entropia a distribuição de Ising surge como a distribuição de menor estrutura que é consistente com as médias de atividade e correlações observadas. Isto é, a distribuição de Ising não é usado como uma analogia, mas sim um mapeamento.

Poderíamos ter calculado uma infinidade de médias a partir dos M estados obtidos experimentalmente, mas queremos usar o menor número de médias possível que consiga capturar o comportamento coletivo da rede neural. Para verificar se as médias escolhidas são suficientes para descrever a rede, podemos calcular outras quantidades a partir da distribuição encontrada e compará-las com aquelas calculadas a partir dos dados experimentais. Tkacik *et al.* mostraram em [20] que a distribuição encontrada pelo princípio da máxima entropia com apenas as médias $\langle \sigma_i \rangle$, conhecida como modelo independente, falha drasticamente em calcular a probabilidade $P(K)$ de K neurônios estarem em atividade ao mesmo tempo. Enquanto que a distribuição de Ising consegue reproduzir o comportamento de $P(K)$. O Gráfico 3 mostra $P(K)$ calculada pelos dados experimentais e pelos modelos independente e de Ising.

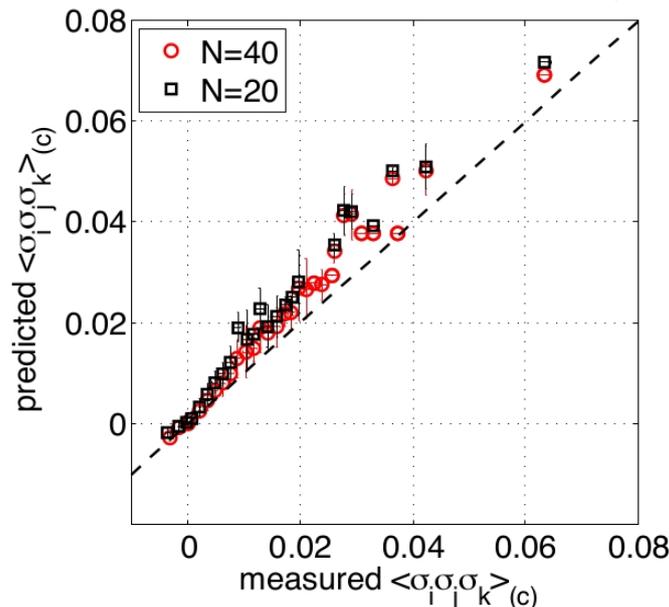
Gráfico 3 – Probabilidade $P(K)$.

Fonte: Gráfico retirado de [20]. Probabilidade de observar K neurônios em atividade ao mesmo tempo calculada pelos dados experimentais (azul), modelo independente (preto) e modelo de Ising (vermelho). Mostrando que devemos procurar modelos com maior estrutura que o independente.

Outro teste considerado por Tkacik *et al.* em [20] foi calcular a correlação entre tripletos de neurônios,

$$\langle \sigma_i \sigma_j \sigma_k \rangle = \langle (\sigma_i - \langle \sigma_i \rangle)(\sigma_j - \langle \sigma_j \rangle)(\sigma_k - \langle \sigma_k \rangle) \rangle, \quad (3.11)$$

onde o modelo independente prevê que todos os $N(N-1)(N-2)/(3!)$ tripletos distintos são iguais a zero, enquanto que a distribuição de Ising consegue determinar esses tripletos com um erro médio de aproximadamente 7% para $N = 40$ [20]. Se observamos que para uma rede com $N = 40$, temos $40(39)(38)/(3!) \approx 10^4$ tripletos distintos e a distribuição 3.9 consegue determiná-los com apenas 820 médias (40 médias $\langle \sigma_i \rangle$ e 780 correlações C_{ij}), então 7% é um erro aceitável. O gráfico 4 compara os valores de $\langle \sigma_i \sigma_j \sigma_k \rangle_{(c)}$ calculados pela distribuição de Ising e pelos dados experimentais.

Gráfico 4 – Teste para os valores de $\langle \sigma_i \sigma_j \sigma_k \rangle_{(c)}$.

Fonte: Gráfico retirado de [20]. Comparação dos valores $\langle \sigma_i \sigma_j \sigma_k \rangle_{(c)}$ previstos pelo modelo de Ising e aqueles obtidos dos dados experimentais para $N = 20$ (vermelho) e $N = 40$ (preto). Vemos que o modelo de Ising é suficiente para descrever as interações de tripletos.

3.2 Problema inverso de Ising

Uma vez que encontramos a distribuição de menor estrutura que está de acordo com as médias $\langle \sigma_i \rangle$ e correlações C_{ij} , e vimos que tal distribuição é semelhante à distribuição do modelo de Ising quando $\beta = 1$, resta agora determinar os valores dos acoplamentos $\mathbf{J} = \{J_{ij}\}$ e dos campos $\mathbf{h} = \{h_i\}$. Tal processo é conhecido como problema inverso de Ising ou *Boltzmann Machine Learning*, como é chamado na ciência da computação [46,47].

Podemos resolver o conjunto de Equações 2.29 para \mathbf{J} e \mathbf{h}

$$\langle \sigma_i \rangle = \frac{\partial}{\partial h_i} \ln Z(\mathbf{J}, \mathbf{h}), \quad (3.12)$$

$$\langle \sigma_i \sigma_j \rangle = \frac{\partial}{\partial J_{ij}} \ln Z(\mathbf{J}, \mathbf{h}), \quad (3.13)$$

mas tais equações tornam-se inviáveis quando N é grande, pois o número de estados que devem ser somados na função de partição $Z(\mathbf{J}, \mathbf{h})$ cresce com 2^N . Devemos, então, usar método aproximados. Observe que queremos usar a distribuição de Ising, $P_{ising}(\sigma)$, para descrever a distribuição observada experimentalmente, $P_{exp}(\sigma)$. Assim, queremos encontrar os valores de \mathbf{J} e \mathbf{h} que minimizem a informação perdida por usar $P_{ising}(\sigma)$, ou seja, queremos \mathbf{J} e \mathbf{h} tais que minimizem a distância Kullback-Leibler definida pela Equação 2.38

$$D_{KL}(P_{exp}||P_{ising}) = \sum_{\{\sigma\}} P_{exp}(\sigma) \ln \frac{P_{exp}(\sigma)}{P_{ising}(\sigma)}. \quad (3.14)$$

Diferenciando $D_{KL}(P_{exp}||P_{ising})$ com relação a J_{ij}

$$\begin{aligned} \frac{\partial D_{KL}(P_{exp}||P_{ising})}{\partial J_{ij}} &= \frac{\partial}{\partial J_{ij}} \sum_{\{\sigma\}} P_{exp}(\sigma) [\ln P_{exp}(\sigma) - \ln P_{ising}(\sigma)] \\ &= - \sum_{\{\sigma\}} P_{exp}(\sigma) \left[\frac{\partial}{\partial J_{ij}} \ln P_{ising}(\sigma) \right] \\ &= - \sum_{\{\sigma\}} P_{exp}(\sigma) \left(\sigma_i \sigma_j - \frac{1}{Z} \sum_{\{\sigma\}} \sigma_i \sigma_j e^{\sum_{i<j} J_{ij} \sigma_i \sigma_j + \sum_{i=1}^N h_i \sigma_i} \right) \\ &= - (\langle \sigma_i \sigma_j \rangle_{exp} - \langle \sigma_i \sigma_j \rangle_{Ising}), \end{aligned} \quad (3.15)$$

onde $\langle \sigma_i \sigma_j \rangle_{exp}$ significa média sobre os dados experimentais e $\langle \sigma_i \sigma_j \rangle_{Ising}$ média com respeito a distribuição de Ising. A Equação 3.15 leva à seguinte regra de atualização [46]:

$$\Delta J_{ij}(t+1) = \epsilon [\langle \sigma_i \sigma_j \rangle_{exp} - \langle \sigma_i \sigma_j \rangle_{Ising}(t)], \quad (3.16)$$

$$J_{ij}(t+1) = J_{ij}(t) + \Delta J_{ij}(t+1), \quad (3.17)$$

onde ϵ define a taxa de atualização do processo. Analogamente, encontramos a seguinte regra de atualização para os campos

$$\Delta h_i(t+1) = \epsilon [\langle \sigma_i \rangle_{exp} - \langle \sigma_i \rangle_{Ising}(t)], \quad (3.18)$$

$$h_i(t+1) = h_i(t) + \Delta h_i(t+1). \quad (3.19)$$

O que tal regra diz é: dados valores iniciais para \mathbf{J} e \mathbf{h} ; calculamos as médias $\langle \sigma_i \rangle_{Ising}$ e $\langle \sigma_i \sigma_j \rangle_{Ising}$ geradas por esses valores; usamos tais médias para atualizar os novos acoplamentos e campos; usamos esses novos valores para obter novas médias e atualizar novamente os \mathbf{J} e \mathbf{h} ; repetimos esse processo até obtermos um estado auto consistente onde as médias inferidas pelo modelo de Ising coincidam com aquelas obtidas experimentalmente. Pode-se mostrar que, se não houver spins escondidos, D_{KL} é uma função convexa dos acoplamentos e campos [46]. Isso garante que o método descrito acima irá atingir o mínimo global.

3.3 Métodos aproximativos

Existem na literatura alguns métodos aproximativos para inferir os acoplamentos e campos no problema inverso de Ising [48–56]. Usaremos dois desses métodos para encontrar \mathbf{J} e \mathbf{h} , a fim de utilizar esses valores como condição inicial para o processo de atualização descrito na seção anterior. Os dois métodos escolhidos foram o *Naive Mean Field* (**nMF**) [57] e as Equações de *Thouless, Anderson e Palmer* (**TAP**) [58].

3.3.1 Naive Mean Field (nMF)

Na aproximação **nMF** consideramos que cada spin é independente de todos os outros, isto é, todos os acoplamentos \mathbf{J} são nulos e cada spins sente um campo $\mathbf{H} = \{H_i\}$ efetivo composto de dois termos

$$H_i = \bar{h}_i + h_i, \quad (3.20)$$

onde h_i é o campo real do modelo de Ising e \bar{h}_i é a contribuição dos demais spins. A distribuição de probabilidades para **nMF** é

$$P_{nMF}(\sigma) = \prod_i \frac{e^{\beta H_i \sigma_i}}{2 \cosh(\beta H_i)}. \quad (3.21)$$

A principal vantagem de ter todos os spins desacoplados é que as magnetizações são calculadas facilmente:

$$\begin{aligned} m_i \equiv \langle \sigma_i \rangle_{nMF} &= \sum_{\{\sigma\}} \sigma_i \prod_k \frac{e^{\beta H_k \sigma_k}}{2 \cosh(\beta H_k)} \\ &= \sum_{\sigma_i} \sigma_i \frac{e^{\beta H_i \sigma_i}}{2 \cosh(\beta H_i)} \\ &= \frac{e^{\beta H_i \sigma_i} - e^{-\beta H_i \sigma_i}}{2 \cosh(\beta H_i)} \\ &= \tanh(\beta H_i). \end{aligned} \quad (3.22)$$

O que devemos fazer agora é procurar os valores dos \bar{h}_i 's que reduzam a informação perdida por usar $P_{nMF}(\sigma)$ como aproximação de $P_{Ising}(\sigma)$, ou seja, vamos novamente minimizar a distância Kullback-Leibler:

$$\begin{aligned} D_{KL}(P_{nMF} || P_{Ising}) &= \sum_{\{\sigma\}} P_{nMF}(\sigma) \ln \frac{P_{nMF}(\sigma)}{P_{Ising}(\sigma)} \\ &= \sum_{\{\sigma\}} P_{nMF}(\sigma) [\ln P_{nMF}(\sigma) - \ln P_{Ising}(\sigma)] \end{aligned}$$

$$\begin{aligned}
&= \sum_{\{\sigma\}} P_{nMF}(\sigma) \left\{ \beta \sum_i H_i \sigma_i - \sum_i \ln [2 \cosh(\beta H_i)] - \beta \sum_{i<j} J_{ij} \sigma_i \sigma_j - \right. \\
&\quad \left. - \beta \sum_i h_i \sigma_i + \ln Z \right\} \\
&= \beta \left(\sum_i H_i m_i - \sum_i h_i m_i - \sum_{i<j} J_{ij} m_i m_j \right) - \sum_i \ln [2 \cosh(\beta H_i)] + \\
&\quad + \ln Z \\
&= \beta \left(\sum_i \bar{h}_i m_i - \sum_{i<j} J_{ij} m_i m_j \right) - \sum_i \ln \{2 \cosh[\beta(\bar{h}_i + h_i)]\} + \\
&\quad + \ln Z, \tag{3.23}
\end{aligned}$$

onde fizemos uso das Equações 3.20, 3.21, 3.22 e do fato que

$$\langle \sigma_i \sigma_j \rangle_{nMF} = \langle \sigma_i \rangle_{nMF} \langle \sigma_j \rangle_{nMF} = m_i m_j. \tag{3.24}$$

A condição de minimização de D_{KL} em relação a \bar{h}_i fornece:

$$\frac{\partial D_{KL}}{\partial \bar{h}_i} = (1 - m_i^2) \left(\bar{h}_i - \sum_{j \neq i} J_{ij} m_j \right) = 0. \tag{3.25}$$

Fazendo, então, a contribuição dos demais spins como:

$$\bar{h}_i = \sum_{j \neq i} J_{ij} m_j, \tag{3.26}$$

a Equação 3.22 tornar-se

$$m_i = \tanh \left[\beta \left(h_i + \sum_{j \neq i} J_{ij} m_j \right) \right]. \tag{3.27}$$

O método **nMF** prevê que todas as correlações C_{ij} são nulas, uma vez que $\langle \sigma_i \sigma_j \rangle_{nMF} = m_i m_j$. Entretanto, existe um caminho simples para calcular C_{ij} a partir do teorema da resposta linear, como feito em [48, 59],

$$C_{ij} = \frac{1}{\beta} \frac{\partial m_i}{\partial h_j}. \tag{3.28}$$

Usando as Equações 3.27 e 3.28, encontramos

$$\begin{aligned}
C_{ij} &= \frac{1}{\beta} \frac{\partial}{\partial h_j} \tanh \left[\beta \left(h_i + \sum_{k \neq i} J_{ik} m_k \right) \right] \\
&= (1 - m_i^2) \left(\delta_{ij} + \beta \sum_{k \neq i} J_{ik} C_{kj} \right), \tag{3.29}
\end{aligned}$$

que pode ser reescrita em uma notação matricial,

$$\mathbf{C} = \mathbf{L}(\mathbf{I} + \beta \mathbf{J} \cdot \mathbf{C}), \quad (3.30)$$

onde $L_{ij} = (1 - m_i^2)\delta_{ij}$ e \mathbf{I} é a matriz identidade.

As Equações 3.27 e 3.30 determinam as magnetizações e as correlações, no entanto, podemos invertê-las para os acoplamentos e campos. Invertendo a Equação 3.30 para os acoplamentos, temos

$$\mathbf{J} = \frac{1}{\beta} (\mathbf{L}^{-1} - \mathbf{C}^{-1}) \Rightarrow \quad (3.31)$$

$$J_{ij} = -\frac{1}{\beta} (C^{-1})_{ij}, \quad (i \neq j), \quad (3.32)$$

e a Equação 3.27 para os campos

$$h_i = \frac{1}{\beta} \tanh^{-1}(m_i) - \sum_{j \neq i} J_{ij} m_j. \quad (3.33)$$

Assim, dado as magnetizações e as correlações, podemos resolver a Equação 3.32 para determinar \mathbf{J} e então encontrar os campos \mathbf{h} pela Equação 3.27.

3.3.2 Equações de TAP

Para obter as equações de **TAP** faremos uso da expansão de Plefka [60] que consiste em expandir a energia livre $F(\alpha)$ em uma série de Taylor em torno de um parâmetro $\alpha = 0$ e no final fazer $\alpha = 1$ para recuperar o hamiltoniano original.

Primeiro modificaremos o hamiltoniano de Ising, Equação 3.8, introduzindo o parâmetro α , tal que

$$H(\alpha) = -\alpha \sum_{i < j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (3.34)$$

de modo que o hamiltoniano original é obtido quando $\alpha = 1$. Agora, realizaremos uma transformação de Legendre para que a energia livre fique em função das magnetizações, $\{m_i\}$, e não dos campos, $\{h_i\}$:

$$F(\alpha, \{m_i\}) = -\frac{1}{\beta} \ln \sum_{\{\sigma\}} e^{-\beta H(\alpha)} + \sum_i h_i m_i. \quad (3.35)$$

Devido à transformação, os campos ficam agora em função de α e $\{m_i\}$.

Suprimindo a dependência $\{m_i\}$ em $F(\alpha, \{m_i\})$, a expansão em série até a segunda ordem fica

$$F(\alpha) = F(0) + \alpha \left. \frac{\partial F}{\partial \alpha} \right|_{\alpha=0} + \frac{\alpha^2}{2} \left. \frac{\partial^2 F}{\partial \alpha^2} \right|_{\alpha=0} \quad (3.36)$$

onde

$$F(0) = \frac{1}{\beta} \sum_i \left[\left(\frac{1+m_i}{2} \right) \ln \left(\frac{1+m_i}{2} \right) + \left(\frac{1-m_i}{2} \right) \ln \left(\frac{1-m_i}{2} \right) \right], \quad (3.37)$$

$$\left. \frac{\partial F}{\partial \alpha} \right|_{\alpha=0} = - \sum_{i<j} J_{ij} m_i m_j \quad (3.38)$$

e

$$\left. \frac{\partial^2 F}{\partial \alpha^2} \right|_{\alpha=0} = -\beta \sum_{i<j} J_{ij}^2 (1-m_i^2)(1-m_j^2). \quad (3.39)$$

Fazendo $\alpha = 1$, a energia livre fica

$$\begin{aligned} F &= \frac{1}{\beta} \sum_i \left[\left(\frac{1+m_i}{2} \right) \ln \left(\frac{1+m_i}{2} \right) + \left(\frac{1-m_i}{2} \right) \ln \left(\frac{1-m_i}{2} \right) \right] - \\ &\quad - \sum_{i<j} J_{ij} m_i m_j - \frac{\beta}{2} \sum_{i<j} J_{ij}^2 (1-m_i^2)(1-m_j^2). \end{aligned} \quad (3.40)$$

Para obter o valor de h_i , que é o nosso objetivo, vamos observar a seguinte derivada da Equação 3.35 para $\alpha = 1$

$$\begin{aligned} \frac{\partial F}{\partial m_k} &= h_k + \sum_i \frac{\partial h_i}{\partial m_k} m_i - \frac{1}{Z} \sum_{\{\sigma\}} \sum_i \frac{\partial h_i}{\partial m_k} \sigma_i e^{-\beta H} \\ &= h_k + \sum_i \frac{\partial h_i}{\partial m_k} m_i - \sum_i \frac{\partial h_i}{\partial m_k} \left(\frac{1}{Z} \sum_{\{\sigma\}} \sigma_i e^{-\beta H} \right) \\ &= h_k + \sum_i \frac{\partial h_i}{\partial m_k} m_i - \sum_i \frac{\partial h_i}{\partial m_k} m_i \\ &= h_k. \end{aligned} \quad (3.41)$$

Assim, usando a relação da Equação 3.41 com a energia livre dada pela Equação 3.40, o campo h_i é

$$h_i = \frac{1}{\beta} \tanh^{-1} m_i - \sum_{j \neq i} J_{ij} m_j + \beta m_i \sum_{j \neq i} J_{ij}^2 (1-m_j^2). \quad (3.42)$$

Já para determinar os acoplamentos, vamos, novamente, usar o teorema da resposta linear, onde

$$(C^{-1})_{ij} = \beta \frac{\partial h_i}{\partial m_j}. \quad (3.43)$$

Usando a equação 3.42, encontramos, para $i \neq j$,

$$(C^{-1})_{ij} = -\beta J_{ij} - 2m_i m_j \beta^2 J_{ij}^2 \Rightarrow \quad (3.44)$$

$$J_{ij} = -\frac{1 - \sqrt{1 - 8m_i m_j (C^{-1})_{ij}}}{4m_i m_j \beta}, \quad (i \neq j). \quad (3.45)$$

Como no caso do **nMF**, dado as magnetizações e as correlações, podemos resolver a Equação 3.45 para determinar **J** e então encontrar os campos **h** pela Equação 3.42. Observe que não só obtemos as Equações de **TAP**, mas também recuperamos o método **nMF**, uma vez que o mesmo é dado se considerarmos a expansão até primeira ordem.

4 RESULTADOS

Investigaremos em quais situações os métodos descritos no capítulo anterior conseguem efetivamente determinar os acoplamentos e campos do sistema subjacente que produziu os dados. Para isso, vamos gerar conjuntos de dados sintéticos, obtidos aplicando Monte Carlo a um hamiltoniano tipo Ising Equação 3.8

$$H(\sigma) = - \sum_{i < j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (4.1)$$

portanto seguindo a distribuição da Equação 3.9:

$$P(\sigma) = \frac{1}{Z} e^{\beta \left(\sum_{i < j} J_{ij} \sigma_i \sigma_j + \sum_{i=1}^N h_i \sigma_i \right)}. \quad (4.2)$$

Nosso objetivo é testar diferentes formas para o hamiltoniano da Equação 4.1 e verificar em que situações as técnicas já discutidas conseguem determinar os acoplamentos e campos do sistema subjacente. Esses resultados permitirão estabelecer limites de aplicabilidade a técnica empregada.

4.1 Dados Sintéticos

Uma primeira limitação que impomos ao modelo é considerarmos que nem todos os spins estão acoplados. Portanto, vamos considerar que a rede de acoplamentos é estabelecida segundo o modelo de redes aleatórias de Erdős–Rényi [61, 62].

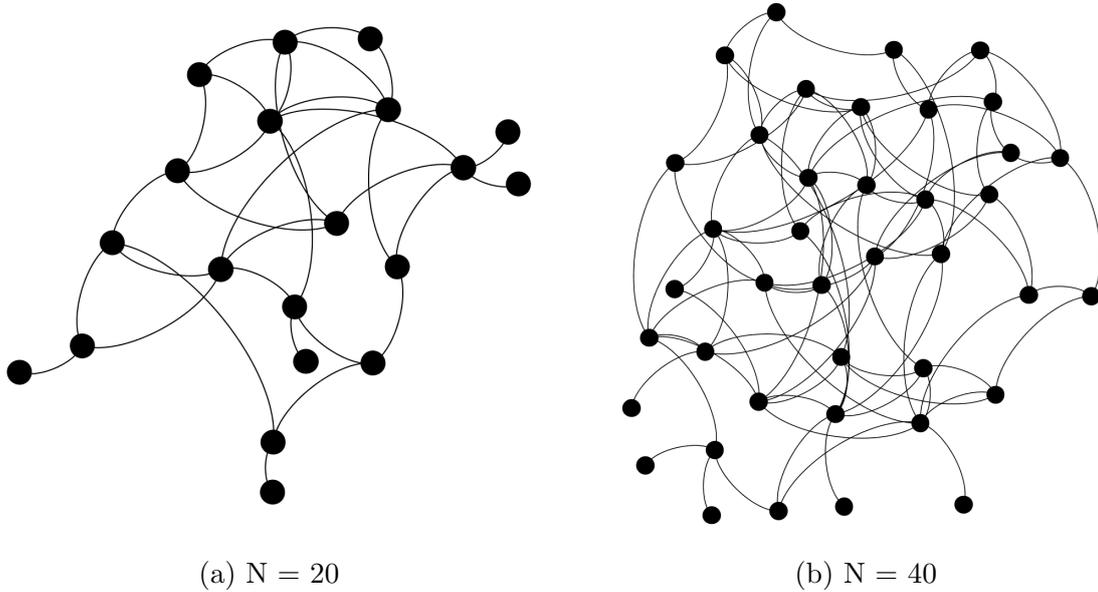
O conjunto de redes aleatórias $G(N, p)$ é obtido construindo redes com N vértices, onde a probabilidade de criar ligações pode ser determinada por:

$$p = \frac{k}{N-1}, \quad (4.3)$$

onde k é a conectividade média dos vértices da rede. Para $k = 0$, não teremos nenhuma ligação e, para $k = N - 1$, teremos uma rede completamente conectada, isto é, todas as ligações são criadas. A Figura 4 mostra dois exemplos de redes aleatórias.

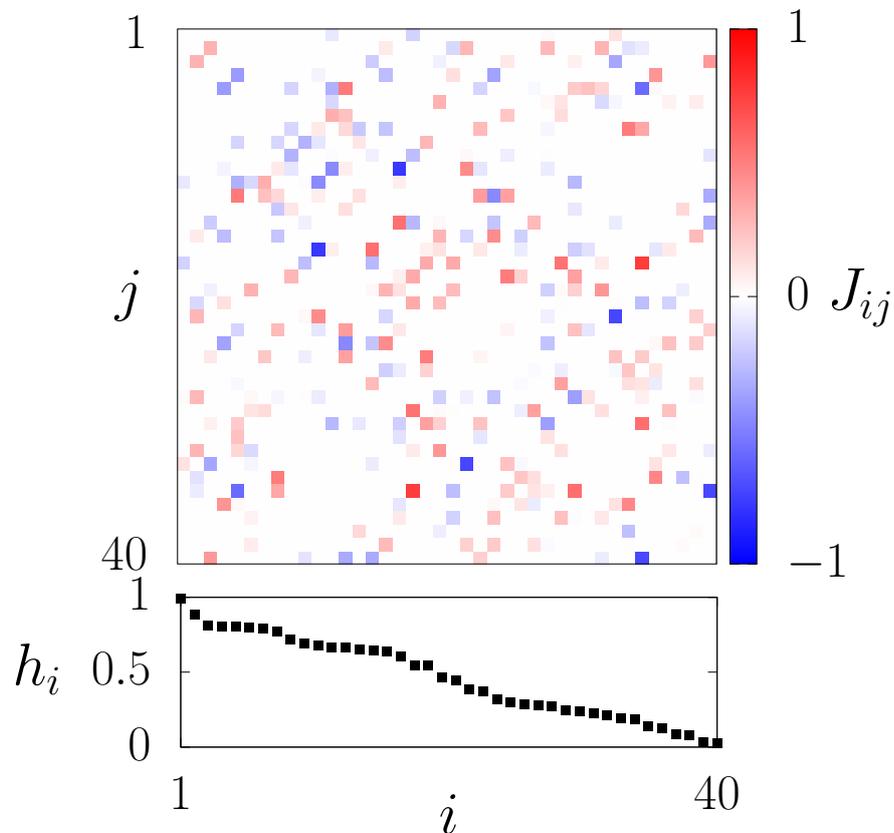
Uma vez que sabemos quais spins estão ligados, podemos atribuir valores aleatórios para os acoplamentos $\{J_{ij}\}$ e os campos $\{h_i\}$, e fazer $J_{ij} = 0$ para aqueles casos que não há ligação entre os spins σ_i e σ_j . Os acoplamentos são distribuídos de acordo com uma gaussiana de média μ e desvio padrão $\bar{\sigma}$ (não confundamos o desvio padrão $\bar{\sigma}$ com o σ usado para descrever o estado do sistema):

Figura 4 – Exemplos de redes aleatórias.



Fonte: Elaborada pelo autor. Dois exemplos de redes aleatórias criadas com conectividade média $k = 4$ para $N = 20$ (a) e $N = 40$ (b).

Figura 5 – Acoplamentos e campos.



Fonte: Figura elaborada pelo autor. Cada ponto ij no gráfico superior representa um acoplamento J_{ij} , onde sua intensidade é dada de acordo com sua cor. No gráfico inferior, ilustramos o valor do campo h_i atuando em cada spin i .

$$P(J_{ij}) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{J_{ij} - \mu}{\bar{\sigma}}\right)^2\right]. \quad (4.4)$$

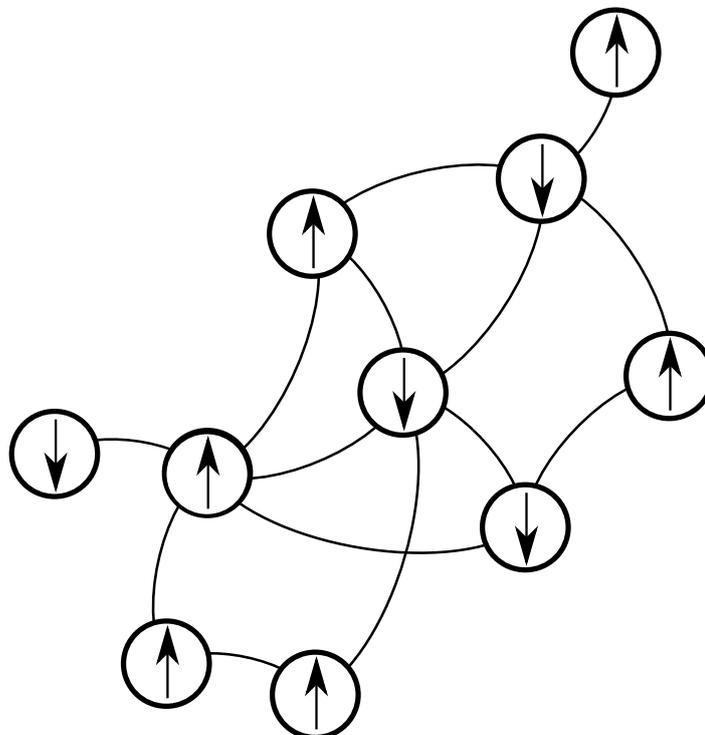
Já os campos são distribuídos uniformemente entre um intervalo determinado.

A Figura 5 mostra um exemplo da matriz dos valores dos acoplamentos e os campos atuando em cada spin para um sistema com $N = 40$. Os acoplamentos são gerados segundo a distribuição da Equação 4.4, usando $\mu = 0$, $\bar{\sigma} = 0.25$ e $J_{ij} \in [-1, 1]$. Os valores dos campos são distribuído uniformemente dentro do intervalo $h_i \in [0, 1]$. Os spins foram numerados de forma decrescente com relação ao campo.

Uma vez que temos todos os parâmetros necessários para escrever a distribuição da Equação 4.2, nos resta gerar um conjunto de estados produzidos por essa distribuição e então realizar as medidas. Para isso usaremos o método de Monte Carlo (**MC**) [63, 64], especificamente o algoritmo de Metropolis [63, 65].

Aqui vale introduzir uma linguagem bastante utilizada. Dizemos que o spin no estado $\sigma_i = +1$ está apontando para cima (\uparrow), enquanto que o spin no estado $\sigma_i = -1$ está apontando para baixo (\downarrow). A Figura 6 mostra o exemplo de um estado, $\sigma = \{\sigma_i\}$, de uma rede com 10 spins na linguagem descrita, onde cada nó representa um spin e seu estado é dado pela direção de sua seta.

Figura 6 – Exemplo de um estado.



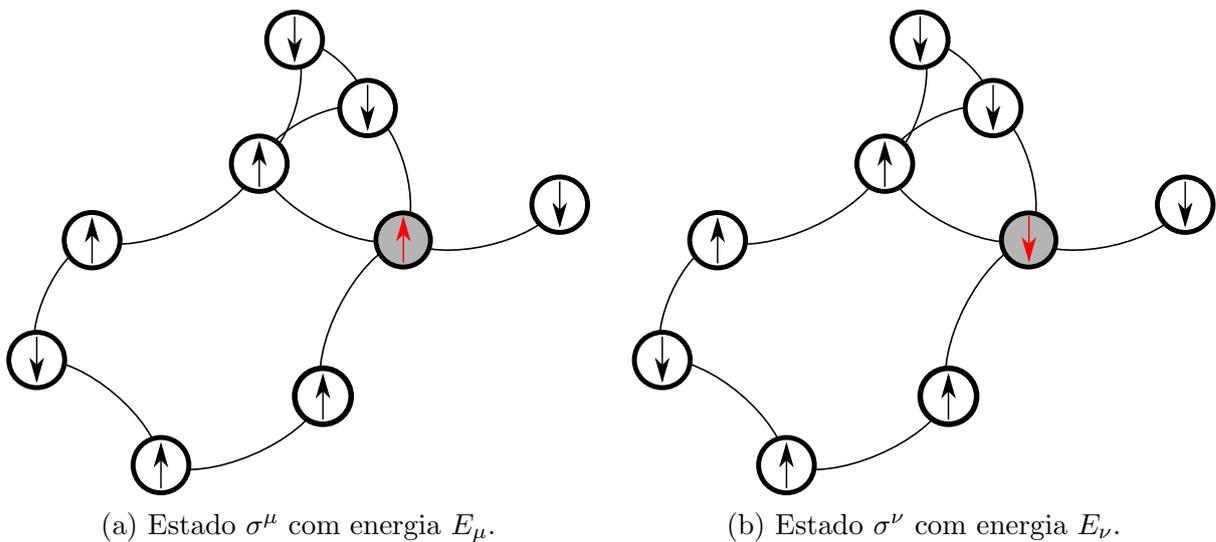
Fonte: Figura elaborada pelo autor. Exemplo de um estado de uma rede com 10 spins, onde cada nó representa um spin e seu estado é dado pela direção de sua seta (\uparrow indica spin no estado +1 e \downarrow no estado -1).

No algoritmo de Metropolis um novo estado é gerado pelo seguinte processo: dado um estado σ^μ com energia E_μ , um novo estado σ^ν com energia E_ν é gerado selecionando um spin aleatório e invertendo seu sentido. Esse novo estado pode ou não ser aceito de acordo com a seguinte probabilidade:

$$P = \begin{cases} e^{-\beta(E_\nu - E_\mu)} & \text{se } E_\nu - E_\mu > 0 \\ 1 & \text{se } E_\nu - E_\mu \leq 0 \end{cases} . \quad (4.5)$$

Assim, se o novo estado possuir energia menor que o estado anterior, esse estado é aceito. Mas se o novo estado possuir energia maior do que o estado anterior, então ele só será aceito segundo a probabilidade dada pela Equação 4.5. O processo de tentar gerar um novo estado caracteriza um passo de Monte Carlo. A Figura 7 mostra o exemplo de um passo.

Figura 7 – Exemplo de uma passo de Monte Carlo.



Fonte: Elaborada pelo autor. Exemplo de uma passo de Monte Carlo segundo o algoritmo de Metropolis. Um estado σ^μ com energia E_μ gera um novo estado σ^ν com energia E_ν mudando o sentido e um spin aleatório. Esse novo estado é aceito segundo a probabilidade da Equação 4.5.

Com o algoritmo de Metropolis descrito acima somos capazes de gerar um conjunto de estados, a uma certa temperatura, e calcular as médias que desejamos. Partindo de um estado inicial (aqui usaremos como estado inicial aquele onde todos os spins são alinhados aleatoriamente) realizamos τ_{eq} passos de Monte Carlo para o sistema atingir o equilíbrio e mais τ_{MC} passos para obter o conjunto de estados. Para gerar estados menos correlacionados, usaremos estados gerados a cada $2N$ passos (N é o número de spins). Assim, para obter M estados independentes devemos realizar, após o equilíbrio, um total de $\tau_{MC} = 2NM$ passos.

Com os M estados obtidos para uma temperatura fixa β podemos calcular as

magnetizações $m_i = \langle \sigma_i \rangle$ e as correlações $C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$, onde

$$\langle \sigma_i \rangle = \frac{1}{M} \sum_{k=1}^M \sigma_i^k \quad (4.6)$$

e

$$\langle \sigma_i \sigma_j \rangle = \frac{1}{M} \sum_{k=1}^M \sigma_i^k \sigma_j^k. \quad (4.7)$$

Utilizando a energia dada pelo hamiltoniano da Equação 3.8,

$$E(\sigma) = - \sum_{i < j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (4.8)$$

uma rede aleatória com $N = 40$ e conectividade $k = 4$, onde os acoplamentos \mathbf{J} são distribuídos de acordo com a gaussiana da Equação 4.4 com $\mu = 1$, $\bar{\sigma} = 0.25$ e $J_{ij} \in [0, 2]$, e os campos \mathbf{h} distribuídos uniformemente dentro do intervalo $h_i \in [0, 1]$, calculamos as correlações e magnetizações com o método descrito acima para $\beta = 0.2$. Realizamos um processo de Monte Carlo, onde deixamos o sistema evoluir por $\tau_{eq} = 10^3$ passos até atingir o equilíbrio e usamos um total de $M = 10^6$ estados independentes. A Figura 8 mostra a matriz de correlações e as magnetizações obtidas com o sistema descrito. Os spins foram numerados de forma decrescente com relação a magnetização.

Vale ressaltar que empregando o método de Monte Carlo para obter as magnetizações e correlações obtemos aproximações que serão melhores quanto maior for o número de amostras independentes M . Obter os valores exatos dependeria de somar-se 2^N termos na função de partição, o que é impraticável mesmo para tamanhos moderados como $N = 40$.

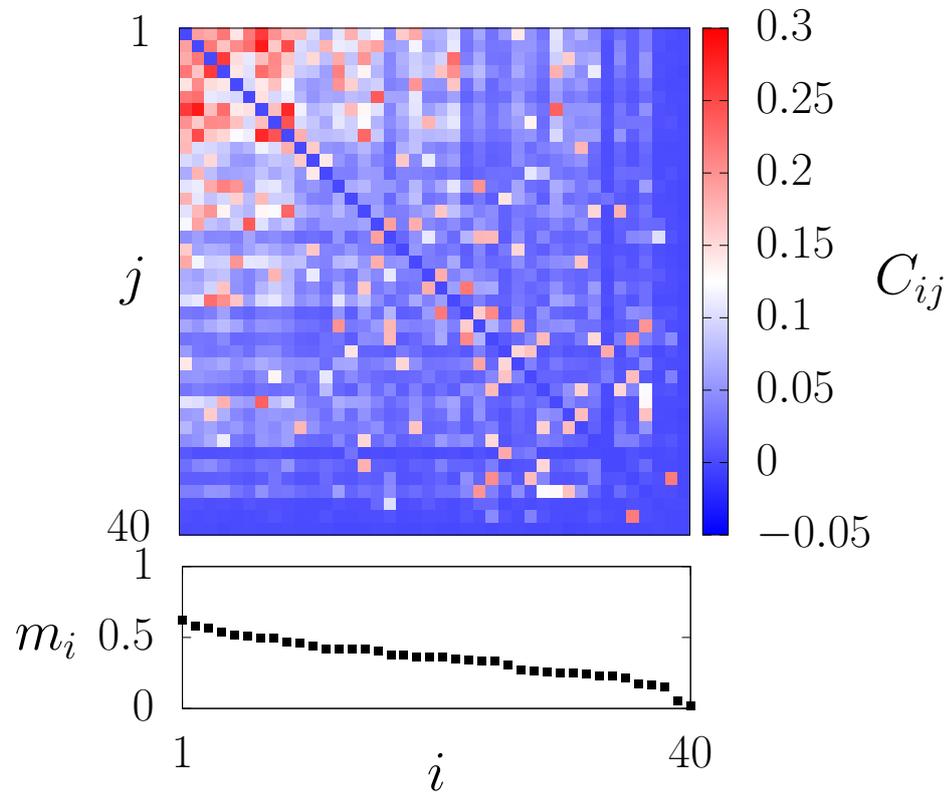
Uma quantidade importante que se pode calcular e nos diz muito sobre o sistema é o calor específico. Podemos calculá-lo pelo método de Monte Carlo usando a seguinte equação [63]:

$$c = \frac{\beta}{N} \left(\langle E^2 \rangle - \langle E \rangle^2 \right). \quad (4.9)$$

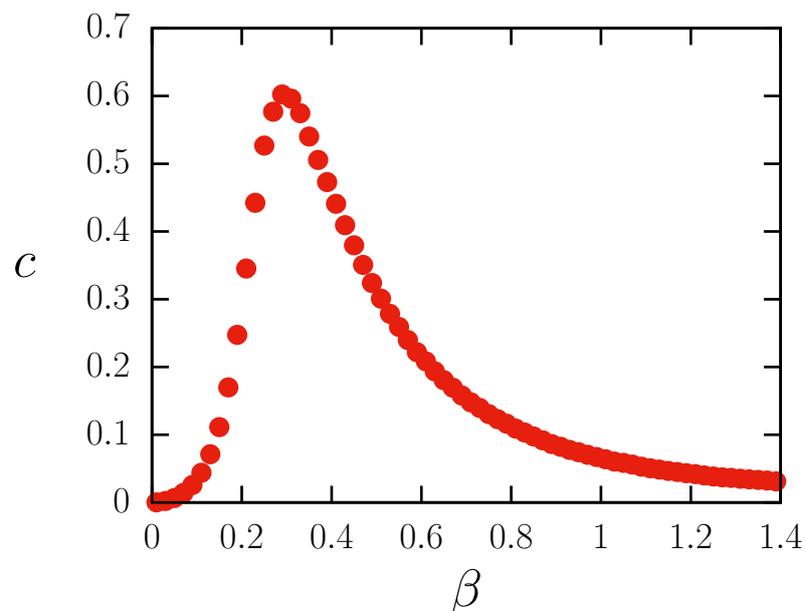
Usando a Equação 4.9 e os mesmos parâmetros usados para gerar os dados da figura 8, vamos calcular o calor específico em diferentes temperaturas. O Gráfico 5 mostra o resultado obtido para c versus o inverso da temperatura β .

Em modelos que podem ser definidos em diferentes escalas, o pico da curva do calor específico diverge no limite termodinâmico, indicando a ocorrência de uma transição de fase. O valor onde essa divergência ocorre é chamado de ponto crítico. A presença desse pico em escala finita, portanto, indica a ocorrência de algo similar a uma transição nesse sistema limitado. Vamos nos referir ao ponto onde ocorre o valor de máximo calor

Figura 8 – Correlações e magnetizações.



Fonte: Figura elaborada pelo autor. Cada ponto ij no gráfico superior representa uma correlação C_{ij} , onde sua intensidade é dada de acordo com sua cor. No gráfico inferior, mostramos o valor da magnetização m_i de cada spin i .

Gráfico 5 – Calor específico versus β .

Fonte: Figura elaborada pelo autor. Calor específico calculado pelo método de Monte Carlo usando uma rede com $N = 40$ spins com acoplamentos e campos todos positivos.

específico por β^* (inverso da temperatura) ou por $T = T^*$ (temperatura).

Na próxima seção, mostraremos os resultados obtidos para diferentes regiões de T^* (β^*).

4.2 Teste de validade

Agora que somos capazes de obter correlações e magnetizações geradas a partir de um sistema subjacente, para uma certa temperatura, vamos dar início a uma série de análises.

Analisaremos a eficiência do método de atualização, desenvolvido no capítulo 3, para os acoplamentos e campos em três regimes para J_{ij} : Ferromagnético, onde todos os acoplamentos são positivos; Antiferromagnético, onde todos os acoplamentos são negativos; e *Spin Glass*, nesse regime os J_{ij} 's podem ser tanto positivo como negativo. Em cada regime os acoplamentos são distribuídos segundo a gaussiana da Equação 4.4:

$$P(J_{ij}) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{J_{ij} - \mu}{\bar{\sigma}}\right)^2\right]. \quad (4.10)$$

Já os campos h_i 's são distribuídos uniformemente entre -1 e 1 em todos os regimes.

Para obter esses acoplamentos e campos, partimos de um sistema modelo construído a partir do método aproximativo **nMF** ou das Equações de **TAP**, dados pelas Equações 3.32, 3.33, 3.42 e 3.45, repetidas abaixo

- **nMF:**

$$J_{ij} = -\frac{1}{\beta} (C^{-1})_{ij}, \quad (i \neq j), \quad (4.11)$$

$$h_i = \frac{1}{\beta} \tanh^{-1}(m_i) - \sum_{j \neq i} J_{ij} m_j. \quad (4.12)$$

- **TAP:**

$$J_{ij} = -\frac{1 - \sqrt{1 - 8m_i m_j (C^{-1})_{ij}}}{4m_i m_j \beta}, \quad (i \neq j). \quad (4.13)$$

$$h_i = \frac{1}{\beta} \tanh^{-1} m_i - \sum_{j \neq i} J_{ij} m_j + \beta m_i \sum_{j \neq i} J_{ij}^2 (1 - m_j^2). \quad (4.14)$$

Então usamos as equações de atualização 3.16, 3.17, 3.18 e 3.19

- Para os acoplamentos J_{ij} :

$$\Delta J_{ij}(t+1) = \epsilon [\langle \sigma_i \sigma_j \rangle_{exp} - \langle \sigma_i \sigma_j \rangle_{Ising}(t)] \quad (4.15)$$

e

$$J_{ij}(t+1) = J_{ij}(t) + \Delta J_{ij}(t+1). \quad (4.16)$$

- Para os campos h_i :

$$\Delta h_i(t+1) = \epsilon [\langle \sigma_i \rangle_{exp} - \langle \sigma_i \rangle_{Ising}(t)] \quad (4.17)$$

e

$$h_i(t+1) = h_i(t) + \Delta h_i(t+1). \quad (4.18)$$

Lembrando que as médias em cada atualização são calculadas pelas Equações 4.6 e 4.7:

$$\langle \sigma_i \rangle = \frac{1}{M} \sum_{k=1}^M \sigma_i^k \quad (4.19)$$

e

$$\langle \sigma_i \sigma_j \rangle = \frac{1}{M} \sum_{k=1}^M \sigma_i^k \sigma_j^k. \quad (4.20)$$

Como mencionado no capítulo anterior, a presença de um pico no calor específico em escala finita indica a ocorrência de algo similar a uma transição de fase. Portanto, em cada um dos regimes citados acima, investigaremos três tipos de temperatura (ou inverso da temperatura β): temperatura acima do ponto de máximo calor específico, $T > T^*$ ($\beta < \beta^*$); temperatura próxima do ponto de máximo calor específico, $T \approx T^*$ ($\beta \approx \beta^*$); e temperatura abaixo do ponto de máximo calor específico, $T < T^*$ ($\beta < \beta^*$).

Podemos assim dar início à série de testes. Em todos os regimes consideramos um sistema com $N = 20$ spins e uma rede com conectividade $k = 4$. O parâmetro da taxa de atualização foi escolhido como $\epsilon = 0.1$ em todos os casos e o número de estados independentes utilizado para calcular as médias em cada atualização foi $M = 10^6$. Para cada passo, calculamos as médias e medimos as distâncias entre a atualização atual e a anterior da seguinte forma:

$$\Delta_{\langle \sigma \rangle}(t) = \frac{1}{N} \sum_{i=1}^N (\langle \sigma_i \rangle(t) - \langle \sigma_i \rangle(t-1))^2 \quad (4.21)$$

e

$$\Delta_{\langle \sigma \sigma \rangle}(t) = \frac{2}{N(N-1)} \sum_{i < j} (\langle \sigma_i \sigma_j \rangle(t) - \langle \sigma_i \sigma_j \rangle(t-1))^2. \quad (4.22)$$

Assim, quando as distâncias $\Delta_{\langle \sigma \rangle}(t)$ e $\Delta_{\langle \sigma \sigma \rangle}(t)$ atingiam valores menores que 10^{-5} , nós cessamos as atualizações. Observe que a precisão na convergência dos parâmetros depende

da correta estimativa das médias calculadas pelo método de Monte Carlo. Verificamos que para $M = 10^6$ amostras, podemos usar como teste de convergência distâncias menores que 10^{-5} .

4.2.1 Ferromagnético

O primeiro regime que analisamos foi o ferromagnético. Para criar o sistema subjacente que gera as correlações e as magnetizações, distribuimos os acoplamentos segundo a gaussiana da Equação 4.10 com os seguintes parâmetros: média $\mu = 1$, desvio padrão $\bar{\sigma} = 0.25$ e $J_{ij}^{sub} \in [0, 2]$.

Com as correlações e magnetizações, C_{ij}^{sub} e $\langle \sigma_i \rangle^{sub}$, geradas usando Monte Carlo, com $M = 10^6$ estados independentes, a partir do sistema subjacente criado, usamos o método das Equações de **TAP** para encontrar os valores iniciais para os acoplamentos e campos, J_{ij}^{mod} e h_i^{mod} , e, então, usamos o método de atualização para tentar encontrar um modelo que consiga reproduzir os acoplamentos e campos do sistema subjacente.

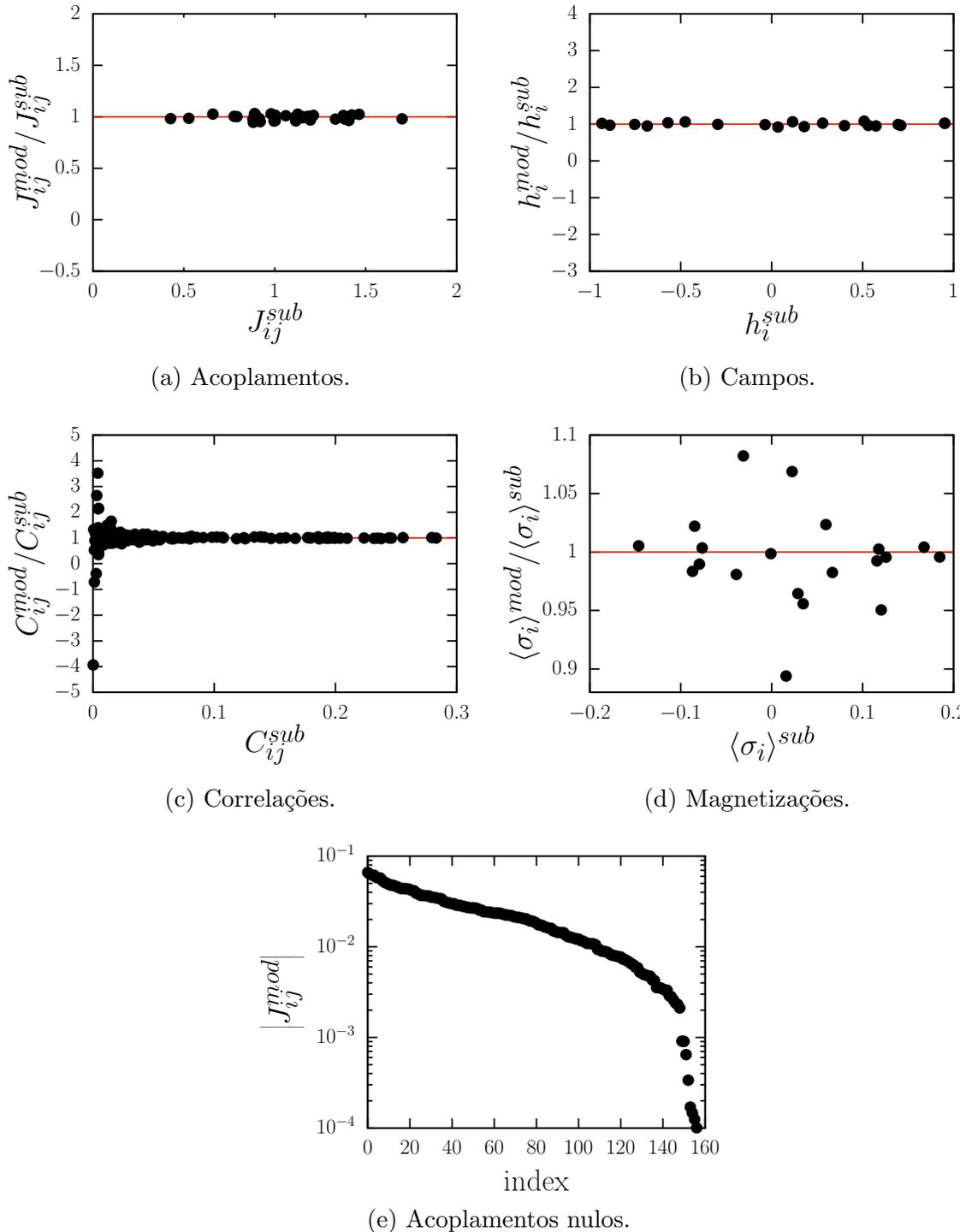
O Gráfico 6 mostra os resultados obtidos para temperatura acima do ponto de máximo calor específico. Nos Gráficos 6a e 6b, ilustramos a comparação dos acoplamentos e campos, J_{ij}^{mod} e h_i^{mod} , do modelo encontrado com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . Observe que os resultados estão normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1, melhores são os resultados. Já o Gráfico 6e mostra o valor absoluto dos acoplamentos do modelo encontrado que são nulos no sistema subjacente, onde os acoplamentos foram ordenados de forma decrescente. Assim, podemos dizer que o modelo encontrado se aproxima do sistema subjacente. Os Gráficos 6c e 6d comparam as correlações e magnetizações geradas pelo modelo encontrado com as do sistema subjacente. Observando o Gráfico 6c, vemos que o efeito relativo das flutuações, causado pelo método de Monte Carlo, é exacerbado quando tomamos o caso de correlações próximas de zero.

O Gráfico 7 mostra as mesmas análises para temperatura próxima do ponto de máximo calor específico. Vemos pelos Gráficos 7a, 7b e 7e que o modelo conseguiu se aproximar dos parâmetros do sistema subjacente com menos de precisão do que na temperatura mais alta. Já nos Gráficos 7c e 7d vemos que as correlações e magnetizações do modelo encontrado foram melhores do que para temperatura mais alta. Ainda vemos a discrepância perto de zero na correlações causada pelo efeito relativo das flutuações.

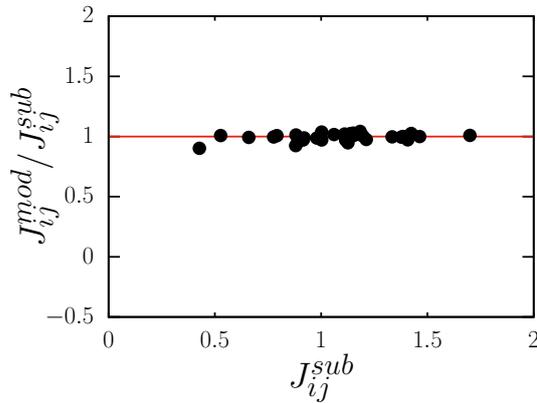
Já para temperatura baixo do ponto de máximo calor específico os resultados são diferentes, como mostra o Gráfico 8. Vemos, pelos Gráficos 8a, 8b e 8e, que o modelo encontrado não foi capaz de se aproximar dos parâmetros do sistema subjacente. Nos Gráficos 8c e 8d, porém, observamos que as correlações e magnetizações produzidas pelo

modelo são mais precisas.

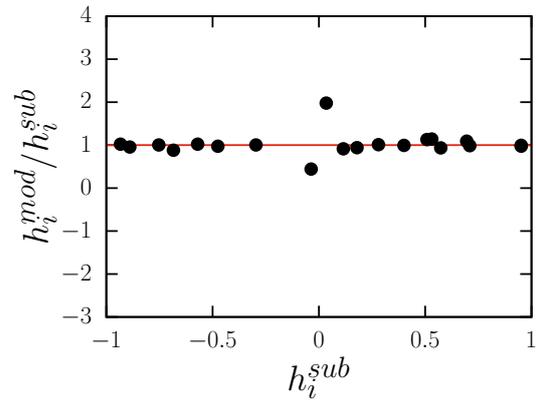
Assim, podemos concluir que, no regime ferromagnético, o método utilizado foi capaz de encontrar um modelo que se aproxima do sistema subjacente em temperatura próxima e acima da temperatura de máximo calor específico. Já para correlações e magnetizações geradas por sistemas subjacentes em temperatura abaixo do máximo de calor específico, o método empregado foi capaz de produzir modelos cuja correlações e magnetizações são consistentes com as observadas. Curiosamente, os parâmetros do sistema modelo não são consistentes com os do sistema subjacente.

Gráfico 6 – Ferromagnético para $T > T^*$.

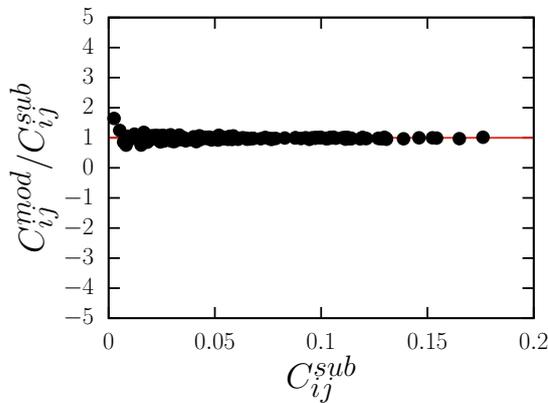
Fonte: Elaborada pelo autor. Regime ferromagnético para $T > T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle_i^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle_i^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 7 – Ferromagnético para $T \approx T^*$.

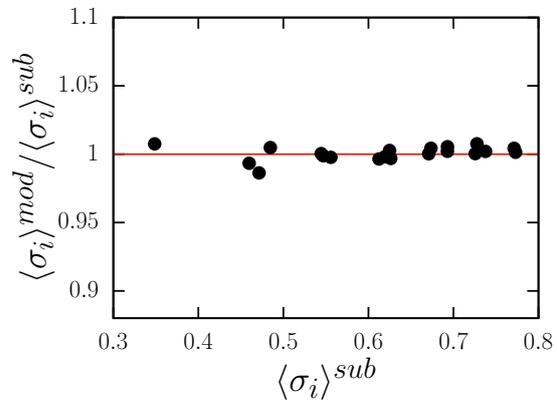
(a) Acoplamentos.



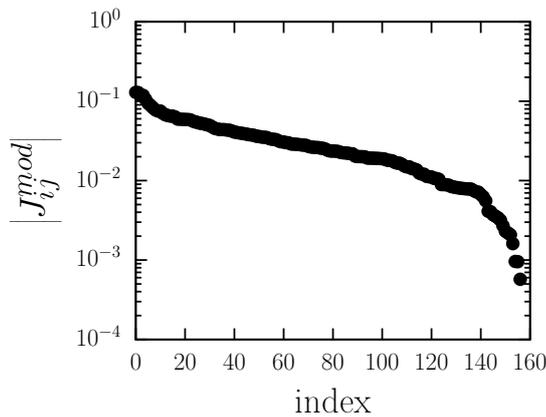
(b) Campos.



(c) Correlações.

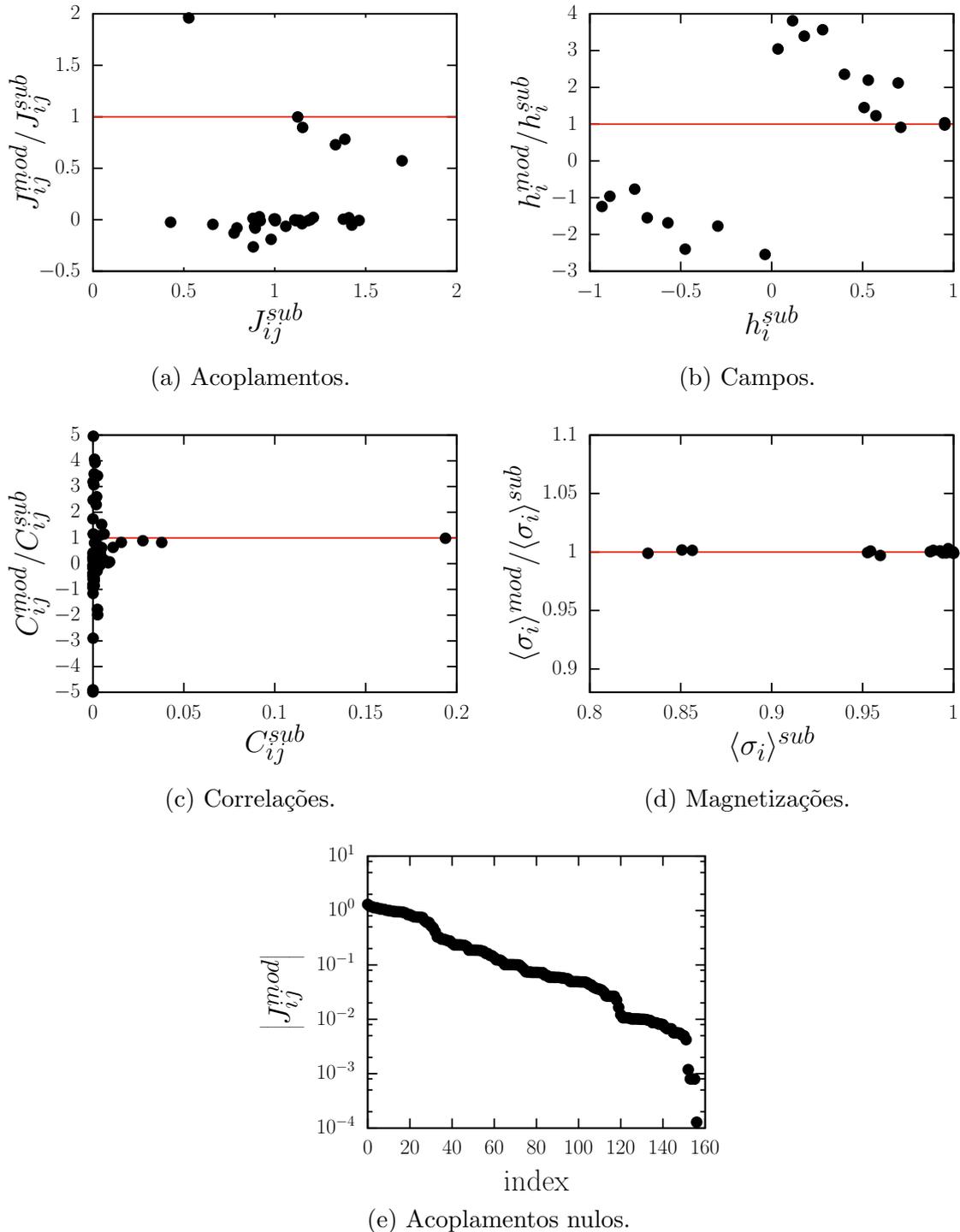


(d) Magnetizações.



(e) Acoplamentos nulos.

Fonte: Elaborada pelo autor. Regime ferromagnético para $T \approx T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 8 – Ferromagnético para $T < T^*$.

Fonte: Elaborada pelo autor. Regime ferromagnético para $T < T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle_i^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle_i^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

4.2.2 Antiferromagnético

O segundo regime que analisamos foi o antiferromagnético. Para criar o sistema subjacente que gera as correlações e magnetizações nesse regime, distribuimos os acoplamentos segundo a gaussiana da Equação 4.10 com os seguintes parâmetros: média $\mu = -1$, desvio padrão $\bar{\sigma} = 0.25$ e $J_{ij}^{sub} \in [-2, 0]$.

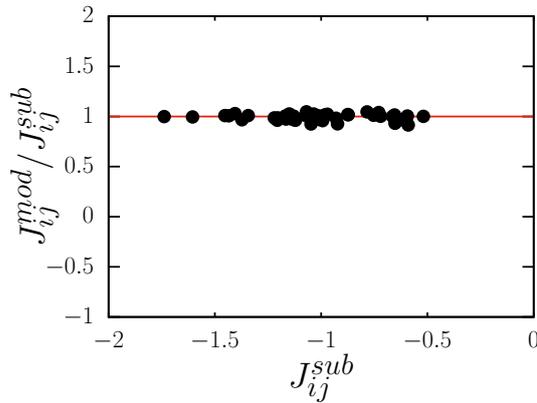
Com as correlações e magnetizações, C_{ij}^{sub} e $\langle \sigma_i \rangle^{sub}$, geradas pelo sistema subjacente, usamos o método das Equações de **TAP** para encontrar os valores iniciais para os acoplamentos e campos, J_{ij}^{mod} e h_i^{mod} , para a temperatura acima da temperatura de máximo calor específico e para a temperatura próxima e abaixo da temperatura de máximo calor específico usamos o método **nMF**. Em seguida, usamos o método de atualização para tentar encontrar um modelo que consiga se aproximar dos valores do sistema subjacente.

Os resultados obtidos para temperatura acima da temperatura de máximo calor específico foram semelhantes aos resultados do regime ferromagnético, como mostra o Gráfico 9. Pelos Gráficos 9a, 9b e 9e vemos que o método conseguiu encontrar um modelo que reproduz o sistema subjacente. Os Gráficos 9c e 9d mostram as correlações e magnetizações obtidas. Dessa vez, vemos tanto correlações positivas como negativas. Mais uma vez, vemos o efeito relativo das flutuações quando consideramos correlações próximas de zero.

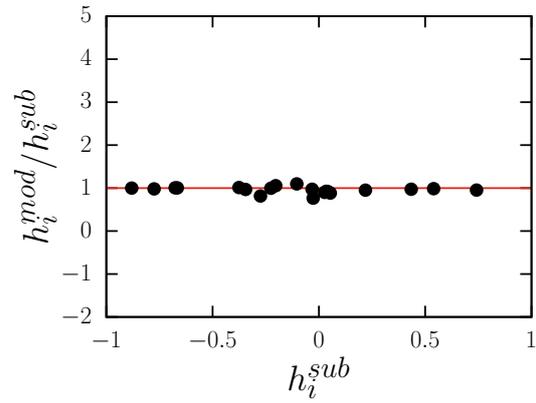
Para temperatura próxima a temperatura de máximo calor específico, os resultados também foram semelhantes aos do regime ferromagnético, como mostra o Gráfico 10. Os resultados para os acoplamentos e campos podem ser vistos nos Gráficos 10a, 10b e 10e. Vemos que os valores encontrados não foram tão precisos como em altas temperaturas. Isso é mais evidente nos resultados para os campos, Gráfico 10b. Nos Gráficos 10c e 10d, mostramos as correlações e magnetizações produzidas pelo modelo encontrado em comparação com as do subjacente. Observamos que, assim como no regime ferromagnético, as correlações e magnetizações encontradas foram melhores.

A última temperatura analisada nesse regime foi abaixo da temperatura de máximo calor específico. Assim como no regime ferromagnético, o método não foi capaz de encontrar um modelo que se aproximasse do sistema subjacente. Esse fato é mostrado no Gráfico 11. Pelos Gráficos 11a, 11b e 11e vemos a discrepância entre os valores do modelo encontrado com relação aos do sistema subjacente. Principalmente, nos campos e nos acoplamentos que deveriam ser nulos.

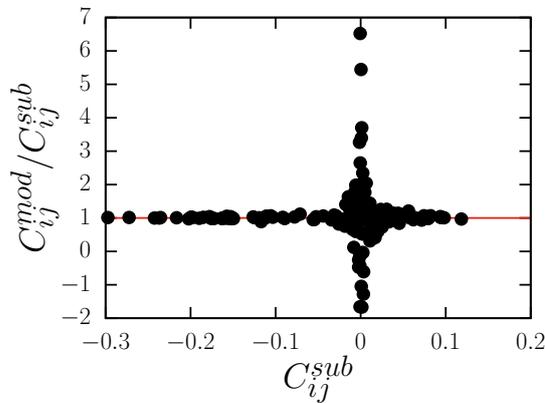
Tiramos a mesma conclusão do regime ferromagnético para o regime antiferromagnético. Quanto maior a temperatura melhores são os J_{ij}^{mod} 's e h_i^{mod} 's e piores são as C_{ij}^{mod} 's e $\langle \sigma_i \rangle^{mod}$'s. E o inverso ocorre para temperaturas cada vez menores.

Gráfico 9 – Antiferromagnético para $T > T^*$.

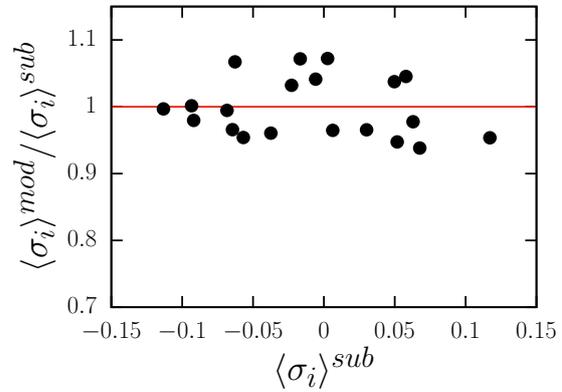
(a) Acoplamentos.



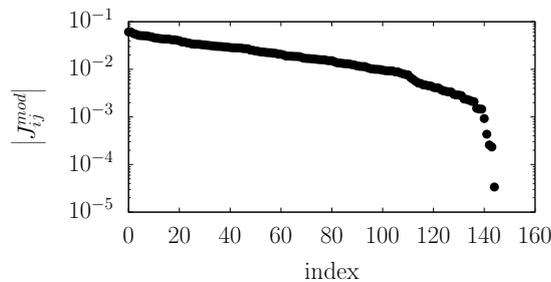
(b) Campos.



(c) Correlações.

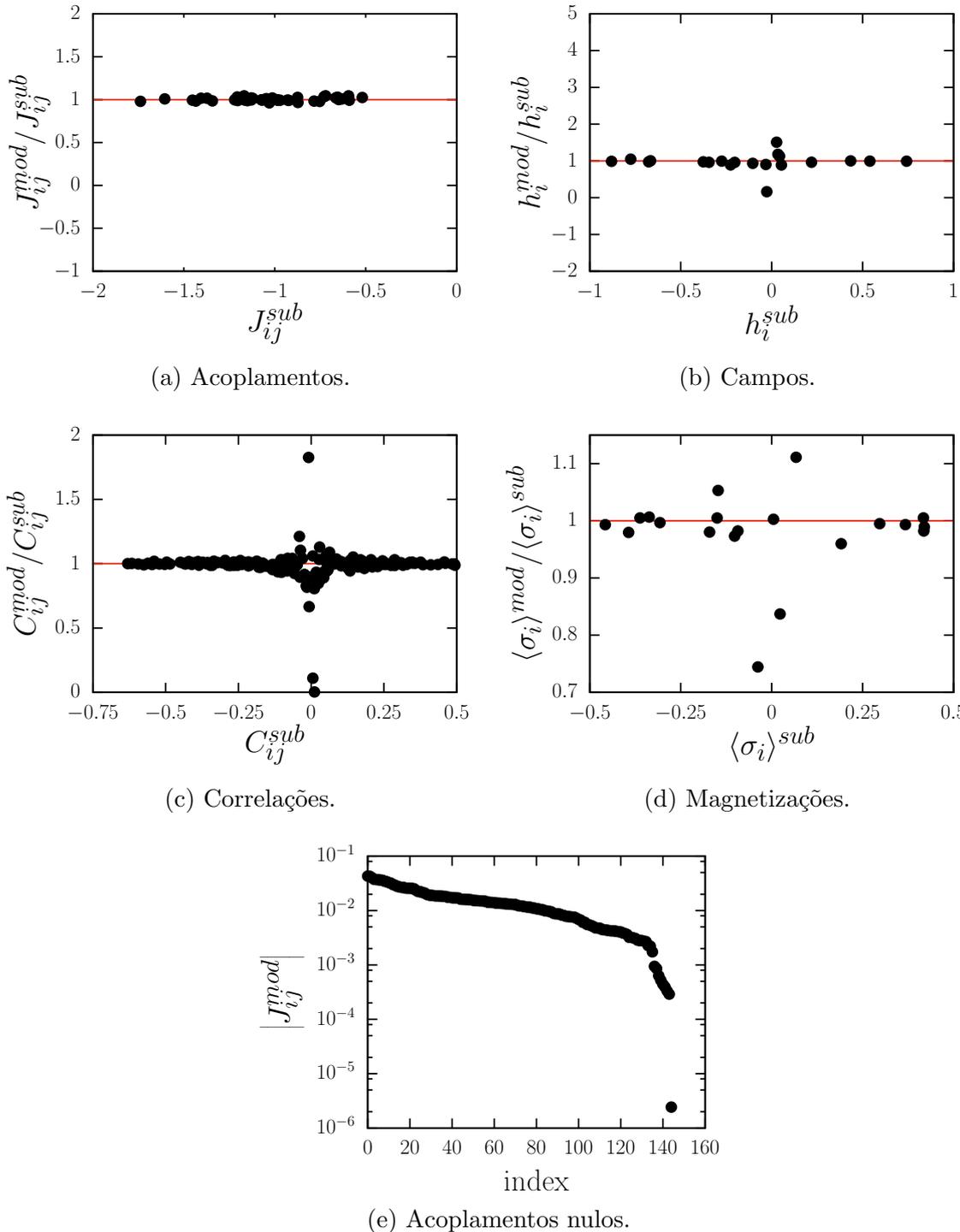


(d) Magnetizações.

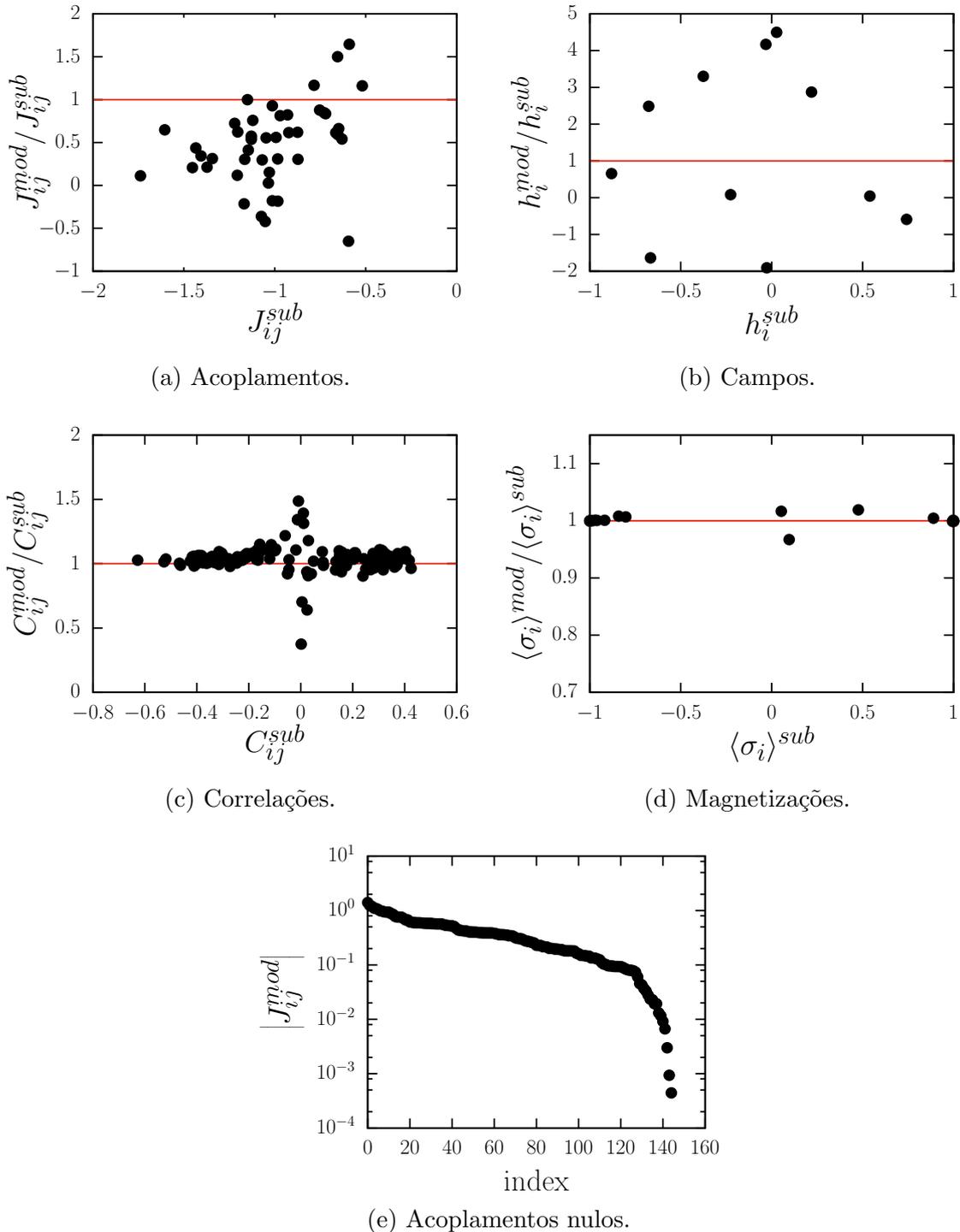


(e) Acoplamentos nulos.

Fonte: Elaborada pelo autor. Regime antiferromagnético para $T > T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 10 – Antiferromagnético para $T \approx T^*$.

Fonte: Elaborada pelo autor. Regime antiferromagnético para $T \approx T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle_i^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle_i^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 11 – Antiferromagnético para $T < T^*$.

Fonte: Elaborada pelo autor. Regime antiferromagnético para $T < T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle_i^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle_i^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

4.2.3 *Spin Glass*

O terceiro e último regime estudo foi o chamado *spin glass*. Esse regime é uma combinação do ferromagnético e antiferromagnético, onde os acoplamentos, agora, podem ser tanto positivos como negativos. Para criar o sistema subjacente que gera as correlações e magnetizações nesse regime, distribuímos os acoplamentos segundo a gaussiana da Equação 4.10 com os seguintes parâmetros: média $\mu = 0$, desvio padrão $\bar{\sigma} = 0.25$ e $J_{ij}^{sub} \in [-1, 1]$.

Com as correlações e magnetizações, C_{ij}^{sub} e $\langle \sigma_i \rangle^{sub}$, geradas pelo sistema subjacente, usamos o método das Equações de **TAP** para encontrar os valores iniciais para os acoplamentos e campos, J_{ij}^{mod} e h_i^{mod} , para a temperatura acima da temperatura de máximo calor específico e para a temperatura próxima e abaixo da temperatura de máximo calor específico usamos o método **nMF**. Em seguida, usamos o método de atualização para tentar encontrar um modelo que consiga se aproximar dos valores do sistema subjacente.

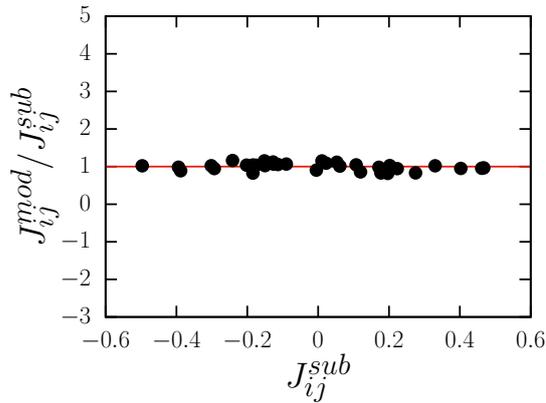
Os resultados obtidos para esse regime não foram diferentes dos regimes ferromagnético e antiferromagnético. Assim como nos regimes citados, o métodos empregado conseguiu encontrar um modelo que se aproxima dos parâmetros do sistema subjacente para temperatura acima do ponto de máximo calor específico, como mostra o Gráfico 12. Nos Gráficos 12a, 12b e 12e vemos os resultados obtidos para os acoplamentos e campos do modelo encontrado. Já nos Gráficos 12c e 12d, mostramos as correlações e magnetizações geradas pelo modelo em comparação com aquelas geradas pelo sistema subjacente. Mais uma vez vemos o efeito relativo das flutuações, causadas pelo Monte Carlo, quando as correlações são muito próximas de zero.

No Gráfico 13 temos os resultados para temperatura próxima da temperatura do máximo calor específico. Tal como nos demais regimes, o método encontrou um modelo que se aproxima do sistema subjacente, mas com menos precisão, como mostra os Gráficos 13a, 13b e 13e. Já nos Gráficos 13c e 13d vemos as correlações e magnetizações calculadas pelo modelo encontrado em comparação às do sistema subjacente. Assim como nos regimes precedentes, as correlações e magnetizações encontradas foram melhores nessa temperatura.

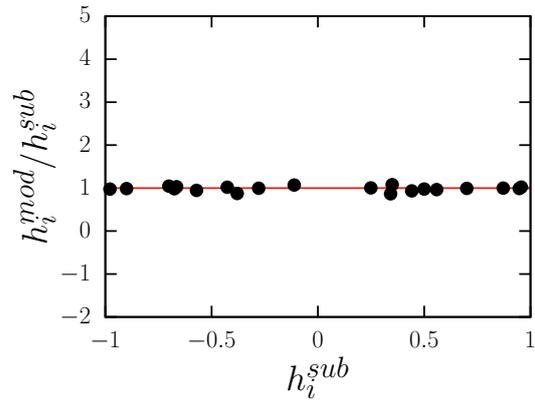
Por último temos os resultados para temperatura baixo da temperatura de máximo calor específico, presente no Gráfico 14. Como esperado, os resultados encontrados foram os mesmos dos regimes anteriores. Nos Gráficos 14a, 14b e 14e vemos que o método não foi capaz de encontrar acoplamentos e campos próximos aos do sistema subjacente. Enquanto que os Gráficos 14c e 14d mostram os resultados para as correlações e magnetizações, que mostram o mesmo comportamento dos regimes já discutidos.

Concluimos, portanto, que o método empregado é capaz de encontrar um modelo que se aproxima do sistema subjacente apenas em regiões próximas e acima da temperatura de máximo calor específico em todos os regimes estudados. Em temperaturas abaixo, apesar do método encontrar um modelo que reproduza melhor as correlações e magnetizações do sistema subjacente, os parâmetros encontrados são totalmente diferentes daquelas do subjacente. Assim, deve-se ter cautela ao aplicar tal método em dados reais, pois encontrar um modelo que reproduza as correlações e magnetizações observadas não quer dizer que o modelo esteja próximo do sistema real.

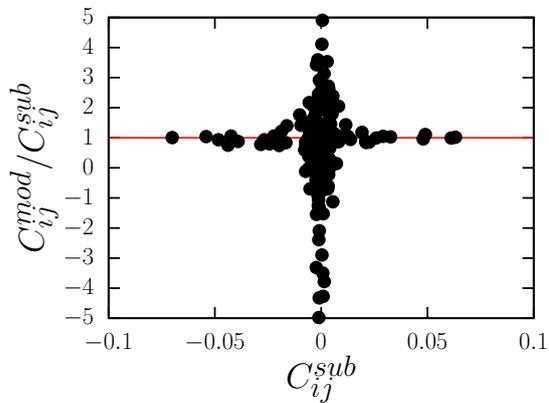
Tais resultados corroboram com os resultados obtidos por Tkacik *et al.* em [20], onde os mesmos conjecturam que a rede neural estudada por eles opera no ponto de máximo calor específico. Para verificar tal fato, Tracik *et al.* encontraram diversos modelos de tamanhos diferentes que reproduzissem as correlações e magnetizações observadas experimentalmente. Assim, os autores criaram redes sintéticas, compatíveis com as distribuições para os acoplamentos e os campos encontradas dos diversos modelos previamente obtidos, e viram que o pico do calor específico se aproximava de $T = 1$ (ponto onde o sistema real é observado, como mencionado no capítulo 3) quando aumentava o número de neurônios observado. O Gráfico 15 mostra tal resultado.

Gráfico 12 – *Spin Glass* para $T > T^*$.

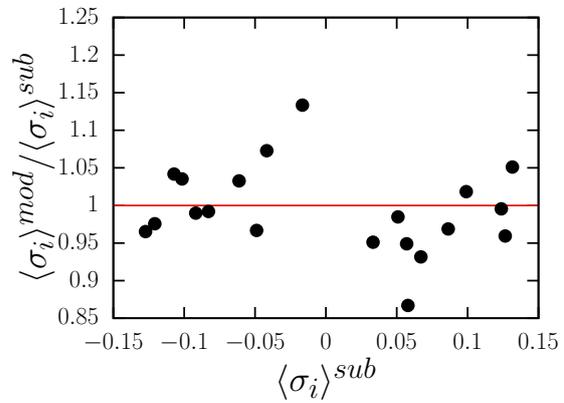
(a) Acoplamentos.



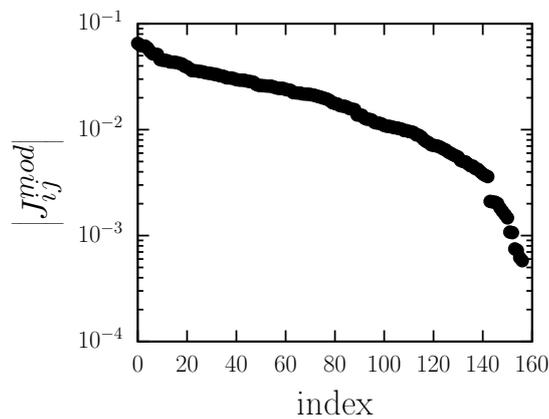
(b) Campos.



(c) Correlações.

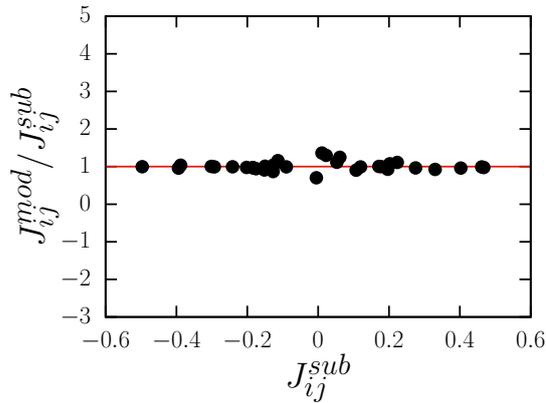


(d) Magnetizações.

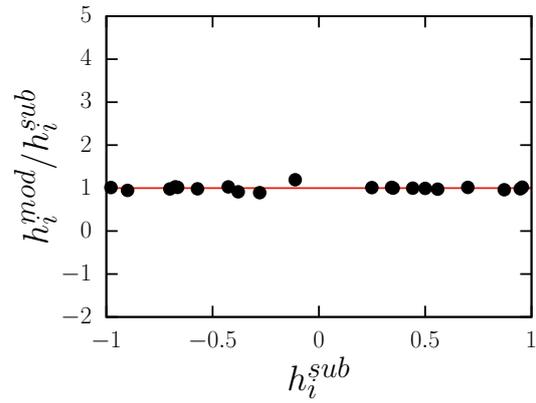


(e) Acoplamentos nulos.

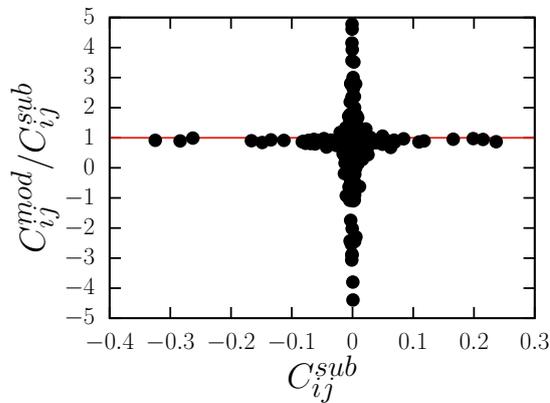
Fonte: Elaborada pelo autor. Regime *spin glass* para $T > T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle_i^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle_i^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 13 – *Spin Glass* para $T \approx T^*$.

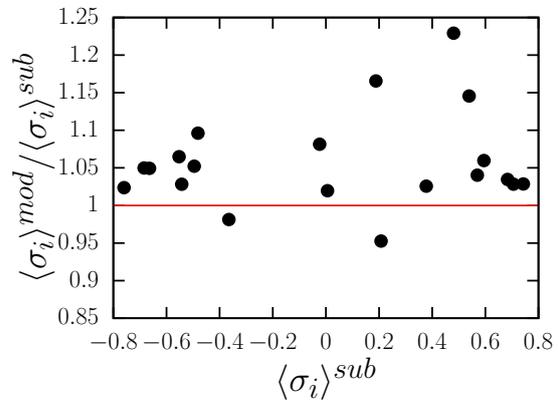
(a) Acoplamentos.



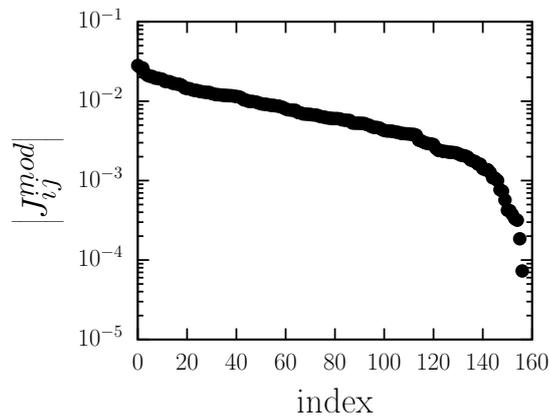
(b) Campos.



(c) Correlações.

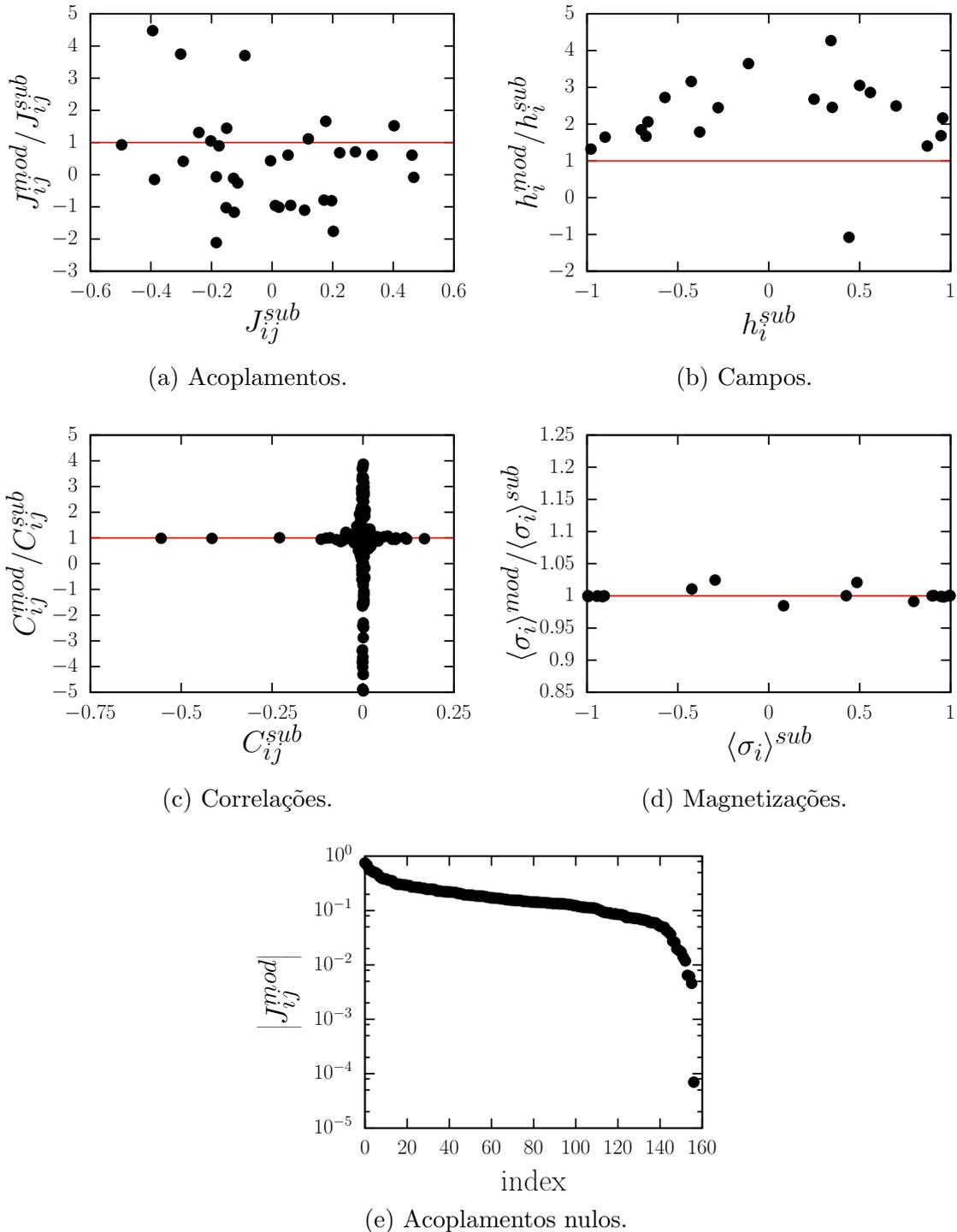


(d) Magnetizações.

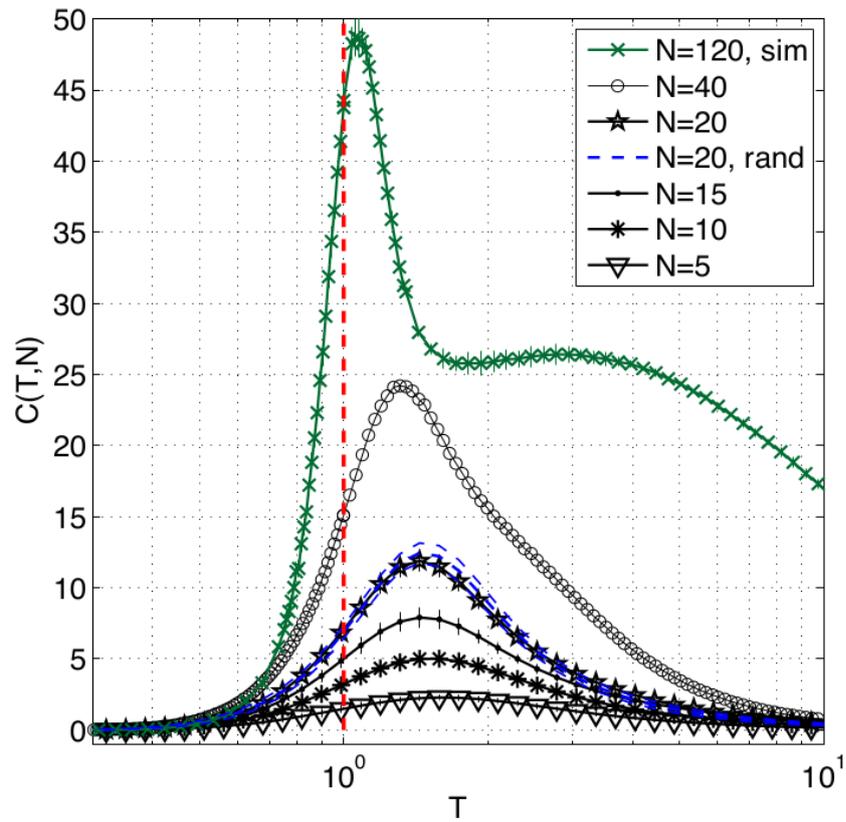


(e) Acoplamentos nulos.

Fonte: Elaborada pelo autor. Regime *spin glass* para $T \approx T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 14 – *Spin Glass* para $T < T^*$.

Fonte: Elaborada pelo autor. Regime *spin glass* para $T < T^*$. (a) e (b): Comparação dos acoplamentos e campos do modelo encontrado, J_{ij}^{mod} e h_i^{mod} , pelo método de atualização com relação aos acoplamentos e campos do sistema subjacente, J_{ij}^{sub} e h_i^{sub} . (c) e (d): Comparação das correlações e magnetizações geradas pelo modelo encontrado, C_{ij}^{mod} e $\langle \sigma_i \rangle_i^{mod}$, com relação as correlações e magnetizações do sistema subjacente, C_{ij}^{sub} e $\langle \sigma_i \rangle_i^{sub}$. (e): Magnitude dos acoplamentos do modelo encontrado que são originalmente nulos no sistema subjacente. Salientamos que os valores encontrados pelo modelo foram normalizados pelos valores do sistema subjacente, logo, quanto mais próximo da linha horizontal em 1 melhores são os resultados.

Gráfico 15 – Calor específico versus temperatura T .

Fonte: Gráfico retirado de [20]. Calor específico calculado para diversos tamanhos N . Quanto maior o número de neurônios, mais próximo o pico de máximo calor específico se aproxima de 1.

5 CONCLUSÃO E PERSPECTIVA

Para entender a natureza de sistema que geram quantidades de dados (informação) exorbitantes, usamos a teoria da informação, desenvolvida por Claude Shannon, e o princípio da máxima entropia para, assim, encontrar uma distribuição de probabilidades que consiga descrever o comportamento coletivo do sistema. Como exemplo, discutimos os resultados do grupo de pesquisadores dos Estados Unidos [20] que estudaram atividades em rede de neurônios. É mostrado que essas redes de neurônios podem ser bem descritas tendo apenas as médias de atividade de cada neurônio e as correlações par-a-par entre os mesmos. Com essas informações encontramos que a distribuição que maximiza a entropia e está de acordo com essas informações é idêntica à distribuição de Ising. Encontrar os parâmetros dessa distribuição é conhecido como *Boltzmann Machine Learning*. A conclusão de que o modelo de Ising pode descrever bem uma rede de neurônios pode ser estendida para vários outros sistemas, desde que as mesmas condições sejam obedecidas. Especificamente, cada elemento do sistema só pode estar em dois estados bem distintos (ativo e inativo) e todas as interações são devidas a efeitos externos (campos) e a interações entre pares (acoplamentos).

Nessa dissertação, desenvolvemos algoritmos para encontrar os parâmetros do modelo de Ising tendo um conjunto de dados sobre as atividades e correlações par-a-par de um sistema. Para testar os métodos desenvolvidos, criamos um conjunto de dados sintéticos a partir de um sistema subjacente e, então, testamos sua eficiência em encontrar os acoplamentos e os campos em diferentes regimes e temperaturas. Observamos que há uma limitação para o método empregado, em que o mesmo foi capaz de encontrar modelos que se aproximavam do sistema subjacente em temperaturas próximas e acima da temperatura de máximo calor específico, enquanto que em temperaturas mais baixas o método encontrou modelos que reproduzem as medidas observadas, mas com parâmetros diferentes do sistema subjacente. Concluimos, então, que devemos ter cautela ao utilizar tal método, uma vez que encontrar os acoplamentos e os campos que reproduzem os resultados observados experimentalmente não significa encontrar a estrutura do sistema.

Em trabalhos futuros pretendemos obter um banco de dados real para aplicarmos o princípio da máxima entropia e testar, com o método desenvolvido nessa dissertação, se o sistema em questão pode ser descrito por uma distribuição do tipo Ising.

Tentaremos também melhorar o modelo encontrado utilizando o critério de informação de Akaike, onde o mesmo mede a qualidade do modelo em função do *likelihood* do mesmo e do número de parâmetros usados. Desse modo, com a técnica melhorada

conjecturamos encontrar comunidades em redes complexas.

REFERÊNCIAS

- [1] CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. *Mobile Networks and Applications*, Springer, v. 19, n. 2, p. 171–209, 2014.
- [2] CHEN, C. P.; ZHANG, C.-Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, Elsevier, v. 275, p. 314–347, 2014.
- [3] HILBERT, M. Big data for development: A review of promises and challenges. *Development Policy Review*, Wiley Online Library, v. 34, n. 1, p. 135–174, 2016.
- [4] DEWDNEY, P. E. et al. The square kilometre array. *Proceedings of the IEEE*, IEEE, v. 97, n. 8, p. 1482–1496, 2009.
- [5] BRYANT, R. E. Data-intensive scalable computing for scientific applications. *Computing in Science & Engineering*, IEEE, v. 13, n. 6, p. 25–33, 2011.
- [6] EINAV, L.; LEVIN, J. Economics in the age of big data. *Science*, American Association for the Advancement of Science, v. 346, n. 6210, p. 1243089, 2014.
- [7] VARIAN, H. R. Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, American Economic Association, v. 28, n. 2, p. 3–27, 2014.
- [8] BRUMFIEL, G. et al. Down the petabyte highway. *Nature*, Macmillan Publishers Ltd., London, England, v. 469, n. 20, p. 282–283, 2011.
- [9] TURK-BROWNE, N. B. Functional interactions as big data in the human brain. *Science*, American Association for the Advancement of Science, v. 342, n. 6158, p. 580–584, 2013.
- [10] CHOUDHURY, S. et al. Big data, open science and the brain: lessons learned from genomics. *Frontiers in human neuroscience*, Frontiers, v. 8, p. 239, 2014.
- [11] STEPHENS, Z. D. et al. Big data: astronomical or genetical? *PLoS Biol*, Public Library of Science, v. 13, n. 7, p. e1002195, 2015.
- [12] LILLO, F. et al. Specialization and herding behavior of trading firms in a financial market. *New Journal of Physics*, IOP Publishing, v. 10, n. 4, p. 043019, 2008.
- [13] MORO, E. et al. Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, APS, v. 80, n. 6, p. 066102, 2009.
- [14] LACHAPPELLE, D. M. de; CHALLET, D. Turnover, account value and diversification of real traders: evidence of collective portfolio optimizing behavior. *New Journal of Physics*, IOP Publishing, v. 12, n. 7, p. 075039, 2010.
- [15] BRAUNSTEIN, A. et al. Inference algorithms for gene networks: a statistical mechanics analysis. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2008, n. 12, p. P12001, 2008.

- [16] BAILLY-BECHET, M. et al. Inference of sparse combinatorial-control networks from gene-expression data: a message passing approach. *BMC bioinformatics*, BioMed Central, v. 11, n. 1, p. 1, 2010.
- [17] SOCOLICH, M. et al. Evolutionary information for specifying a protein fold. *Nature*, Nature Publishing Group, v. 437, n. 7058, p. 512–518, 2005.
- [18] WEIGT, M. et al. Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 106, n. 1, p. 67–72, 2009.
- [19] SCHNEIDMAN, E. et al. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, Nature Publishing Group, v. 440, n. 7087, p. 1007–1012, 2006.
- [20] TKACIK, G. et al. Spin glass models for a network of real neurons. *arXiv preprint arXiv:0912.5409*, 2009.
- [21] SHLENS, J. et al. The structure of multi-neuron firing patterns in primate retina. *The Journal of neuroscience*, Soc Neuroscience, v. 26, n. 32, p. 8254–8266, 2006.
- [22] COCCO, S.; LEIBLER, S.; MONASSON, R. Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 106, n. 33, p. 14058–14062, 2009.
- [23] SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, ACM, v. 5, n. 1, p. 3–55, 2001.
- [24] GIBBS, J. W. *Elementary principles in statistical mechanics*. [S.l.]: Courier Corporation, 2014.
- [25] JAYNES, E. T. Information theory and statistical mechanics. *Physical review*, APS, v. 106, n. 4, p. 620, 1957.
- [26] JAYNES, E. T. Information theory and statistical mechanics. ii. *Physical review*, APS, v. 108, n. 2, p. 171, 1957.
- [27] GRUNWALD, P.; VITÁNYI, P. Shannon information and kolmogorov complexity. *arXiv preprint cs/0410002*, 2004.
- [28] MACKAY, D. J. *Information theory, inference and learning algorithms*. [S.l.]: Cambridge university press, 2003.
- [29] MARKOWITZ, H. Portfolio selection. *The journal of finance*, Wiley Online Library, v. 7, n. 1, p. 77–91, 1952.
- [30] COVER, T. M.; THOMAS, J. A. *Elements of information theory*. [S.l.]: John Wiley & Sons, 2012.
- [31] ROBERT, B. *Ash. Information Theory*. [S.l.]: Dover Publications Inc., New York, 1990.

- [32] KLENKE, A. *Probability theory: a comprehensive course*. [S.l.]: Springer Science & Business Media, 2013.
- [33] KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The annals of mathematical statistics*, JSTOR, v. 22, n. 1, p. 79–86, 1951.
- [34] MEZARD, M.; MONTANARI, A. *Information, physics, and computation*. [S.l.]: Oxford University Press, 2009.
- [35] ANDERSON, D. R. *Model selection and multi-model inference: a practical information-theoretic approach*. [S.l.]: Springer, 2002.
- [36] ANDERSEN, B. B.; KORBO, L.; PAKKENBERG, B. A quantitative study of the human cerebellum with unbiased stereological techniques. *Journal of Comparative Neurology*, Wiley Online Library, v. 326, n. 4, p. 549–560, 1992.
- [37] HERCULANO-HOUZEL, S.; LENT, R. Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *Journal of Neuroscience*, Soc Neuroscience, v. 25, n. 10, p. 2518–2521, 2005.
- [38] ALBERTS, B. et al. *Biologia molecular da célula*. [S.l.]: Artmed Editora, 2009.
- [39] DISPONÍVEL em: < <https://www.lecturio.com/magazine/membrane-potential/> >. Acesso em 02 de fevereiro de 2017.
- [40] RIEKE, F. *Spikes: exploring the neural code*. [S.l.: s.n.], 1999.
- [41] MCCOY, B. M.; WU, T. T. *The two-dimensional Ising model*. [S.l.]: Courier Corporation, 2014.
- [42] DOMINICIS, C. D.; GIARDINA, I. *Random fields and spin glasses: a field theory approach*. [S.l.]: Cambridge University Press, 2006.
- [43] FISCHER, K. H.; HERTZ, J. A. *Spin glasses*. [S.l.]: Cambridge university press, 1993. v. 1.
- [44] HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 79, n. 8, p. 2554–2558, 1982.
- [45] AMIT, D. J. *Modeling brain function: The world of attractor neural networks*. [S.l.]: Cambridge University Press, 1992.
- [46] ACKLEY, D. H.; HINTON, G. E.; SEJNOWSKI, T. J. A learning algorithm for boltzmann machines. *Cognitive science*, Wiley Online Library, v. 9, n. 1, p. 147–169, 1985.
- [47] HINTON, G. E.; SEJNOWSKI, T. J. Learning and relearning in boltzmann machines. *Parallel Distributed Processing*, v. 1, 1986.
- [48] KAPPEN, H. J.; RODRÍGUEZ, F. d. B. Efficient learning in boltzmann machines using linear response theory. *Neural Computation*, MIT Press, v. 10, n. 5, p. 1137–1156, 1998.

- [49] TANAKA, T. Mean-field theory of boltzmann machine learning. *Physical Review E*, APS, v. 58, n. 2, p. 2302, 1998.
- [50] SESSAK, V.; MONASSON, R. Small-correlation expansions for the inverse ising problem. *Journal of Physics A: Mathematical and Theoretical*, IOP Publishing, v. 42, n. 5, p. 055001, 2009.
- [51] ROUDI, Y.; AURELL, E.; HERTZ, J. Statistical physics of pairwise probability models. *arXiv preprint arXiv:0905.1410*, 2009.
- [52] RICCI-TERSENGHI, F. The bethe approximation for solving the inverse ising problem: a comparison with other inference methods. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2012, n. 08, p. P08015, 2012.
- [53] MÉZARD, M.; MORA, T. Constraint satisfaction problems and neural networks: A statistical physics perspective. *Journal of Physiology-Paris*, Elsevier, v. 103, n. 1, p. 107–113, 2009.
- [54] AURELL, E.; OLLION, C.; ROUDI, Y. Dynamics and performance of susceptibility propagation on synthetic data. *The European Physical Journal B*, Springer, v. 77, n. 4, p. 587–595, 2010.
- [55] MARINARI, E.; KERREBROECK, V. V. Intrinsic limitations of the susceptibility propagation inverse inference for the mean field ising spin glass. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2010, n. 02, p. P02008, 2010.
- [56] AURELL, E.; EKEBERG, M. Inverse ising inference using all the data. *Physical review letters*, APS, v. 108, n. 9, p. 090201, 2012.
- [57] OPPER, M.; SAAD, D. *Advanced mean field methods: Theory and practice*. [S.l.]: MIT press, 2001.
- [58] THOULESS, D. J.; ANDERSON, P. W.; PALMER, R. G. Solution of 'solvable model of a spin glass'. *Philosophical Magazine*, Taylor & Francis, v. 35, n. 3, p. 593–601, 1977.
- [59] PARISI, G.; SHANKAR, R. *Statistical field theory*. [S.l.]: AIP, 1988.
- [60] PLEFKA, T. Convergence condition of the tap equation for the infinite-ranged ising spin glass model. *Journal of Physics A: Mathematical and general*, IOP Publishing, v. 15, n. 6, p. 1971, 1982.
- [61] ERDÖS, P.; RÉNYI, A. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, v. 6, p. 290–297, 1959.
- [62] BOLLOBÁS, B. Random graphs. In: *Modern Graph Theory*. [S.l.]: Springer, 1998. p. 215–252.
- [63] NEWMAN, M.; BARKEMA, G. *Monte Carlo Methods in Statistical Physics chapter 1-4*. [S.l.]: Oxford University Press: New York, USA, 1999.
- [64] ROBERT, C. P. *Monte carlo methods*. [S.l.]: Wiley Online Library, 2004.
- [65] METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, AIP, v. 21, n. 6, p. 1087–1092, 1953.