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

WAGNER RODRIGUES DE SENA

NOVO MODELO HIERÁRQUICO PARA DECOMPOSIÇÃO DO *BACKBONE* DE PERCOLAÇÃO REVELA NOVAS LEIS DE ESCALA DA DISTRIBUIÇÃO DE CORRENTES

FORTALEZA
2023

WAGNER RODRIGUES DE SENA

NOVO MODELO HIERÁRQUICO PARA DECOMPOSIÇÃO DO *BACKBONE* DE PERCOLAÇÃO REVELA NOVAS LEIS DE ESCALA DA DISTRIBUIÇÃO DE CORRENTES

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

Orientador: Prof. Dr. André Auto Moreira.

FORTALEZA
2023

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

S477n Sena, Wagner Rodrigues de.
Novo modelo Hierárquico para decomposição do backbone de percolação revela novas leis de escala da distribuição de correntes / Wagner Rodrigues de Sena. – 2023.
93 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Física , Fortaleza, 2023.
Orientação: Prof. Dr. André Auto Moreira.

1. Percolação. 2. Rede de resistores aleatórios. 3. Decomposição de grafos. 4. Componente biconectadas. 5. Componente triconectada. I. Título.

CDD 530

WAGNER RODRIGUES DE SENA

NOVO MODELO HIERÁRQUICO PARA DECOMPOSIÇÃO DO *BACKBONE* DE PERCOLAÇÃO REVELA NOVAS LEIS DE ESCALA DA DISTRIBUIÇÃO DE CORRENTES

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

Aprovada em 24/02/2023.

BANCA EXAMINADORA

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

Prof. Dr. Hans Jürgen Herrmann
Universidade Federal do Ceará (UFC)

Prof. Dr. César Ivan Nunes Sampaio Filho
Universidade Federal do Ceará (UFC)

Prof. Dr. César Menezes Vieira
Instituto Federal do Ceará (IFCE - Acaraú)

Prof. Dr. Rilder de Sousa Pires
Universidade de Fortaleza (UNIFOR)

Aos Meus Pais

e

Amigos

AGRADECIMENTOS

Agradeço principalmente aos meus pais por sempre me incentivarem e me darem condições para sempre focar nos estudos desde criança, a minha irmã que esteve presente quando precisei e a minha namorada por todo apoio, conversas e incentivo na minha jornada.

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

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, Pedro Henrique e Laura Barth por sempre estarem presentes em todas as dificuldades e alegrias encontradas ao longo do caminho da graduação ao doutorado.

Agradeço aos meus amigos do laboratório de sistemas complexos, Marciel Carvalho, Israel Nascimento, Samuel Morais, Emanuel Fontelles, Felipe Operti, Edson Ares e Débora Torres pelas conversas e trocas de ideias ao longo dos projetos e desafios encontrados.

E por fim, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (**CNPq**) pelo apoio financeiro.

RESUMO

Neste trabalho, realizamos dois estudos. Primeiro, aplicamos o algoritmo de *Boltzmann Machine Learning* para analisar os dados de movimentos oculares durante a leitura de diferentes tipos de textos. Encontramos que podemos descrever a complexidade dos textos por meio da magnetização média e que a distância entre a temperatura de leitura e crítica ($T_o - T_c$) é capaz de refletir a coerência dos mesmos. Segundo, estudamos as propriedades do agregado de percolação, mais especificamente, a aplicação do modelo de percolação de rede de resistores aleatórios, no qual cada ligação do sistema possui uma resistência. Quando uma corrente é introduzida na rede, ela é distribuída por cada ligação de acordo com as leis de *Kirchhoff*. Muitas das propriedades do sistema de resistores aleatórios no ponto crítico de percolação são obtidas a partir da distribuição de probabilidades das correntes, $P(i)$. Antigamente, acreditava-se que a distribuição de correntes no *backbone* no ponto crítico de percolação seguia uma distribuição log-normal, como um modelo hierárquico simples. No entanto, posteriormente, foi observado que a distribuição de correntes no *backbone* do agregado crítico de percolação não seguia uma log-normal. Devido à autossimilaridade do *backbone* crítico de percolação, criamos um modelo hierárquico pela decomposição do *backbone* em componentes triconectadas. Isso nos permitiu descobrir que a distribuição de correntes é formada por duas distribuições: uma corresponde à distribuição do número de ligações em cada nível, e a outra corresponde às distribuições dos fatores multiplicativos que compõem as correntes em cada nível. Nossa metodologia de decomposição do *backbone* também nos permitiu medir correntes pequenas com precisão de até 10^{-35} para sistemas de tamanho $L = 8192$.

Palavras-chave: percolação; rede de resistores aleatórios; decomposição de grafos; componente biconectadas; componente triconectada.

ABSTRACT

In this work, we conducted two studies. First, we applied the Boltzmann Machine Learning algorithm to analyze eye movement data during the reading of different types of texts. We found that we can describe the complexity of texts through the average magnetization, and that the distance between the reading temperature and critical temperature ($T_o - T_c$) is capable of reflecting their coherence. Second, we studied the properties of the percolation aggregate, more specifically, the application of the percolation model in the random resistors network, where each bond in the network has a resistance. When a current is introduced into the network, it is distributed through each bond according to Kirchhoff's laws. Many of the properties of the random resistor system are obtained from the probability distribution $P(i)$. Previously, it was believed that the distribution of currents in the backbone at the critical percolation point followed a log-normal distribution, like a simple hierarchical model. However, later it was observed that the distribution of currents in the backbone of the critical percolation aggregate did not follow a log-normal distribution. Due to the self-similarity of the critical percolation backbone, we created a hierarchical model by decomposing the backbone into triconnected components. This allowed us to discover that the distribution of currents is formed by two distributions: one corresponds to the distribution of the number of connections at each level, and the other corresponds to the distributions of the multiplicative factors that make up the currents at each level. Our decomposition methodology also allowed us to accurately find small currents up to 10^{-35} for systems up to $L = 8192$.

Keywords: percolation; random resistor network; graph decomposition; biconnected component; triconnected component.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Durante o período de pesquisa de doutorado, trabalhei em dois projetos diferentes. No primeiro, em colaboração com D. Torres e outros, aplicamos o algoritmo de *Boltzmann Machine Learning* para estudar a complexidade e coerência de textos [1]. Já no segundo projeto (principal foco dessa tese), desenvolvemos uma nova metodologia para determinar a distribuição de correntes em um *backbone* crítico de percolação.

1.1 Motivação

As propriedades do agregado de percolação é objeto de inúmeros estudos [2, 3]. No caso de percolação de ligações, cada ligação de uma rede regular é considerada ocupada ou desocupada de acordo com uma probabilidade p. Uma característica desse modelo é a existência de um ponto crítico, p_c, onde é formado um agregado de vértices ligados que se estende por toda a rede. Sabe-se que abaixo do ponto crítico existe uma escala característica para o tamanho dos agregados observados. Isto é, a probabilidade de encontrar um agregado maior que essa escala decai exponencialmente com o tamanho do agregado. Acima do ponto crítico existe um agregado que se estende por todo o sistema ocupando homogeneamente uma fração da rede. Exatamente no ponto crítico o maior agregado é um fractal. Uma aplicação interessante do modelo de percolação é a rede de resistores aleatórios [4–6], onde cada ligação da rede possui uma resistência. Uma corrente é introduzida na rede e distribuída por cada ligação, de acordo com as leis de Kirchhoff. Resolvendo o conjunto de equações, obtidas pela a aplicação da lei de Kirchhoff em cada vértice, podemos encontrar a distribuição de correntes $p(i)$. Muitas das propriedades do sistema de resistores aleatórios são obtidas a partir da distribuição de probabilidade $p(i)$ [7–16]. Na Figura 1, vemos o exemplo do *backbone* de um agregado no ponto crítico de percolação, onde cada ligação possui uma resistência e transporta uma corrente. O *backbone* é o conjunto de ligações do agregado percolante por onde passa corrente. Veremos adiante que o *backbone* condutor é um componente biconectado do agregado crítico de percolação.

Arcangelis *et al.* [15, 16] sugeriram que a distribuição de correntes, do *backbone* no ponto crítico de percolação, seguia uma distribuição log-normal, tal como de um modelo hierárquico. Eles mostraram que a dimensão fractal, $\phi(x)$, do *backbone* estava relacionada com a distribuição do número de correntes, $n(i)$, da seguinte forma:

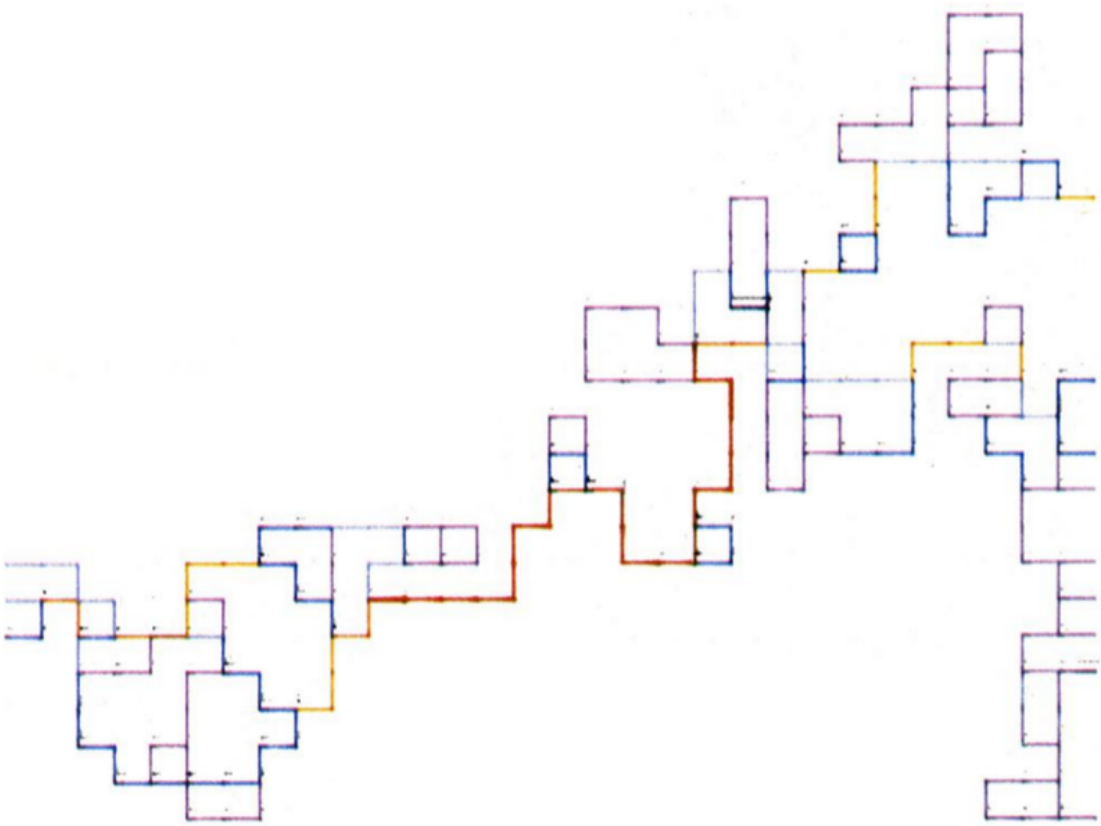
$$n(i) = L^{\phi(x)}, \tag{1.1}$$

onde $x = \ln i / \ln L$ e L é a dimensão da rede. Resolvendo para a dimensão fractal, vemos que

$$\phi(x) = \frac{\ln n(i)}{\ln L}. \tag{1.2}$$

No Gráfico 1, observamos o resultado de Arcangelis *et al.* para a dimensão fractal de uma rede

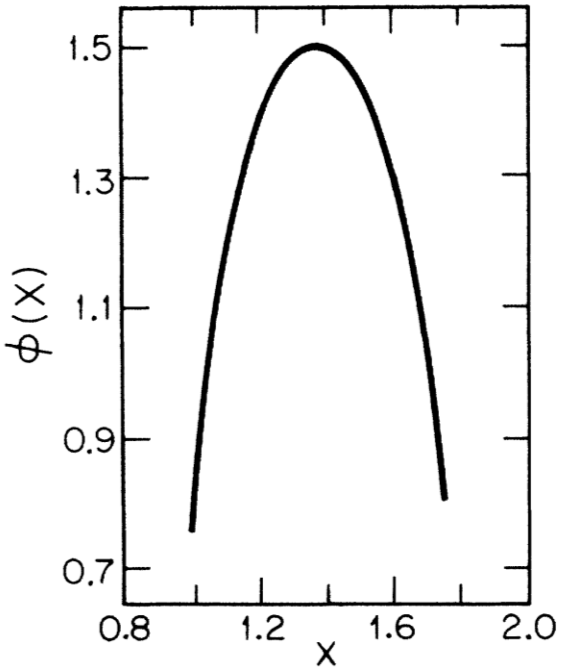
Figura 1 – *Backbone* de uma rede de resistores aleatórios.



Fonte: Figura retirada de [16]. *Backbone* (ligações por onde há passagem de corrente) de um agregado de percolação, onde cada ligação possui uma resistência e por ela passa uma corrente. As ligações em vermelho são aquelas em que se passa toda a corrente do circuito. As cores em ordem da maior corrente para a menor são: vermelho, marrom, laranja, amarelo, azul claro, azul escuro e violeta.

hierárquica, que segue o logaritmo de uma distribuição log-normal. Logo, pela Equação 1.2, vemos que a distribuição de correntes, $n(i)$, segue uma log-normal.

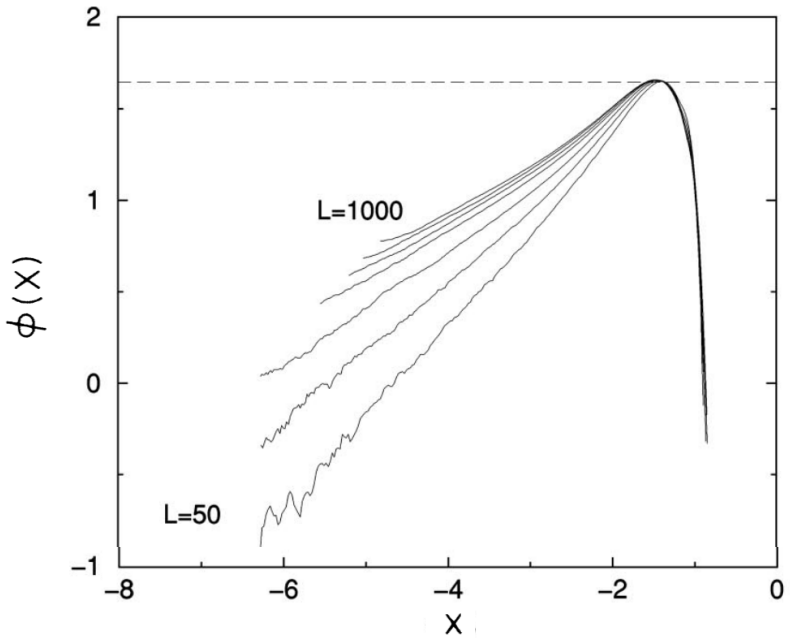
Gráfico 1 – Dimensão fractal obtida por Arcangelis *et al.*.



Fonte: Figura retirada de [15]. Dimensão fractal de uma rede hierárquica. A dimensão fractal está relacionada com a distribuição do número de correntes por $\phi(x) = \ln n(i)/\ln L$, onde $x = \ln i/\ln L$.

Posteriormente, observou-se que a distribuição de correntes no *backbone* do agregado crítico de percolação não seguia uma log-normal. De fato, para pequenas correntes, a distribuição segue uma lei de potência, $p(i) \sim i^{b-1}$ [17, 18]. No Gráfico 2, mostramos o resultado obtido por Barthélémy *et al.* [19], onde vemos que a dimensão fractal do *backbone*, dado pela Equação 1.2, claramente não segue uma distribuição normal. Logo, a de correntes não é log-normal.

Gráfico 2 – Dimensão fractal obtida por Barthélémy *et al.*.



Fonte: Gráfico retirado de [19] com alteração do autor. Dimensão fractal de uma rede de resistores aleatórios no ponto crítico de percolação. A dimensão fractal está relacionada com a distribuição do número de correntes por $\phi(x) = \ln n(i)/\ln L$, onde $x = \ln i/\ln L$. Vemos, agora, que a distribuição do número de correntes não segue uma log-normal. Percebemos que o número de baixas correntes fica mais evidente quando aumentamos o tamanho do sistema.

1.2 Estrutura da tese

No Capítulo 2, apresentamos uma breve introdução sobre o princípio da máxima entropia e o algoritmo de *Boltzmann Machine Learning*. Em seguida, aplicamos o algoritmo nos dados de movimento ocular e encontramos uma relação entre a magnetização das ativações de palavras durante a leitura e a complexidade dos textos. Também vimos uma relação entre a temperatura crítica e a coerência dos textos, sugerindo que essas duas quantidades podem ser usadas para medir a complexidade e coerência de textos.

O Capítulo 3 é dedicado para a introdução da decomposição de grafos em componentes biconectadas e triconectadas. Mostramos que esses dois objetos possuem dimensão fractal no ponto crítico de percolação de uma rede quadrada de ligação. Isso permitiu criar um modelo hierárquico de níveis, onde cada ligação do *backbone* possui um nível associado.

Já no Capítulo 4, apresentamos como a decomposição do *backbone* crítico de percolação em componentes triconectadas e o modelo hierárquico proposto pode ser utilizado para determinar a distribuição de correntes com alta precisão, assim, revelando uma nova lei de escala das distribuições.

For fim, no Capítulo 5, trazemos as principais conclusões ao aplicar a nova metodologia para determinar a distribuição de correntes no *backbone* crítico de percolação. No Anexo 1, temos o artigo "*Decomposing the Percolation Backbone Reveals Novel Scaling Laws of the*

Current Distribution" produzido mostrando os resultados obtidos pela nova metodologia de determinação da distribuição de correntes. O Anexo 2 é o artigo "*Eye-tracking as a proxy for coherence and complexity of texts*", desenvolvido com D. Torres e outros sobre a complexidade e coerência de textos a partir de dados de movimento ocular, publicado na revista *PLOS ONE* [1].

2 MODELO DE MÁXIMA ENTROPIA E COMPLEXIDADE EM TEXTOS

2.1 Princípio da Máxima Entropia

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

$$\langle f(x) \rangle = \sum_{i=1}^{n} p_i f(x_i), \tag{2.1}$$

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 2.1 e a condição de normalização

$$\sum_{i}^{n} p_i = 1, \tag{2.2}$$

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 [20], 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 $\{p_i\}$, mas sim, maximizar a entropia dada por:

$$H(p_1, \ldots, p_n) = -c \sum_{i=1}^{n} p_i \ln p_i \tag{2.3}$$

Para maximizar a entropia (Equação 2.3) sujeita as restrições 2.1 e 2.2, podemos fazer uso dos multiplicadores de Lagrange $\{\lambda_0, \lambda_1\}$, de tal modo que a lagrangiana fica:

$$L = -\sum_{i=1}^{n} p_i \ln p_i + \lambda_0 \left(\sum_{i}^{n} p_i - 1 \right) + \lambda_1 \left(\sum_{i=1}^{n} p_i f(x_i) - \langle f(x) \rangle \right), \tag{2.4}$$

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

$$\frac{\partial L}{\partial p_i} = 0 \Rightarrow \tag{2.5}$$

$$-(\ln p_i + 1) + \lambda_0 + \lambda_1 f(x_i) = 0 \Rightarrow \tag{2.6}$$

$$p_i = e^{-1+\lambda_0+\lambda_1 f(x_i)} = p_0 e^{\lambda_1 f(x_i)}, \tag{2.7}$$

onde $p_0 = e^{1-\lambda_0}$ pode ser determinado pela restrição (Equação 2.2). Utilizando a condição de normalização (Equação 2.2) e sabendo que p_0 é constante, temos:

$$\sum_i p_0 e^{\lambda_1 f(x_i)} = 1 \Rightarrow \tag{2.8}$$

$$p_0 = \frac{1}{Z(\lambda_1)}, \tag{2.9}$$

onde,

$$Z(\lambda_1) = \sum_i e^{\lambda_1 f(x_i)}, \tag{2.10}$$

é chamada de função de partição. Já λ_1 é determinado mediante a restrição 2.1, que, em termos da função de partição $Z(\lambda_1)$, pode ser reescrita como:

$$\langle f(x) \rangle = \frac{\partial}{\partial \lambda_1} \ln Z(\lambda_1). \tag{2.11}$$

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

$$\langle f_j(x) \rangle = \sum_i p_i f_j(x_i), \quad (j = 1, 2, \ldots, m), \tag{2.12}$$

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

$$p_i = \frac{1}{Z(\lambda_1, \lambda_2, \ldots, \lambda_m)} e^{\lambda_1 f_1(x_i) + \lambda_2 f_2(x_i) + \cdots + \lambda_m f_m(x_i)}, \tag{2.13}$$

com a função de partição da seguinte forma:

$$Z(\lambda_1, \lambda_2, \ldots, \lambda_m) = \sum_i e^{\lambda_1 f_1(x_i) + \lambda_2 f_2(x_i) + \cdots + \lambda_m f_m(x_i)}. \tag{2.14}$$

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

$$\langle f_j(x) \rangle = \frac{\partial}{\partial \lambda_j} \ln Z(\lambda_1, \lambda_2, \ldots, \lambda_m), \quad (j = 1, 2, \ldots, m). \tag{2.15}$$

2.2 Modelos com correlação par-a-par de dois estados

Seja um sistema com N elementos binários $\{\sigma_1, \sigma_2, ..., \sigma_N\}$, onde os estados permitidos para cada elemento são $\sigma_i = +1$ e $\sigma_i = -1$. Para um número de M amostras experimentais do sistema, podemos dizer que um estado t é dado pelo conjunto de variáveis binárias $\sigma^{(t)} = \{\sigma_i\}$:

$$\sigma^{(1)} = \{+1,-1,-1,\ldots,+1,+1,-1\}, \tag{2.16}$$
$$\sigma^{(2)} = \{-1,-1,+1,\ldots,+1,-1,-1\}, \tag{2.17}$$
$$\vdots$$
$$\sigma^{(M)} = \{-1,+1,-1,\ldots,+1,+1,+1\}. \tag{2.18}$$

Queremos então encontrar a distribuição de probabilidades $P_{exp}(\sigma)$ que reproduz os estados observados experimentalmente. Com o procedimento desenvolvido na seção anterior de máxima entropia, 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 elemento, $\langle\sigma_i\rangle$, e a correlação entre os mesmos, $C_{ij} = \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle$, onde

$$\langle\sigma_i\rangle = \sum_{\{\sigma\}} \sigma_i P_{exp}(\sigma) = \frac{1}{M}\sum_{k=1}^{M} \sigma_i^{(k)} \tag{2.19}$$

e

$$\langle\sigma_i\sigma_j\rangle = \sum_{\{\sigma\}} \sigma_i\sigma_j P_{exp}(\sigma) = \frac{1}{M}\sum_{k=1}^{M} \sigma_i^{(k)}\sigma_j^{(k)}. \tag{2.20}$$

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 atividades 2.19 e correlações 2.20, o princípio da máxima entropia implica que a distribuição de estados é da forma:

$$P(\sigma) = \frac{1}{Z} e^{\sum_{i<j}\alpha_{ij}\sigma_i\sigma_j + \sum_i \lambda_i\sigma_i}, \tag{2.21}$$

$$Z = \sum_{\{\sigma\}} e^{\sum_{i<j}\alpha_{ij}\sigma_i\sigma_j + \sum_i \lambda_i\sigma_i}, \tag{2.22}$$

onde o conjunto $\{\alpha_{ij},\lambda_i\}$ são os multiplicadores de Lagrange. Vemos que a distribuição obtida é igual a do modelo de Ising [21–23] para spins que interagem par-a-par sob a influência de um campo magnético com energia:

$$H_{Ising}(\sigma) = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i, \tag{2.23}$$

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:

$$P_{Ising}(\sigma) = \frac{1}{Z_{Ising}} e^{\beta\left(\sum_{i<j} J_{ij}\sigma_i\sigma_j + \sum_{i=1}^{N} h_i\sigma_i\right)}, \tag{2.24}$$

$$Z_{Ising} = \sum_{\{\sigma\}} e^{\beta\left(\sum_{i<j} J_{ij}\sigma_i\sigma_j + \sum_{i=1}^{N} h_i\sigma_i\right)}, \tag{2.25}$$

onde β é o inverso da temperatura, $\beta = (k_B T)^{-1}$. Então, a distribuição encontrada, Equação 2.21, é o modelo de Ising para $\beta = 1$. 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 usada como uma analogia, mas sim um mapeamento.

2.3 Problema inverso de Ising

Vimos que a distribuição de menor estrutura que está de acordo com as médias experimentais $\langle\sigma_i\rangle$ e correlações C_{ij}, é semelhante à distribuição do modelo de Ising quando $\beta = 1$. Precisamos, então, encontrar uma forma de determinar os valores dos acoplamentos $\mathbf{J} = \{J_{ij}\}$ e dos campos $\mathbf{h} = \{h_i\}$. O processo de encontrar os acoplamentos e campos é conhecido como problema inverso de Ising ou *Boltzmann Machine Learning*, como é chamado na ciência da computação [24, 25].

Precisamos resolver o conjunto de Equações 2.15 para $\mathbf{J}$ e $\mathbf{h}$

$$\langle\sigma_i\rangle = \frac{\partial}{\partial h_i} \ln Z(\mathbf{J},\mathbf{h}) \tag{2.26}$$

e

$$\langle\sigma_i\sigma_j\rangle = \frac{\partial}{\partial J_{ij}} \ln Z(\mathbf{J},\mathbf{h}). \tag{2.27}$$

Resolver tais equações de forma exata se torna inviável quando N é muito grande, pois o número de estados que devem ser somados na função de partição $Z(\mathbf{J},\mathbf{h})$ cresce exponencialmente com 2^N. Assim, vamos buscar por métodos aproximativos. 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 [26–28]:

$$D_{KL}(P_{exp}||P_{ising}) = \sum_{\{\sigma\}} P_{exp}(\sigma) \ln \frac{P_{exp}(\sigma)}{P_{ising}(\sigma)}. \tag{2.28}$$

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

$$\begin{aligned}
\frac{\partial D_{KL}(P_{exp}||P_{ising})}{\partial J_{ij}} &= \frac{\partial}{\partial J_{ij}} \sum_{\{\sigma\}} P_{exp}(\sigma) \left[\ln P_{exp}(\sigma) - \ln P_{Ising}(\sigma)\right] \\
&= -\sum_{\{\sigma\}} P_{exp}(\sigma) \left[\frac{\partial}{\partial J_{ij}} \ln P_{Ising}(\sigma)\right] \\
&= -\sum_{\{\sigma\}} P_{exp}(\sigma) \left(\sigma_i\sigma_j - \frac{1}{Z}\sum_{\{\sigma\}} \sigma_i\sigma_j e^{\sum_{i<j} J_{ij}\sigma_i\sigma_j + \sum_{i=1}^{N} h_i\sigma_i}\right) \\
&= -\left(\langle\sigma_i\sigma_j\rangle_{exp} - \langle\sigma_i\sigma_j\rangle_{Ising}\right), \qquad (2.29)
\end{aligned}$$

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 2.29 leva à seguinte regra de atualização [24]:

$$\Delta J_{ij}(t+1) = \varepsilon\left[\langle\sigma_i\sigma_j\rangle_{exp} - \langle\sigma_i\sigma_j\rangle_{Ising}(t)\right], \qquad (2.30)$$

$$J_{ij}(t+1) = J_{ij}(t) + \Delta J_{ij}(t+1), \qquad (2.31)$$

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

$$\Delta h_i(t+1) = \varepsilon\left[\langle\sigma_i\rangle_{exp} - \langle\sigma_i\rangle_{Ising}(t)\right], \qquad (2.32)$$

$$h_i(t+1) = h_i(t) + \Delta h_i(t+1). \qquad (2.33)$$

O que a regra diz é: dados os 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 e usamos tais médias para atualizar os novos acoplamentos e campos. Após isso, utilizamos 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 [24]. Isso garante que o método descrito acima irá atingir o mínimo global.

2.4 Complexidade e coerência em textos

Em colaboração com D. Torres e outros, aplicamos o modelo inverso de Ising para estudar a complexidade e coerência de textos [1]. A complexidade e a coerência de um texto são consideradas atributos linguísticos cruciais na avaliação da compreensão da leitura e das dificuldades de aprendizagem, como relatado em [29, 30]. Nas últimas décadas, pesquisadores linguísticos têm se concentrado muito na medição da complexidade dos textos [31]. Isso levou ao desenvolvimento de expressões matemáticas e métricas para quantificar a complexidade dos textos e classificar o material de leitura [28]. Algumas das variáveis comumente usadas incluem

o comprimento médio das palavras e sua frequência na língua, que contribuem para a dificuldade semântica e o comprimento da frase, que está associado à complexidade sintática.

Acredita-se que os padrões de movimento dos olhos de indivíduos durante a leitura de trechos de texto possam espelhar características como gênero e estilo do mesmo. Como resultado, diferentes tipos de textos poderiam apresentar diferentes respostas de leitura em termos de configurações de fixação e reações cognitivas. Para examinar essa interação, uma abordagem de modelagem que utiliza dados de movimento dos olhos, padrões de fixação, é utilizada para explorar os processos cognitivos internos envolvidos na leitura.

Na Figura 2, mostramos o fluxo utilizado para caracterizar a complexidade $\langle \pi \rangle$ e a coerência $\langle \psi \rangle$ de textos de forma quantitativa, tomando uma abordagem dupla. Na primeira abordagem, realizamos experimentos de rastreamento ocular (*eye-tracking*) [32–37] com um grupo limitado de 20 pessoas brasileiras entre graduandos de Física, Engenharia e pós graduação com idade entre 17 e 34 anos para coletar diretamente seus dados de fixação enquanto liam diferentes textos, incluindo histórias infantis, obras literárias e textos gerados aleatoriamente com palavras. A Tabela 1 mostra mais detalhes de quais textos foram utilizados. Aplicar o modelo inverso de Ising nos dados coletados de fixação nos permitiu revelar dois índices que podem ser usados para medir a complexidade e coerência de textos: a magnetização e a distância entre a "temperatura de operação"do sistema e sua temperatura crítica. Já na segunda abordagem, nosso experimento é validado através de um questionário quantitativo, com acesso a uma vasta amostra de respondentes, para categorizar os mesmos textos de acordo com diferentes níveis de complexidade e coerência, permitindo assim uma comparação direta com os índices obtidos pelos dados coletados de rastreamento ocular.

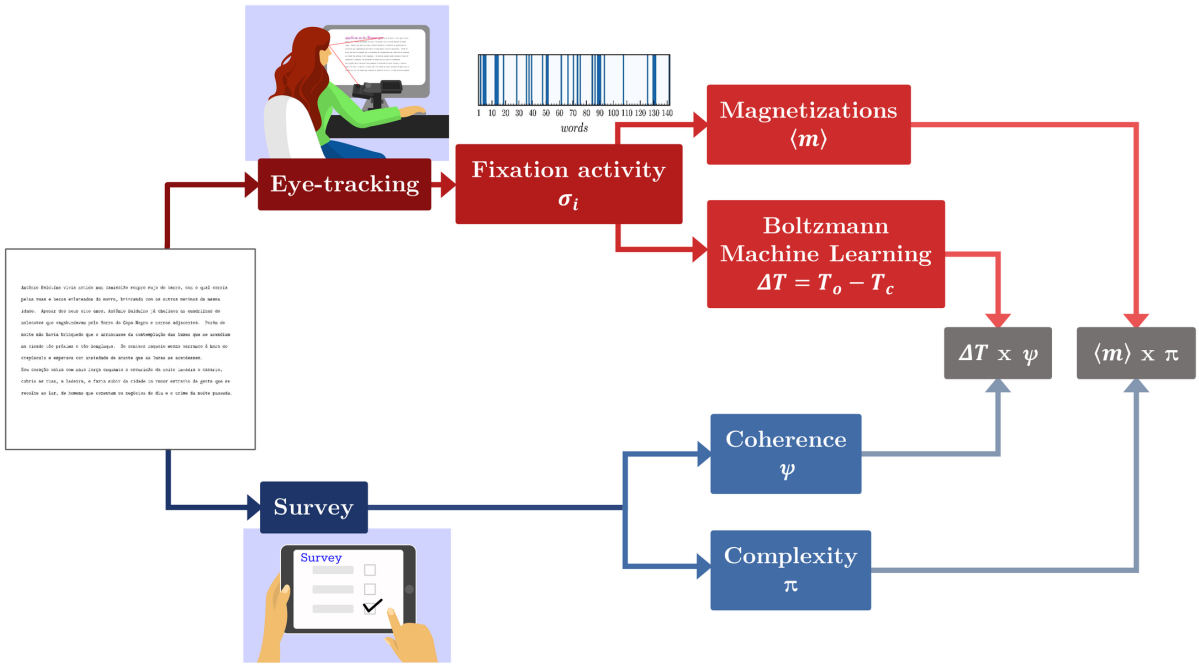
Tabela 1 – Textos utilizados no experimento

Símbolo	**Título**	**Autor**	**Ano**	**País**
GAU	O Gaúcho	José de Alencar	1870	Brasil
GSV	Grande Sertão: Veredas	João Guimarães Rosa	1956	Brasil
HCL	História do Cerco de Lisboa	José Saramago	1989	Portugal
JUB	Jubiabá	Jorge Amado	1935	Brasil
MEL	A Mão e a Luva	Machado de Assis	1874	Brasil
QUI	O Quinze	Rachel de Queiroz	1930	Brasil
RT1	Random text 1	-	-	-
RT2	Random text 2	-	-	-
ST1	Story 1: A patinha Esmeralda	-	-	Brasil
ST2	Story 2: A menina do leite	-	-	Brasil

Fonte: Tabela retirada de [1]. Tabela mostra os 10 textos utilizados no experimento. Todos os textos são escritos em português brasileiro. Os textos ST1 e ST2 são textos populares de conto de fadas para crianças. Os textos RT1 e RT2 foram gerados aleatoriamente.

Para criar um paralelo com o modelo de Ising, definimos a seguinte regra para a atividade de fixação $\sigma_i = \{\sigma_i^{(1)}, \sigma_i^{(2)}, ..., \sigma_i^{(M)}\}$ de cada participante i lendo um texto com M palavras:

Figura 2 – Fluxo do experimento para coleta de dados.



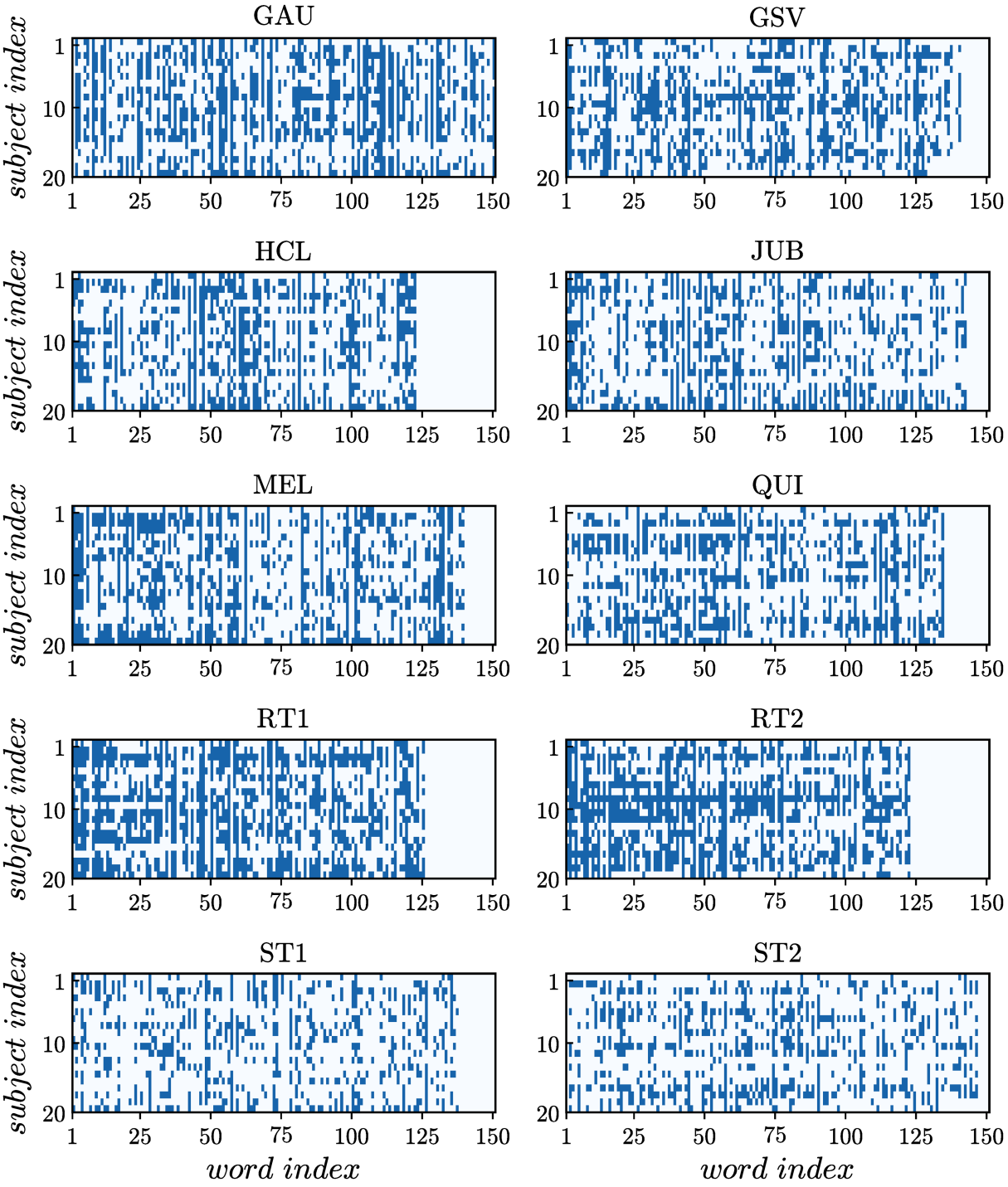
Fonte: Figura retirada de [1]. O experimento foi conduzido em dois estágios para quantificar a complexidade e coerência de textos. Primeiro, realizamos um experimento de rastreamento ocular para coletar dados de fixações de um grupo de 20 participantes, convertendo o estado de cada palavra em uma representação binária, com um valor de +1 se a palavra foi fixada pelo menos duas vezes e -1 se a palavra foi fixada apenas uma vez ou nenhuma vez. Em seguida, determinamos a "magnetização"do texto para cada participante, tirando a média de suas leituras individuais e obtendo o valor de $\langle m_i \rangle$. Além disso, aplicamos o problema inverso de Ising, ou *Boltzmann Machine Learning*, para determinar a proximidade de cada texto com o ponto crítico do "calor específico". Em paralelo, também coletamos dados a partir de um questionário quantitativo com 400 participantes sobre a complexidade $\langle \pi \rangle$ e coerência $\langle \psi \rangle$ dos textos.

$$\sigma_i^{(r)} = \begin{cases} +1 & se \;\; n_i^r \geq 2 \\ -1 & se \;\; n_i^r > 2, \end{cases} \tag{2.34}$$

onde n_i^r é o número de vezes que a pessoa i fixou na palavra r durante a leitura do texto. O *threshold* de 2 fixações por palavra foi estabelecido como critério para determinar se uma palavra representa um estado ativo ou não no texto. Isso se baseia nos resultados de nossos experimentos de rastreamento ocular, onde foi observado que quase todas as palavras em qualquer texto foram fixadas pelo menos uma vez por todas as pessoas durante as leituras. A Figura 3 mostra o mapa de ativações para cada pessoa em cada um dos texto do experimento. Em azul são as palavras ativadas ($\sigma_i^{(r)} = +1$) com mais de duas fixações e em branco as palavras que não foram ativadas ($\sigma_i^{(r)} = -1$).

Uma vez que encontramos os valores dos campos h_i e acoplamento J_{ij} para todos os participantes que melhor reproduzem as magnetizações m_i e covariancas C_{ij} a partir do algoritmo *Boltzmann Machine Learning*, definimos a "temperatura de operação"ou "temperatura de leitura"de cada texto como sendo $T_o = 1$. Então calculamos o "calor específico"de cada um

Figura 3 – Mapa de ativações de cada participante e textos.



Fonte: Figura retirada de [1]. Mapa de ativação das palavras para cada participante em cada um dos textos do experimento. Para cada participante i, o estado da palavra r é considerado ativo ($\sigma_i^{(r)} = +1$) se $n_i^r \geq 2$ ou inativo ($\sigma_i^{(r)} = +1$) se $n_i^r \geq 2$. As palavras ativadas são representadas em azul e as inativas em branco.

dos textos

$$C_v = \frac{\partial \langle E \rangle}{\partial T}, \tag{2.35}$$

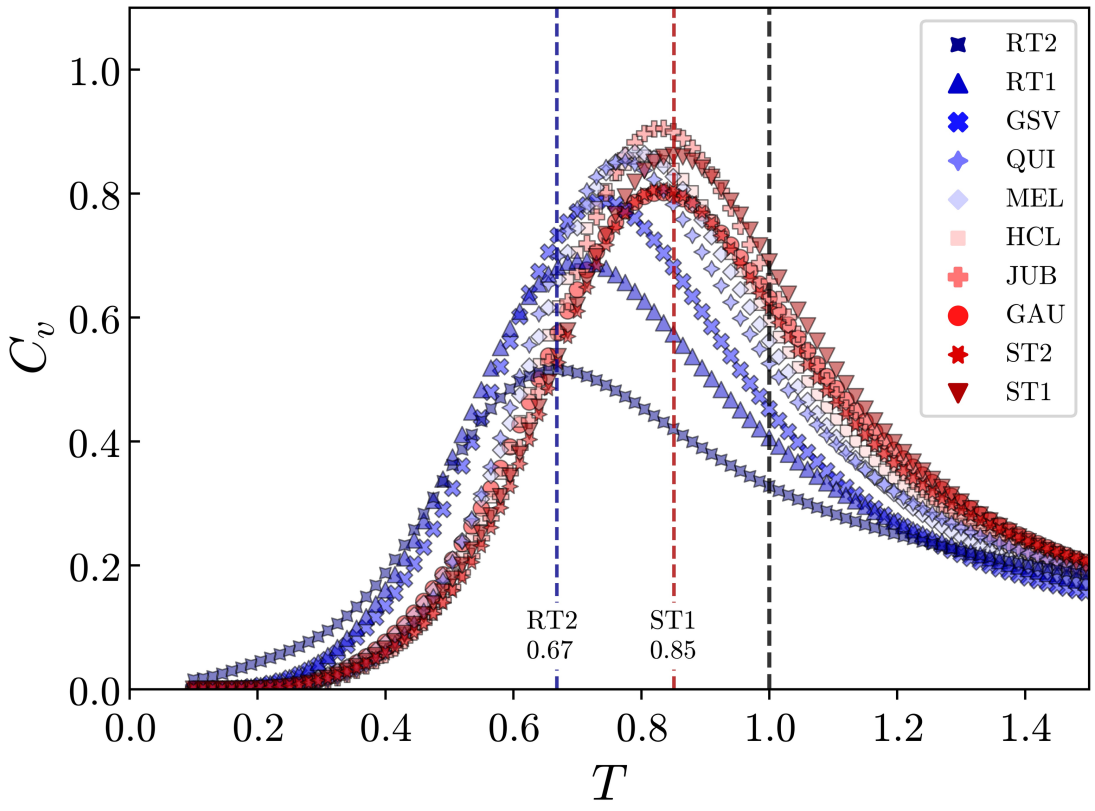
onde T é a temperatura e E a energia do sistema dada por

$$E = H_{Ising}(\sigma) = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i. \tag{2.36}$$

O calor específico mede quanto de energia o sistema pode absorver à medida que a temperatura T aumenta. Em uma temperatura crítica T_c, o calor específico (C_v) atinge seu máximo, significando uma transição de fase. Se T_c é menor que T_o, o sistema está em um estado desordenado ou aleatório. Por outro lado, se T_c é maior que T_o, o sistema está em um estado mais organizado.

No Gráfico 3, temos a relação entre o calor específico C_v e a temperatura T para todos os textos analisados. Independentemente do texto, a temperatura de operação ($T_o = 1$) é sempre maior do que a temperatura crítica (T_c). No entanto, a diferença entre T_o e T_c varia muito entre os textos. Os textos aleatórios, RT1 e RT2, apresentam valores maiores de diferença do que os outros textos, sugerindo que essa diferença pode ser usada para diferenciar textos significativos de textos aleatórios. Essa diferença referida como $T_o - T_c$, pode ser relacionada à coerência percebida no processamento da linguagem durante a leitura de um texto.

Gráfico 3 – Calor específico em função da temperatura para cada um dos textos.



Fonte: Gráfico retirado de [1]. Calor específico em função da temperatura para todos os textos do experimento. A temperatura de leitura (ou temperatura de operação) é dada por $T = T_o = 1$. Podemos ver que para todos os textos a temperatura de leitura é acima da temperatura crítica e que os textos aleatórios, RT1 e RT2, são os que possuem maior distância entre a temperatura crítica e de leitura.

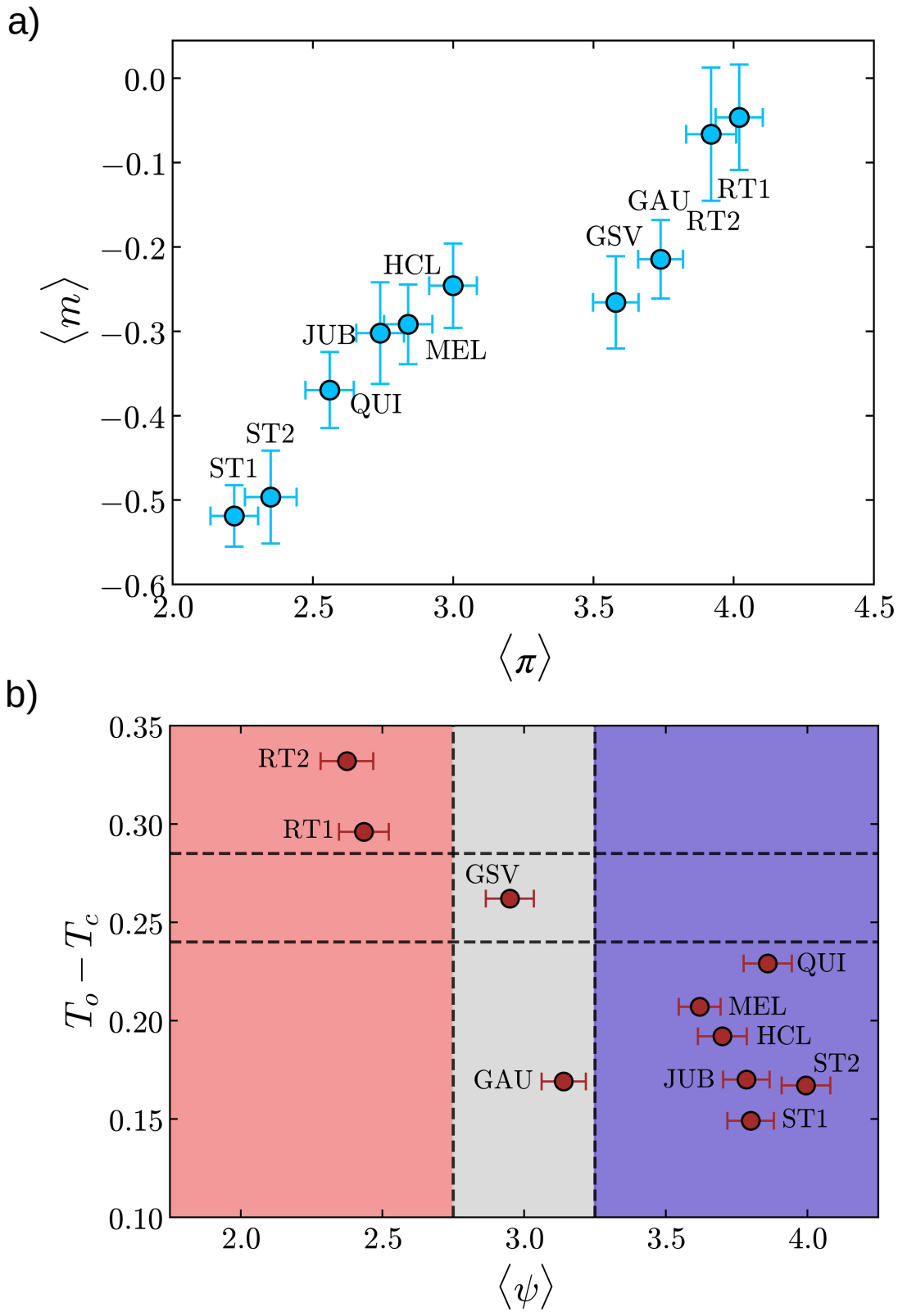
A relação entre a magnetização média ($\langle m \rangle$) obtida das ativações de fixações em função da complexidade média ($\langle \pi \rangle$) medida a partir do questionário quantitativo para os tex-

tos é mostrada no Gráfico 4.a. Percebemos que ambas as quantidades são altamente correlacionadas, onde a magnetização cresce quase que monotonicamente em função da complexidade. É possível que textos mais simples como ST1 e ST2 possuam menos complexidade e baixa magnetização, ou seja, menos fixações. Já no outro extremo, temos os textos aleatórios RT1 e RT2 com alta magnetização e complexidade. Esses resultados são esperados e sugerem que a magnetização pode refletir a complexidade dos textos. Podemos justificar essa relação com o fato de que textos mais complexos requerem mais tempo de análise do leitor, logo, mais fixações por palavras.

O Gráfico 4.b mostra como a distância $T_o - T_c$ está relacionada com a coerência obtida pelo questionário para os textos analisados. Os textos randômicos (RT1 e RT2) são consistentes com uma baixa coerência e alta distância da criticalidade. Já os textos ditos coerentes (ST1, ST2, JUB, HCL, MEL e QUI) formam um grupo onde a temperatura de leitura é mais próxima da temperatura crítica, isto é, baixos valores para $T_o - T_c$. Os textos GAU e GSV ficaram em uma região intermediária de coerência, mas com distância de temperatura bem diferentes. Isso pode ser justificado pelo de fato que o texto GSV (Grande Sertão: Veredas), escrito pelo Guimarães Rosa, ser reconhecido como um texto de estilo de escrita singular [38] e o texto GAU (O Gaúcho), escrito por José de Alencar, ser caracterizado por uma escrita demasiadamente filosófica da descrição do cenário [39].

Os resultados do estudo mostram que os dados de rastreamento ocular podem ser processados e analisados para determinar a complexidade e coerência de diferentes textos. Incluindo textos como histórias infantis, textos de palavras gerados aleatoriamente e obras literárias. Isso foi confirmado por meio de um questionário com um grande número de respondentes. Os resultados mostraram uma relação quase monotônica entre a magnetização média das atividades de fixação e a complexidade média dos textos e a distância entre a temperatura de leitura e a temperatura crítica ($T_o - T_c$) para separação de textos coerentes de textos aleatórios.

Gráfico 4 – Relação entre magnetização, complexidade, temperatura e coerência.



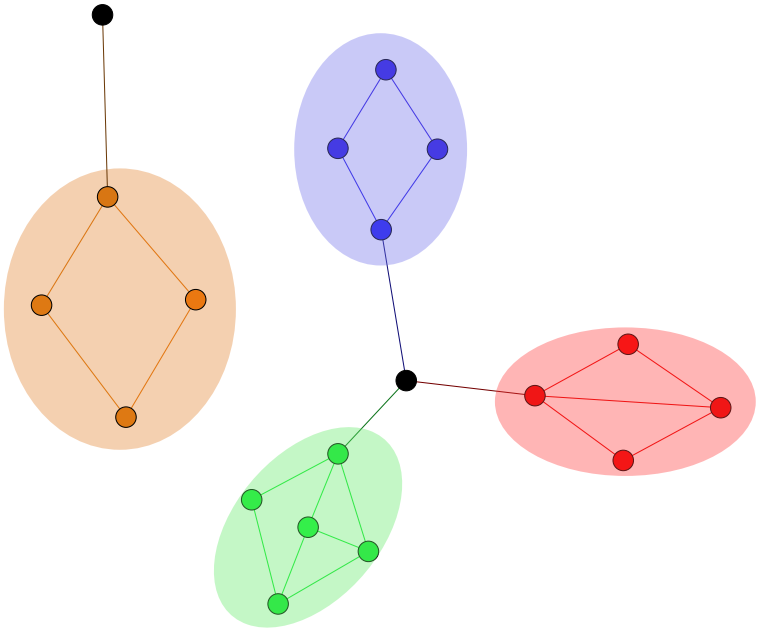
Fonte: Gráfico retirado de [1]. (a) Média da magnetização ($\langle m \rangle$)obtida das ativações de fixações em função da complexidade média ($\langle \pi \rangle$) medida a partir do questionário quantitativo para os textos. Podemos ver que a magnetização cresce quase monotonicamente com relação a complexidade. (b) Relação da distância entre a temperatura do ponto de leitura e a temperatura crítica ($T_o - T_c$) em função da coerência média do questionário quantitativo. Os textos aleatórios, RT1 e RT2, possuem baixa coerência e alto valor de $T_o - T_c$. Já os textos com alta coerência (ST1, ST2, JUB, HCL, MEL e QUI) apresentam baixo valor para $T_o - T_c$. Os gráficos sugerem que a magnetização e a distância $T_o - T_c$ são medidas para a complexidade e coerência em textos, respectivamente.

3 DECOMPOSIÇÃO DE GRAFOS EM COMPONENTES MENORES

3.1 Componentes biconectadas

Dizemos que um grafo $\overline{G}(\overline{\mathbf{V}},\overline{\mathbf{E}})$ é um subgrafo de $G = (\mathbf{V},\mathbf{E})$, se o conjunto de vértices $\overline{\mathbf{V}}$ é um subconjunto de $\mathbf{V}$ e $\overline{\mathbf{E}}$ é um subconjunto de $\mathbf{E}$. Dizemos que um subgrafo $\overline{G}(\overline{\mathbf{V}},\overline{\mathbf{E}})$ é uma componente biconectada de $G = (\mathbf{V},\mathbf{E})$ quando o mesmo se mantém conectado com a remoção de qualquer um de seus vértices. Ou seja, para qualquer trio de vértices v, w e a contidos em $\overline{\mathbf{V}}$, existe um caminho $p : v \Rightarrow w$ que não passa pelo vértice a. Na Figura 4, temos o exemplo de um grafo não direcionado desconectado e suas componentes biconectadas.

Figura 4 – Componentes biconectadas.



Fonte: Elaborada pelo autor. Exemplo de um grafo desconectado e suas componentes biconectadas. Se removermos qualquer vértice de algum subconjunto mostrado, a componente em questão se mantém conectada, mas podemos observar que os vértices em cor preta (que não fazem parte de nenhum subconjunto biconectado) e seus vizinhos são chamados pontos de separação ou pontos de articulação do grafo completo.

Uma componente biconectada pode ser vista como um ”bloco”de um grafo, de tal forma que se for removida, o grafo remanescente não é mais biconectado. Um grafo pode possuir múltiplas componentes biconectadas, onde cada uma é considerada um subgrafo independente, como exemplificado na Figura 4.

Uma forma de identificar as componentes biconectadas em um grafo, se existir, é utilizando algoritmos baseados em Busca em Profundidade (*Depth-first search* - DFS) [40–42] para percorrer todos os vértices no grafo. Isso permite a detecção de vértices de separação (vértices de articulação), que são vértices que, se removidos, desconectariam o grafo, bem como a identificação das ligações que compõem as diferentes componentes biconectadas. Uma

das principais vantagens da busca em profundidade é sua simplicidade. É um método fácil de implementar e de entender, podendo ser usado para resolver diferentes problemas em grafos. Entretanto, os algoritmos baseados em DFS possuem suas limitações também, o método pode não visitar todos os vértices se o grafo não for conectado e o alto uso de memória RAM se o grafo possui muitos vértices e ligações.

Um dos principais métodos que é eficiente para encontrar as componentes biconectadas em um grafo é o algoritmo desenvolvido por Tarjan [43]. Nessa tese, foi usado esse método para determinar as componentes biconectadas.

3.1.1 Componentes biconectadas e percolação

A probabilidade de ligação em um grafo de rede quadrada leva ao surgimento do cluster de percolação, um objeto fractal [2, 3]. Esse objeto em questão possui dimensão fractal $d_C = 1.89 \pm 0.01$ [2, 3] quando a probabilidade de ocupação de ligação é $p^* = 0.5$.

Outro objeto importante em percolação no ponto crítico são os chamados *blobs*. Os *blobs* são descritos pela teoria da percolação como componentes (subgrafos) que não podem ser separados removendo apenas uma ligação [44]. Eles são descritos por possuírem múltiplas conexões, ou seja, existem pelo menos dois caminhos disjuntos entre cada vértice de um *blob* e todos os demais vértices. Essa é a definição de componentes biconectadas que descrevemos anteriormente, logo podemos dizer que os *blobs* são as componentes biconectadas de um grafo. A decomposição do cluster de percolação em componentes biconectadas (*blobs*) já foi extensivamente estudada [45].

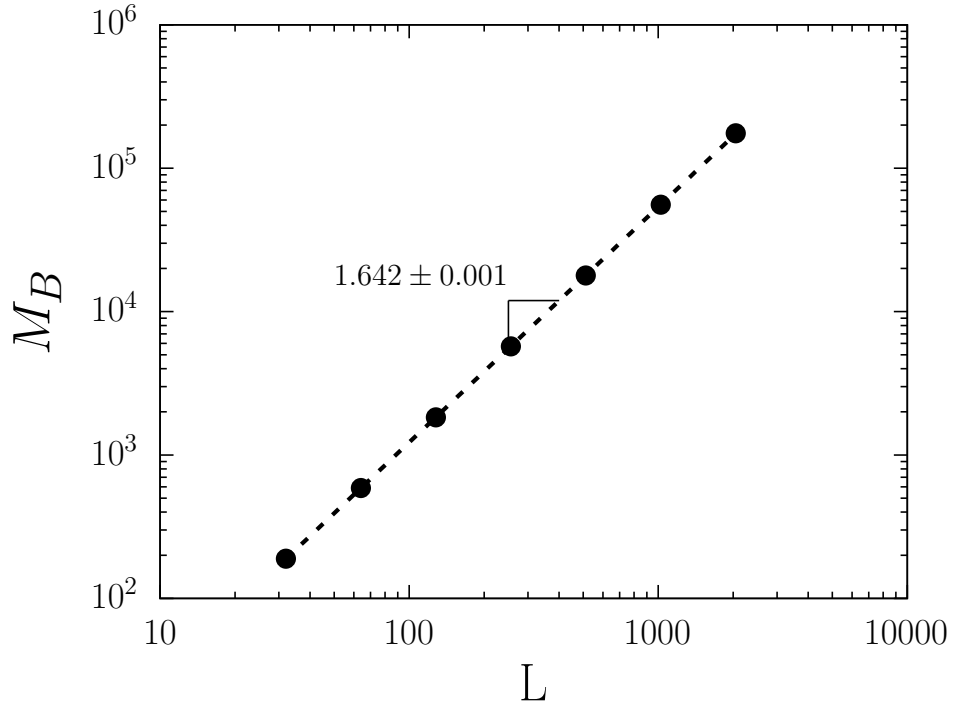
Assim como o maior agregado de percolação (cluster de percolação) de uma rede quadrada possui dimensão fractal, o maior *blob* ou a maior componente biconectada do cluster percolante também é um objeto fractal com dimensão $d_B = 1.642 \pm 0.001$. O Gráfico 5 mostra como a massa, M_B, da maior componente biconectada cresce com o tamanho do sistema L. Vale ressaltar o fato que a dimensão fractal do maior *blob* é menor que a dimensão fractal do cluster de percolação.

Essa dimensão fractal do maior *blob* é a mesma para outro objeto importante na teoria da percolação, o chamado *backbone*. O *backbone* é um conjunto de ligações pertencentes ao cluster percolante que conecta uma extremidade à outra da rede, ou seja, é um subgrafo do maior agregado no ponto crítico que possui um caminho de uma ponta a outra. A dimensão fractal de ambos é a mesma [46].

3.2 Componentes triconectadas

Vimos na seção anterior que podemos decompor um grafo em componentes menores chamadas de componentes biconectadas e essas componentes são estruturas importantes dentro da teoria da percolação. Além disso, observamos que a dimensão fractal no ponto crítico de percolação da maior componente biconectada é igual a dimensão fractal do *backbone* e menor

Gráfico 5 – Comportamento do tamanho da maior componente biconectada em função do tamanho do sistema *L*.



Fonte: Elaborada pelo autor. Comportamento da massa da maior componente biconectada do *backbone* crítico de percolação em função do tamanho do sistema *L*. Pela inclinação da reta, observamos que a dimensão fractal da maior componente biconectada é $d_B = 1.642 \pm 0.001$. Assim, vemos que a dimensão fractal das componentes biconectadas é menor que a dimensão fractal do cluster percolante.

que a dimensão do cluster percolante.

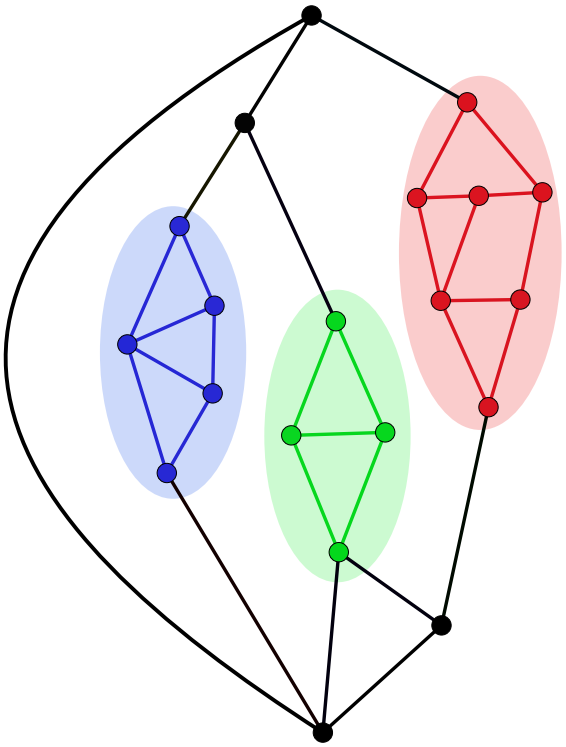
Seguindo a lógica de decompor um grafo em componentes cada vez menores, podemos decompor as componentes biconectadas ou o *backbone* em outras componentes menores chamadas de componentes triconectadas. As componentes triconectadas são subgrafos onde a remoção de quaisquer par de vértices (v,w) não desconecta o subgrafo em questão. Na Figura 5, temos um exemplo de um grafo biconectado e suas três componentes triconectadas.

3.2.1 Decompondo componentes biconectadas em triconectadas

Observando a estrutura das componentes triconectadas na Figura 5, percebemos que as mesmas possuem um par de vértices (v,w) que mantém a conexão com o restante do grafo. Esse par de vértice é chamado de par de separação, onde podemos substituir a componente triconectada por uma ligação virtual que conecta o par em questão. Podemos não apenas substituir componentes triconectadas por uma ligação virtual, mas também de ligações em série e em paralelo.

O processo de substituir componentes que são formadas por conjunto de ligações em série, paralelo e triconectadas por ligações virtuais, permite encontrar novas outras componentes e, então, repetir o processo de substituição até que todo o grafo seja decomposto, onde não é mais possível realizar mais substituições por ligações virtuais. Esse processo é conhecido

Figura 5 – Exemplos de Componentes Triconectadas.



Fonte: Elaborada pelo autor. Exemplo de um grafo biconectado e suas componentes triconectadas. Se removermos qualquer par de vértices de algum subconjunto mostrado (azul, verde ou vermelho), a componente em questão se mantém conectada. O par de vértices que conecta cada componente triconectada no grafo é chamado de par de separação, onde podemos substituir cada componente triconectada por uma ligação virtual que conecta esse par de vértices.

como decomposição de grafos em árvore SPQR (SPQR-*tree*).

SPQR-*tree* é uma estrutura de dados desenvolvida por Di Battista *et al.* [47–49] para decomposição de um grafo biconectado em suas componentes triconectadas. Essa estrutura de dados tem sido muito utilizada na teoria de grafos para o estudo de planaridade em grafos [50–53]. Cada letra em SPQR-*tree* representa um tipo de componente:

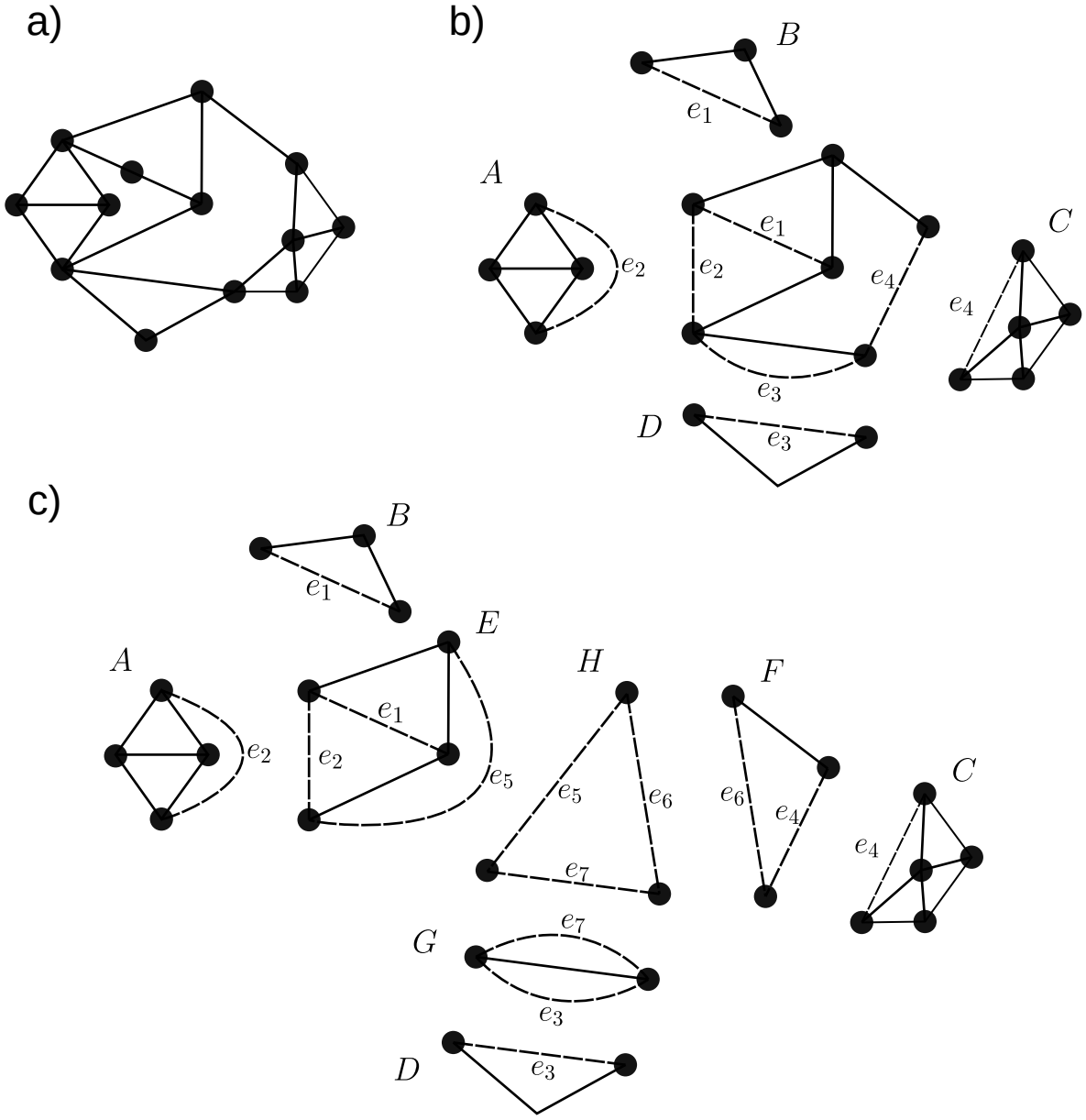
- S: Conjunto de vértices e ligações que formam um ciclo;
- P: Par de vértices com três ou mais ligações em paralelo;
- Q: Caso trivial onde a componente biconectada é formada por apenas dois vértices e duas ligações;
- R: Conjunto de vértices e ligações que formam componentes triconectadas.

Na Figura 6, temos um exemplo do processo de decomposição de um grafo biconectado na estrutura SPQR-*tree*. A Figura 6.b mostra o primeiro passo da decomposição do grafo biconectado da Figura 6.a, onde são identificadas duas componentes do tipo cíclica (tipo S) B e D e duas componentes do tipo triconectada (tipo R) A e C. Essas componentes são substituídas por ligações virtuais (linhas tracejadas) e_1, e_2, e_3 e e_4. O grafo resultante pode ser decomposto

mais uma vez. A Figura 6.c mostra o segundo passo da decomposição, no qual é identificada outra componente E triconectada, uma componente F cíclica e uma componente paralela G. Essas três componentes são substituídas pelas ligações virtuais e_5, e_6 e e_7, formando, assim, a componente cíclica final H.

O procedimento mostrado na Figura 6, resultou na decomposição do grafo em oito componentes menores. Algumas dessas componentes estão contidas em outas componentes, sendo assim, podemos definir um sistema hierárquico, no qual componentes contidas em outras formam níveis hierárquicos. Para tanto, vamos definir que um nível é formado quando uma componente triconectada se encontra contida em outra componente triconectada. Dessa forma, podemos dizer que as componentes C e E são componentes de nível 1, enquanto que a componente A é de nível 2, pois está contida na componente E. Seguindo essa definição, podemos dizer também que cada ligação possui nível igual ao nível de sua componente triconectada e que ligações que não fazem parte de nenhuma componente triconectada, como por exemplo as ligações das componentes D, F, G e H, são de nível 0. Esse modelo hierárquico será utilizado no próximo capítulo.

Figura 6 – Exemplo de decomposição de grafo na estrutura SPQR-*tree*.

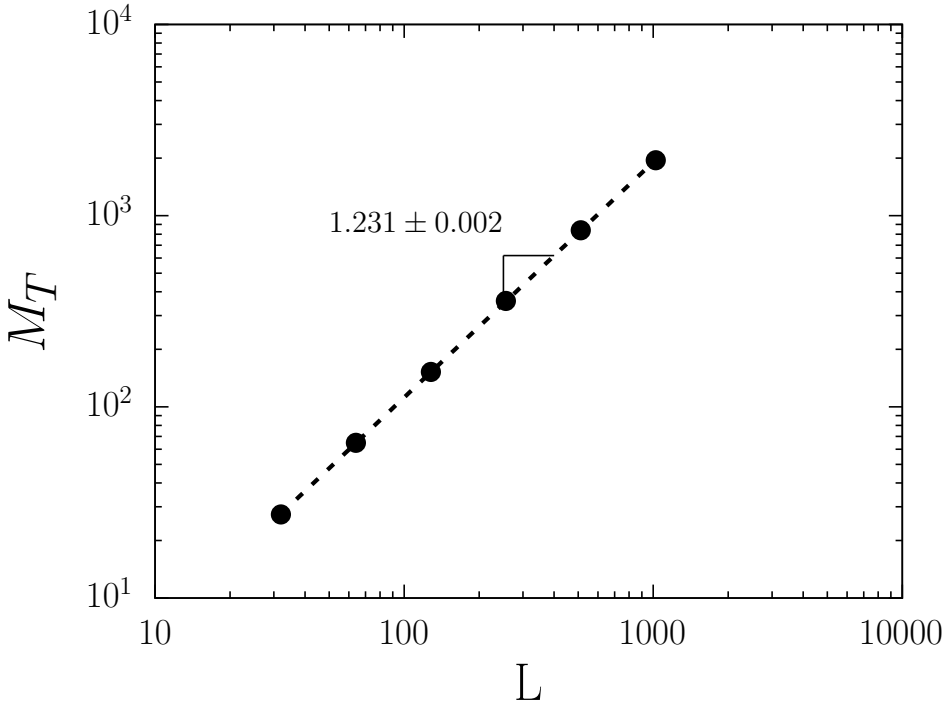


Fonte: Elaborada pelo autor. Decomposição de um grafo biconectado na estrutura SPQR-*tree*. (a) Exemplo de uma grafo biconectado. A remoção de um vértice qualquer não desconecta o grafo. (b) Primeira etapa na decomposição do grafo, onde são identificadas duas componentes do tipo S (cíclica) *B* e *D* e duas componentes do tipo R (triconectadas) *A* e *C*. As componentes são substituídas pelas ligações virtuais e_1, e_2, e_3 e e_4. (c) Segunda etapa de decomposição, onde são identificadas e substituídas por ligações virtuais e_5, e_6 e e_7 uma componente triconectada *E*, uma cíclica *F* e outra paralela (tipo P) *G*. A decomposição completa do grafo resulta na componente cíclica final *H*. Podemos definir níveis hierárquicos para cada componente triconectadas. As componente *E* e *C* são componentes de nível um e a componente *A* é de nível dois, pois a mesma está contida na componente *E*.

Um dos algoritmos mais conhecidos e utilizados para decompor um grafo em suas componentes triconectadas é o algoritmo desenvolvido por Hopcroft e Tarjan em [54]. Segundo Gutwenger e Mutzel em [55], há alguns erros no algoritmo proposto por Tarjan e que são corrigidos no artigo em questão. Para essa tese, foi desenvolvido o *script* segundo o algoritmo proposto por Tarjan, juntamente com as correções de Mutzel que foram constatadas durante o desenvolvimento.

Como era de se esperar, as componentes triconectadas também possuem dimensão fractal no *backbone* crítico de percolação. Paul *et al.* [56], seguindo o algoritmo proposto por Tarjan, mostraram que a dimensão fractal da maior componente triconectada no *backbone* crítico de percolação em uma rede quadrada é $d_T = 1.15 \pm 0.1$. No Gráfico 6, podemos ver como a massa da maior componente triconectada, M_T, cresce com relação ao tamanho L do sistema. Foi utilizado o *script* desenvolvido para determinar as componentes triconectadas e sua massa, onde, pela inclinação da reta no Gráfico 6, podemos constatar que a dimensão fractal das componentes triconectadas é $d_T = 1.231 \pm 0.002$. O valor encontrado nessa tese está dentro da margem de erro do valor encontrado por Paul *et al.*, mas vale destacar que no artigo de Paul *et al.*, não é citado as correções de Mutzel para o algoritmo de Tarjan.

Gráfico 6 – Comportamento do tamanho da maior componente triconectada em função do tamanho do sistema L.

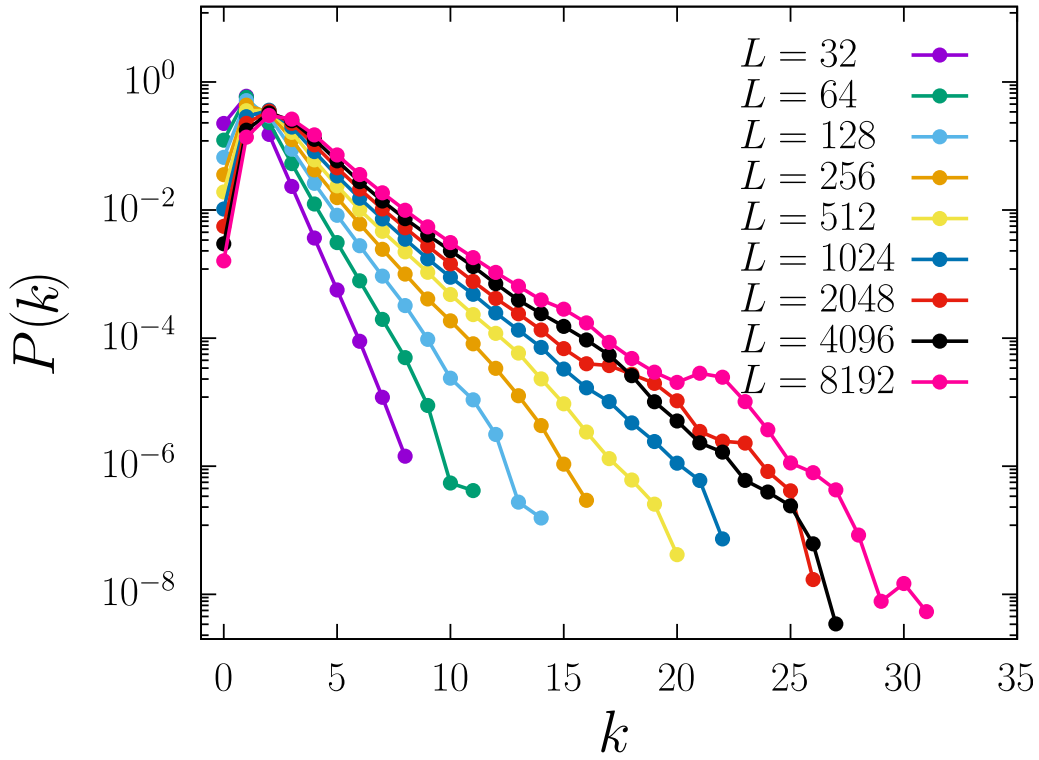


Fonte: Elaborada pelo autor. Comportamento da massa da maior componente triconectada do *backbone* crítico de percolação em função do tamanho do sistema L. Pela inclinação da reta, observamos que a dimensão fractal das componentes triconectadas é $d_T = 1.231 \pm 0.002$. O valor é consistente com o encontrado por Paul *et al.* [56]. Assim, vemos que a dimensão fractal das componentes triconectadas é menor que a dimensão fractal das componentes biconectadas.

Utilizando o sistema hierárquico de níveis criado, o Gráfico 7 mostra a distribuição

do número de ligações de nível k do *backbone* crítico de percolação em rede quadrada. Vemos que quanto maior é o tamanho do sistema, mais ligações de níveis superiores são presentes na rede, onde, para $L = 8192$, temos ligações que pertencem a níveis superiores a 30 e que ligações de níveis 0 se tornam menos representativas.

Gráfico 7 – Distribuição de níveis das ligações do *backbone* crítico.



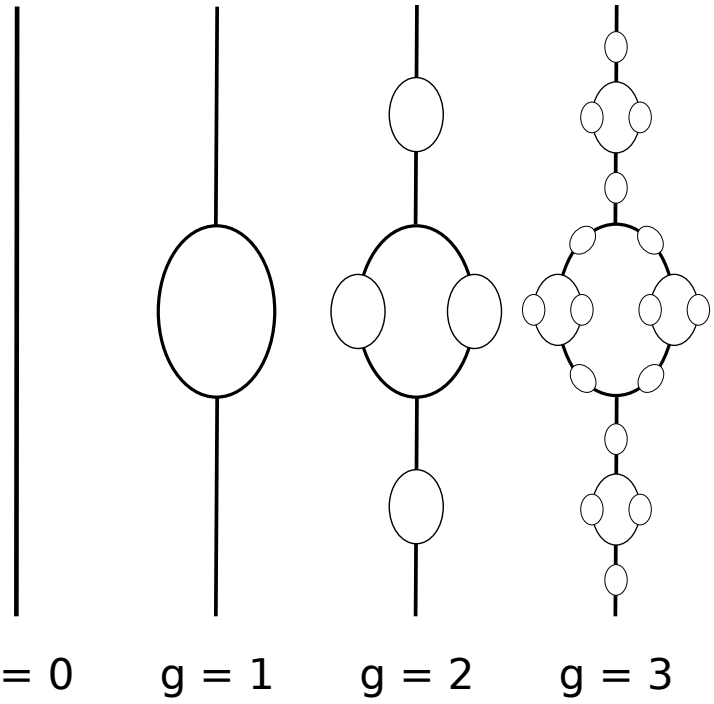
Fonte: Elaborada pelo autor. Distribuição do número de ligações de nível k no *backbone* crítico de percolação em uma rede quadrada para sistemas de tamanho $L = 32,\ 64,\ 128, ...,\ 4096,\ 8192$. Quanto maior o tamanho do sistema, mais níveis hierárquicos surgem e o número de ligações de nível 0 se tornam menos representativas.

4 REDE DE RESISTORES NO BACKBONE DE PERCOLAÇÃO

4.1 Modelo hierárquico simples

Para descrever a geometria e a distribuição de correntes no *backbone* do agregado de percolação, vamos começar estudando um modelo hierárquico simples como mostrado na Figura 7. Para obter a estrutura hierárquica mostrada, realizamos um processo iterativo, onde cada ligação da geração atual é substituída pela célula unitária (estrutura da primeira geração, $g = 1$). Assim, para obter a estrutura de geração N deve-se substituir todas as ligações da geração $N-1$ pela estrutura unitária ($g = 1$). Faremos uma análise para o caso dessa estrutura unitária simples e em seguida generalizaremos para qualquer estrutura.

Figura 7 – Modelo hierárquico.

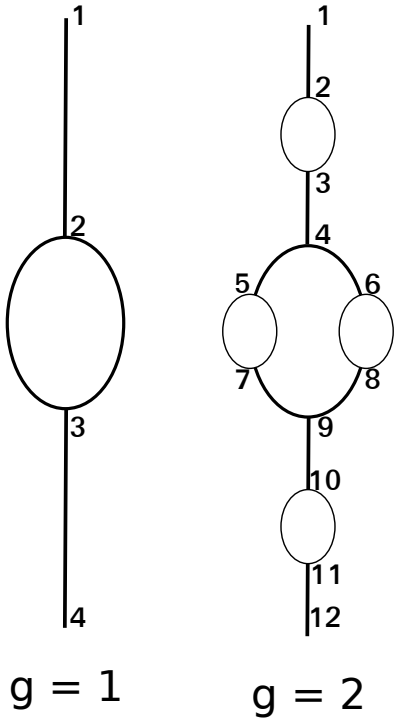


Fonte: Elaborada pelo autor. Os primeiros níveis de um modelo de rede hierárquica. Na geração zero ($g = 0$) a rede possui apenas uma ligação. Já na primeira geração ($g = 1$) temos o que chamamos de célula unitária da estrutura. Para obter as gerações futuras, substituímos cada ligação da geração atual pela célula unitária.

Suponhamos que cada ligação possui uma resistência de valor unitário e que aplicamos uma corrente, também unitária, em uma das extremidades saindo completamente pela outra extremidade. Nosso primeiro passo é encontrar a distribuição de correntes a cada geração da nossa estrutura hierárquica. Vamos primeiro calcular o número de correntes i na geração g, $n_g(i)$.

Na geração $g = 0$, o número de correntes é trivial uma vez que existe apenas uma ligação. Já na geração $g = 1$, que é nossa célula unitária, o número de correntes é facilmente

Figura 8 – Primeira e segunda geração.



Fonte: Elaborada pelo autor. Vemos as duas primeira gerações do modelo hierárquico em questão. Em $g = 1$ a estrutura possui apenas 4 ligações, ja na segunda geração possui um total de 16 ligações.

determinado. Observando a Figura 8, vemos que a ligação entre os pontos $1-2$ passa a corrente completa, $i_{1-2} = 1$, essa corrente é então dividida igualmente no ponto 2, uma vez que estamos considerando cada ligação possuindo resistência unitária, ou seja, possuímos 2 correntes iguais entre os pontos $2-3$, cujo valor é $i_{2-3} = \frac{1}{2}$. E por fim, essa corrente se junta no ponto 3, assim temos a última corrente da primeira geração entre os pontos $3-4$, $i_{3-4} = 1$. Vemos que o número de correntes na célula unitária é:

$$n_1(i=1) = 2, \tag{4.1}$$

$$n_1\left(i=\frac{1}{2}\right) = 2. \tag{4.2}$$

Analisaremos a rede de segunda geração ($g = 2$) da Figura 8, onde cada ligação da primeira geração foi substituída pela célula unitária. Realizando o mesmo procedimento anterior, de contar as correntes, vemos que temos quatro correntes $i = 1$, ligações entre os pontos $1-2$, $3-4$, $9-10$ e $11-12$. São oito correntes de $i = \frac{1}{2}$, duas entre os pontos $2-3$ e $10-11$, e uma entre os pontos $4-5$, $4-6$, $7-9$ e $8-9$. Por fim, temos mais quatro ligações com corrente $i = \frac{1}{4}$, duas entre os pontos $5-7$ e outras duas entre os pontos $6-8$. Observe que a corrente foi dividida no ponto 4 e depois cada metade foi dividida novamente nos pontos 5 e 6. A distribuição de correntes na segunda geração é, então:

$$n_2(i=1) = 4, \tag{4.3}$$
$$n_2\left(i=\frac{1}{2}\right) = 8, \tag{4.4}$$
$$n_2\left(i=\frac{1}{4}\right) = 4. \tag{4.5}$$

Aplicando o mesmo procedimento para a estrutura de terceira geração encontramos que a distribuição de correntes é:

$$n_3(i=1) = 8, \tag{4.6}$$
$$n_3\left(i=\frac{1}{2}\right) = 24, \tag{4.7}$$
$$n_3\left(i=\frac{1}{4}\right) = 24, \tag{4.8}$$
$$n_3\left(i=\frac{1}{8}\right) = 8. \tag{4.9}$$

Percebemos que a cada geração a distribuição de correntes segue um padrão, dado por:

$$n_g(i_j) = 2^g \binom{g}{j}, \tag{4.10}$$

onde

$$i_j = \frac{1}{2^j}, \quad j = 0,1,\ldots,g. \tag{4.11}$$

Como o número de ligações da nossa célula unitária é 4, o número total de correntes na geração g é dado por 4^g. Com isso, concluímos que a densidade de probabilidade das correntes na geração g é dada por uma distribuição semelhante à binomial com q e p dados por $1/2$.:

$$p_g(i) = \frac{1}{2^g}\sum_{j=0}^{g}\binom{g}{j}\delta\left(i-\frac{1}{2^j}\right). \tag{4.12}$$

Note que, se os termos da delta fossem igualmente espaçados, essa seria a binomial. Como são geometricamente espaçados, essa distribuição é log-binomial. Para g muito grande, portanto, se aproxima bem da log-normal.

Pelo processo de construção da nossa rede hierárquica e analisando a Equação 4.11, vemos que a corrente é uma multiplicação de fatores que aumenta a cada geração. Para a célula unitária usada, temos apenas dois fatores, $f_1 = 1$ e $f_2 = \frac{1}{2}$, que são combinados para determinar a corrente

$$i_j = f_1^{g-j} f_2^j, \quad j = 0, 1, \ldots, g. \tag{4.13}$$

Podemos generalizar esse resultado para uma célula unitária arbitrária com L ligações e m fatores multiplicativos $\{f_1, f_2, \ldots, f_m\}$ distribuídos sobre as ligações. Eventualmente, mais de uma ligação da célula unitária pode carregar a mesma fração da corrente. Por exemplo, no caso anterior tínhamos na célula unitária duas ligações carregando a corrente total, $f = 1$, e duas carregando a metade, $f = 1/2$. Em nosso caso geral, vamos supor que temos q_j ligações com o fator f_j, de forma que

$$\sum_{j=1}^{m} q_j = L. \tag{4.14}$$

O número total de correntes (ligações) na geração g para esse sistema hierárquico é

$$\begin{aligned} N_g &= L^g = (q_1 + q_2 + \cdots + q_m)^g \\ &= \sum_{\substack{x_1=0 \\ x_1+x_2+\cdots+x_m=g}}^{g} \sum_{x_2=0}^{g} \cdots \sum_{x_m=0}^{g} \frac{g!}{x_1!x_2!\ldots x_m!} q_1^{x_1} q_2^{x_2} \cdots q_m^{x_m} \\ &= \sum_{\substack{\{\mathbf{x}\} \\ x_1+x_2+\cdots+x_m=g}} \binom{g}{x_1, x_2, \ldots, x_m} \prod_{j=1}^{m} q_j^{x_j}, \end{aligned} \tag{4.15}$$

onde

$$\binom{g}{x_1, x_2, \ldots, x_m} = \frac{g!}{x_1!x_2!\ldots x_m!} \tag{4.16}$$

e a somatória em $\{\mathbf{x}\}$ é feita sobre todas as configurações possíveis, respeitando $\sum_j x_j = g$. Esses x_j representam o número de vezes que cada um dos m fatores aparecem no produtório que vai determinar a corrente.

O número de correntes, $i(\mathbf{x})$, na geração g é dado por cada termo da somatória em 4.15

$$n_g[i(\mathbf{x})] = \binom{g}{x_1, x_2, \ldots, x_m} \prod_{j=1}^{m} q_j^{x_j}, \tag{4.17}$$

onde

$$i(\mathbf{x}) = \prod_{j=1}^{m} f_j^{x_j}, \tag{4.18}$$

e a distribuição de probabilidades para as correntes fica

$$p_g[i(\mathbf{x})] = \frac{1}{L^g} \sum_{\substack{\{\mathbf{k}\} \\ k_1+k_2+\cdots+k_m=g}} \binom{g}{k_1,k_2,\ldots,k_m} \prod_{j=1}^{m} q_j^{k_j} \delta \left[i(\mathbf{x}) - \prod_{j=1}^{m} f_j^{y_j} \right]. \tag{4.19}$$

Ou, definindo $\rho_j = q_j/L$, ficamos com a distribuição conhecida como log-Multinomial

$$p_g[i(\mathbf{x})] = \sum_{\substack{\{\mathbf{k}\} \\ k_1+k_2+\cdots+k_m=g}} \binom{g}{k_1,k_2,\ldots,k_m} \prod_{j=1}^{m} \rho_j^{k_j} \delta \left[i(\mathbf{x}) - \prod_{j=1}^{m} f_j^{y_j} \right]. \tag{4.20}$$

A Equação 4.18 mostra que a corrente é uma multiplicação de fatores e o nosso objetivo é determinar a distribuição dessas correntes, mas para isso, vamos utilizar o truque do logaritmo para transformar um produtório em uma somatória,

$$\ln i(\mathbf{x}) = \sum_{j=1}^{m} x_j \ln f_j, \tag{4.21}$$

ao invés de determinar a distribuição de $i(\mathbf{x})$, vamos em busca da distribuição do logaritmo das correntes.

4.1.1 Teorema da convolução

Seja a variável $z = x + y$, onde x e y são variáveis independentes e identicamente distribuídas com densidade de probabilidade $f_x(x)$ e $f_y(y)$, respectivamente. Para encontrar a densidade de probabilidade $f_z(z)$, observe que a distribuição acumulativa de z, $F_z(z)$, é

$$\begin{aligned} F_z(z) &= \int\int_{x+y\leq z} f(x,y)dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f(x,y)dxdy \\ &= \int_{-\infty}^{+\infty} f_x(x)dx \int_{-\infty}^{z-x} f_y(y)dy, \end{aligned} \tag{4.22}$$

onde fizemos o uso do fato que $f(x,y) = f_x(x)f_y(y)$, pois assumimos que as variáveis são independentes. Sabemos que a densidade de probabilidade é dada pela derivada da distribuição acumulativa, então:

$$\begin{aligned} f_z(z) = \frac{d}{dz}F_z(z) &= \int_{-\infty}^{+\infty} f_x(x)dx \frac{d}{dz}\int_{-\infty}^{z-x} f_y(y)dy \\ &= \int_{-\infty}^{+\infty} f_x(x)f_y(z-x)dx. \end{aligned} \tag{4.23}$$

Vemos que a função densidade de probabilidade $f_z(z)$ é dada pela convolução de $f_x(x)$ e $f_y(y)$.

Podemos enunciar o teorema da convolução. Sejam as funções características de $f_x(x)$ e $f_y(y)$, transformadas de Fourier,

$$\phi_x(x) = \int_{-\infty}^{+\infty} f_x(x) e^{ixt} dx, \tag{4.24}$$

$$\phi_y(y) = \int_{-\infty}^{+\infty} f_y(y) e^{iyt} dy \tag{4.25}$$

com inversas dadas por

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_x(t) e^{-ixt} dt, \tag{4.26}$$

$$f_y(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_y(t) e^{-iyt} dt. \tag{4.27}$$

Substituindo a Equação 4.27 na definição de convolução 4.23

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{+\infty} f_x(x) \left[\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-it(z-x)} \phi_y(t) dt \right] dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f_x(x) e^{ixt} dx \right] \phi_y(t) e^{-itz} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_x(t) \phi_y(t) e^{-itz} dt. \end{aligned} \tag{4.28}$$

Aplicando a transformada de Fourier em ambos os lados da Equação 4.28 encontramos que a função característica da densidade de probabilidade de z é

$$\phi_z(t) = \phi_x(t) \phi_y(t). \tag{4.29}$$

Assim, o teorema da convolução diz que a função característica da variável z é o produto das funções características das variáveis x e y.

O teorema pode ser generalizado para n variáveis independentes

$$z = x_1 + x_2 + \cdots + x_n. \tag{4.30}$$

Temos que:

$$f(z)dz = \int\int \cdots \int_{z < \sum x_i \leq z + dz} f_1(x_1) f_2(x_2) \cdots f_n(x_n) dx_1 dx_2 \cdots dx_n. \tag{4.31}$$

Para remover a restrição nos limites das integrais vamos utilizar uma delta de Dirac e garantir a igualdade $\sum x_i = z$, ou seja, $\delta\left(\sum x_i - z\right) dz$, para a desigualdade $z < \sum x_i \leq z + dz$. Ficamos com

$$f(z)dz = \int \int_{-\infty}^{+\infty} \cdots \int f_1(x_1) f_2(x_2) \cdots f_n(x_n) \delta\left(\sum x_i - z\right) dx_1 dx_2 \cdots dx_n dz. \tag{4.32}$$

Usando

$$\delta\left(\sum x_i - z\right) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i(\sum x_i - z)t} dt \tag{4.33}$$

Obtemos

$$\begin{aligned} f(z) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f_1(x_1) e^{ix_1 t} dx_1 \right] \cdots \left[\int_{-\infty}^{+\infty} f_n(x_n) e^{ix_n t} dx_n \right] e^{-izt} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_1(t) \phi_2(t) \cdots \phi_n(t) e^{-izt} dt \end{aligned} \tag{4.34}$$

Aplicando a transformada de Fourier em ambos os lados obtemos que

$$\phi_z(t) = \phi_1(t) \phi_2(t) \ldots \phi_n(t). \tag{4.35}$$

Assim, concluímos que dado uma variável z onde a mesma é a soma de n variáveis independentes, a função característica de z é dada pelo produto das funções características das n variáveis independentes.

4.2 Modelo hierárquico no *backbone* de percolação

Uma vez que entendemos como a distribuição de correntes é formada em um sistema hierárquico simples, iremos agora generalizar o modelo hierárquico para uma rede quadrada no ponto crítico de percolação, onde cada ligação possui uma resistência/condutância aleatória uniformemente distribuída. No Capítulo 3, vimos que podemos decompor o *backbone* percolante em componentes triconectadas e essas podem formar um sistema hierárquico de níveis, onde uma componente triconectada contida em outra componente triconectada cria um novo nível mais profundo.

Usando o sistema hierárquico de componentes triconectadas, vamos estudar a distribuição de correntes e suas propriedades realizando simulações adotando os seguintes passos:

1. Criamos uma rede quadrada de ligações com tamanho L x L, onde a probabilidade de ocupação de cada ligação é $p^* = 1/2$ e atribuímos para cada uma resistência uniformemente distribuída entre 1 e 2 (ou uma condutância uniformemente distribuída entre 0.5 e 1). Foi utilizado o algoritmo de Newman–Ziff [57] para construção da rede;

2. Com a rede construída, identificamos e extraímos o cluster percolante da rede, então, selecionamos aleatoriamente um par de vértices (v_t, w_t) tal que a distância Manhattan

entre eles seja igual a L. O par de vértices (v_t, w_t) configurará o terminal do sistema, por onde uma corrente $I = 1$ é introduzida na rede. Se não existir um par de vértices com essa condição, a rede é descartada e voltamos para o passo 1;

3. Criamos uma ligação entre os terminais (v_t, w_t) para representar uma bateria de condutância igual a 0, ou seja, garantimos que nenhuma corrente irá passar por essa nova ligação e nos asseguramos que os terminais façam parte de uma componente biconectada;

4. Extraímos o *backbone* que conecta os terminais utilizando o algoritmo de Tarjan [43] para a identificação de componentes biconectadas;

5. Uma vez com as ligações do *backbone* extraídas, realizamos a decomposição do mesmo em componentes triconectadas, isto é, decompomos o *backbone* na estrutura SPQR-*tree*, tal como exemplificado na Figura 6. A cada passo, calculamos a resistência/condutância equivalente da ligação virtual de cada componente. Para componente com dois, três e quarto vértices, isto é, componentes em série, paralela e ponte de *Wheatstone*, calculamos a resistência/condutância de forma exata. Já para componentes triconectadas maiores, usamos o método iterativo gradiente biconjugado esparso [58] para resolver o conjunto de equações de *Kirchhoff*;

6. Com o *backbone* decomposto, realizamos o processo inverso de reconstruir o mesmo e, assim, determinamos a corrente em cada ligação real pela multiplicação de fatores.

Realizamos o processo acima repetidas vezes para diferentes tamanhos de rede L. A Tabela 2 mostra o número de amostras que foram geradas para realizar as medidas em cada tamanho de sistema.

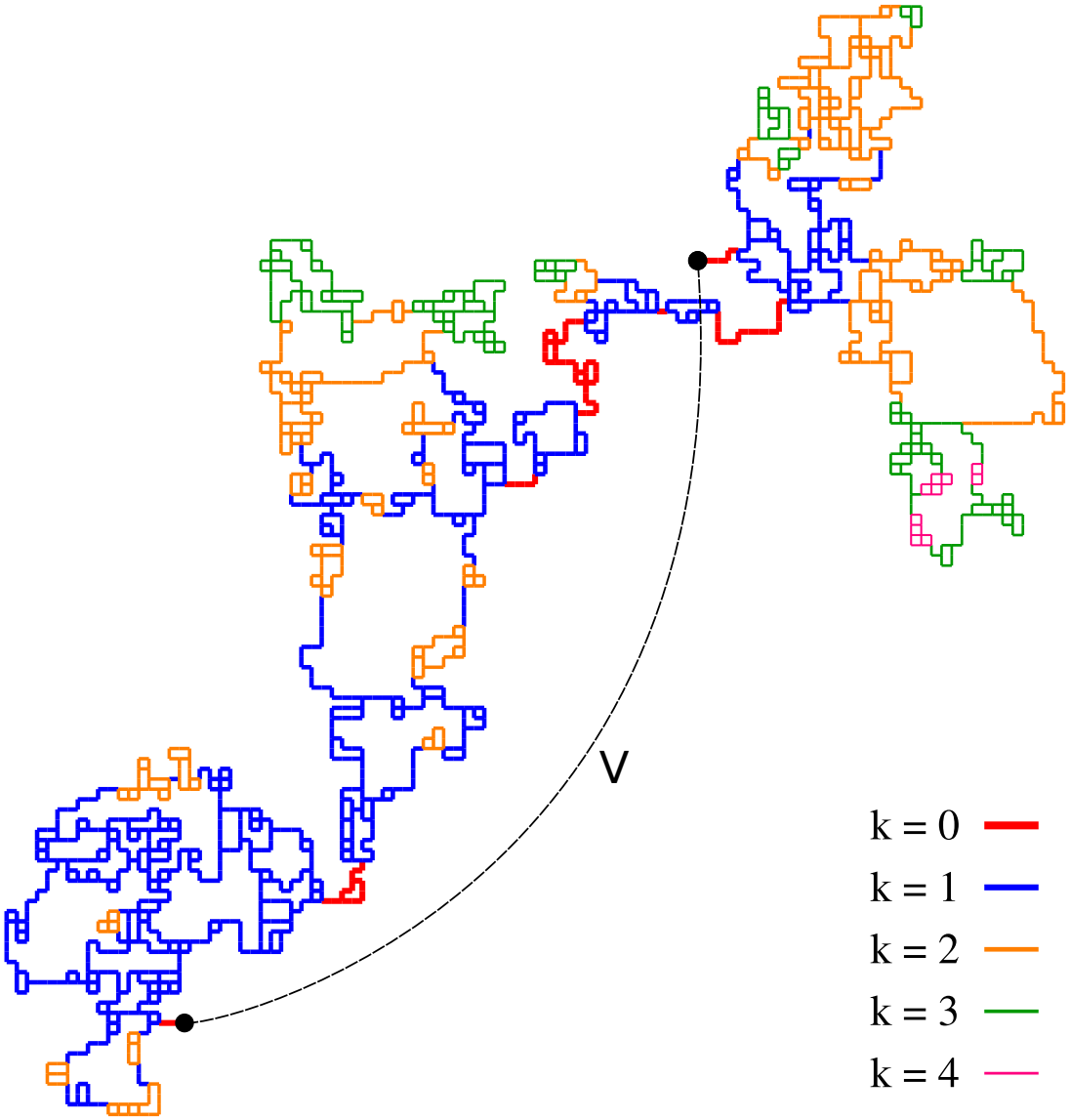
Tabela 2 – Número de amostras geradas para cada tamanho de rede

L	**# Amostras**
32	100000
64	100000
128	100000
256	100000
512	50000
1024	50000
2048	20000
4096	5000
8192	2000

Fonte: Elaborada pelo autor. Tabela com o número de amostras que foram geradas seguindo o processo descrito para realização de medidas em cada tamanho L de rede.

A Figura 9 mostra o exemplo de um *backbone* e os níveis de cada ligação após o precedimento descrito acima.

Figura 9 – Exemplo de um *backbone* de uma rede de tamanho $L = 128$ e os níveis de cada ligação.



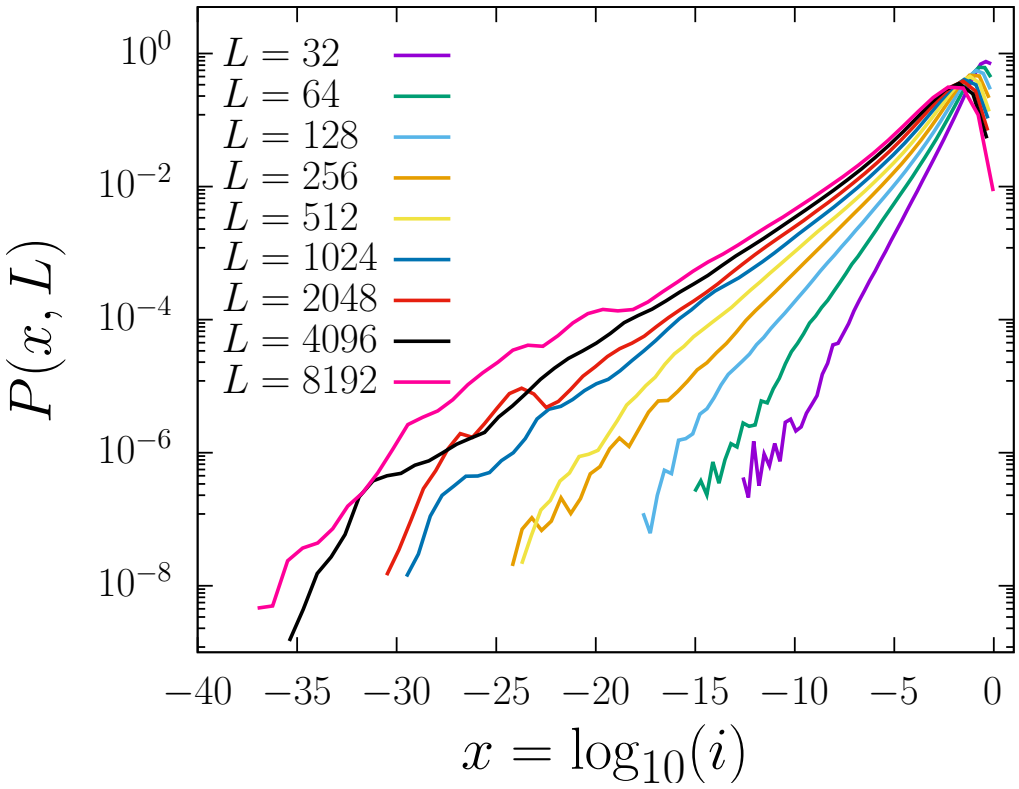
Fonte: Elaborada pelo autor. Ligações do *backbone* crítico de uma rede de tamanho $L = 128$, no qual a cor de cada ligação representa o seu nível. Ligações de nível k são aquelas dentro de uma componente triconectada de mesmo nível, mas não de componente triconectada de nível $k+1$ ou superior. A ligação tracejada representa a bateria V de condutância zero entre os terminais representados pelos pontos em preto, por onde uma corrente I é introduzida no sistema. O *backbone* em questão possui um total de quatro níveis, onde as ligações *red bonds* são presentes apenas no nível $k = 0$.

4.2.1 Distribuição de correntes

A metodologia desenvolvida para determinar as correntes do *backbone* crítico de percolação possui duas grandes vantagens: a primeira é que não precisamos resolver um sistema de equações de Kirchhoff muito grande, mas sim vários sistemas de equações menores. Por exemplo, seja um sistema de tamanho $L = 1024$, o número de equações de Kirchhoff para todo o *backbone* é aproximadamente $M_B = 1024^{1.64} \sim 86000$, mas para as componentes triconectadas é de apenas $M_T = 1024^{1.23} \sim 5000$; a segunda grande vantagem é que conseguimos determinar

correntes extremamente pequenas e com alta precisão. O Gráfico 8 mostra a distribuição de correntes para diferentes tamanhos de sistema seguindo o procedimento descrito anteriormente, observamos que podemos determinar correntes com precisão de até 10^{-35}.

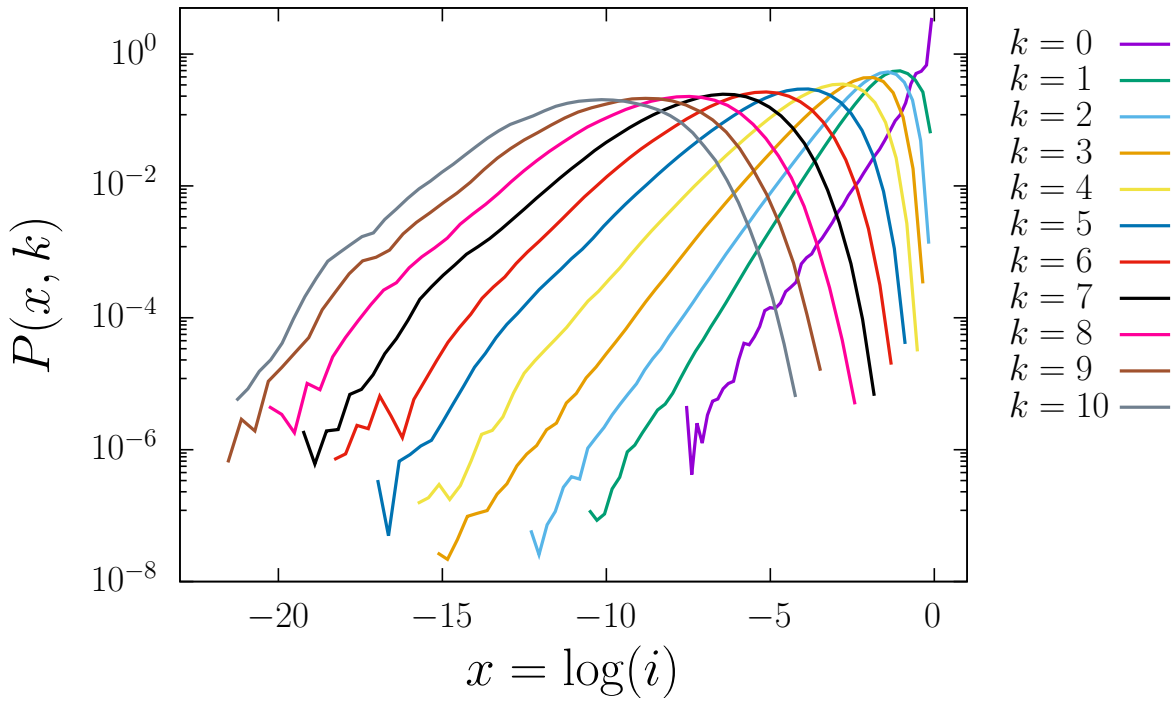
Gráfico 8 – Distribuição de correntes no *backbone* crítico de percolação para diferentes tamanhos de sistema.



Fonte: Elaborada pelo autor. Distribuição de correntes no *backbone* crítico de percolação de rede quadrada em diferentes tamanhos de sistemas L. Para cada tamanho L, foram simulados amostras de acordo com a Tabela 2, onde calculamos a corrente i passando por cada ligação. Cada ligação possui uma resistência uniformante distribuída entre 1 e 2. Utilizando o método desenvolvido pela decomposição do *backbone* em componentes triconectadas, permitiu que obtivéssemos correntes com precisão da ordem de 10^{-35}.

Devido ao modelo hierárquico que criamos, podemos calcular a distribuição de correntes para as ligações em cada nível. No Gráfico 9, temos a distribuição de correntes para ligações separadas por níveis k para um sistema de tamanho $L = 2048$, no qual vemos que a curva para o nível $k = 0$ possui um pico em $x = 0$, pois é no nível zero que temos ligações que carregam todas a corrente do sistema (ligações *red bonds*). Já para os outros níveis, nenhuma ligação carrega toda corrente do sistema, justificando o decaimento exponencial na proximidade de $x = 0$. As distribuições para $k > 0$ apresentam um formato semelhante entre elas, mas podemos notar que quanto maior for o nível, menor são as correntes. Isso é o esperado, uma vez que para níveis maiores mais vezes a corrente se divide.

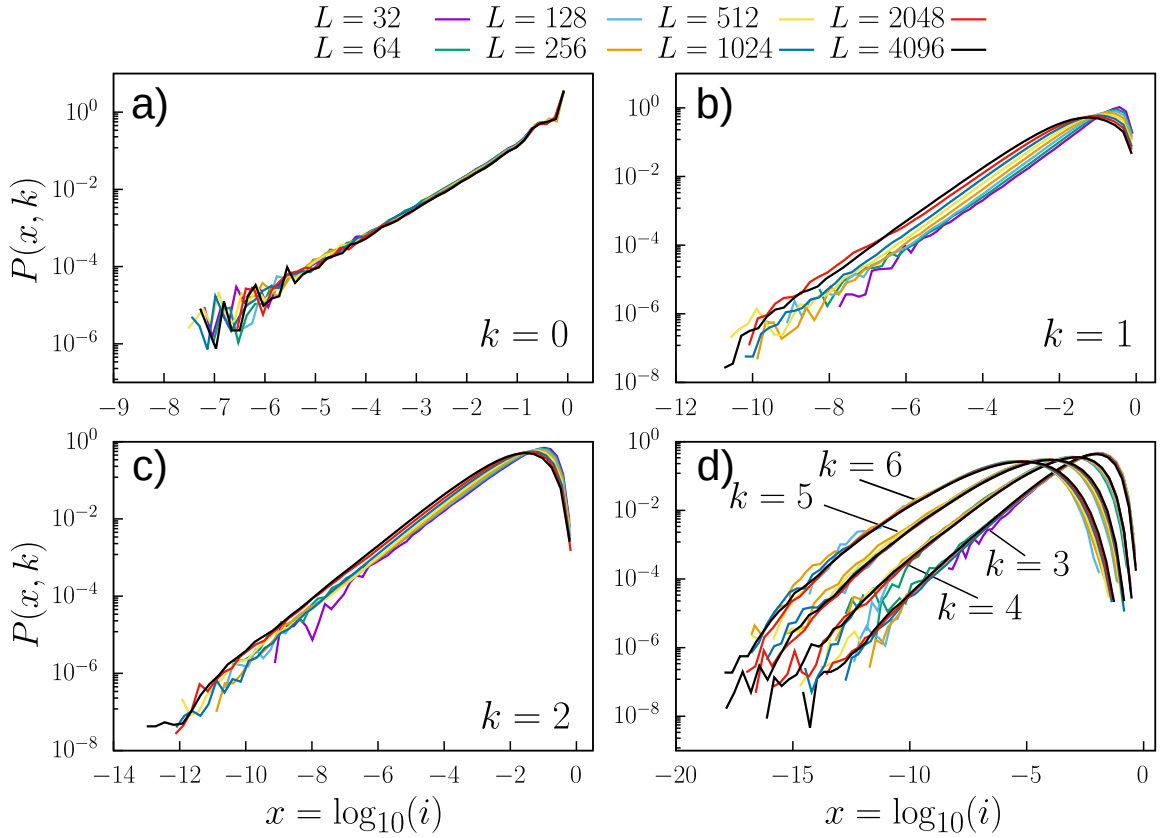
Gráfico 9 – Distribuição de correntes por nível k para sistema de tamanho $L = 2048$.



Fonte: Elaborada pelo autor. Distribuição de correntes das ligações do *backbone* crítico de percolação separadas até o nível dez para sistema de tamanho $L = 2048$. Como esperado, a distribuição para $k = 0$ possui um pico em $x = 0$, pois é nesse nível que encontramos as *red bonds*, ligações por onde passa toda corrente. Já para os outros níveis, nenhuma ligação carrega toda a corrente do sistema, justificando o decaimento exponencial na proximidade de $x = 0$. Para níveis cada vez maiores, a corrente se divide cada vez mais, resultando em correntes menores, explicando a translação para a esquerda.

Nos Gráficos 10.a, 10.b, 10.c e 10.d, podemos ver como a distribuição das correntes de cada nível muda com o tamanho do sistema L. Para o nível $k = 0$ (Gráfico 10.a), onde estão presentes as *red bonds*, a distribuição de correntes é invariante com relação ao tamanho do sistema. Já para os níveis $k = 1$ e $k = 2$ (Gráficos 10.b e 10.c respectivamente) a distribuição de correntes é ligeiramente diferente para cada valor de L, onde, quanto maior o sistema, mais frequentes as pequenas correntes se tornam. A partir do nível $k = 3$ e superior, as distribuições voltam a ser invariantes com relação ao tamanho do sistema, como podemos ver no Gráfico 10.d.

Gráfico 10 – Distribuição de correntes por nível k para diferentes tamanhos de sistema.



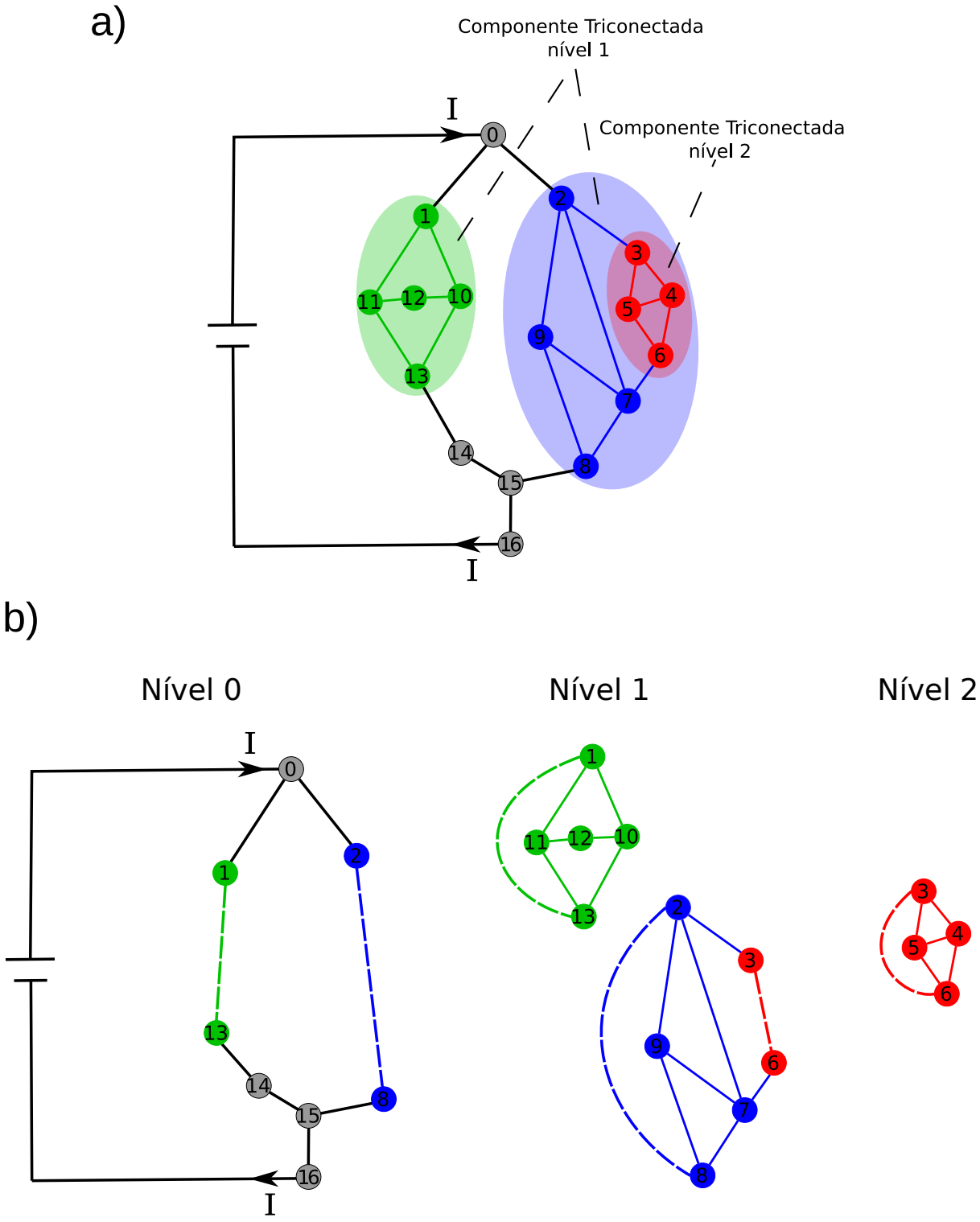
Fonte: Elaborada pelo autor. Distribuição de correntes das ligações do *backbone* crítico de percolação para os níveis $k = 1,2,3,4,5,6$ e em diferentes tamanhos de sistema L. O nível de cada ligação é definida de acordo com o nível da componente triconectada que a mesma pertence. Pelas curvas das distribuições dos níveis $k = 1$ e 2 em (b) e (c), percebemos o efeito do tamanho do sistema na distribuição, mas para os demais níveis, a distribuição de correntes é invariante com relação ao tamanho do sistema.

4.2.2 Distribuição de fatores

Assim como no modelo hierárquico simples discutido anteriormente, podemos pensar nas correntes em cada ligação l como a multiplicação de fatores, ou ainda, o logaritmo das correntes como a soma do logaritmo de fatores. A Figura 10 mostra um sistema com três componentes triconectadas, duas de nível $k = 1$ e uma de nível $k = 2$. Cada ligação possui uma resistência/condutância e uma corrente I entra no sistema pelo vértice 1 e tem como ponto de saída o vértice 16. O par $(v_t = 1, w_t = 16)$ é o terminal do sistema e a ligação entre os mesmos representa uma bateria com condutância igual a zero. Observe que a ligação da bateria garante que o grafo seja uma componente biconectada, caso contrário, a remoção do vértice 15 desconectaria o grafo.

Para estudar como os fatores compõem as correntes, vamos considerar dois tipos de fatores: fatores reais, f_r, sendo esses representando uma ligação real do sistema; e fatores virtuais, f_v, fatores que representam ligações virtuais de componentes triconectadas. Na Figura 10.b, temos a substituição das componentes triconectadas por ligações virtuais, onde as ligações contínuas são ligações reais do sistema e as tracejadas são as ligações virtuais de componentes

Figura 10 – Exemplo de um sistema de dois níveis e suas componentes triconectadas.



Fonte: Elaborada pelo autor. Sistema de dois níveis e suas componentes triconectadas. Em (a), temos um sistema com 16 vértices e três componentes triconectadas, duas de nível $k = 1$ e uma de nível $k = 2$, por onde uma corrente I é introduzida na rede pelo vértice 0. A substituição das componentes triconectadas por ligações virtuais (ligações tracejadas), permite que identifiquemos um total de três níveis de ligações, como mostrado em (b). Ligações em preto sólidas são de nível $k = 0$, ligações sólidas em verde e azul são de nível $k = 1$ e ligações sólidas vermelhas são de nível $k = 2$.

triconectadas. Tal como fizemos no sistema hierárquico simples, vamos analisar a composição das correntes de cada ligação l em cada nível:

- Nível $k = 0$

  Pela Figura 10.b, vemos que temos no total oito ligações de nível $k = 0$ das quais seis são reais e duas virtuais. Para as ligações reais, temos que o logaritmo das correntes é composta por:

$$\log[i_{rl}^{(0)}] = \log I + \log[f_{rl}^{(0)}], \tag{4.36}$$

  onde $i_{rl}^{(0)}$ é a corrente que passa pela ligação real l de nível 0 e $f_{rl}^{(0)}$ é o fator real multiplicativo de nível 0 da ligação real l. Já para as duas ligações virtuais, o logaritmo das correntes é:

$$\log[i_{vl}^{(0)}] = \log I + \log[f_{vl}^{(0)}], \tag{4.37}$$

  onde $i_{vl}^{(0)}$ é a corrente que passa pela ligação virtual l de nível 0 e $f_{vl}^{(0)}$ é o fator virtual multiplicativo de nível 0 da ligação virtual l.

- Nível k = 1

  Para o nível $k = 1$, temos duas componentes triconectadas com um total de 13 ligações reais de nível um e uma ligação virtual de mesmo nível. Para este nível, as correntes passando pelas ligações reais são dadas por:

$$\log[i_{rl}^{(1)}] = \log I + \log[f_{vl}^{(0)}] + \log[f_{rl}^{(1)}], \tag{4.38}$$

  onde vemos que o fator virtual de nível $k = 0$ é presente na composição da corrente. Já a corrente que passa pela ligação virtual é:

$$\log[i_{vl}^{(1)}] = \log I + \log[f_{vl}^{(0)}] + \log[f_{vl}^{(1)}], \tag{4.39}$$

  vemos assim a influência dos fatores virtuais dos níveis anteriores.

- Nível $k = 2$

  Por último, temos quatro ligações reais que formam a componente triconectada de nível $k = 2$. O logaritmo da corrente em cada ligação l é dada por:

$$\log[i_{rl}^{(2)}] = \log I + \log[f_{vl}^{(0)}] + \log[f_{vl}^{(1)}] + \log[f_{rl}^{(2)}], \tag{4.40}$$

Percebemos que os fatores das ligações virtuais de nível k das componentes triconectadas são parte das correntes das ligações reais de níveis superiores. Realizando a con-

tagem dos fatores reais e virtuais pelo número de vezes que ele contribui nas ligações reais, a Tabela 3 mostra a contagem de fatores para o exemplo discutido.

Tabela 3 – Contagem de fatores reais e virtuais

Tipo	**Fator**	**Total**
Real	$f_r^{(0)}$	6
	$f_r^{(1)}$	13
	$f_r^{(2)}$	4
Virtual	$f_v^{(0)}$	18
	$f_v^{(1)}$	5

Fonte: Elaborada pelo autor. Contagem do número de vezes que cada fator (real e virtual) contribuem para cada corrente de ligações reais do sistema da Figura 10.

Podemos generalizar que as correntes $i^{(k)}$ em cada ligação real de nível k é dada por:

$$\log[i^{(k)}] = \log I + \sum_{j=0}^{k-1} \log[f_v^{(j)}] + \log[f_r^{(k)}]. \tag{4.41}$$

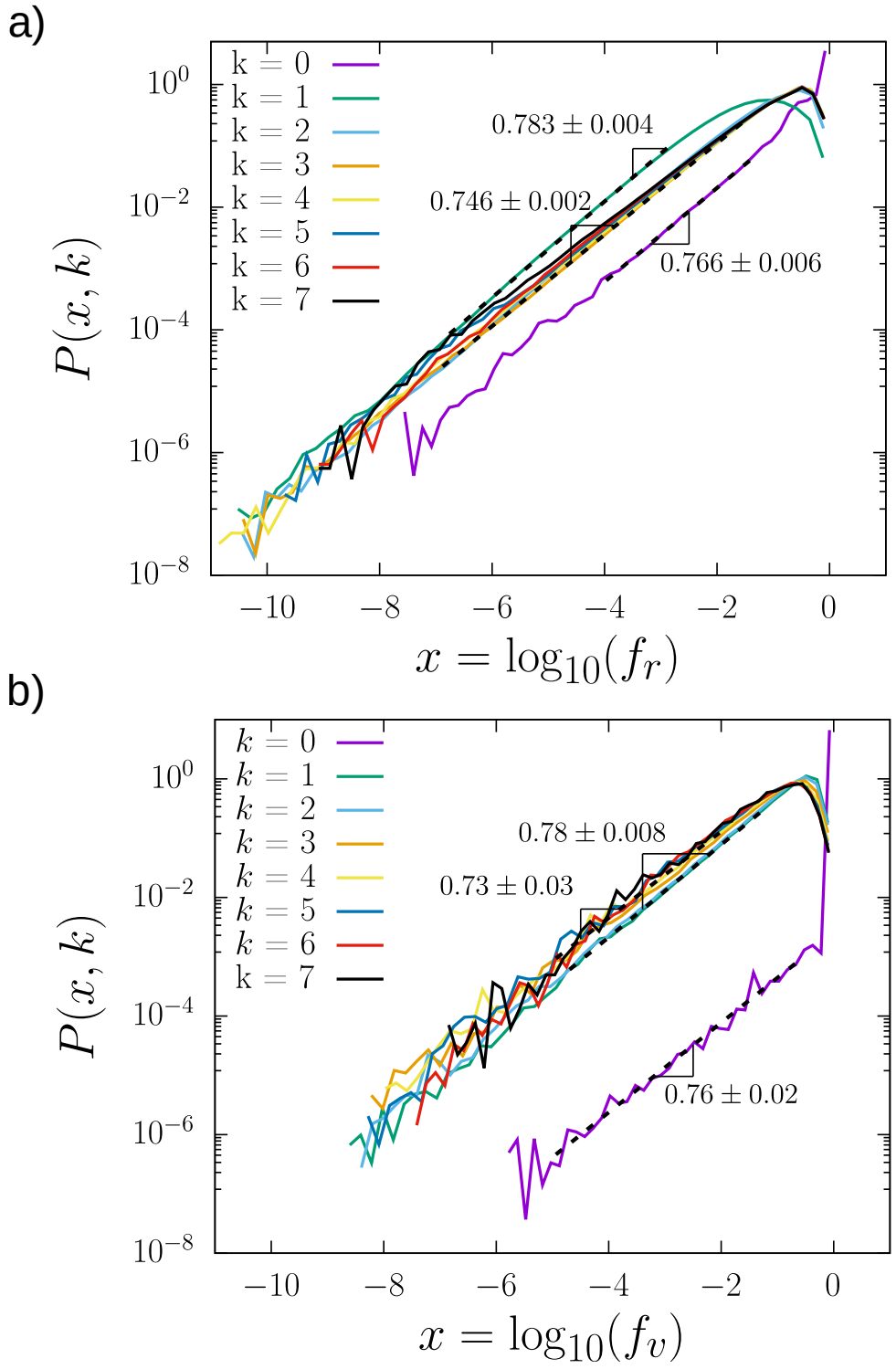
Essa metodologia de separar em dois tipos de fatores e de contagem, nos fornece diferentes distribuições para os dois tipos de fatores e diferentes níveis. Nos Gráficos 11.a e 11b, temos as distribuições de fatores reais e virtuais, respectivamente, em diferentes níveis para sistema de tamanho $L = 2048$. Ambos os tipos de distribuição de fatores seguem uma lei de potência $P(f) \sim f^{\alpha}$ (onde f está representando tanto fatores reais, f_r, como fatores virtuais, f_v), em todos os níveis por mais de sete ordens de grandeza. Percebemos que o expoente α para todos os níveis é, considerando as flutuações estatísticas, aproximadamente $\alpha = 3/4$. Para $k > 0$, todas as distribuições (reais e virtuais) possuem um corte exponencial próximo de $f = 1$. Isso é esperado, pois dentro de qualquer componente triconectada não existe nenhuma ligação carregando toda corrente do sistema, ou seja, nenhuma ligação *red bond* e fatores para essas ligações são presentes apenas no nível $k = 0$. Outra característica das distribuições de fatores que percebemos é que para as distribuições reais de níveis $k > 1$ mostram-se ser independentes do nível, enquanto que as de fatores virtuais mostram-se independentes para $k > 0$.

A fim de verificar se as distribuições de fatores são de fato independentes, isto é, se os fatores de diferentes níveis não possuem correlações e que a Equação 4.41 para o logaritmo das correntes pode ser interpretada como a soma do logaritmo de fatores independentes, tal como a Equação 4.21 para o modelo hierárquico simples, reconstruímos a distribuição de correntes realizando amostragens aleatórias e independentes dos fatores de acordo com as distribuições mostradas nos Gráficos 11.a e 11b.

Como podemos ver pelo Gŕafico 12, a distribuição de correntes reconstruída a partir das distribuições de fatores, é muito próxima das distribuições reais simuladas em cada nível para sistema de tamanho $L = 2048$. Assim, indicando que os fatores são distribuições indepen-

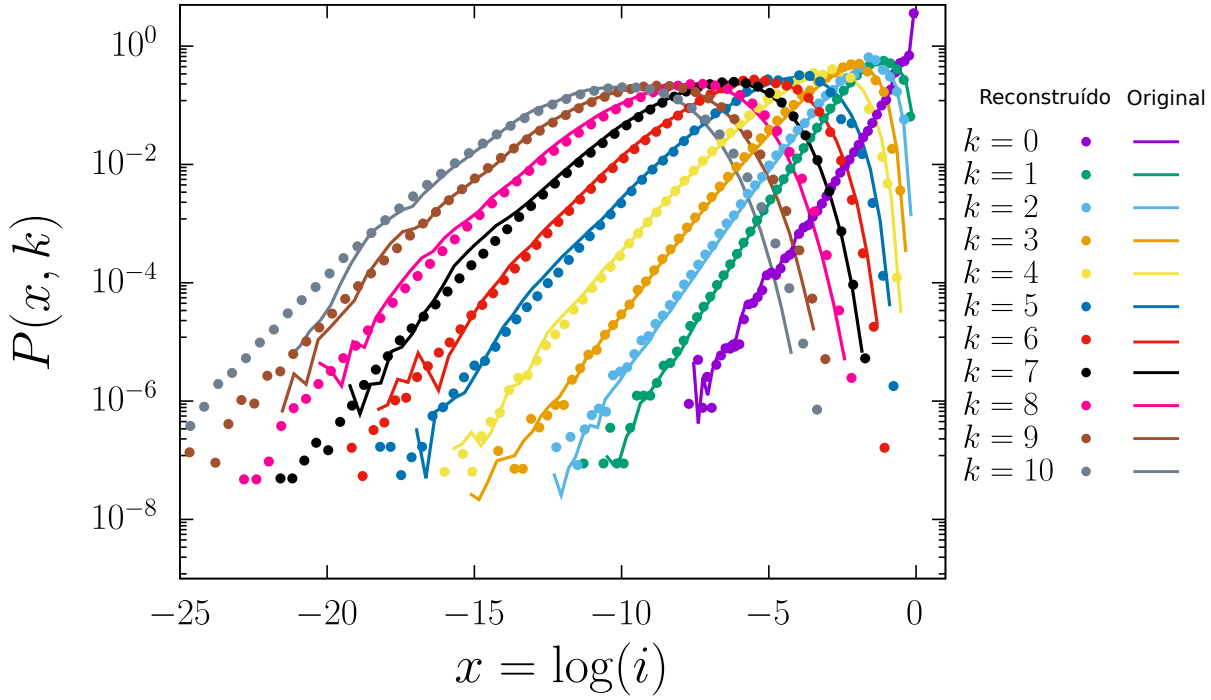
dentes.

Gráfico 11 – Distribuição de fatores reais e virtuais por nível para sistema de tamanho $L = 2048$.



Fonte: Elaborada pelo autor. Distribuição de fatores reais e virtuais separados por níveis k no modelo hierárquico do *backbone* crítico de percolação. Para ambos os tipos de fatores, as distribuições apresentam um corte exponencial na proximidade de $x = 0$, com exceção do nível $k = 0$, onde então presentes as ligações *red bonds*. Em (a) temos a distribuição de fatores reais, onde, considerando flutuações estatísticas, as caldas das distribuições colapsam umas nas outras para níveis $k > 1$, e em (b) são as distribuições de fatores virtuais, onde a contagem é feita pelo número de vezes que cada fator virtual contribui para as correntes em ligações reais. Percebemos que todas as distribuições seguem uma lei de potência, $P(f) \sim f^{\alpha}$, sendo f fatores reais (f_r) ou virtuais (f_v), por mais de sete ordens de grandeza, cujo expoente é consistente com $\alpha = 3/4$.

Gráfico 12 – Comparação das distribuições de correntes por níveis originais e reconstruídas para $L = 2048$.



Fonte: Elaborada pelo autor. Distribuição de correntes separadas por níveis para *backbone* crítico de percolação para sistema de tamanho $L = 2048$. As linhas sólidas correspondem a distribuição de correntes originais simuladas. Já as distribuições representadas pelos pontos são das correntes reconstruídas a partir da Equação 4.41 e da amostração aleatória dos fatores dos Gráficos 11.a e 11b. Percebemos que as distribuições reconstruídas são próximas das distribuições originais, sugerindo que os fatores de diferentes níveis são estatisticamente independentes.

5 CONCLUSÃO

No primeiro projeto com D. Torres e outros, os resultados do estudo mostraram que os dados de rastreamento ocular podem ser analisados para determinar a complexidade e coerência de diferentes textos. Os resultados mostraram uma relação quase monotônica entre a magnetização média das atividades de fixação e a complexidade média dos textos e a distância entre a temperatura de leitura e a temperatura crítica ($T_o - T_c$) para separação de textos coerentes de textos aleatórios. Isso foi confirmado por meio de um questionário com 400 respondentes.

Já no segundo projeto, ao explorar a estrutura auto-similar do *backbone* crítico de percolação, fomos capazes de criar uma estrutura hierárquica usando a decomposição do mesmo em componentes triconectadas, onde uma componente triconectada contida em outra cria um novo nível superior. Isso permitiu que resolvêssemos uma série de equações de *Kirchhoff* menores ao invés de um maior. Por exemplo, para um sistema de tamanho $L = 1024$, teríamos que resolver um conjunto de equações de *Kirchhoff* da ordem de $1024^{1.64} \sim 86000$ equações, mas pelo método desenvolvido, precisamos resolver aproximadamente conjuntos de $1024^{1.23} \sim 5000$ equações.

Em nosso modelo hierárquico, cada ligação possui um nível, no qual definimos que ligações de nível k são aquelas que pertencem a uma componente triconectada de mesmo nível, mas que não fazem parte de componentes de nível $k+1$ e superior. O primeiro nível dessa hierarquia corresponde a ligações que estão em série ou em paralelo e é apenas nesse nível que se encontram as ligações *red bonds*. Vimos pela distribuição de nível que quanto maior o tamanho do sistema, níveis mais profundos são formados.

Ao calcularmos as correntes em cada ligação do *backbone* crítico de percolação pela metodologia desenvolvida, fomos capazes de determinar correntes muito pequenas com precisão de até 10^{-35}. Isso proporcionou que observássemos com precisão a distribuição de correntes em cada nível, mostrando que as distribuições de nível $k = 0$ e $k > 2$ são invariantes com relação ao tamanho do sistema, onde apenas nos níveis $k = 1$ e $k = 2$ o efeito finito do tamanho do sistema é presente.

Tal como no modelo hierárquico simples, exploramos a composição das correntes como a multiplicação de fatores. Em nosso modelo hierárquico, definimos dois tipos de fatores: fatores reais, aqueles que representam ligações reais do sistema; e fatores virtuais representando fatores de ligações virtuais de componentes triconectadas. Para determinar a distribuição dos fatores, contamos o número de vezes que cada fator contribuía para as correntes nas ligações, mas os fatores virtuais de nível k contribuíam para todas as ligações k e superior. Assim, obtivemos distribuições de fatores para cada nível e tipo, onde cada distribuição segue uma lei de potência por mais de sete ordens de grandeza com expoentes que são consistentes com o valor de $3/4$.

A fim de demonstrar que as distribuições de fatores são independentes, recon-

struímos as correntes por um processo de *reshuffling* de dados, e assim mostramos que as distribuições de fatores para ligações reais e virtuais são não correlacionadas. Isso significa que o comportamento finito complexo da distribuição de corrente pode ser modelado multiplicando aleatoriamente fatores extraídos de suas distribuições de potência.

REFERÊNCIAS

[1] TORRES, D. *et al.* Eye-tracking as a proxy for coherence and complexity of texts. *Plos one*, Public Library of Science San Francisco, CA USA, v. 16, n. 12, p. e0260236, 2021.

[2] STAUFFER, D.; AHARONY, A. *Introduction to percolation theory: revised second edition*. [S.l.]: CRC press, 2014.

[3] SAHINI, M.; SAHIMI, M. *Applications of percolation theory*. [S.l.]: CRC Press, 2014.

[4] HERRMANN, H. J.; ROUX, S. *Statistical models for the fracture of disordered media*. [S.l.]: Elsevier, 2014.

[5] LANDAUER, R. Electrical conductivity in inhomogeneous media. In: AIP. *AIP conference proceedings*. [S.l.], 1978. v. 40, n. 1, p. 2–45.

[6] BERGMAN, D. J.; STROUD, D. Physical properties of macroscopically inhomogeneous media. In: *Solid state physics*. [S.l.]: Elsevier, 1992. v. 46, p. 147–269.

[7] ARCANGELIS, L. D.; REDNER, S.; CONIGLIO, A. Anomalous voltage distribution of random resistor networks and a new model for the backbone at the percolation threshold. *Physical Review B*, APS, v. 31, n. 7, p. 4725, 1985.

[8] RAMMAL, R. *et al.* Flicker (1 f) noise in percolation networks: A new hierarchy of exponents. *Physical review letters*, APS, v. 54, n. 15, p. 1718, 1985.

[9] KAHNG, B. *et al.* Electrical breakdown in a fuse network with random, continuously distributed breaking strengths. *Physical Review B*, APS, v. 37, n. 13, p. 7625, 1988.

[10] ARCANGELIS, L. de; CONIGLIO, A. Infinite hierarchy of exponents in a two-component random resistor network. *Journal of statistical physics*, Springer, v. 48, n. 3-4, p. 935–942, 1987.

[11] FOURCADE, B.; TREMBLAY, A.-M. Anomalies in the multifractal analysis of self-similar resistor networks. *Physical Review A*, APS, v. 36, n. 5, p. 2352, 1987.

[12] MEIR, Y.; AHARONY, A. Averaging of multifractals. *Physical Review A*, APS, v. 37, n. 2, p. 596, 1988.

[13] HARRIS, A. B.; MEIR, Y.; AHARONY, A. Resistance distributions of the random resistor network near the percolation threshold. *Physical Review B*, APS, v. 41, n. 7, p. 4610, 1990.

[14] NAGATANI, T. Renormalisation group approach to an infinite set of exponents of random resistor networks at the percolation threshold. *Journal of Physics A: Mathematical and General*, IOP Publishing, v. 20, n. 6, p. L417, 1987.

[15] ARCANGELIS, L. D.; REDNER, S.; CONIGLIO, A. Multiscaling approach in random resistor and random superconducting networks. *Physical Review B*, APS, v. 34, n. 7, p. 4656, 1986.

[16] ARCANGELIS, L. D.; CONIGLIO, A.; REDNER, S. Multifractal structure of the incipient infinite percolating cluster. *Physical Review B*, APS, v. 36, n. 10, p. 5631, 1987.

[17] BATROUNI, G. G.; HANSEN, A.; ROUX, S. Negative moments of the current spectrum in the random-resistor network. *Physical Review A*, APS, v. 38, n. 7, p. 3820, 1988.

[18] AHARONY, A.; BLUMENFELD, R.; HARRIS, A. B. Distribution of the logarithms of currents in percolating resistor networks. i. theory. *Physical Review B*, APS, v. 47, n. 10, p. 5756, 1993.

[19] BARTHÉLÉMY, M. *et al.* Multifractal properties of the random resistor network. *Physical Review E*, APS, v. 61, n. 4, p. R3283, 2000.

[20] JAYNES, E. T. Information theory and statistical mechanics. *Physical review*, APS, v. 106, n. 4, p. 620, 1957.

[21] MCCOY, B. M.; WU, T. T. *The two-dimensional Ising model*. [S.l.]: Courier Corporation, 2014.

[22] DOMINICIS, C. D.; GIARDINA, I. *Random fields and spin glasses: a field theory approach*. [S.l.]: Cambridge University Press, 2006.

[23] FISCHER, K. H.; HERTZ, J. A. *Spin glasses*. [S.l.]: Cambridge university press, 1993. v. 1.

[24] 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.

[25] HINTON, G. E.; SEJNOWSKI, T. J. Learning and releaming in boltzmann machines. *Parallel Distrilmted Processing*, v. 1, 1986.

[26] KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The annals of mathematical statistics*, JSTOR, v. 22, n. 1, p. 79–86, 1951.

[27] COVER, T. M.; THOMAS, J. A. *Elements of information theory*. [S.l.]: John Wiley & Sons, 2012.

[28] MEZARD, M.; MONTANARI, A. *Information, physics, and computation*. [S.l.]: Oxford University Press, 2009.

[29] MCNAMARA, D. S. *et al.* Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, Taylor & Francis, v. 14, n. 1, p. 1–43, 1996.

[30] ROTHMAN, R. The complex matter of text complexity. *Harvard Education Letter*, v. 28, n. 5, p. 1–2, 2012.

[31] FISHER, D.; FREY, N. Text complexity and close readings. *International Reading Association: Newark DE*, 2012.

[32] JR, C. C. *et al.* Eye movements in reading and information processing: Keith rayner's 40 year legacy. *Journal of Memory and Language*, Elsevier, v. 86, p. 1–19, 2016.

[33] RAYNER, K.; DUFFY, S. A. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, Springer, v. 14, n. 3, p. 191–201, 1986.

[34] KLIEGL, R.; NUTHMANN, A.; ENGBERT, R. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of experimental psychology: General*, American Psychological Association, v. 135, n. 1, p. 12, 2006.

[35] RAYNER, K. *et al.* Eye movements and on-line language comprehension processes. *Language and Cognitive processes*, Taylor & Francis, v. 4, n. 3-4, p. SI21–SI49, 1989.

[36] RAYNER, K.; RANEY, G. E. Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, Springer, v. 3, n. 2, p. 245–248, 1996.

[37] EHRLICH, S. F.; RAYNER, K. Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, Elsevier, v. 20, n. 6, p. 641–655, 1981.

[38] ROSA, A. J. G. Do sertão às fronteiras. *Revista Brasileira (Academia Brasileira de Letras)*, 2018;.

[39] ARARIPE, T. A. *José de Alencar: perfil literario*. [S.l.]: Rio de Janeiro : Typ. da Escola de Serafim José Alves, 1880.

[40] TARJAN, R. Depth-first search and linear graph algorithms. *SIAM journal on computing*, SIAM, v. 1, n. 2, p. 146–160, 1972.

[41] CORMEN, T. H. *et al. Introduction to algorithms*. [S.l.]: MIT press, 2009.

[42] EVEN, S. *Graph algorithms*. [S.l.]: Cambridge University Press, 2011.

[43] HOPCROFT, J.; TARJAN, R. Algorithm 447: efficient algorithms for graph manipulation. *Communications of the ACM*, ACM, v. 16, n. 6, p. 372–378, 1973.

[44] CONIGLIO, A. Cluster structure near the percolation threshold. *Journal of Physics A: Mathematical and General*, IOP Publishing, v. 15, n. 12, p. 3829, 1982.

[45] GYURE, M. F. *et al.* Mass distribution on clusters at the percolation threshold. *Physical Review E*, APS, v. 51, n. 3, p. 2632, 1995.

[46] HERRMANN, H. J.; STANLEY, H. E. Building blocks of percolation clusters: Volatile fractals. *Physical review letters*, APS, v. 53, n. 12, p. 1121, 1984.

[47] BATTISTA, G. D.; TAMASSIA, R. Incremental planarity testing. In: IEEE COMPUTER SOCIETY. *30th Annual Symposium on Foundations of Computer Science*. [S.l.], 1989. p. 436–441.

[48] BATTISTA, G. D.; TAMASSIA, R. On-line maintenance of triconnected components with spqr-trees. *Algorithmica*, Springer, v. 15, n. 4, p. 302–318, 1996.

[49] BATTISTA, G. D.; TAMASSIA, R. On-line planarity testing. *SIAM Journal on Computing*, SIAM, v. 25, n. 5, p. 956–997, 1996.

[50] BERTOLAZZI, P.; BATTISTA, G. D.; DIDIMO, W. Computing orthogonal drawings with the minimum number of bends. *IEEE Transactions on Computers*, IEEE, v. 49, n. 8, p. 826–840, 2000.

[51] BIENSTOCK, D.; MONMA, C. L. On the complexity of embedding planar graphs to minimize certain distance measures. *Algorithmica*, Springer, v. 5, p. 93–109, 1990.

[52] GUTWENGER, C.; MUTZEL, P.; WEISKIRCHER, R. Inserting an edge into a planar graph. *Algorithmica*, Springer, v. 41, p. 289–308, 2005.

[53] MUTZEL, P.; WEISKIRCHER, R. Optimizing over all combinatorial embeddings of a planar graph. In: SPRINGER. *Integer Programming and Combinatorial Optimization: 7th International IPCO Conference Graz, Austria, June 9–11, 1999 Proceedings 7*. [S.l.], 1999. p. 361–376.

[54] HOPCROFT, J. E.; TARJAN, R. E. Dividing a graph into triconnected components. *SIAM Journal on Computing*, SIAM, v. 2, n. 3, p. 135–158, 1973.

[55] GUTWENGER, C.; MUTZEL, P. A linear time implementation of spqr-trees. In: SPRINGER. *International Symposium on Graph Drawing*. [S.l.], 2000. p. 77–90.

[56] PAUL, G.; STANLEY, H. E. Beyond blobs in percolation cluster structure: The distribution of 3-blocks at the percolation threshold. *Physical Review E*, APS, v. 65, n. 5, p. 056126, 2002.

[57] NEWMAN, M. E.; ZIFF, R. M. Fast monte carlo algorithm for site or bond percolation. *Physical Review E*, APS, v. 64, n. 1, p. 016706, 2001.

[58] PRESS, W. H. *et al. Numerical recipes 3rd edition: The art of scientific computing*. [S.l.]: Cambridge university press, 2007.

ANEXO A – ARTIGO: DECOMPOSING THE PERCOLATION BACKBONE REVEALS NOVEL SCALING LAWS OF THE CURRENT DISTRIBUTION

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 20 December 2023
DOI 10.3389/fphy.2023.1335339

OPEN ACCESS

EDITED BY
Fernando A. Oliveira,
University of Brasilia, Brazil

REVIEWED BY
Mikko Alava,
Aalto University, Finland
Vaughan Voller,
University of Minnesota Twin Cities, United States

*CORRESPONDENCE
André A. Moreira,
auto@fisica.ufc.br

RECEIVED 08 November 2023
ACCEPTED 28 November 2023
PUBLISHED 20 December 2023

CITATION
Sena WRd, Andrade JS Jr., Herrmann HJ and Moreira AA (2023), Decomposing the percolation backbone reveals novel scaling laws of the current distribution. *Front. Phys.* 11:1335339. doi: 10.3389/fphy.2023.1335339

Decomposing the percolation backbone reveals novel scaling laws of the current distribution

Wagner R. de Sena[1], José S. Andrade Jr.[1], Hans J. Herrmann[1,2] and André A. Moreira[1]*

[1]Departamento de Física, Universidade Federal do Ceará, Fortaleza, Brazil, [2]PMMH, ESPCI, CNRS UMR 7636, Paris, France

The distribution of currents on critical percolation clusters is the fundamental quantity describing the transport properties of weakly connected systems. Nevertheless, its finite-size extrapolation is still one of the outstanding open questions concerning disordered media. By hierarchically decomposing the 3-connected components of the backbone, we disclose that the current distribution is determined from two distributions, namely, the one corresponding to the number of bonds in each level and another one corresponding to the factors by which the current is reduced, when going from one level to the next. The first distribution follows a finite-size scaling, while the second is a power law with an exponent consistent with 3/4 in two dimensions. The standard hierarchical model for the backbone is too simple to reproduce this complex scenario. Our new decomposition method of the backbone also allows to calculate much smaller currents than before, attaining a precision of 10^{-35} and systems of size $L = 8192^2$. Moreover, our method is not restricted to electric currents on critical percolation clusters but could also be applied to other transport problems on sparse graphs including fluid flow and car traffic.

KEYWORDS

percolation, multifractal, transport phenomena, finite-size scaling analysis, self-similar (fractal) systems

1 Introduction

Percolation was originally proposed to model the flow through a porous medium [1] but has since turned out to be a fundamental model in statistical physics [2], serving as a geometrical template for phase transitions with multiple applications in physics and beyond, including the sol–gel transition, the onset of fluid flow, the conductivity of random media, opinion dynamics, and the outbreak of epidemics. An important issue in percolation theory is the solution of linear transport at criticality [3]. Under such a framework, one replaces the bonds of a percolation cluster by Ohmic resistors and applies a potential difference between two distant sites on this cluster. Solving the set of linear equations given by Kirchhoff's nodal rule at each node yields the currents at each bond. This linear transport problem in percolation has many applications. Examples include flow through porous materials [4–8], oil production [9], and conductivity of semiconducting materials or metal–insulator mixtures [10].

The distribution of the currents on the percolation cluster at criticality has been found to be multifractal [11] since different moments exhibit unrelated scaling exponents. However, its multifractal spectrum $f(\alpha)$ strongly depends on the system size [12]. Despite the great

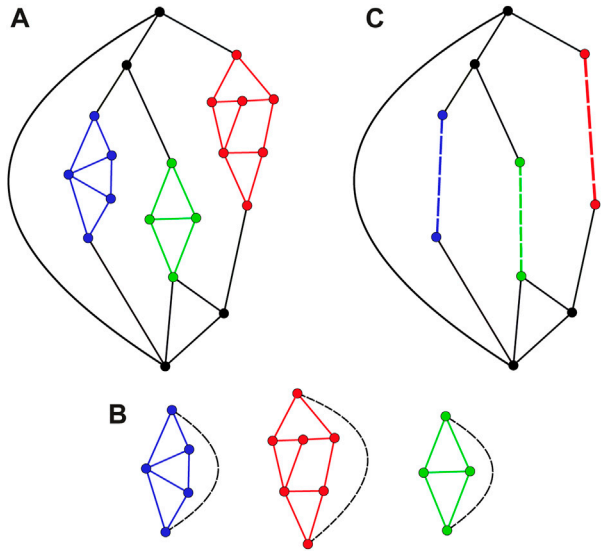


FIGURE 1
Formally, a 3-connected component (3CC) is the set of nodes in a graph that will remain in the same component after any two bonds are removed [16]. For a physicist, however, an intuitive definition may be any subset of the graph that connects to the rest only at a pair of articulating nodes and is not formed by simple parallel and/or series conformations. The hierarchical partition of a graph in 3CCs can be used to solve efficiently the Kirchhoff problem [16]. In panel **(A)**, we show a graph with four 3CCs. After solving the Kirchhoff problem for the three 3CCs shown in **(B)**, these components can be replaced by effective resistances, simplifying the solution of 3CC on a larger scale, as shown in **(C)**.

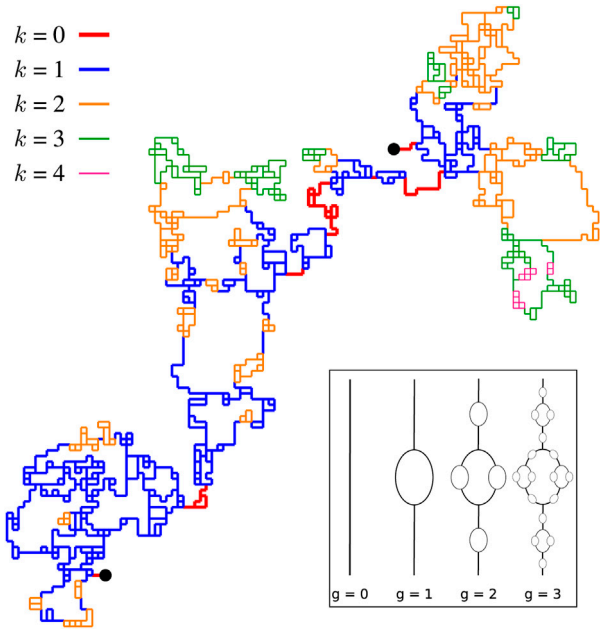


FIGURE 2
Typical critical backbone with bonds colored by level. Bonds of level k are those inside a 3CC of level k but outside 3CCs of level $k + 1$. The black dots indicate the pole nodes, between which a potential difference is applied. The inset shows the hierarchical model proposed in [11]. Inspired by the self-similarity property of the conducting backbone, this simple model can be solved analytically to compute the current distribution. Higher generations g reveal smaller scales. Considering that every bond has the same resistance, the current passing through resistor ℓ will be given by $I_\ell = I_t 2^{-n_\ell}$, where I_t is the total current, and the numbers n_ℓ follow a binomial distribution, resulting in a log-normal distribution for the currents [20]. However, although the critical backbone in the main figure is self-similar and presents a hierarchical structure [18], the current distribution does not follow a log-normal relation [12].

controversy generated about the asymptotic behavior of the distribution for weak currents and large systems [13–15], no satisfactory solution has been found yet [12].

By definition, a 3-connected component (3CC) is the set of nodes of a graph that remains connected after any two bonds are removed [16]. It should be noted that simple parallel and series conformations cannot form 3CCs. This concept can be directly associated with physical partitioning and has been very useful in the solution of several problems in graph theory [16, 17]. For instance, the partition of the critical conducting backbone in 3CCs has been successfully used to demonstrate that these components are also fractal [18], like other structures in critical percolation [19].

Here, we will focus precisely on weak currents and large system sizes to introduce a new way to calculate the current distribution in the critical conducting backbone based on its hierarchical partition in terms of 3-connected components. The backbone is the part of the infinite cluster that takes part in the conduction. Formally, the backbone is defined as the union of the sets of self-avoiding paths that connect two extremes of the cluster. By taking advantage of the partition of the backbone on 3CCs, we are able to solve subsets of coupled linear equations in sequence, starting from 3CCs on the smallest scale. As illustrated in Figure 1, the solution of the electrical transport problem on these components allows determining their effective resistances. By replacing these 3CCs to their corresponding effective conductances, the Kirchhoff problem can then be sequentially solved on larger and larger 3CC scales, up to the scale of the critical backbone itself.

1.1 Hierarchical structure of the conducting backbone

It has been proposed that the conducting backbone can be separated in blobs and red bonds, with the red bonds being the connections that are removed to split the backbone into two separate parts, while the blobs are the parts of the conducting backbone that are multi-connected, that is, that remain connected after the removal of any bond. Here, we expand on this idea using the concept of 3CCs. In our definition, 3CCs at the largest scale, namely, the blobs of the backbone, are of level 1, and components of level 2 are those that are replaced by effective bonds in components of level 1, and so on, with components of one level replacing effective bonds in the components of the level below. It should be reminded that sets of bonds in simple parallel and series conformations, despite being connected to the rest of the graph at just two points, do not form 3CCs. Therefore, one needs to include factors of level 0, accounting for splits of the current that take place outside all 3CCs. A typical example of the critical conducting backbone is shown in Figure 2, where the bonds are colored according to their 3CC levels, clearly indicating the underlying hierarchical structure of the partition.

The idea that the structure of critical percolation clusters can be described in terms of a hierarchical model to determine the current

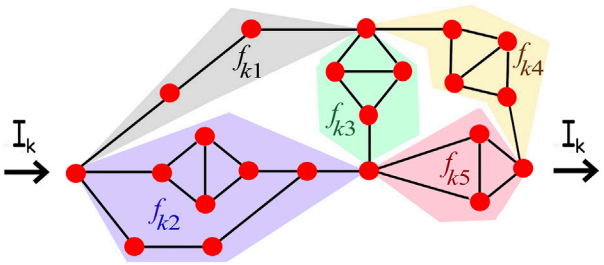


FIGURE 3
Recovering the currents in a typical 3-connected component of level k through which passes a current I_k. After its internal 3CCs have been replaced by effective bonds, this component turns into a simple Wheatstone bridge that can be readily solved to obtain the fraction f_{ks} of the current I_k passing through each segment s of the Wheatstone bridge. The current of every bond in segment s is then reduced in level k by this factor f_{ks}. We consider two different ways of sampling the factors f_{ks}. In the first sampling, we count the so-called *actual* bonds of level k. These are bonds in segment s that are inside level k but outside level $k + 1$. In the second sampling, we count the *internal bonds*, namely, all other bonds in the segment s that are not actual bonds of level k, that is, any bond in s found inside 3CCs of level $k + 1$. In this example, the factor f_{k1} is sampled three times in the distribution for actual bonds and is not included in the distribution for internal bonds as there are no 3CCs in this segment. The factor f_{k5} is sampled five times in the distribution of internal bonds and is not included in the distribution of actual bonds as there is no bond carrying the whole current passing through this segment. The factor f_{k2} is sampled 10 times by internal bonds and just one by actual bonds.

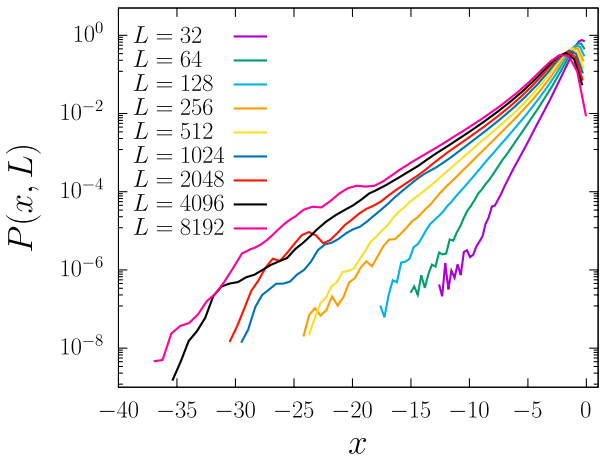


FIGURE 4
Current distribution on critical conducting backbones for different system sizes $L \times L$. In this graph, $x = \log_{10}(i_\ell)$, with i_ℓ, being the currents in each bond. The results were obtained for square lattice subjected to bond percolation at the critical point $p_c = 1/2$. For each size, we simulated at least 2,000 samples. For each sample, we recovered the largest cluster, applied a potential difference at a pair of nodes separated by $L/2$ lattice units, and extracted the backbone [22]. Our approach allows obtaining with precision currents of the order of 10^{-35}.

distribution has been originally suggested in [20]. In this model, as shown in the inset of Figure 2, each bond ℓ has a current $I_\ell = I_t 2^{-n_\ell}$, where I_t is the total current through the lattice, and the exponents n_ℓ are distributed according to a binomial distribution. As a result, the currents follow a log-normal distribution and exhibit multifractal properties [20]. Unfortunately, the clever approach proposed in [20] does not succeed in describing the current distribution on critical conducting backbones since their distribution is not log-normal [12]. This deviation from a log-normal may seem surprising since the critical conducting backbone is self-similar and presents a hierarchical structure [18]. Even more surprising is the fact that the current distribution on the critical backbone does not appear to follow consistent finite-size scaling laws since the distribution assumes different shapes at different scales [12].

2 Methods

Taking advantage of the partition in 3-connected components, the complex problem of a large conducting backbone, with a large number of unknown variables, is replaced by a series of steps. At each step, a single Kirchhoff problem is solved for a single 3CC, which will be replaced by an effective bond when solving the component at a larger scale. In this way, at each step, the number of variables is much smaller. Since the complexity of solving these sets of coupled equations grows super-linearly, for large system sizes, the time gained by reducing the rank of the equations should, therefore, compensate the pre-processing step to obtain the partition. To give an idea of the advantage of employing this decomposition, in two dimensions, the number of nodes in the backbone scales with the system size L as $M_b \sim L^{1.64}$ [21], while the largest 3CC scales as $M_3 \sim L^{1.15}$ [18].

After using the partition in 3-connected components to solve the Kirchhoff problem on the conducting backbone, we proceed with the recovering of the current distribution. To obtain the current through a given bond ℓ, we need to collect the information from every component containing this bond. Each bond will carry a fraction $f_\ell = \prod_{j=0}^{k} f_{j\ell}$ of the total current, where the factors $f_{j\ell}$ must be determined by solving the Kirchhoff problem on the 3CC at level j, as shown in Figure 3. The number of factors in this product is equal to the level k of the bond, that is, the number of 3CCs where bond ℓ is nested. At each step of the hierarchical solution, the current in each effective bond is typically a significant portion of the current passing through the component. However, when the factors of several nested components are multiplied to obtain the final current of a given bond, the result can be many orders of magnitude smaller than the total current. As a consequence, our hierarchical approach allows in accurately obtaining a current distribution that spans over nearly 16 orders of magnitude, as seen in Figure 4.

2.1 Distribution of current factors

At this point, we make use of our method to study the distribution of the multiplying factors, f_{ks}. In order to fully describe their statistics, two different sampling ways are adopted, as shown in Figure 3. In the first way, a factor f_{ks} is sampled by the number of *actual* bonds of level k in segment s. A bond is of level k when it is inside level k but outside level $k + 1$. Alternatively, we sample the same factor f_{ks} by the number of *internal* bonds. The internal bonds correspond to all other bonds in segment s that are not actual bonds, namely, all bonds in the segment s that are nested inside 3CCs of level $k + 1$ or higher.

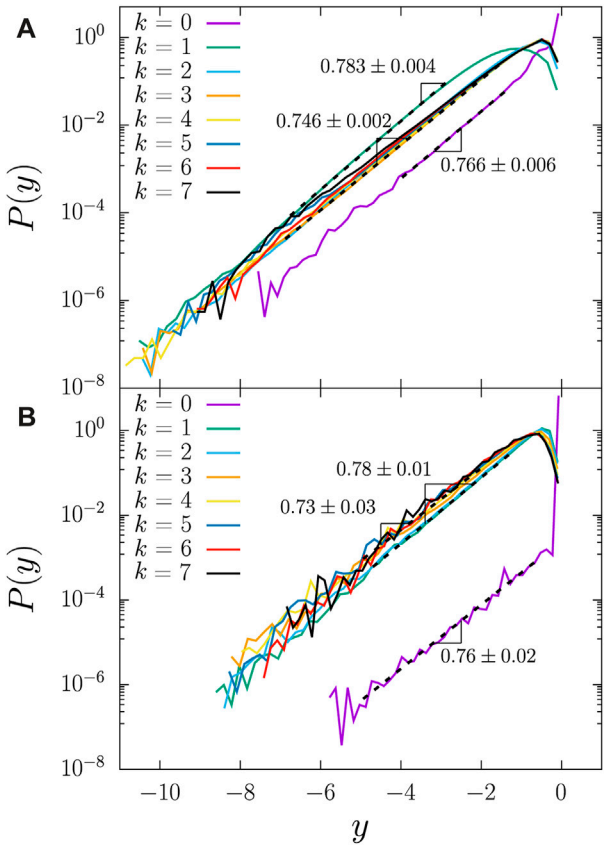


FIGURE 5
Distribution of factors for each level k of the hierarchical structure of the conducting backbone at the critical bond percolation point, $p_c = 1/2$. In this graph, $y = \log_{10}(\mathbf{F})$, with f being the factor by which the currents are reduced when split inside a given 3CC. These results were obtained from numerical simulations using 10,000 realizations of two-dimensional square lattices with size $L = 2048$. In **(A),** we show the distribution of factors for *actual* bonds, and in **(B),** the distribution for factors on *effective* bonds. As explained in the main text and in the caption of Figure 3, both distributions are obtained from the same set of factors but with different sampling weights. The distributions closely follow power laws, $P(\mathbf{F}) \sim f^{\alpha}$, over up to seven orders of magnitude with exponents that are consistent with the value $\alpha = 3/4$. Except for $k = 0$, they all display an exponential cut-off in the proximity of $f = 1$. Within the statistical fluctuations, the tail of the curves falls on top of each other for $k > 1$.

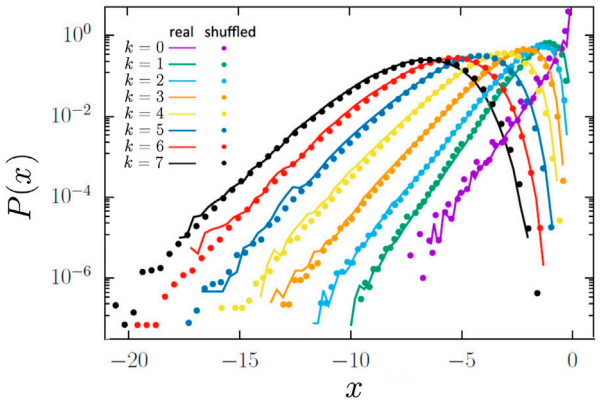


FIGURE 6
Current distribution segregated by the bond level. In this graph $x = \log_{10}(i_\ell)$, with i_ℓ being the currents in a given bond. The solid lines correspond to the current distributions on the bonds within a given level k. As expected, the bonds at higher levels tend to carry a smaller fraction of the total current. In order to test for the presence of correlations between values of factors at different levels, we also generate current distributions from randomly shuffled data (dots). As can be seen, the shuffled data follow closely the numerical results, suggesting that the factors at different levels are statistically independent.

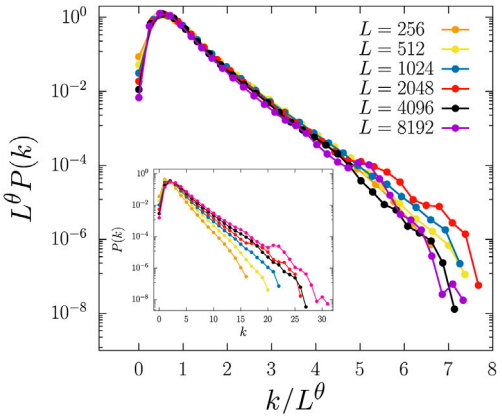


FIGURE 7
Distribution of bond levels k. As shown in the main panel, the distribution of levels seems to obey a scaling relation since the variable k/L^{θ}, with $\theta = 0.16 \pm 0.01$, allows collapsing the data for all system sizes. Some deviation at the tail of the distribution is most likely due to fluctuations resulting from the small frequency of extremely small currents. It should be noted, however, that the same scaling does not hold for bonds of level 0, which seem to collapse with an exponent consistent with 3/4 (not shown). In the inset, we show the distribution without rescaling the axis.

As shown in Figure 5, the distributions of factors generated in both sampling ways closely follow power-law behavior, $P(f) \sim f^{\alpha}$, for over more than seven orders of magnitude. Moreover, the least-squares fits to the datasets of this power-law for all levels k give similar exponents that, within the statistical error bars, are consistent with the value $\alpha = 3/4$. Approaching the limit $f = 1$, all distributions above level $k = 0$ display an exponential cut-off to a vanishing probability. This is to be expected since inside a 3CC, there are no bonds carrying the whole current. For $k > 1$, the distributions appear to be independent on the level. Level 0 contains the so-called red bonds that carry the total current. For bonds at level 0, very few bonds have $f \neq 1$, indicating that only a very small fraction of the blobs consists of parallel bonds, like the ones present in the hierarchical model, as shown in the inset of Figure 2.

Since from level 1 upwards factors are strictly smaller than unity and higher levels correspond to a multiplication of more factors, the current distribution for bonds of higher levels should move toward smaller values. This is confirmed in Figure 6, where we present the current distributions separated by the bond level. The curves move toward weak currents by over one order of magnitude per level, becoming less skewed.

In order to see if there are correlations between factors of different levels, we also compute the distributions after randomly shuffling the data. Precisely, we construct a shuffled current of level k in the following way. Precisely, to construct a shuffled current of level k, for each level from 0 to $k-1$, we chose randomly a factor from the distribution of factors for *internal* bonds (Figure 5B), while for the level k, we chose randomly a factor from the distribution of factors for *actual* bonds (Figure 5A). As can be seen in Figure 6, the distributions for the shuffled data follow closely the distributions obtained for the real current, suggesting that factors at each level are drawn from independent distributions. The current distribution of the whole network should be obtained by summing the expected distributions for each level weighted by the fraction of bonds of each level in the backbone.

2.2 Distribution of hierarchical levels

Figure 7 shows that the distributions of bond levels k display exponential decays for sufficiently large values of k, indicating that the appearance of a 3CC is a Poissonian process. Moreover, the larger the system size L, the less abrupt the decay becomes. As shown in the main panel of Figure 7, for $k \neq 0$, this size dependence is suppressed for levels at the interval $1 \leq k \leq 10$ when both axes are re-scaled by a factor L^{θ}, where the exponent $\theta = 0.16 \pm 0.01$. The fraction of bonds in level 0 systematically decreases with system size like $L^{-5/4} = L^{d_r-2}$, where $d_r = 1/\nu = 3/4$ is the fractal dimension of the red bonds [23].

3 Discussion

By exploiting the self-similarity of a critical percolation backbone, we disclosed a hierarchical structure in its 3-connected components, which ends up allowing an extremely efficient decomposition of the whole system. Level 0 of this hierarchy corresponds to bonds that are just in series or in parallel, and their number increases with system size like the red bonds. The occupancy of higher levels follows a Poisson distribution scaling with the fractal dimension of 3CCs, while the fractions of the current at each level are power-law-distributed with exponents consistent with the value 3/4. Finally, through data reshuffling, we showed that the distributions of factors for internal and actual bonds are uncorrelated. In this way, the complex finite-size behavior of the current distribution can be recovered by multiplying factors randomly drawn from their power law distributions, according to the Poisson distribution of levels.

Another important outcome of our work is the development of a very efficient solver for the local currents in the critical backbone. We also implemented our algorithm on triangular and hexagonal lattices obtaining the same scaling relations and exponents observed for the square lattice. The generalization of our algorithm to higher dimensions is straightforward, and we are presently working on three-dimensional lattices.

Our new way of evaluating current distributions on fractal graphs and the huge gain in precision that we could achieve with this method will allow not only to gain insights on the multifractality of percolation clusters, as shown in the present work, but also to analyze with higher precision than before, for instance, traffic on sparse networks, fluid flow in capillary systems, or the effect of weak bonds in incipient gels.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material; further inquiries can be directed to the corresponding author.

Author contributions

WS: writing–original manuscript and writing–review and editing. JA: writing–original manuscript and writing–review and editing. HH: writing–original manuscript and writing–review and editing. AM: writing–original manuscript and writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors thank the Brazilian agencies CNPq, CAPES, FUNCAP, the National Institute of Science and Technology for Complex Systems of Brazil (INCT-SC), Petrobras ("Física do Petróleo em Meios Porosos," Project Number: F0185), and the PRONEX-FUNCAP/CNPq Award PR2-0101-00050.01.00/15 for financial support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Broadbent SR, Hammersley JM. Percolation processes: I. Crystals and mazes. *Proc Cambridge Phil Soc* (1957) 53:629–41. doi:10.1017/s0305004100032680

2. Stauffer D, Aharony A. *Introduction to percolation theory*. United Kingdom: Taylor & Francis (1992).

3. Kirkpatrick S. Classical transport in disordered media: scaling and effective-medium theories. *Phys Rev Lett* (1971) 27:1722–5. doi:10.1103/physrevlett.27.1722

4. Sahimi M. *Flow and transport in porous media and fractured rock*. United Kingdom: John Wiley and Sons (2011).

5. Hunt AG, Ewing R, Ghanbarian B. Percolation theory for flow in porous media. *Lecture Notes Phys* (2014) 771. doi:10.1007/978-3-319-03771-4

6. Andrade JS, Jr, Street DA, Shinohara T, Shibusa Y, Arai Y. Percolation disorder in viscous and nonviscous flow through porous media. *Phys Rev E* (1995) 51:5725–31. doi:10.1103/physreve.51.5725

7. Andrade JS, Jr, Almeida MP, Mendes Filho J, Havlin S, Suki B, Stanley HE. Fluid flow through porous media: the role of stagnant zones. *Phys Rev Lett* (1997) 79:3901–4. doi:10.1103/physrevlett.79.3901

8. King PR, Andrade JS, Jr, Buldyrev SN, Dokholyan N, Lee Y, Havlin S. Distribution of shortest paths in percolation. *Physica A: Stat Mech its Appl* (1999) 266:55–61. doi:10.1016/s0378-4371(98)00574-3

9. King PR, Masihi M. *Percolation theory in reservoir engineering*. Singapore: World Scientific (2018).

10. Tremblay AMS, Fourcade B, Breton P. Multifractals and noise in metal-insulator mixtures. *Physica A* (1989) 157:89–100. doi:10.1016/0378-4371(89)90282-3

11. de Arcangelis L, Redner S, Coniglio A. Anomalous voltage distribution of random resistor networks and a new model for the backbone at the percolation threshold. *Phys Rev B* (1985) 31:4725–7. doi:10.1103/physrevb.31.4725

12. Barthelemy M, Buldyrev SV, Havlin S, Stanley HE. *Fractals* (2003) 11:19–27. doi:10.1142/s0218348x03001689

13. Batrouni G, Hansen A, Roux S. Negative moments of the current spectrum in the random-resistor network. *Phys Rev A* (1988) 38:3820–3. doi:10.1103/physreva.38.3820

14. Aharony A, Blumenfeld R, Harris AB. Distribution of the logarithms of currents in percolating resistor networks. I. Theory. *Phys Rev B* (1993) 47:5756–69. doi:10.1103/physrevb.47.5756

15. Barthelemy M, Buldyrev SV, Havlin S, Stanley HE. Multifractal properties of the random resistor network. *Phys Rev E* (2000) 61:R3283–6. doi:10.1103/physreve.61.r3283

16. Hopcroft JE, Tarjan RE. Dividing a graph into triconnected components. *SIAM J Comput* (1973) 2:135–58. doi:10.1137/0202012

17. Gutwenger C, Mutzel P. *Int. Symp. On graph drawing*. Berlin, Germany: Springer (2000). p. 77.

18. Paul G, Stanley HE. Beyond blobs in percolation cluster structure: the distribution of 3-blocks at the percolation threshold. *Phys Rev E* (2002) 65:056126. doi:10.1103/physreve.65.056126

19. Herrmann HJ, Stanley HE. Building blocks of percolation clusters: volatile fractals. *Phys Rev Lett* (1984) 53:1121–4. doi:10.1103/physrevlett.53.1121

20. de Arcangelis L, Redner S, Coniglio A. Multiscaling approach in random resistor and random superconducting networks. *Phys Rev B* (1986) 34:4656–73. doi:10.1103/physrevb.34.4656

21. Rintoul MD, Nakanishi H. A precise determination of the backbone fractal dimension on two-dimensional percolation clusters. *J Phys A* (1992) 25:945–8. doi:10.1088/0305-4470/25/15/008

22. Hopcroft J, Tarjan R. Algorithm 447: efficient algorithms for graph manipulation. *Commun ACM* (1973) 16:372–8. doi:10.1145/362248.362272

23. Coniglio A. Cluster structure near the percolation threshold. *J Phys A* (1982) 15:3829–44. doi:10.1088/0305-4470/15/12/032

ANEXO B – ARTIGO: EYE-TRACKING AS A PROXY FOR COHERENCE AND COMPLEXITY OF TEXTS

PLOS ONE

RESEARCH ARTICLE

Eye-tracking as a proxy for coherence and complexity of texts

Débora Torres[1], Wagner R. Sena[1], Humberto A. Carmona[1], André A. Moreira[1], Hernán A. Makse[2], José S. Andrade Jr.[1]*

1 Departamento de Física, Universidade Federal do Ceará, Fortaleza, Ceará, Brazil, **2** Levich Institute and Physics Department, The City College of New York, New York City, New York, United States of America

* soares@fisica.ufc.br

OPEN ACCESS

Citation: Torres D, Sena WR, Carmona HA, Moreira AA, Makse HA, Andrade JS, Jr. (2021) Eye-tracking as a proxy for coherence and complexity of texts. PLoS ONE 16(12): e0260236. https://doi.org/10.1371/journal.pone.0260236

Editor: Haroldo V. Ribeiro, Universidade Estadual de Maringa, BRAZIL

Received: September 1, 2021

Accepted: November 4, 2021

Published: December 13, 2021

Data Availability Statement: The data underlying the results presented in the study are available from the Kaggle database: https://www.kaggle.com/debtorres/eyetracking-reading-experiment.

Funding: Authors JSA, HAC and AAM thank the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) for financial support. JSA also acknowledges support from the National Institute

Abstract

Reading is a complex cognitive process that involves primary oculomotor function and high-level activities like attention focus and language processing. When we read, our eyes move by primary physiological functions while responding to language-processing demands. In fact, the eyes perform discontinuous twofold movements, namely, successive long jumps (saccades) interposed by small steps (fixations) in which the gaze "scans" confined locations. It is only through the fixations that information is effectively captured for brain processing. Since individuals can express similar as well as entirely different opinions about a given text, it is therefore expected that the form, content and style of a text could induce different eye-movement patterns among people. A question that naturally arises is whether these individuals' behaviours are correlated, so that eye-tracking while reading can be used as a proxy for text subjective properties. Here we perform a set of eye-tracking experiments with a group of individuals reading different types of texts, including children stories, random word generated texts and excerpts from literature work. In parallel, an extensive Internet survey was conducted for categorizing these texts in terms of their complexity and coherence, considering a large number of individuals selected according to different ages, gender and levels of education. The computational analysis of the fixation maps obtained from the gaze trajectories of the subjects for a given text reveals that the average "magnetization" of the fixation configurations correlates strongly with their complexity observed in the survey. Moreover, we perform a thermodynamic analysis using the Maximum-Entropy Model and find that coherent texts were closer to their corresponding "critical points" than non-coherent ones, as computed from the Pairwise Maximum-Entropy method, suggesting that different texts may induce distinct cohesive reading activities.

Introduction

Understanding how people capture and assimilate written information involves multiple fields of science, from human anatomy and neurology to linguistics. In particular, the eye movement has always been of interest for the comprehension of reading behavior, since it represents an observable link between the mechanics of vision and its cognitive activity.

of Science and Technology for Complex Systems in Brazil (INCT-SC). DT and WRS received support from CNPq through awarded fellowships for research. HAM acknowledges funding from the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institute of Mental Health (NIMH) through the NIH BRAIN Initiative Grant R01 EB028157 and NSF DMR-1945909. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

In the 19th century, the founders of visual-behavior research examined eye movement in elementary experiments and described how the eyes capture visual information for processing in the brain [1]. In these experiments, it was noticed that the eyes do not register information through smooth continuous movements; instead, they make successive jerks (saccades), between events in which the gaze is briefly maintained on small confined regions (fixations) [2–6]. Moreover, we move our eyes in such a way that, through fixations of precise duration and location, they can efficiently capture pieces of visual data that our brain then puts together in order to create a complete neat image. In this way, the mental demands in processing the image also influences where we sequentially direct the gaze. With the development of eye-tracking devices, scientists were able to observe gaze trajectories and it became clear that the eye movement also depends on the attention focus and examination strategies [6–9].

In the late 20th century, a new area of research in linguistics focused on studying eye movements while reading. In these studies, different strategies were applied in order to analyze how words are fixated depending on specific linguistic factors. It was found that both the number and duration of the fixations on each word are plausible measures to quantify word processing and language comprehension [10–13]. It has been generally accepted that the properties of a given word which are significantly correlated with read processing are its frequency in the language [11, 12, 14–17], length [12, 15, 18] and predictability in context [12, 15, 17, 19, 20]. Under the same framework, eye-tracking experiments have been successfully combined with mathematical models to formally describe how individuals control eye movement while reading, with attention shifting from word to word [21–24].

Complexity and coherence of a text are considered main linguistic attributes to evaluate reading comprehension and learning difficulties [25, 26]. Linguistic researchers have been focusing on measuring texts complexity over the last decades [27]. Mainly, the issue has gained importance due to the need to select appropriate texts for different scholarly levels that would allow students to progressively develop reading and text comprehension skills [25]. For this reason, mathematical expressions for readability and metrics have been developed in order to quantify the complexity of texts and categorize reading material [28]. The variables frequently used are the average length of the words and their frequency in the language, both accounting for semantic difficulty, as well as sentence length, which is closely related to syntactic complexity. The relevant premise behind these empirical expressions is rather obvious, namely, that texts with unusual, long words and extensive sentences are more difficult to process than texts with familiar vocabulary and short sentences. In contrast, the coherence of a text is related to its meaningfulness, a notion associated with semantics rather than grammatical structure [29]. A coherent text makes sense in such a way that the ideas in it are continually connected and the text is consistent as a whole. Its sentences not only have meaning on their own, but, more importantly, they successively build on the meaning of the text.

It is expected that features of a text such as genre and style may be reflected in the eye movement patterns of individuals when reading text passages. Different types of texts may therefore prompt different reading responses in terms of fixation configurations and, consequently, different cognitive reactions. Thus, in order to study this interplay, a phenomenological modeling approach based on eye-movement data (*i.e.*, fixation patterns) should be able to capture the inner cognitive processes underlying reading. In this regard, models from Statistical Physics such us the Maximum-Entropy Model (MEM) developed in information theory can provide a statistical conceptual framework to understand a given natural process in terms of the "interactions" among its many elementary units using statistical data obtained experimentally [30–32]. The principle of maximum entropy states that the probability distribution that best represents the state of a given system is the one that maximizes its entropy, being also in conformity with one or a set of specific constraints. This principle, by itself, contains the essence of the so called

Inverse Ising Problem solution, in which the Hamiltonian (*i.e.*, the interactions) in a given complex system can be inferred from observed statistical correlations among its components. This statistical analysis is frequently referred to as the *Boltzmann-machine*, since it uses the Boltzmann distribution in its core.

The MEM approach has been applied to a wide variety of systems which can be mapped to Ising-like models. In this representation, the interacting elements can be in an active or inactive state, analogously to an Ising type system (*i.e.*, a lattice of dipole moments in which the spins are in either up, +1, or down, −1, states that can be under the action of an external field). In the case of neuronal networks, for example, the interactions between neurons subjected to some stimuli are inferred from the collected data of their firing patterns [33–36]. In a larger scale, the interactivity among regions of the human brain, for example, has been investigated from data of nuclear magnetic resonance [37]. An important application of the MEM is the characterization of protein-protein interaction benefiting from large protein databases [38, 39]. In particular, this strategy has been used to infer genetic interaction networks from known gene expression patterns [40–42]. The collective response exhibited by flocks of birds was also studied by means of the MEM [43, 44] as well as the emergence of collective behavior from the eye movement patterns of a group of people while watching commercial videos [45]. In the last case, pairwise correlations among series of instantaneous eye's velocities were utilized to capture the collective response of the individuals and relate it to video popularity. Finally, the MEM approach has also proved to be effective in other fields outside biology. In [46], for example, the intricate network micro-structure of interactions in the stock market has been captured through pairwise correlations calculated from big data bases of stocks variability.

As shown in Fig 1, here we address the problem of characterizing the complexity and coherence of diverse texts quantitatively by taking a twofold approach. First, we perform eye-tracking experiments with a limited group of people to directly analyze their fixation data while reading different texts, namely, children stories, excerpts from literary works and random word generated texts (see Table 1). This is achieved by expressing the experimental results in terms of a binary model for fixation sequences (analogous to an Ising system) which is duly embedded in the MEM. It enables us to disclose two indexes that can be used as potential proxies for complexity and coherence: the magnetization (a measure of the density of the fixation sequences) and the distance between the "operating temperature" of the system and its critical temperature (a measure of cohesion among the fixation sequences). Second, our experimental approach is then validated through an extensive Internet reading survey, with access to a vast respondent sample, to categorize the same texts according to different complexity and coherence levels, therefore permitting a direct comparison with the obtained eye-tracking indexes.

Materials and methods

Eye-tracking reading experiment

Methodology. The experiments were conducted using a SR Research EyeLink 1000 eye tracker, with the Desktop Mount Participant Setup. It operates at a sampling frequency of 1 kHz using a monocular device and an infrared video-based eye tracker [48]. This equipment is based on the Pupil Center Corneal Reflection system (PCCR) [49, 50], one of the most accurate, non-intrusive eye-tracking techniques. When a stimuli is presented to the subject on a display monitor, near infrared light is shined onto the subjects' eyes and the reflections are recorded with a special camera. Part of the light is reflected in the cornea, appearing as a small, sharp glint (known as the "first Purkinje image"), and another part reaches the retina and reflects back making the pupil appear as a bright, well defined disc ("bright pupil" effect). The

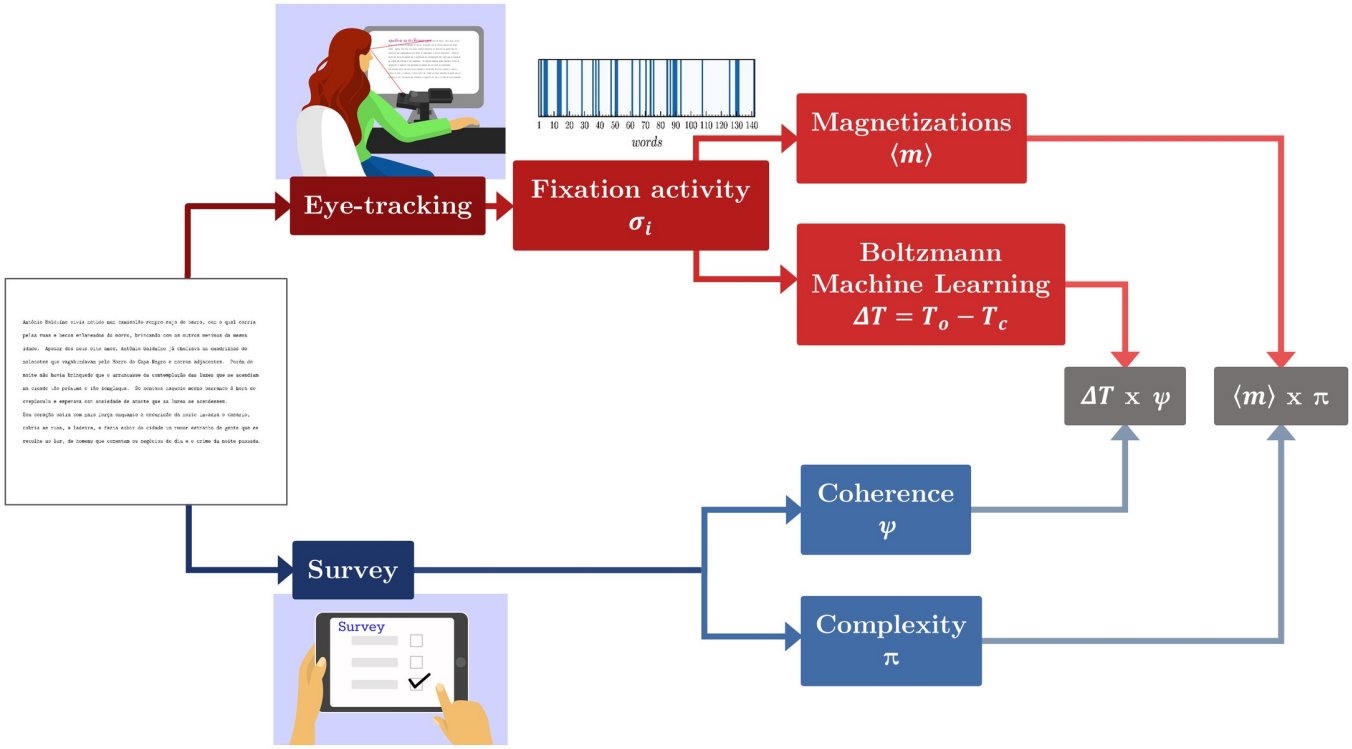


Fig 1. Diagram of the research route. We take a twofold approach to characterize through the cognitive reading activity the complexity and coherence of texts. On one side, we perform an eye-tracking experiment to collect fixation data from a group of people. The fixation activity associated to each subject while reading a given text is computed by binarizing the states of each word, defined positive +1, if the subject fixates it at least twice, or negative −1 if the subject does not fixate or fixates it only once. We then compute the reading "magnetization" of a given text for each subject and average it over all subjects to obtain $\langle m \rangle$. From the pairwise cross-correlations between the fixation sequences of the subjects, we also infer a "Hamiltonian" for each text by means of the Maximum Entropy principle using a Boltzmann machine-learning algorithm. A thermodynamic analysis of the energy fluctuations allows us to determine whether the text is near a "critical point". In parallel, we collected reading-comprehension data from an Internet extensive survey performed with 400 people in an attempt to quantify the complexity $\langle \pi \rangle$ and coherence $\langle \psi \rangle$ of the texts.

https://doi.org/10.1371/journal.pone.0260236.g001

Table 1. Texts information.

Symbol	Title	Author	Year	Country
GAU	O Gaúcho	José de Alencar	1870	Brazil
GSV	Grande Sertão: Veredas	João Guimarães Rosa	1956	Brazil
HCL	História do Cerco de Lisboa	José Saramago	1989	Portugal
JUB	Jubiabá	Jorge Amado	1935	Brazil
MEL	A Mão e a Luva	Machado de Assis	1874	Brazil
QUI	O Quinze	Rachel de Queiroz	1930	Brazil
RT1	Random text 1	-	-	-
RT2	Random text 2	-	-	-
ST1	Story 1: A patinha Esmeralda	-	-	Brazil
ST2	Story 2: A menina do leite	-	-	Brazil

Basic information about the 10 texts used in the eye-tracking experiments. All texts are written in Portuguese. RT1 and RT2 texts are generated with an online random word generator [47]. ST1 and ST2 texts are popular children stories of unknown author and year.

https://doi.org/10.1371/journal.pone.0260236.t001

reflected images are captured by the camera and are then processed by the EyeLink software. The vector between the pupil and corneal reflections is used to calculate the exact gaze location of each sample.

We selected 20 participants among physics and engineering graduate and postgraduate students with ages from 17 to 34, all Brazilian Portuguese native speakers. The reading material consisted of 10 different types of texts (all in Portuguese), including two children stories, two random word generated texts (with standard grammatical structure, but random content) [47] and six excerpts from literature work (see Table 1). All texts were in 12 point size mono-spaced font. For our equipment setup, these characteristics are compatible with the condition that a visual angle of 1˚ spans a length of 3 characters, which gives word position accuracy [50]. Letters were light cyan and the background dark gray, which provides high color contrast and moderate brightness in order to ensure readability while improving the eye-tracking accuracy.

The eye-tracking calibration process consists in collecting raw eye data when the subject fixates at target points, presented one by one at the display monitor. Next, the information is processed and the gaze positions are calculated. The offset between the sampled gaze and the displayed point positions determines the quality of calibration. This protocol was followed before each reading for every participant. A validation test was then performed after calibration to confirm that its accuracy was always within the error range from 0.25˚ to 1˚.

Before starting the experiment, with the purpose of motivating the participants to read consciously, they were warned that we would ask them to answer a simple question after reading each text. To run the experiment, the subject seated in front of a display screen and the head was stabilized by the use of an adjustable head and chin rest. Without imposing any time limit for reading, the texts were sequentially shown on the screen, intercalated by their corresponding questions. During the reading session, the eye tracker collected gaze location data at a sample rate of 1000 Hz, which gives an average temporal error of 0.5 ms (approximately half the duration of the time between samples) [50]. The collected data was de-identified in order to preserve the participants' privacy. The experiment was designed using the SR Research Experiment Builder program (version 1.10.1630) and the collected data was displayed and filtered with the Eye-Link Data Viewer software (version 1.10.1). The reading data is processed by first delimiting each word in the texts with rectangles. For each subject reading a given text, we obtain the spatial coordinates of the fixations from the eye tracker and count how many of them fall into each box, as exemplified in Fig 2. At the end of the experiment, an array can be associated to each text, whose elements n_i^r correspond to the number of fixations that a given word r received from the subject i. In addition, the time duration of each reading was simultaneously registered during every eye-tracking experiment.

Ethics. The eye-tracking experiment procedures were approved by the Research Ethics Committee of the Federal University of Ceará (COMEPE, Universidade Federal do Ceará, Brasil). All subjects gave written informed consent. Also, parental consent was obtained from the parents of the minors included in the study.

Fixation activity model. In order to make possible the analogy with the Ising system, we define the fixation activity $\sigma_i = \{\sigma_i^1, \ldots, \sigma_i^M\}$ for each subject i reading a given text with M words in terms of the state of each word r $\sigma_i^r = \pm 1$ according to the following rule,

$$\sigma_i^r = \begin{cases} +1 & \text{if } n_i^r \geq 2 \\ -1 & \text{if } n_i^r < 2 \ , \end{cases} \tag{1}$$

where n_i^r represents the number of times the subject i fixates on word r during the reading. The value of 2 fixations per word has been adopted here as threshold parameter to define whether a word is active or not in the text due to the fact that, from our eye-tracking

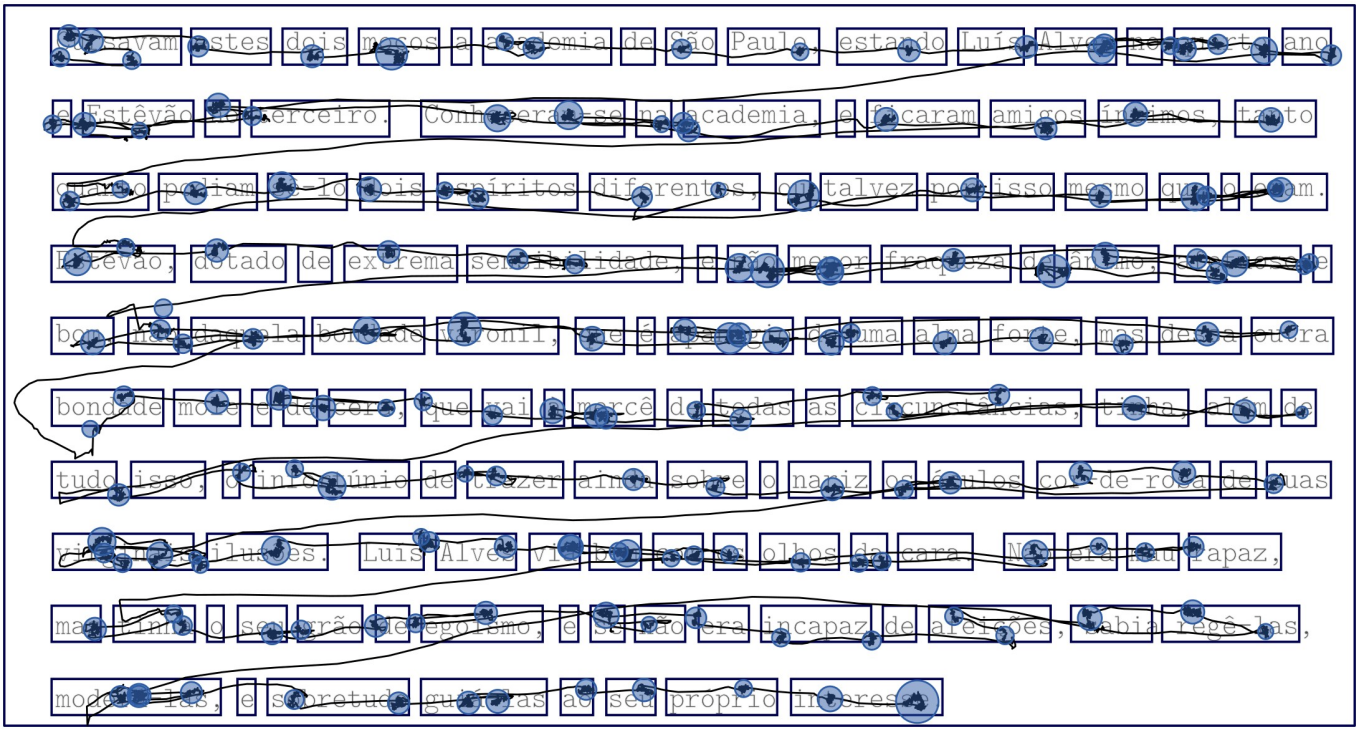


Fig 2. Eye-tracking reading pattern. Plot showing the sequences of gazes and fixations during a typical eye-tracking experiment. In this particular case, the data was collected while the subject was reading the MEL text. The blue circles represent the fixations and their sizes stand for the corresponding duration times. The solid lines between circles indicate the gaze trajectory along the text. For a given text, we measure the number of times n_i^r during the entire reading that a fixation of subject i falls into the rectangle box delimiting a word r.

https://doi.org/10.1371/journal.pone.0260236.g002

experiments, almost every word in any text was fixated at least once during the readings of all subjects. This reading pattern is compatible with observations reported previously [10]. Thus, relevant variations among the fixation activities would be detected considering the words with one fixation and those with two or more.

The raster plots corresponding to the fixation activities obtained from the eye-tracking experiments with all subjects are shown in Fig 3 for all the texts. Clear differences in the reading activity patterns can be observed. In particular, we notice that the density of active states observed for ST1 and ST2 is significantly lower than for RT1 and RT2. The fixation activity density is quantified here in terms of the "magnetization" m_i for each subject $i = 1, \ldots, N$, defined as the average of the fixation states σ_i^r over the M words of the text,

$$m_i = \langle \sigma_i \rangle = \frac{1}{M} \sum_{r=1}^{M} \sigma_i^r \ . \tag{2}$$

In this way, for every text, we can define an overall magnetization as,

$$\langle m \rangle = \frac{1}{N} \sum_{i=1}^{N} m_i \ . \tag{3}$$

A relevant measure that could be readily obtained from the experiments performed here, being certainly indicative of text processing demand, is the average reading time per word of

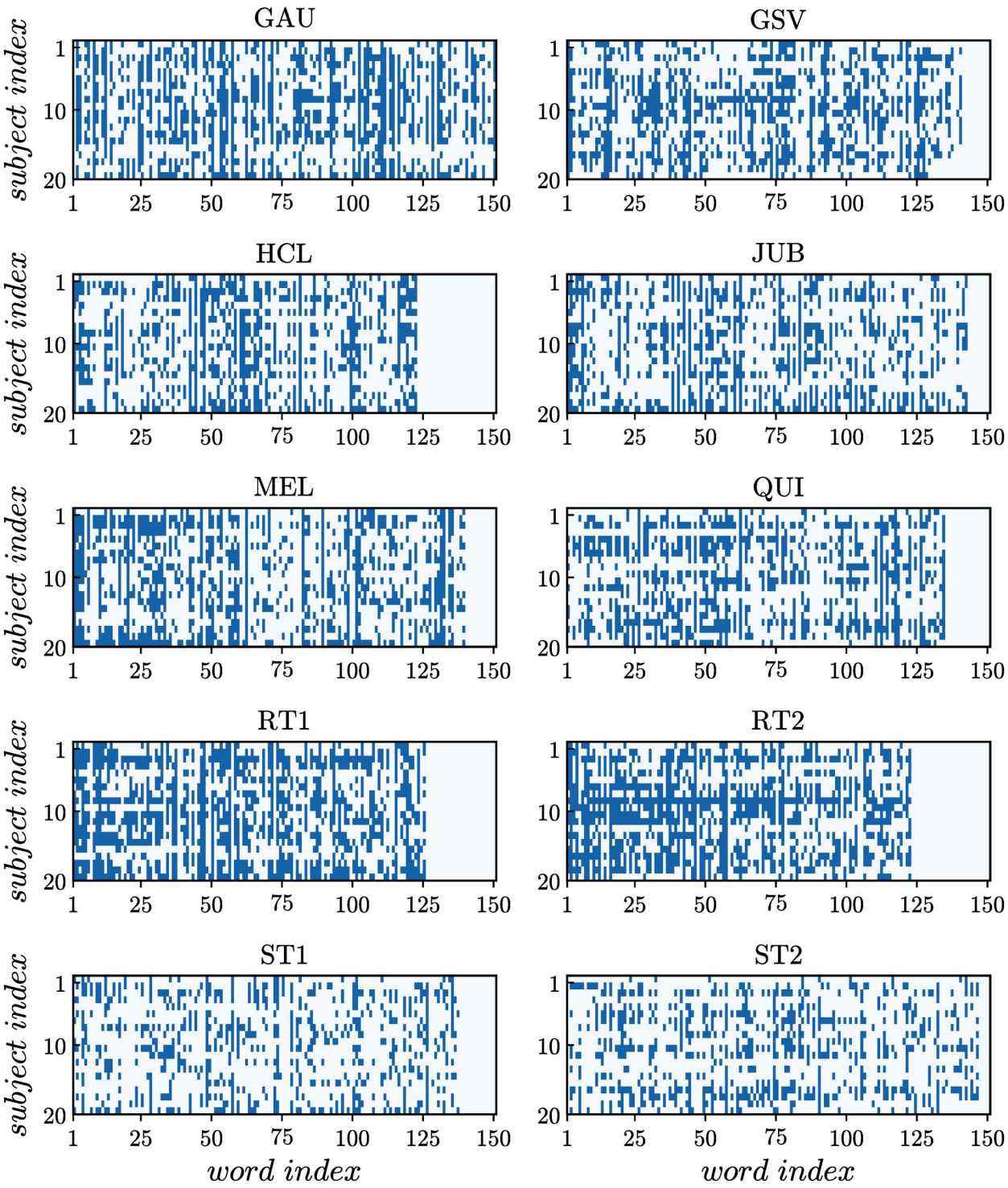


Fig 3. Fixation activities. Raster plots of the fixation activities obtained for all subjects while reading the texts. Accordingly, for each subject *i*, the state σ_i^r of a word is active (+1) if $n_i^r \geq 2$ (blue) or inactive (−1) if $n_i^r < 2$ (white).

https://doi.org/10.1371/journal.pone.0260236.g003

Table 2. Average magnetizations and reading times per word.

Text	$\langle m \rangle$	$\langle t \rangle$ (ms)
GAU	-0.2147	744.82
GSV	-0.2657	611.12
HCL	-0.2918	595.08
JUB	-0.3697	552.34
MEL	-0.2460	641.70
QUI	-0.3022	513.37
RT1	-0.0464	923.98
RT2	-0.0664	793.13
ST1	-0.5190	398.38
ST2	-0.4966	394.21

The magnetization $\langle m \rangle$ and reading time per word $\langle t \rangle$ of each text correspond to the average values of the reading fixation activity and time per word, respectively, also averaged over all subjects.

https://doi.org/10.1371/journal.pone.0260236.t002

the text,

$$\langle t \rangle = \frac{1}{N} \sum_{i=1}^{N} t_i / M \ , \tag{4}$$

where t_i is the reading time of subject i. The values of the average magnetization $\langle m \rangle$ and average reading time per word $\langle t \rangle$ obtained from the experiments for every text are reported in Table 2. Accordingly, the reading of ST1 and ST2 texts resulted in the two lowest values of both measures. Also, the similar values obtained for RT1 and RT2 that are, however, somewhat higher than for most of the other texts demonstrates some degree of correlation between the two measures. Despite the similarities, $\langle m \rangle$ and $\langle t \rangle$ are not perfectly compatible. As a matter of fact, in what follows we will show that $\langle m \rangle$ captures information on the cognitive activity while reading more subtly than $\langle t \rangle$ does.

Maximum Entropy Model

We model the data obtained from our eye-tracking experiments following the Maximum Entropy principle [31] considering a system of binary variables (the fixation activities) with pairwise couplings [32]. Let us denote $\sigma = \{\sigma_1^r, \ldots, \sigma_N^r\}$ the state of the system consisting of N subjects reading word r in a given text. Since every subject can only be in one of two states (+1 or −1), overall we have a set $\{\sigma\}$ of 2^N possible states that the system can occupy, for each word in the text. Next, we calculate the covariance C_{ij} between the fixation activities, for every pair of subjects i and j along the M words of the text,

$$C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \ , \tag{5}$$

where

$$\langle \sigma_i \sigma_j \rangle = \frac{1}{M} \sum_{r=1}^{M} \sigma_i^r \sigma_j^r \ , \tag{6}$$

and $\langle \sigma_i \rangle$ is given by Eq (2). The minimal probability distribution $P(\{\sigma\})$ that represents our system is the one that maximizes the entropy while reproducing our observations, *i.e.*, the average m_i and covariance C_{ij} for all i and j. Subject to these constraints, the form of P is the

Boltzmann's probability distribution [31] (see S1 Appendix),

$$P(\{\sigma\}) \sim \mathrm{e}^{-E/T}, \tag{7}$$

where T is analogous to a temperature and E to a Hamiltonian. This distribution results as the least biased representation for an Ising-type system like ours, with known first and second moments. Specifically, as a first approximation, the energy term has the same form of the Ising model [32],

$$E = -\sum_{i=1}^{N} h_i \sigma_i - \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j. \tag{8}$$

This mathematical correspondence naturally lead us to interpret h_i as the action of a local external stimulus (text) on subject i, analogous to a "random field", and J_{ij} as "coupling coefficients" between subjects i and j. Although the participants never really communicate with each other in our eye-tracking experiments, we can think of the text as a medium through which subjects i and j "interact". This means that, although the subjects read the texts individually, their fixation activities may relate to each other given similarities in their cognitive responses induced by the characteristics of the texts. These pairwise couplings or interactions between the subjects reading activities give rise to the observed correlations among them. Consequently, the correlations may lead to emergent collective effects, which can be of importance for the study of the system.

At this point, we seek compute the local fields h_i and the interactions J_{ij} by directly solving the inverse problem given by Eq (8) (see S1 Appendix). Once we infer the values of h_i and J_{ij} for all subjects that better reproduce the experimentally observed magnetizations m_i and covariances C_{ij}, while maximizing the entropy, the Boltzmann probability distribution of Eq (7) characterizes the statistics of each text. For simplicity, here we arbitrarily set the "operating temperature", namely, the reading temperature, $T_o = 1$. By doing so, from Eq (7) it is possible to compute the rate in which the average energy of a given text changes with T,

$$C_v = \frac{\partial \langle E \rangle_P}{\partial T} \tag{9}$$

This rate of change is analogous to a heat capacity, *i.e.*, a measure of how much energy the system can absorb as the temperature T increases. Moreover, at a "critical temperature" T_c, C_v is maximal, which is interpreted as a phase transition: if $T_c < T_o$, the system is in a "liquid" or random state. On the other hand, $T_c > T_o$ is indicative of a more "ordered" condition.

In Fig 4 we show the variation of C_v as a function of T for all texts. As depicted, regardless of the text, the operating temperature $T_o = 1$ is always above the critical point T_c. The distance to criticality $T_o - T_c$, however, notably depends on the text. By simply considering the fact that larger values have been found for RT1 and RT2, as compared to the other texts (see Table 3), we can obviously anticipate that this distance can be used as an index to distinguish meaningful texts from random ones. Indeed, we will show next that $T_o - T_c$ can be related to language processing in terms of the perceived coherence from reading a text.

Survey for measuring texts complexity and coherence

Quantifying the complexity and coherence of the texts from a survey. In order to validate our eye-tracking results, an Internet survey was conducted by request to the MindMiners services company [51], of São Paulo, Brazil. The agency provides the usage of a digital platform that enables to develop research projects, from questionnaire creation to data collection, through a respondents panel with more than 400 thousand engaged users distributed all over

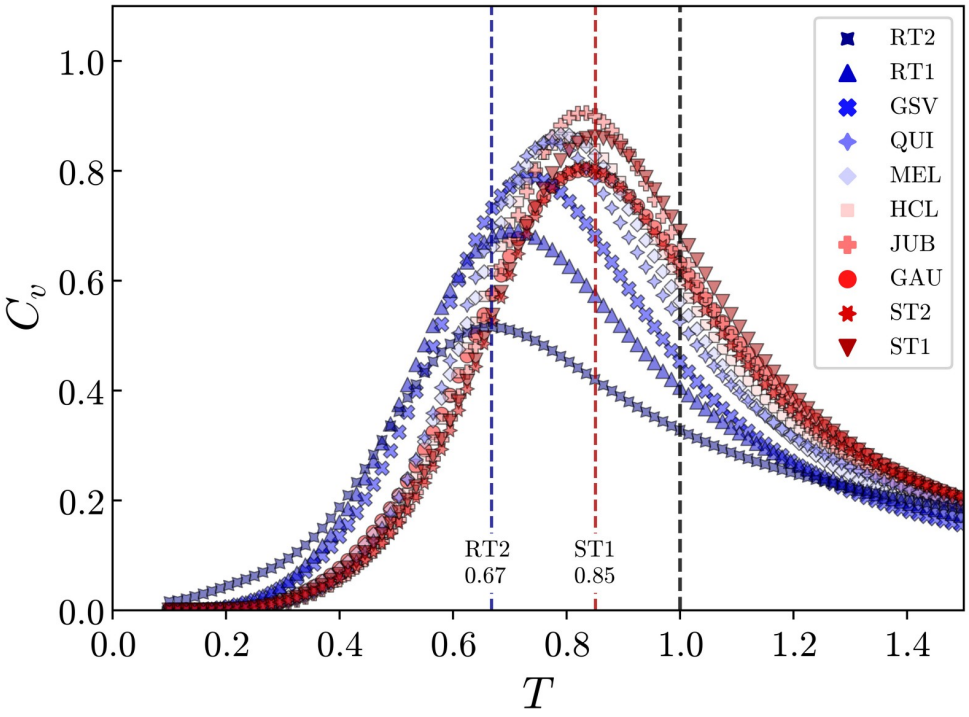


Fig 4. Heat capacity as a function of temperature for the system of fixation activities. Heat capacity curves for all texts, with C_v maximal at the critical temperature T_c. The temperature at which the texts are being read is the operating temperature $T = T_o = 1$. It can be seen that the system is above and near the critical point for all texts, and the RT1 and RT2 texts are clearly the furthest.

https://doi.org/10.1371/journal.pone.0260236.g004

Brazil (MeSeems [52]). The work methodology consists of two main stages, namely, the selection of respondents according to specific requirements of the study, and the production and revision of the questionnaire.

In our study, two groups of 200 people of diverse age, gender and place of residence were selected from the 400 thousand respondents constituting the panel of the survey agency.

Table 3. Distance to criticality.

Text	$T_o - T_c$
GAU	0.169
GSV	0.262
HCL	0.192
JUB	0.170
MEL	0.207
QUI	0.229
RT1	0.296
RT2	0.332
ST1	0.149
ST2	0.167

The table reports the distances to criticality $T_o - T_c$ calculated for different texts from eye-tracking experiments using the MEM. $T_o = 1$ is the reading operating temperature and the critical temperature T_c corresponds to the value of T where the heat capacity C_v for a given text is maximal.

https://doi.org/10.1371/journal.pone.0260236.t003

Table 4. Respondents panel data.

	GroupA		GroupB	
	Respondents	**Percentage**	**Respondents**	**Percentage**
Gender				
Male	86	43.0%	90	45.0%
Female	114	57.0%	110	55.0%
Age				
≤ 17	2	1.0%	3	1.5%
18–24	48	24.0%	46	23.0%
25–30	26	13.0%	35	17.5%
31–40	70	35.0%	61	30.5%
≥ 41	54	27.0%	55	27.5%
Education				
High school	99	49.5%	83	41.5%
University	101	50.5%	117	58.5%
Place of residence				
Central-West	20	10.0%	20	10.0%
Northeast	46	23.0%	46	23.0%
North	10	5.0%	9	4.5%
Southeast	94	47.0%	98	49.0%
South	30	15.0%	27	13.5%

Respondents stratification based on gender, age, education level (high school degree was required) and place of residence in Brazil (indicated by regions).

https://doi.org/10.1371/journal.pone.0260236.t004

Details of the stratification are shown in Table 4. A minimum of high school degree was requested to ensure mature reading-comprehension skills. We did not draw a distinction between socio-economic classes. Each group of respondents was given a questionnaire, related to 5 of the total 10 texts. The texts were divided as follows:

- Group A: O Quinze, Grande Sertão: Veredas, História do Cerco de Lisboa, Story 2, Random text 2
- Group B: Jubiabá, O Gaúcho, A mão e a Luva, Story 1, Random text 1

The questionnaire is divided into three parts and all questions are multiple-choice. In the first part, the respondent was asked to read the texts one by one and answer a simple question related to the text content, with the purpose of merely motivating a conscious reading. In the second part, each text was presented again and the respondent was asked to rate the text complexity level in a scale of 1 to 5, ranging from a "very simple text" to a "very complex text", respectively. Lastly, the texts were again presented and the respondent was asked to assess the text coherence level in a scale from 1 to 5, ranging from a "text not coherent at all" to a "very coherent text", respectively. The texts were shown in a randomized order to each respondent to minimize bias in the responses.

The MindMiners company relies on manual and automatic validation of each user's information through external data sources such as social networks and the Brazilian department of revenue. In addition, they develop algorithms to identify people with atypical behaviors such as non-compliance with questionnaire instructions and abnormal response speed. Identifiers are embedded in all of the respondent's devices, so that the respondents panel is composed exclusively of unique users.

Survey main results. In Fig 5 we show the distribution $P(\pi)$ of fractions of individuals who rated a given value of complexity π for each text. For instance, we can see that almost 40% of the respondents rated the ST1 and ST2 texts as "very simple", while up to 50% of the respondents rated the RT1 and RT2 texts as "very complex". The other texts were rated with varied ranges of intermediate complexity values. The mean values of complexity, $\langle\pi\rangle$ are shown in Table 5. The distributions $P(\psi)$ of fraction of individuals who rated the texts with a value ψ of coherence are shown in Fig 6. The results for ST1 and JUB, for example, show a very similar type of response. Most of the people ($\approx$ 60%) ranked these texts as coherent and very coherent (grades of 4 and 5), approximately 25% ranked with an intermediate level of coherence (grade 3), and only a small fraction (less than 15%) ranked the texts with a low level of coherence (grades 1 and 2). For the random texts RT1 and RT2, on the other hand, the ratings for both are mostly in favor of a "not coherent" opinion ($\approx$ 50%), while approximately 20% thinks that they possess an intermediate value of coherence, and little more than 15% ranked them as coherent or very coherent. Interestingly, the reading of the GSV text led to a distinctive type of response, namely, the larger group of respondents ($\approx$ 40%) ranked the text with the intermediate grade 3. In this case, it suggests that many individuals were undecided about their evaluations on the coherence of the text.

As shown in Table 5, the coherence mean values, $\langle\psi\rangle$, obtained from the survey for all texts suggest they can be sorted into three groups, namely, RT1 and RT2 have low levels of coherence ($\langle\psi\rangle < 2.75$), whereas the reading of ST1, ST2, JUB, HCL, MEL, QUI resulted in high coherence ratings ($\langle\psi\rangle > 3.25$). Finally, the GAU and GSV texts were rated with intermediate coherence levels ($2.75 \leq \langle\psi\rangle \leq 3.25$).

Results

The average magnetization of the fixation activity reflects the level of text complexity

As we already pointed out, the reading time span is certainly not a dispensable measure of text processing. Indeed, a correlation between reading time and the text complexity level seems then straightforward. In Fig 7(A), we plot the average reading times per word against the average values of the complexity. Although $\langle t\rangle$ generally increases with $\langle\pi\rangle$, the relation is hardly monotonic. This becomes more evident in Fig 7(B), where the crescent relative ranks of these variables are plotted against each other, and several discordant pairs in the rank order are observed, out of a total of ten pairs that can be compared with each other. Here we use a non-parametric statistic, namely, the Kendall rank correlation coefficient τ, to measure the rank correlation between $\langle t\rangle$ and $\langle\pi\rangle$ of the texts (see S3 Appendix). In spite of the discordant pairs, we find a value of $\tau = 0.87$ ($p = 0.0001$), which indicates a reasonably high degree of correlation with positive monotonicity trend between the two variables.

The situation is rather different when we plot the average magnetization, $\langle m\rangle$, against $\langle\pi\rangle$ for all texts, as shown in Fig 8(A). The two measures are highly correlated, although in a non-linear fashion, with $\langle m\rangle$ increasing almost monotonically with $\langle\pi\rangle$, except for a capricious local minimum at the average complexity of GSV. Moreover, by plotting the relative ranks of $\langle m\rangle$ and $\langle\pi\rangle$ for a given text against each other (see Fig 8(B)), we notice that, out of ten texts, eight of them occupy identical positions in both lists. As compared to the reading time per word, the higher Kendall rank correlation coefficient found in this case, $\tau = 0.96$ ($p = 5 \times 10^{-6}$), confirms that the measure $\langle m\rangle$ certainly represents a better proxy to rank the complexities $\langle\pi\rangle$ of the texts.

One may argue that more complex texts are expected to require more time for analysis, and therefore more fixations overall. However, when considering the average reading times per

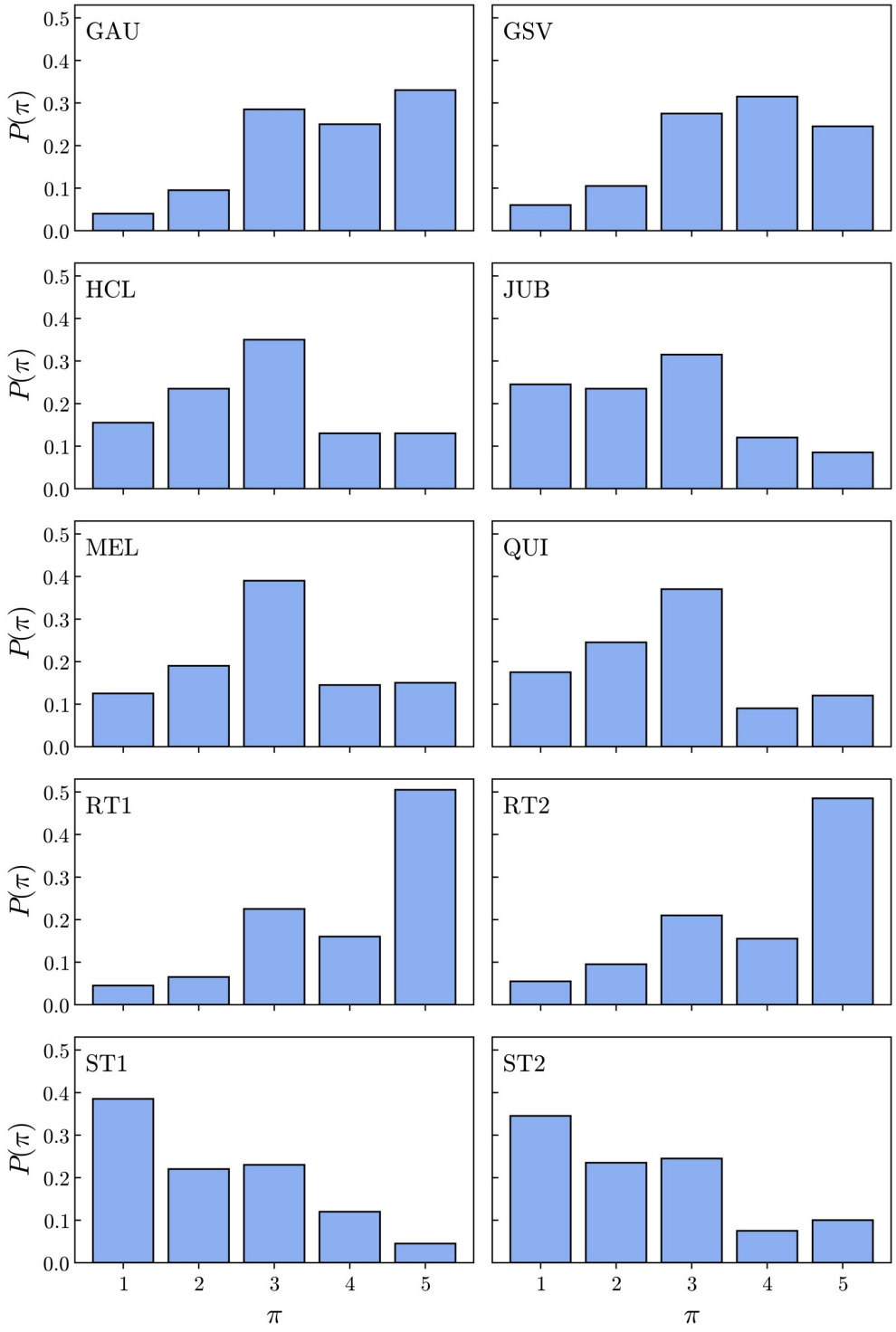


Fig 5. Distributions of complexity ratings. Distributions of complexity ratings among individuals for all texts read in the survey. The values $\pi = 1, 2, 3, 4, 5$ correspond to a scale ranging from a "very simple" text ($\pi = 1$) to a "very complex" text ($\pi = 5$).

https://doi.org/10.1371/journal.pone.0260236.g005

Table 5. Complexity and coherence mean values.

Text	$\langle \pi \rangle$	$\langle \psi \rangle$
GAU	3.74 ± 0.08	3.14 ± 0.08
GSV	3.58 ± 0.08	2.95 ± 0.09
HCL	2.85 ± 0.09	3.70 ± 0.09
JUB	2.57 ± 0.09	3.79 ± 0.08
MEL	3.01 ± 0.09	3.62 ± 0.07
QUI	2.74 ± 0.09	3.86 ± 0.09
RT1	4.02 ± 0.08	2.44 ± 0.09
RT2	3.92 ± 0.09	2.38 ± 0.09
ST1	2.22 ± 0.09	3.80 ± 0.08
ST2	2.35 ± 0.09	4.00 ± 0.09

Complexity $\langle \pi \rangle$ and coherence $\langle \psi \rangle$ mean values obtained for all texts from the survey.

https://doi.org/10.1371/journal.pone.0260236.t005

word, we observe that, for example, it takes around 80 more milliseconds per word on average to read the HCL text than to read the QUI text (a 15% difference), in spite of their similar levels of complexity, according to the survey. The same happens with the GSV and GAU texts, with the subjects spending an additional 133 milliseconds per word on average to read the latter (a difference of 20%), even though their relative difference in average complexities is smaller than 5% (see Table 5). In this regard, a comparison between Figs 7(B) and 8(B), and between their corresponding Kendall coefficients, unambiguously indicate that the average magnetization represents a more reliable indicator of the perceived complexity than the reading time per word.

Text coherence perception evidenced by distance to criticality

The results shown in Fig 9 reveal that large distances to criticality $(T_o - T_c)$ are consistent with the low coherent nature ($\langle \psi \rangle < 2.75$) of both random texts (RT1 and RT2). Moreover, all texts rated with high coherence $\langle \psi \rangle > 3.25$ (ST1, ST2, JUB, HCL, MEL, and QUI) group at the bottom-left corner of the plot due to their correspondingly small values of $(T_o - T_c)$. Coherent texts therefore prompt higher correlated responses in the fixation activity of the readers, suggesting implicit cohesive interactions among them. Interestingly, the two texts ranked with intermediate values of coherence $2.75 < \langle \psi \rangle < 3.25$ in the survey (GAU and GSV), however, induced very different responses in terms of the distance to criticality obtained from the eye-tracked readings. In order to understand this difference, a more detailed analysis is required with respect to the literary styles and linguistic aspects of the books from where these texts have been extracted, as we present in the Discussion.

It is important to stress that one can only rely on the particular features of the cross-correlations from the fixation activity series (see Fig 3) between pairs of readers for a given text to justify the clear numerical differences found among the values of $(T_o - T_c)$. In order to test for this hypothesis, we perform additional calculations with the fixation activities of the subjects for a given text, preserving the mean magnetization $\langle \sigma_i \rangle$, but shuffling the values of σ_i among randomly chosen pairs of words in the text. In this way, strong correlations, if present between the fixation activities, should disappear. Once we have shuffled the data, we follow the same sequence of calculations as before, namely, we find the pairwise correlations C_{ij}, compute the fields h_i and couplings J_{ij}, and determine the heat capacity C_v at different temperatures T. The effect of suppressing strong correlations is to substantially reduce the interactions J_{ij}, therefore

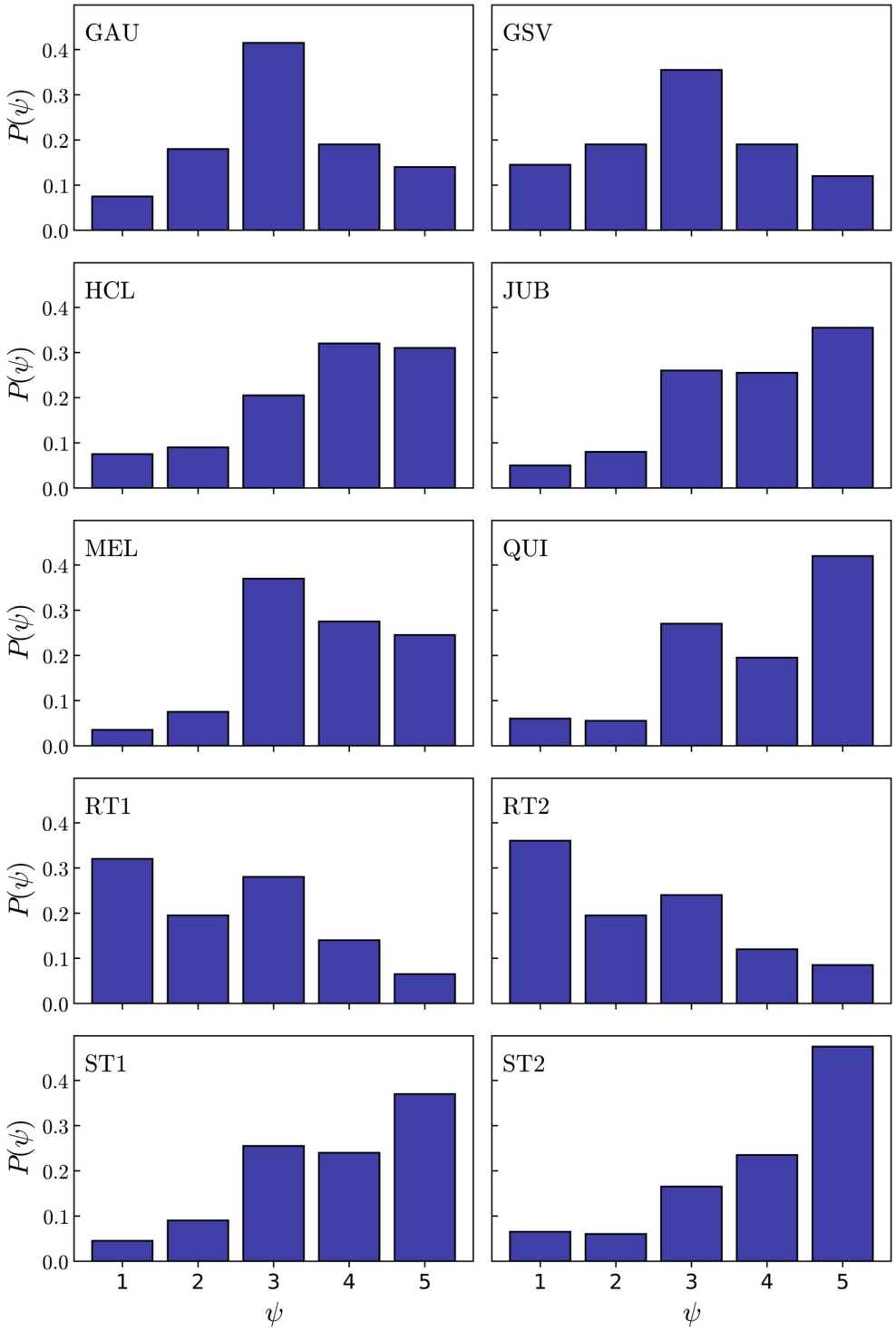


Fig 6. Distributions of coherence ratings. Distributions of coherence ratings among individuals for all texts read in the survey. The values $\psi = 1, 2, 3, 4, 5$ correspond to a scale ranging from a "not coherent" text ($\psi = 1$) to a "very coherent" text ($\psi = 5$).

https://doi.org/10.1371/journal.pone.0260236.g006

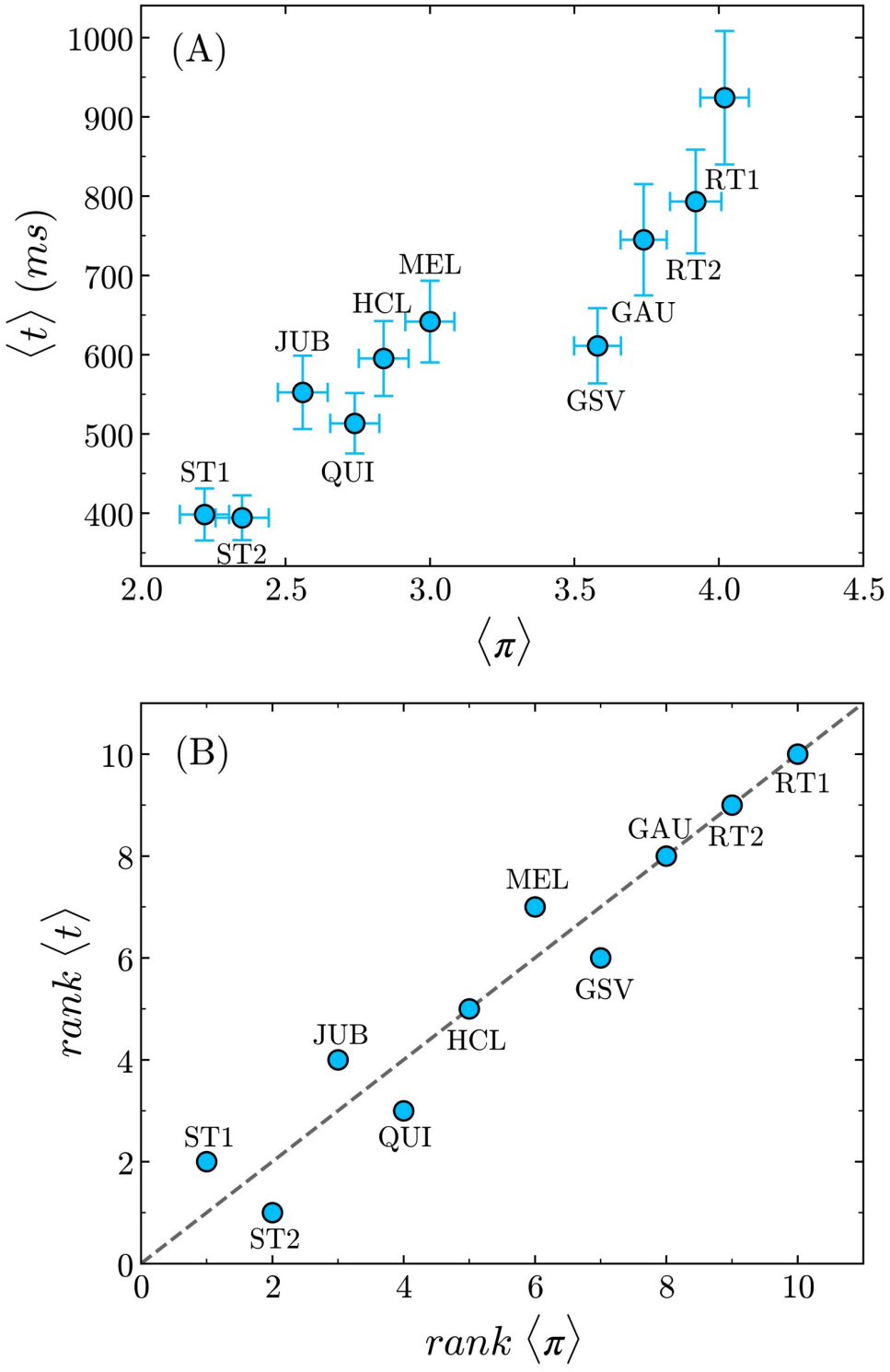


Fig 7. Reading times against text complexity. (A) The average reading time per word $\langle t \rangle$ generally increases with $\langle \pi \rangle$, although the relation is not monotonic. **(B)** The rank of $\langle t \rangle$ plotted against the rank of $\langle \pi \rangle$ shows that several discordant pairs are observed between the two variables. The dashed line corresponds to the function $y = x$. The Kendall rank correlation coefficient is $\tau = 0.87$ ($p = 0.0001$).

https://doi.org/10.1371/journal.pone.0260236.g007

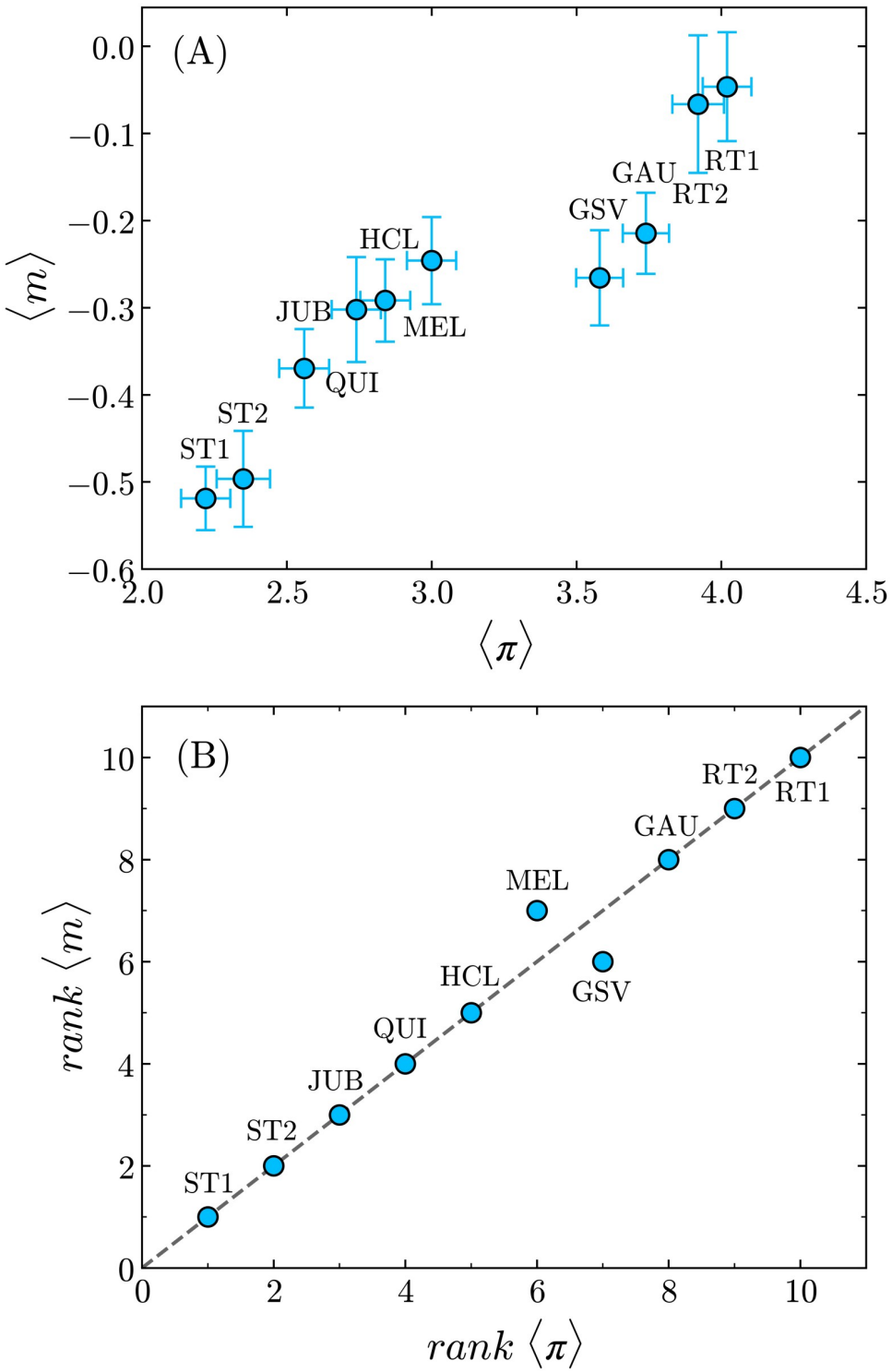


Fig 8. Average magnetization against text complexity. (A) The average magnetization, $\langle m \rangle$, of the fixation activities increases almost monotonically with $\langle \pi \rangle$, except for a local minimum at the complexity of GSV. **(B)** By ranking both measures in crescent order and plotting the ranks of $\langle m \rangle$ against the ranks of $\langle \pi \rangle$ for all texts, we can see that eight out of ten texts occupy exactly the same positions in the two lists. The dashed line corresponds to the function $y = x$. The Kendall rank correlation coefficient $\tau = 0.96$ ($p = 5 \times 10^{-6}$) indicates the very high trend of monotonicity between the two variables.

https://doi.org/10.1371/journal.pone.0260236.g008

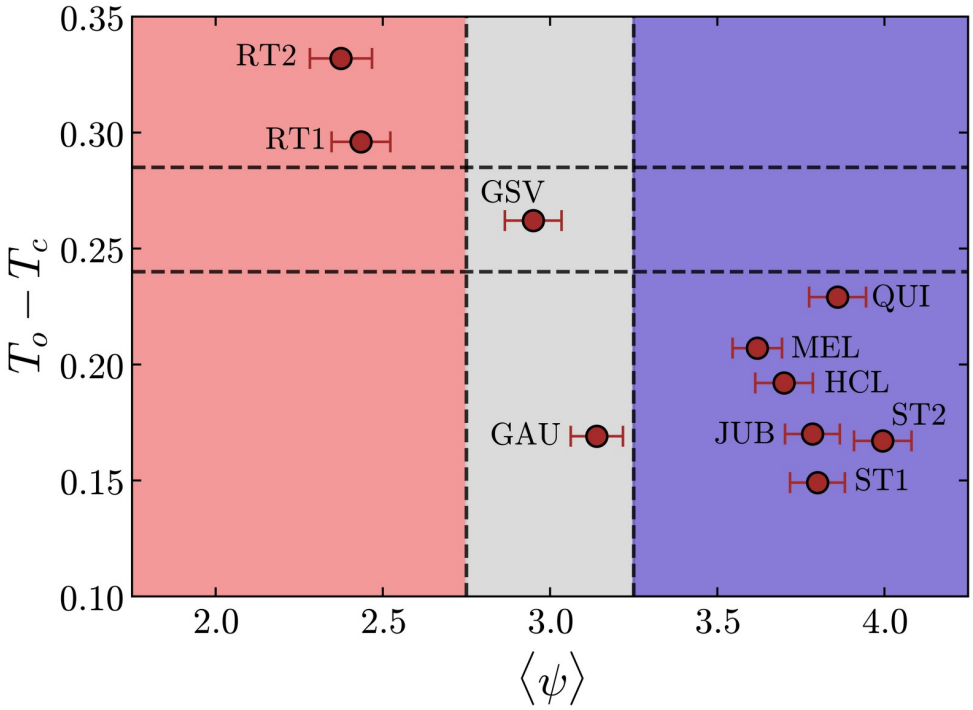


Fig 9. Distance to criticality and text coherence. Relation between the distance to criticality $T_o - T_c$ and the average coherence $\langle\psi\rangle$ of the texts. Texts rated with low coherence $\langle\psi\rangle < 2.75$ are associated with large values of $T_o - T_c$ (RT1 and RT2), while texts considered to be coherent $\langle\psi\rangle > 3.25$ are close to criticality (ST1, ST2, JUB, HCL, MEL, QUI), suggesting an implicit cohesive reading response among individuals. The two texts rated with intermediate values of $\langle\psi\rangle$ (GAU and GSV), however, induced rather distinct responses in terms of $T_o - T_c$.

https://doi.org/10.1371/journal.pone.0260236.g009

decreasing the value of T_c for all texts and increasing their distance ($T_o - T_c$) (see S2 Appendix).

Discussion

Linguists have long studied the notions of text complexity and coherence. More recently, attempts to measure these quantities have been made through mathematical expressions, known as readability formulas, that mainly rely in metrics of word and sentence lengths and word frequency. However, it is arguable whether these formulas are sufficient or not to determine the complexity of a text, mostly for two sound reasons. First, there are uncomplicated pieces of writing that use many infrequent words (an indicator of high complexity for these formulas), like informational texts, as the one suggested in [25], "Any text on raccoons would use 'raccoon' a lot, as well as 'nocturnal' and 'foraging'". This is a typical example for which most readability expressions would overrate the complexity score of the text. Second, texts encompassing elaborated ideas can nevertheless be written with words from a simple vocabulary and constituted of short sentences. This is the case, for example, of some texts containing abstract narratives, usage of metaphors and obscure allusions, for which those complexity scores would be underrated. A well-known example in the literature is Ernest Hemingway's book "The Old Man and the Sea" that could be easily underrated in complexity by readability formulas, despite its profoundness and story-rich writing. As a consequence, in addition to using these metrics, a qualitative analysis is usually recommended by linguists in order to more adequately categorize the reading material.

Here, instead of applying empirical mathematical expressions, we approached the problem of quantifying the complexity and coherence of texts by using data from an extensive survey with a group of 400 people, and comparing it with properties from reading patterns obtained with eye-tracking experiments. We calculated the magnetizations, the most elemental measure of our system that represents the density of fixation activity that the readers had for each text. We hypothesized and confirmed with the survey that the more complex the text the greater the average magnetization. These results therefore suggest that the fixation acitivity, as we defined here, contains the sufficient cognitive information to characterize the reading patterns in terms of active and inactive states. The adopted threshold for the number of fixations in a given word r ($n_i^r = 2$) succeeded in delimiting the minimum to characterize the words in the text that need further cognitive processing. Furthermore, it somehow accounts for the effect of repetitive fixation due to word size and low frequency, while relativizing the underlying variations among the fixation patterns of individuals related to their particular reading skills and capability to predict the occurrence of words in context. This was verified by testing larger values of threshold. Already for a threshold of 3, the fixation activities become heavily dominated by the words size and/or their frequency, showing significant dissimilarities among subjects. In this framework, the threshold adopted in our work acts as a simple, but surprisingly effective way to somehow attenuate differences among subjects.

The fact that the randomly generated texts (RT1 and RT2) were rated with low levels of coherence, while most of literary texts were considered to be coherent by the readers in the survey could be anticipated. The intermediate ratings of coherence for the literary texts GSV and GAU, however, are certainly worth of a more detailed analysis. The celebrated Brazilian author, Guimarães Rosa, who wrote "Grande Sertão: Veredas", from which the fragment GSV was extracted, is well-known for his distinctive writing style, frequently compared to that of James Joyce, in what concerns to the astonishing linguistic work and experimentation [53].

In the literary work of Rosa, we often find unconventional punctuation and grammar in story-rich writing, while creating neologisms from erudite and popular expressions, regionalisms, archaic words and inventive use of prefixes and suffixes [53, 54]. In fact, many of these linguistic features are found in the GSV excerpt used here, with which the book opens. The first word of the text, "nonada" is already an unusual term that, although existing in the Portuguese language, is old-fashioned and hardly used in literature. Even in context, the expression seems so enigmatic that it has led to different interpretations over the years [55–57]. The second sentence has a non-traditional syntactic structure, typical of regional orality, as we learn from the study of Garcia [58]. We see that this linguistic resource is found in the fifth and fifteenth sentences as well. The author also makes use of incomplete suggestive expressions in three sentences of the text (second, fifth and seventh sentences), a linguistic construction typical of Rosa's writing. In addition to this, there are two neologisms in the text created by the author, namely, "erroso" and "prascovio" (in the eighth and thirteenth sentences, respectively). Without entering into a denser analysis, it is fair to say that the GSV fragment is not part of a conventional literary work, being quite difficult to grasp, especially when removed from the global context of the book's narrative. We therefore conjecture that the reader might feel confused and undecided when processing the text, finding it hard to qualify the narrative as coherent.

The text GAU, on its turn, is a transcription from the novel "O Gaúcho", written by José de Alencar in the year 1870. The fragment was extracted from the end of chapter one, where the setting in which the story takes place is described. The writing is characterized by an overwhelming, philosophical representation of the scenario [59]. The abstract tone in the narrative possibly gives the reader an impression that the text is somehow vague, leading to the uncertainty in qualifying it as coherent. In direct contrast with GSV and GAU, the other excerpts of

literary works employed here correspond to passages of descriptive, straightforward writings (HCL, MEL), or linear, plain storytelling (QUI, JUB, ST1, ST2), which very likely make them easy to interpret and therefore be considered as coherent.

Our results from a very simple statistical model and from the analysis with the Pairwise Maximum-Entropy method revealed that the distance to the critical point was capable to segregate the texts into three main groups. The random generated texts (RT1 and RT2) are the farthest from the critical point, with the operating temperatures T_o significantly higher than T_c, the GSV text follows just behind, and the rest of the texts fall much closer to T_c. As we argued previously, in the physical context of critical phenomena, when $T_o > T_c$, the interactions between the component elements are weak and the system is in a disordered state. Our results then suggest that the fixation activities for texts with low coherence (RT1 and RT2) are random to a certain degree, meaning that the reading response to the text stimuli does not promote strong virtual connections among different individuals. When T_o approaches T_c, the "interactions" among elements increase and local effects can propagate over the entire group. The system then becomes susceptible to global changes, and a collective behavior may emerge. This effect has been observed in the texts that were rated with high levels of coherence (ST1, ST2, JUB, HCL, MEL, and QUI) and also with the GAU text, although it was rated with intermediate average coherence. We reason that a high degree of coherence in a text is likely to induce a cohesive reading response, here manifested in terms of a proximity to its critical point. Although the readers never interact with one another, we can think of them responding with a similar cognitive behavior when the content of the text is consistent. A question that naturally arises is why the relation between the average coherence rating $\langle\psi\rangle$ and $T_o - T_c$ is ambiguous for the GAU and GSV texts, given that the former falls into the cluster of texts with operating temperatures close to T_c, while the latter is far from T_c, and still both of them were rated on average with intermediate levels of coherence. Previously we elaborated on the characteristics of these texts, and referred to the linguistic features that made them stand out from the other literary fragments investigated here. On the case of the GAU excerpt, the reader has to process an intelligible text, and can yet have a dubious interpretation due to the abstract style of the writing. Perhaps rating this type of text with a specific value of coherence is equivocal, but we can fairly state that the content of the text is coherent, *i.e.*, its narrative is consistent. In contrast, the GSV text appears atypical to an average reader because of its highly technical writing and uncommon linguistic elements. In a sense, we can think of the GSV as an intermediate type of text, in between a concrete narrative and a random incongruous one. Taking this and the fact that GAU otherwise induced a low value for $T_o - T_c$, the results shown in Fig 9 evidence that the distance to the critical point is actually segregating the texts according to some coherence measure. Such a measure, which originates from an inner cognitive mechanism, is perhaps less subjected to the influence of extrinsic factors than the response to a questionnaire within the protocol of a digital survey.

Conclusion

In summary, the results presented here show that eye-tracking data can be duly processed and analyzed to produce consistent proxies for complexity and coherence of diverse texts. The same texts, including children stories, random word generated texts and excerpts from literature work, have been used to validate this hypothesis by means of an extensive Internet survey with a large number of readers. Our results were substantiated by (*i*) the nearly monotonic relation between the average magnetization $\langle m\rangle$ of the fixation activities and the average complexity $\langle\pi\rangle$ of the texts and (*ii*) the suitability of the distance $(T_o - T_c)$ to segregating random texts (meaningless in content, but with preserved grammar structures) from coherent ones.

We recall that the curve $T - T_c$ for each text is computed from the "energy" of the system, which we obtain by applying the maximum-entropy learning algorithm to the fixation activities of all eye-tracked readers. This finding is particularly significant for several reasons. For one thing, it is another example of how learning algorithms are efficient in extracting relevant information out of large amounts of experimental data, and specifically it supports the maximum-entropy approach as an elementary yet solid method to study complex systems. At the same time, we get to notice how humans respond cohesively to a coherent, consistent text, which is indicative of the advanced language formation and reading prediction mechanisms that we have developed. Instead, when the written information is nonsensical, the collective cognitive response is dispersed.

Supporting information

S1 Appendix. Derivation of the Maximum-Entropy Model and inverse Ising problem.
(PDF)

S2 Appendix. Randomization of fixation activities.
(PDF)

S3 Appendix. Kendall correlation coefficient.
(PDF)

S4 Appendix. Survey raw data.
(PDF)

S1 Fig. Experimental values of magnetizations against theoretical values. The calculated values $\langle \sigma_i \rangle_{th}$ reproduce the experimental values $\langle \sigma_i \rangle$, as evidenced by the plots falling on the red dashed lines corresponding to $y = x$.
(JPG)

S2 Fig. Experimental values of covariances against theoretical values. The calculated values $\langle \sigma_i \sigma_j \rangle_{th}$ reproduce the experimental values $\langle \sigma_i \sigma_j \rangle$, as evidenced by the plots falling on the red dashed lines corresponding to $y = x$.
(JPG)

S3 Fig. Heat capacity as a function of temperature for the system of fixation activities with shuffled data. Average heat capacity curves for all texts, after shuffling the values of fixation states σ_i among randomly chosen pairs of words in the text. The average values are calculated over 100 shuffling trials and the error bars are smaller than the symbols. This suppresses strong correlations, here evidenced by a a significant increase of the distance to the critical point ($T_o - T_c$).
(JPG)

S4 Fig. Distance to criticality and text coherence with shuffled data. Relation between the average distance to criticality $\langle T_o - T_c \rangle$ and the average coherence $\langle \psi \rangle$ of the texts, after shuffling the values of fixation states σ_i among randomly chosen pairs of words in the text.
(JPG)

S1 Table. Distance to criticality with shuffled data. The table reports the average distances to criticality $\langle T_o - T_c \rangle$ calculated using the MEM by shuffling the data from the fixation maps of the eye-tracking experiments (average calculated over 100 trials). $T_o = 1$ is the reading operating temperature and the critical temperature T_c corresponds to the value of T where the heat capacity C_v for a given text is maximal.
(PDF)

S2 Table. Group A and Group B survey raw data.
(PDF)

Author Contributions

Conceptualization: Débora Torres, Humberto A. Carmona, André A. Moreira, Hernán A. Makse, José S. Andrade, Jr.

Data curation: Débora Torres, Wagner R. Sena.

Formal analysis: Débora Torres, Wagner R. Sena, José S. Andrade, Jr.

Funding acquisition: José S. Andrade, Jr.

Investigation: Débora Torres, José S. Andrade, Jr.

Methodology: Débora Torres, Wagner R. Sena, José S. Andrade, Jr.

Project administration: José S. Andrade, Jr.

Resources: José S. Andrade, Jr.

Supervision: José S. Andrade, Jr.

Visualization: Débora Torres, Humberto A. Carmona, Hernán A. Makse, José S. Andrade, Jr.

Writing – original draft: Débora Torres, Humberto A. Carmona, André A. Moreira, Hernán A. Makse, José S. Andrade, Jr.

Writing – review & editing: Débora Torres, Humberto A. Carmona, André A. Moreira, Hernán A. Makse, José S. Andrade, Jr.

References

1. Wade NJ. Pioneers of Eye Movement Research. i-Perception. 2010; 1(2):33–68. https://doi.org/10.1068/i0389 PMID: 23396982
2. Javal E, Ciuffreda KJ, Bassil N. Essay on the physiology of reading. Ophthalmic and Physiological Optics. 1990; 10(4):381–384. https://doi.org/10.1111/j.1475-1313.1990.tb00885.x PMID: 2263372
3. Lamare M. Des mouvements des yeux pendants la lecture. Bulletins et Mémoires de la Société Française d'Ophthalmologie. 1892; 10:354–364.
4. Brown AC. A Lecture ON THE RELATION BETWEEN THE MOVEMENTS OF THE EYES AND THE MOVEMENTS OF THE HEAD. The Lancet. 1895; 145(3743):1293–1298. https://doi.org/10.1016/S0140-6736(01)94423-X
5. Hering E. Spatial Sense and Movements of the Eye. Oxford, England: American Academy of Optometry; 1942.
6. Yarbus AL. Eye Movements and Vision. Boston, MA: Springer; 1967.
7. Credidio HF, Teixeira EN, Reis SDS, Moreira AA, Andrade JS Jr. Statistical patterns of visual search for hidden objects. Scientific Reports. 2012; 2(1). https://doi.org/10.1038/srep00920 PMID: 23226829
8. Amor TA, Reis SDS, Campos D, Herrmann HJ, Andrade JS Jr. Persistence in eye movement during visual search. Scientific Reports. 2016; 6(1). https://doi.org/10.1038/srep20815 PMID: 26864680
9. Amor TA, Luković M, Herrmann HJ, Andrade JS Jr. Influence of scene structure and content on visual search strategies. Journal of The Royal Society Interface. 2017; 14(132). https://doi.org/10.1098/rsif.2017.0406 PMID: 28747401
10. Clifton C, Ferreira F, Henderson JM, Inhoff AW, Liversedge SP, Reichle ED, et al. Eye movements in reading and information processing: Keith Rayner's 40year legacy. Journal of Memory and Language. 2016; 86:1–19. https://doi.org/10.1016/j.jml.2015.07.004
11. Rayner K, Duffy SA. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Mem Cognit. 1986; 14(3):191–201. https://doi.org/10.3758/BF03197692 PMID: 3736392

12. Kliegl R, Nuthmann A, Engbert R. Tracking the mind during reading: The influence of past, present, and future words on fixation durations. Journal of Experimental Psychology. 2006; 135:12–35. https://doi.org/10.1037/0096-3445.135.1.12 PMID: 16478314

13. Gaskell MG, Staub A, Rayner K. 19. In: Eye movements and on-line comprehension processes. Oxford University Press; 2007. p. 327–342.

14. Rayner K, Raney GE. Eye movement control in reading and visual search: Effects of word frequency. Psychonomic Bulletin & Review. 1996; 3:245–248. https://doi.org/10.3758/BF03212426 PMID: 24213875

15. Kliegl R, Grabner E, Rolfs M, Engbert R. Length, frequency, and predictability effects of words on eye movements in reading. European Journal of Cognitive Psychology. 2004; 16(1-2):262–284. https://doi.org/10.1080/09541440340000213

16. Ashby J, Rayner K, Clifton C. Eye movements of highly skilled and average readers: differential effects of frequency and predictability. The Quarterly journal of experimental psychology A, Human experimental psychology. 2005; 58(6):1065–86. https://doi.org/10.1080/02724980443000476 PMID: 16194948

17. Rayner K, Reichle ED, Stroud MJ, Williams CC, Pollatsek A. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. Psychology and Aging. 2006; 21:448–465. https://doi.org/10.1037/0882-7974.21.3.448 PMID: 16953709

18. Juhasz BJ, Liversedge SP, White SJ, Rayner K. Eye Movements and the Use of Parafoveal Word Length Information in Reading. J Exp Psychol Hum Percept Perform. 2008; 34(6):1560–1579. https://doi.org/10.1037/a0012319 PMID: 19045993

19. Ehrlich SF, Rayner K. Contextual effects on word perception and eye movements during reading. J Verb Learn Verb Be. 1981; 20(6):641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

20. Schad DJ, Nuthmann A, Engbert R. Eye movements during reading of randomly shuffled text. Vision Research. 2010; 50(23):2600–2616. https://doi.org/10.1016/j.visres.2010.08.005 PMID: 20719240

21. Reichle ED, Pollatsek A, Fisher DL, Rayner K. Toward a model of eye movement control in reading. Psychological Review. 1998; 105:125–157. https://doi.org/10.1037/0033-295X.105.1.125 PMID: 9450374

22. Engbert R, Kliegl R. Mathematical models of eye movements in reading: a possible role for autonomous saccades. Biological cybernetics. 2001; 85(2):77–87. https://doi.org/10.1007/PL00008001 PMID: 11508778

23. Engbert R, Longtin A, Kliegl R. A dynamical model of saccade generation in reading based on spatially distributed lexical processing. Vision Research. 2002; 42(5):621–636. https://doi.org/10.1016/S0042-6989(01)00301-7 PMID: 11853779

24. Engbert R, Kliegl R, Longtin A. Complexity of eye movements in reading. International Journal of Bifurcation and Chaos. 2004; 14(2):493–503. https://doi.org/10.1142/S0218127404009491

25. Rothman R. The Complex Matter of Text Complexity. Harvard Education Letter. 2012; 28(5).

26. McNamara DS, Kintsch E, Songer NB, Kintsch W. Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning From Text. Cognition and Instruction. 1996; 14(1):1–43. https://doi.org/10.1207/s1532690xci1401_1

27. Fisher D, Frey N, Lapp D. Text Complexity: Raising Rigor in Reading. International Reading Association; 2012.

28. Nelson J, Perfetti C, Liben D, Liben M. Measures of text difficulty: Testing their predictive value for grade levels and student performance; 2012. https://achievethecore.org/page/1196/measures-of-text-difficulty-testing-their-predictive-value-for-grade-levels-and-student-performance.

29. Reinhart T. Conditions for Text Coherence. Poetics Today. 1980; 1(4):161–180. https://doi.org/10.2307/1771893

30. Shannon CE. A Mathematical Theory of Communication. SIGMOBILE Mob Comput Commun Rev. 2001; 5(1):3–55. https://doi.org/10.1145/584091.584093

31. Jaynes ET. Information Theory and Statistical Mechanics. Phys Rev. 1957; 106(4):620–630. https://doi.org/10.1103/PhysRev.106.620

32. Nguyen HC, Zecchina R, Berg J. Inverse statistical problems: from the inverse Ising problem to data science. Advances in Physics. 2017; 66(3):197–261. https://doi.org/10.1080/00018732.2017.1341604

33. Schneidman E, Berry MJ, Segev R, Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature. 2006; 440:1007–1012. https://doi.org/10.1038/nature04701 PMID: 16625187

34. 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. 2009; 106(33):14058–14062. https://doi.org/10.1073/pnas.0906705106 PMID: 19666487

35. Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, Sher A, et al. The Structure of Multi-Neuron Firing Patterns in Primate Retina. Journal of Neuroscience. 2006; 26(32):8254–8266. https://doi.org/10.1523/JNEUROSCI.1282-06.2006 PMID: 16899720

36. Tang A, Jackson D, Hobbs J, Chen W, Smith JL, Patel H, et al. A Maximum Entropy Model Applied to Spatial and Temporal Correlations from Cortical Networks In Vitro. Journal of Neuroscience. 2008; 28 (2):505–518. https://doi.org/10.1523/JNEUROSCI.3359-07.2008 PMID: 18184793

37. Watanabe T, Hirose S, Wada H, Imai Y, Machida T, Shirouzu I, et al. A pairwise maximum entropy model accurately describes resting-state human brain networks. Nature communications. 2013; 4. https://doi.org/10.1038/ncomms2388 PMID: 23340410

38. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proceedings of the National Academy of Sciences. 2011; 108(49):E1293–E1301. https://doi.org/10.1073/pnas.1111471108 PMID: 22106262

39. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein–protein interaction by message passing. Proceedings of the National Academy of Sciences. 2009; 106 (1):67–72. https://doi.org/10.1073/pnas.0805923106 PMID: 19116270

40. Stein RR, Marks DS, Sander C. Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models. PLoS Comput Biol. 2015; 11(7):1–22. https://doi.org/10.1371/journal.pcbi.1004182 PMID: 26225866

41. Lezon TR, Banavar JR, Cieplak M, Maritan A, Fedoroff NV. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. Proceedings of the National Academy of Sciences. 2006; 103(50):19033–19038. https://doi.org/10.1073/pnas.0609152103 PMID: 17138668

42. Locasale JW, Wolf-Yadlin A. Maximum Entropy Reconstructions of Dynamic Signaling Networks from Quantitative Proteomics Data. PLOS ONE. 2009; 4(8):1–10. https://doi.org/10.1371/journal.pone.0006522 PMID: 19707567

43. Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M, et al. Statistical mechanics for natural flocks of birds. Proceedings of the National Academy of Sciences. 2012; 109(13):4786–4791. https://doi.org/10.1073/pnas.1118633109 PMID: 22427355

44. Bialek W, Cavagna A, Giardina I, Mora T, Pohl O, Silvestri E, et al. Social interactions dominate speed control in poising natural flocks near criticality. Proceedings of the National Academy of Sciences. 2014; 111(20):7212–7217. https://doi.org/10.1073/pnas.1324045111 PMID: 24785504

45. Burleson-Lesser K, Morone F, DeGuzman P, Parra LC, Makse HA. Collective Behaviour in Video Viewing: A Thermodynamic Analysis of Gaze Position. PLoS One. 2017; 12(1):1–19. https://doi.org/10.1371/journal.pone.0168995 PMID: 28045963

46. Bury T. Market structure explained by pairwise interactions. Physica A: Statistical Mechanics and its Applications. 2013; 392(6):1375–1385. https://doi.org/10.1016/j.physa.2012.10.046

47. RANDOM TEXT GENERATOR;. http://randomtextgenerator.com/.

48. SR Research Eye Link—Eye tracker;. https://www.sr-research.com/.

49. Ghaoui C. Encyclopedia of Human Computer Interaction. ITPro collection. Idea Group Reference; 2005.

50. Raney GE, Campbell SJ, Bovee JC. Using Eye Movements to Evaluate the Cognitive Processes Involved in Text Comprehension. Journal of Visualized Experiments: JoVE. 2014;(83):641–655. https://doi.org/10.3791/50780 PMID: 24457916

51. MindMiners—Pesquisa Digital;. https://mindminers.com/.

52. MeSeems—Respondents Panel;. https://meseems.com.br/.

53. Almino J. Guimarães Rosa, do Sertão às fronteiras. Revista Brasileira (Academia Brasileira de Letras). 2018; 96:19–36.

54. Silviano S. In: Genealogia da ferocidade: Ensaio sobre Grande Sertão: Veredas. 1st ed. Companhia Editora de Pernambuco (CEPE); 2017. p. 21–23.

55. Zilly B. "Procuro chocar e estranhar o leitor" Grande Sertão: Veredas—a poética da criação e da tradução. Revista do Programa de Estudos Pós-Graduados em Literatura e Crítica Literária da PUC-SP. 2017; 19:4–31.

56. de Castro NL. Universo e vocabulário do Grande sertão. Coleção Documentos brasileiros. J. Olympio; 1970.

57. de Castro MA. In: O homem provisório no Grande Sertão: um estudo de Grande sertão: Veredas. Biblioteca Tempo universitário. Edições Tempo Brasileiro; 1976. p. 44–44.

58. García MS. Grande Sertão: Veredas, de João Guimarães Rosa. Análise textual da obra e duas traduções ao espanhol; 2015. Available from: https://repositorio.ufsc.br/handle/123456789/160576.

59. de Alencar Araripe Júnior T. In: José de Alencar: perfil literario. Rio de Janeiro: Typ. da Escola de Serafim José Alves; circa 1880. p. 140–146. Available from: https://digital.bbm.usp.br/handle/bbm/5206.