

**UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS**

**Bruno Messias Farias de Resende**

**Conectando grafos, grupos e termodinâmica via  
deformações no Laplaciano para análise de dados em  
redes complexas**

**São Carlos**

**2023**



**Bruno Messias Farias de Resende**

**Conectando grafos, grupos e termodinâmica via  
deformações no Laplaciano para análise de dados em  
redes complexas**

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

Área de concentração: Física Aplicada

Opção: Física Computacional

Orientador: Prof. Luciano da Fontoura Costa

**Versão original**

**São Carlos**

**2023**

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

Resende, Bruno Messias Farias de  
Conectando grafos, grupos e termodinâmica via  
deformações no Laplaciano para análise de dados em redes  
complexas / Bruno Messias Farias de Resende; orientador  
Luciano da Fontoura Costa -- São Carlos, 2023.  
104 p.

Tese (Doutorado - Programa de Pós-Graduação em Física  
Aplicada Computacional) -- Instituto de Física de São  
Carlos, Universidade de São Paulo, 2023.

1. Grafos. 2. Laplaciano. 3. Aprendizado de máquina.  
4. Helmholtz-Hodge. 5. Problemas de sincronização de  
grupo. I. Costa, Luciano da Fontoura, orient. II. Título.

## FOLHA DE APROVAÇÃO

Bruno Messias Farias de Resende

Tese apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências. Área de Concentração: Física Aplicada.

Aprovado (a) em: 28/08/2023

Comissão Julgadora

Dr(a): Luciano da Fontoura Costa

Instituição: (IFSC/USP)

Dr(a): Francisco Aparecido Rodrigues

Instituição: (ICMC/USP)

Dr(a): Diego Furtado Silva

Instituição: (ICMC/USP)

Dr(a): Wallace Correa de Oliveira Casaca

Instituição: (UNESP/S. J. do Rio Preto)

Dr(a): Fernando Lucas Metz

Instituição: (UFRGS/Porto Alegre)



*Dedico este doutorado à memória de duas pessoas inesquecíveis, cujas presenças enriqueceram minha vida e cujas partidas em conjunto me marcaram no processo do doutorado: minha amada avó e meu querido amigo Afonso.*

*Minha avó, sempre teve uma xícara de café quente e uma conversa com memórias de tempos passados. Suas histórias e conselhos foram sempre aconchegantes em momentos de incerteza.*

*Ao meu amigo Afonso, uma alma jovem e criativa, cujas ideias futuristas (algumas distópicas) e sonhos grandes nos desafiavam e nos faziam rir. As muitas tardes na cafeteria Kamzu, desfrutando de cookies e aliviando o estresse do doutorado, foram momentos de alívio e alegria.*

*Cada um à sua maneira, moldou minha perspectiva e minha jornada. A perda precoce de Afonso reforçou a fragilidade do tempo, impulsionando-me a agir com propósito. A partida de minha avó me lembrou da nossa conexão com o passado e do valor incalculável da sabedoria adquirida.*

*Dedico a eles este trabalho, um produto da resiliência diante da adversidade e do compromisso com a busca do conhecimento, tanto quanto um reflexo do amor, do apoio e da inspiração que eles me proporcionaram. Agradeço a vocês pela luz que deixaram em minha vida.*



## AGRADECIMENTOS

À minha mãe, irmãos e primos que, embora não tenha convivido tanto durante o processo de doutorado em virtude das distâncias, sempre carreguei um pouco de cada um dentro de mim.

Não poderia deixar de agradecer à minha namorada, Yara, que esteve ao meu lado durante toda esta empreitada. Ao meu querido amigo/irmão, João Lucas e sua família, sou grato por seu apoio constante, pelas músicas e conversas que tornaram essa jornada mais agradável.

Aos amigos Caio e Afonso, por toda a descontração, hambúrgueres, cafés e sessões de cinema e Jojo, que tornaram os dias mais agradáveis.

A Mara Godino, minha professora de inglês, que também durante as conversações sempre agradáveis sempre me deu conselhos valiosos e foi parte importante desta tese.

Gostaria de expressar minha gratidão ao meu orientador, Prof. Dr. Luciano da Fontoura Costa, pelo seu apoio e orientação durante o meu doutorado. Agradeço também por sua compreensão.

Um agradecimento especial à coorientação informal aos colegas de pesquisa e amigos: Thomas K. D. M. Peron, Eric K. Tokuda, Henrique F. de Arruda e Filipi N. Silva. Que me auxiliaram não só no desenvolvimento científico mas também pessoal e, porque não dizer, político. Especialmente a toda dedicação que Thomas teve do início ao fim em me auxiliar no processo do doutorado. Ao Eric, que se juntou a nós no meu segundo ano de doutorado, mas muito me auxiliou em conversas e a contornar adversidades. Ao Filipi, que especialmente durante meu tempo no Google Summer of Code, mostrou que será em um futuro um excelente orientador.

Ao grupo de pesquisa, agradeço por ter sido uma fonte constante de suporte, conhecimento e colaboração. Agradeço a cada um dos membros: Alexandre Benatti, Éverton F. da Cunha, Paulo E. P. Burke, Paulo C. Ventura, Paulo JP Souza e Roberto. Cada momento compartilhado com vocês deixou um impacto significativo em minha jornada.

A USP e ao IFSC pela estrutura fornecida ao longo desses anos assim como a liberdade de poder escolher grade de formação e tópicos de pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) Código de Financiamento 001.



*“Em verdade, em verdade vos digo que se o grão de trigo que cai na terra não morrer, fica só; mas se morrer, produz muito fruto”*

*Evangelho de São João 12, 24-25 - Em Irmãos Karamazóv.*

*Dm7/4, Bb7M, A7(b13)*



## RESUMO

RESENDE, B. M. F. **Conectando grafos, grupos e termodinâmica via deformações no Laplaciano para análise de dados em redes complexas.** 2023. 104p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023.

Nesta tese, exploramos a interseção entre grafos, grupos e termodinâmica, empregando uma variedade de ferramentas analíticas e computacionais recentes. O estudo se concentra na análise de grafos dirigidos por meio de deformações de grupo no Laplaciano combinatório, bem como na análise de tais deformações e na definição de novas medidas e técnicas para maximizar o aproveitamento dessas ferramentas. Nosso foco particular recai sobre a deformação magnética (grupo unitário deformando o Laplaciano combinatório). Por meio do formalismo de matrizes e estatística circulante, descobrimos relações entre as simetrias das medidas termodinâmicas definidas nesta tese e a presença ou ausência de comunidades. Mostramos também como os problemas de sincronização de grupo podem ser utilizados para investigar transições de fase em sistemas físicos discretos, demonstrando que a exploração desses novos operadores deformados é uma via de mão dupla que auxilia tanto a ciência de dados em redes quanto problemas mais fundamentais em física. As investigações teóricas dessas transformações realizadas nesta tese são relevantes, principalmente devido ao recente interesse na utilização de tais transformações em técnicas de aprendizado em grafos direcionados e transformers. Entretanto, consideramos pertinente explorar ainda mais as possíveis direções de aplicações práticas. Dentre essas aplicações, mostramos como a abordagem da teoria de matrizes aleatórias, adaptada para esses operadores, pode ser usada para conjuntos de dados de redes gênicas, identificando elementos de tais conjuntos que são apenas ruído. Além disso, desenvolvemos um arcabouço que, utilizando este formalismo como uma das bases, permite a análise de dados tabulares complexos. Utilizamos como estudo de caso desse arcabouço os dados do Censo Escolar PeNSE de 2015. Propomos e analisamos o conceito de grafos efetivos derivados dessas deformações de grupo, e mostramos como tais grafos efetivos permitem aplicar várias técnicas e medidas restritas a grafos não direcionados em grafos direcionados tal como o conceito de grupo de renormalização.

**Palavras-chave:** Grafos. Laplaciano. Aprendizado de máquina. Helmholtz-Hodge. Problemas de sincronização de grupo.



## ABSTRACT

RESENDE, B. M. F. **Connecting graphs, groups, and thermodynamics through Laplacian deformations for data analysis in complex networks.** 2023. 104p.  
Thesis (Doctor in Science) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023.

In this work, we explore the intersection between graphs, groups, and thermodynamics, employing a variety of recent analytical and computational tools. The study focuses on the analysis of directed graphs through group deformations in the combinatorial Laplacian, as well as on the analysis of such deformations and the definition of new measures and techniques to maximize the use of these tools. Our particular focus is on magnetic deformation (unitary group of the combinatorial Laplacian). Through the formalism of matrices and circulant statistics, we discovered relationships between the symmetries of the thermodynamic measures defined in this thesis and the presence or absence of communities. We also showed how group synchronization problems can be used to investigate phase transitions in discrete physical systems, demonstrating that the investigation of these new deformed operators is a two-way street that helps both network data science and more fundamental problems in physics. The theoretical investigations of these transformations carried out in this thesis are relevant, especially due to the recent interest in the use of such transformations in learning techniques in directed graphs and transformers. However, we consider it pertinent to explore further potential directions of practical applications. Among these applications, we show how the approach of the theory of random matrices, adapted for these operators, can be used for gene network data sets, identifying elements of such data sets that are just noise. In addition, we developed a framework that, using this formalism as one of the bases, allows the analysis of complex tabular data. We used the PeNSE School Census data from 2015 as a case study for this framework. We propose and analyze the concept of effective graphs derived from these group deformations, and show how such effective graphs allow the application of various techniques and measures restricted to undirected graphs on directed graphs. The thesis concludes by emphasizing the importance of edge semantics when using graphs in data analysis, and proposes new research directions, including the exploration of new group deformations and the creation of concepts of effective graphs derived from such deformations.

**Keywords:** Graphs. Laplacian. Machine learning. Helmholtz-Hodge. Group synchronization problems.



## LISTA DE FIGURAS

- Figura 1 – Funcional de energia para um grafo do tipo anel. Energia para uma função sinal  $f_V = \sin(\frac{2\pi v\omega}{N})$  para dois valores de  $\omega$ . Note que a função que possui uma maior frequência possui um valor maior de energia. . . . . 33
- Figura 2 – Figura representando um caminhante aleatório confuso.. Na figura à esquerda, representamos os dois possíveis trajetos para o caminhante aleatório embriagado. Em cada etapa, o caminhante aleatório pode começar a girar no sentido anti-horário ou horário, e um observador externo registra os ângulos de rotação. Na figura à direita, representamos a diferença entre os ângulos de rotação para cada etapa. . . . . 38
- Figura 3 – Figura representando um GSP com independência de caminho. A figura acima ilustra a condição em que as fases do caminhante aleatório são registradas sem a presença de ruído. Neste cenário, a soma acumulada das diferenças de fase ao longo do trajeto independe do caminho escolhido. Isto é, é possível atribuir uma função potencial nos vértices. . . . . 39
- Figura 4 – Figura representando um GSP com a ruídos nas medidas. A ilustração acima representa um cenário no qual tanto o caminhante quanto o observador consomem álcool ao passar pelo bar no caminho superior. Sob a influência de bebida alcoólica, o observador começa a registrar as orientações de maneira imprecisa, introduzindo ruído nas medidas. . . . . 40
- Figura 5 – Devido ao ruído na medição realizada pelo observador, o registro da caminhada agora depende do trajeto escolhido. Isto é, a soma dos giros (fases) não depende mais apenas dos vértices inicial e final, mas também do caminho percorrido por ambos. Neste cenário, a frustração do GSP é maior que zero. . . . . 40
- Figura 6 – Na figura da esquerda, é mostrado um sinal em um grafo geométrico. Na figura central, o mesmo grafo é apresentado com o sinal contaminado por um ruído ‘salt & pepper’. Na figura da direita, é mostrado o resultado da filtragem utilizando a agregação de discrepâncias entre o sinal, ou seja,  $f(u) = f(u) + \frac{1}{D(u)} \sum_{v \in \text{Nei}(u)} (f(v) - f(u))$ . . . . . 45

Figura 7 – Diagrama sequencial da metodologia proposta para análise exploratória de dados tabulares. O dataset tabular é primeiramente transformado em um grafo direcionado e ponderado. Posteriormente, um processo de esparsificação do grafo é aplicado. Este grafo é então empregado para conduzir (b) uma análise de agrupamento utilizando técnicas de análise espectral e identificação de comunidades. Além disso, possibilita a execução de (c) análises individuais (ranqueamento) e comparativas (mergulho das colunas via Laplaciano magnético). . . . .	55
Figura 8 – Representação da construção do grafo direcionado de interpretabilidade. Inicialmente, as colunas dos dados tabulares são mapeadas para um grafo direcionado completo. Em seguida, o peso $w_{(C1,C2)}$ do arco $(C1, C2)$ é calculado a partir do valor SHAP de $C1$ na tarefa de predição de $C2$ . O mesmo procedimento é repetido para atribuir o peso de cada arco no grafo. . . . .	56
Figura 9 – Estrutura de pétalas do calor específico para redes direcionadas modulares. Calor específico em cores em termos da carga $2\pi q$ (coordenadas polares) e temperatura (coordenada radial). . . . .	65
Figura 10 – Calor específico aproximado via KPM para uma rede com $ V  = 3000$ , $N_f = 3$ , $p_c = 0.25$ and $p_d = 0.5$ veja <sup>1</sup> para mais detalhes. As barras de erros representam o desvio obtido entre o valor exato (obtido numericamente via método iterativo) e o valor aproximado obtido via KPM e integração numérica. . . . .	66
Figura 11 – Logo do eMaTe. eMaTe é um pacote python implementado em tensorflow e cupy cujo objetivo principal é fornecer métodos úteis capazes de estimar densidades espectrais e funções traço de matrizes esparsas sejam elas hermitianas (via transformação magnética) ou não. O Desenvolvimento deste pacote foi essencial para publicação do trabalho disponível em. <sup>1</sup> <a href="https://github.com/stdogpkg/emate">https://github.com/stdogpkg/emate</a> . . . . .	66
Figura 12 – Distâncias 1-Wasserstein Médias para Redes BA e ER Utilizando a Estimativa Espectral via KPM. As curvas em (a) e (b) representam a média das distâncias 1-Wasserstein, respectivamente para os modelos de redes BA (Barabási-Albert) e ER (Erdos-Rényi), considerando os parâmetros adotados na geração dessas redes. Esses resultados foram obtidos em redes com cem mil vértices. A estimativa espectral foi realizada utilizando o Método Polinomial do Kernel (KPM), com 100 termos na expansão do KPM e 20 vetores aleatórios para aplicação do “truque de Hutchinson”. <sup>2</sup> Calor específico aproximado via KPM para uma rede com . . . . .	67

Figura 13 – Distribuição da razão de espaçamento de espectros de matrizes de adjacência magnética extraídas de grafos ER para diversos valores de probabilidade de conexão ( $p$ ) e conectividade ( $q$ ). Quando  $q = 0$ , a transformação magnética mapeia o ensemble de Ginibre para um ensemble GOE. Com  $q \neq 0$  o ensemble é mapeado em um com distribuição  $P(S)$  idêntica ao GUE. No limite em que a rede é totalmente desconectada,  $p = 0$ , obtêm-se a distribuição de Poisson, como esperado. 68

Figura 14 – Em (a) é apresentado a função densidade de probabilidade da variável aleatória  $\cos \mathcal{N}(0, 2)$  a qual está relacionada com o termo de fase gerado pela transformação magnética no operador de adjacência combinatório. Tal distribuição pode ser obtida via formalismo de estatística circulante. Em (b) é apresentado a densidade da variável aleatória  $\mathcal{N}(0, 1) \cos \mathcal{N}(0, 1)$  que representa a componente real dos valores da matriz de adjacência magnética. A expressão analítica para tal distribuição foi obtida via teorema de convolução para variáveis aleatórias. . . . . 69

Figura 15 – Distribuição dos autovalores da matriz de adjacência magnética em uma rede gênica com a aplicação da técnica de desdobramento. Esta figura ilustra a distribuição dos autovalores da matriz de adjacência magnética após a aplicação da técnica de desdobramento em uma rede gênica. As curvas teóricas para as distribuições GOE (Gaussian Orthogonal Ensemble), GUE (Gaussian Unitary Ensemble) e Poisson são apresentadas para comparação. A coincidência do formato da distribuição com a curva GUE é uma forte indicação que as interações na rede serem puramente aleatórias. . . . . 70

Figura 16 – Fluxo do espectro do Laplaciano deformado  $\mathbb{Z}_2$  para um modelo de Ising em uma grade periódica 2D de tamanho  $48 \times 48$  em função da temperatura. A primeira linha representa os espectros dos GSPs usando um grafo completo. A segunda linha representa os espectros usando um modelo de Watts-Strogatz 2D. Cada coluna representa uma etapa de renormalização para blocos de  $3 \times 3$ . As linhas verticais tracejadas representam a temperatura crítica associada ao modelo de Ising 2D no limite de  $N$ ,  $T_c \approx 2.269$ . As configurações de spins foram obtidas usando o algoritmo de Metrópolis-Hastings. . . . . 71

Figura 17 – Visualização hierárquica das comunidades de perguntas do questionário PeNSE 2015 obtida por nosso framework. Os vértices no círculo exterior representam as perguntas do questionário e as arestas direcionadas indicam as relações entre elas. A cor de cada vértice representa a classificação original da pergunta. A estrutura hierárquica sobreposta, obtida por meio da aplicação do Nested Stochastic Block Model,<sup>3,4</sup> revela as comunidades no grafo. Destaca-se uma boa concordância entre as comunidades identificadas e a classificação original das questões, sugerindo a eficácia do modelo proposto. Detalhes adicionais podem ser encontrados no artigo,<sup>5</sup> que expande além do escopo desta tese. . . . . 73

Figura 18 – Em (a), apresentamos um grafo direcionado estruturado com nódulos coloridos para facilitar a inspeção visual, com posições definidas manualmente. Em (b), o resultado do mergulho é exibido, com as fases do primeiro autovetor do Laplaciano Magnético ( $q = 1/3$ ) servindo como coordenadas polares e a coordenada radial determinada pelo ranking do Laplaciano de dilatação ( $g = 3/10$ ). Realizamos uma transformação para posicionar o vértice de maior rank no centro. Notavelmente, os nós de alta hierarquia estão mais próximos do centro do círculo e os braços do digrafo são mergulhados de forma a refletir a simetria do grafo original. Este exemplo simples serve para ilustrar como a aplicação de um mergulho usando as soluções dos problemas de frustração (GSP), considerando tanto a hierarquia quanto os ciclos em um grafo, pode ser um recurso útil na análise de estruturas direcionadas. . . . . 74

Figura 19 – Mergulho no espaço toroidal com informação do ranqueamento (radial) do grafo de interpretabilidade do PeNSE. As features com maior rank (conforme definido em<sup>5</sup>) estão mais próximas ao centro axial. À esquerda, observa-se um grupo bem definido de vértices rosa que representa questões da classe *Atividades Físicas*. . . . . 75

Figura 20 – Esta figura ilustra os mergulhos resultantes da pesquisa PeNSE após a remoção de features via Laplaciano Magnético ( $q = 1/10$ ), aprimorando a análise exploratória. As respostas são codificadas por cores de acordo com duas questões específicas: gênero e a posse de carro por alguém na família. As imagens superiores representam os mergulhos obtidos por meio do UMAP ponderado pelo hub score do grafo de interpretabilidade, utilizando todas as questões. As duas imagens inferiores, por outro lado, retratam os mergulhos após a remoção de um subconjunto de questões (excluindo aquelas marcadas em vermelho no meio da figura . . . . . 76

Figura 21 – Decomposição de Helmholtz-Hodge de um grafo direcionado e análise dos seus respectivos laplacianos magnético e de dilatação para cada componente. Na imagem localizada no canto superior esquerdo, é apresentado um grafo direcionado. Nas sequências ilustradas em b, c, d, e e, na mesma linha superior, podemos observar os resultados da decomposição de HodgeRank do referido grafo direcionado, representados nas componentes curl-free, divergence-free, harmônica e rotacional, respectivamente. Na segunda linha, está demonstrado o calor específico magnético, tanto para o grafo original quanto para cada uma de suas componentes. Finalmente, na última linha, é exibido o score obtido pelo laplaciano de dilatação, novamente tanto para o grafo original quanto para cada decomposição do mesmo. É fácil notar que, em algumas componentes, o resultado do laplaciano magnético é informativo enquanto o do laplaciano de dilatação não é, e vice-versa . . . . . 78

Figura 22 – Visualização de uma matriz de adjacência de um grafo dirigido e suas representações na forma simétrica e efetivas. Na figura acima, em (a), é apresentada uma matriz de adjacência de um grafo direcionado amostrado de um modelo de blocos. Em (b), é apresentada a matriz de adjacência obtida através da simetrização ( $\mathbf{A} + \mathbf{A}^T$ ). Em (c), (d) e (e), são apresentadas diferentes matrizes de adjacência efetivas obtidas pela equação 2.32. Nos títulos de (c), (d) e (e), é apresentado o valor de frustração ( $\eta$ ) e o valor da carga para deformação magnética ( $q$ ). Para todos os grafos efetivos, o parâmetro de dilatação foi definido como  $0.1/(|V| - 1)$ , onde  $|V| = 500$ . . . . . 79

Figura 23 – Visualização de um grafo e sua representação no espaço que combina a solução para o problema de frustração magnética e dilatação. Cores representam betweenness obtido no grafo simetrizado e efetivo. A figura ilustra, em (a) e (b), o mergulho definido pela equação 3.12 do grafo direcionado utilizado para gerar os resultados da figura 22. As cores representam a betweenness para a versão simetrizada e efetiva ( $q = 1/10$ ) do grafo mencionado. Conforme pode ser observado, a medida efetiva de betweenness conseguiu captar de maneira efetiva a hierarquia no grafo direcionado. . . . . 81

Figura 24 – Processo de granulação grosseira definida via espaço toroidal e evolução do calor específico. Em (a), (c) e (e), está representado o procedimento de granulação grosseira definido no espaço toroidal, descrito na seção 2.1.5.1, para uma rede dirigida com três módulos. Esta rede foi gerada com o mesmo modelo utilizado para obter os resultados apresentados na figura 9. Os calores específicos para cada etapa do procedimento de granulação grosseira são apresentados em (f), (d) e (b). Observe que a estrutura de pétalas é preservada pelo procedimento de granulação grosseira. . . . .	83
Figura 25 – Evolução da matriz de adjacência obtida pelo processo de RG para o Grafo Simetrizado. O processo de RG eliminou toda a informação de direção. Três passos de renormalização aplicados na versão simetrizada de um grafo direcionado com três blocos em ciclo, conforme discutido em. <sup>1</sup> Note o processo de homogeneização indicando que a abordagem de RG não se aplica adequadamente ao grafo direcionado usando a versão simétrica. . . . .	85
Figura 26 – Evolução da matriz de adjacência obtida pelo processo de RGEG para o Grafo Efetivo. Quatro passos de RGEG aplicados utilizando $q = 1/10$ e $g = 0$ em um grafo direcionado com três blocos em ciclo, conforme discutido em. <sup>1</sup> O processo de RGEG reforça a estrutura de ciclos e preserva a direção. . . . .	86

## LISTA DE TABELAS

Tabela 1 – Relações entre os conceitos nos métodos HodgeRank, SpringRank e Trophic Laplacian . . . . .	104
--	-----



## LISTA DE ABREVIATURAS E SIGLAS

GSP	Problema de Sincronização de Grupo
PeNSE	Pesquisa Nacional de Saúde do Escolar
RMT	Teoria de Matrizes Aleatórias
KPM	Método do Kernel Polinômial
RG	Grupo de Renormalização
RGEG	Grupo de Renormalização para Grafo Efetivo



## LISTA DE SÍMBOLOS

$U(1)$	Grupo Unitário
$\mathbb{Z}_2$	Grupo de adição mod 2
$\mathbb{R}_+$	Grupo de adição dos reais positivos



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>27</b>
<b>2</b>	<b>DESENVOLVIMENTO</b>	<b>31</b>
<b>2.1</b>	<b>Conceitos Básicos</b>	<b>31</b>
2.1.1	Grupos de Simetria	31
2.1.2	Grafos: operadores e cálculo diferencial	32
2.1.3	Problema da sincronização de grupos em grafos	35
2.1.4	GSP, cálculo em grafos e deformações	36
2.1.5	Deformação magnética ( $U(1)$ )	36
2.1.5.1	Mergulho no espaço toroidal	41
2.1.6	Deformação por dilatação e ranqueamento ( $\mathbb{R}_+$ )	41
2.1.7	Grafos com sinais e sincronia de osciladores ( $\mathbb{Z}_2$ )	43
<b>2.2</b>	<b>Novas grandezas e métodos para grafos e fluxos</b>	<b>44</b>
2.2.1	Além da simetrização	44
2.2.1.1	Generalizando o Grau de um Vértice	44
2.2.1.2	Discrepância nas Arestas e Matrizes de Pesos Efetivas	46
2.2.2	Espectro de um Grafo no Contexto de Deformações de Grupo e Grandezas Associadas	47
2.2.2.1	Análogos termodinâmicos e mecânico estatísticos	47
2.2.2.2	Entropia	48
2.2.2.3	Calor Específico	48
2.2.3	Análise de Gases Matriciais e Teoria de Matrizes Aleatórias para Grafos Direcionados	49
<b>2.3</b>	<b>Algumas aplicações para análise de dados e aprendizado de máquina</b>	<b>52</b>
2.3.1	Dados tabulares e clusterização de colunas	52
2.3.2	Clusterização de vértices e imagens	57
2.3.3	Uma perspectiva sobre transições de fase em sistemas físicos via GSPs	58
2.3.3.1	Granulação grosseira e conexão com GSPs via grafos de medidas	59
<b>3</b>	<b>RESULTADOS</b>	<b>63</b>
<b>3.1</b>	<b>Pétalas e simetrias de rotação</b>	<b>63</b>
<b>3.2</b>	<b>Medidas e comparação de grafos dirigidos em web escala</b>	<b>64</b>
<b>3.3</b>	<b>RMT para operadores deformados</b>	<b>67</b>
<b>3.4</b>	<b>Transições de fase e GSPs</b>	<b>70</b>
<b>3.5</b>	<b>Grafos, dados tabulares e remoção de atributos via Laplaciano Magnético</b>	<b>72</b>

<b>3.6</b>	<b>Além da Simetrização</b> . . . . .	<b>77</b>
3.6.1	Decomposição de HodgeRank e o impacto nas deformações magnéticas e de dilatação. . . . .	77
3.6.2	Aplicando propriedades de grafos não dirigidos em dirigidos. . . . .	79
3.6.3	Grupo de Renormalização Deformado . . . . .	80
<b>4</b>	<b>CONCLUSÃO</b> . . . . .	<b>87</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>89</b>
	<b>APÊNDICES</b>	<b>101</b>
	<b>APÊNDICE A – HODGERANK, SPRINGRANK E LAPLACIANO TRÓFICO</b> . . . . .	<b>103</b>

## 1 INTRODUÇÃO

Redes complexas constituem grafos com propriedades que não podem ser efetivamente resumidas em termos de um pequeno número de medidas ou características. Desta forma, este tipo de estrutura de dados constitui uma maneira intrinsecamente adequada para a representação de estruturas e sistemas particularmente intrincados, como é o caso de diversos sistemas tanto do mundo físico quanto do abstrato.

Embora o estudo sistemático de grafos remonte a desenvolvimentos anteriores aos de Leonard Euler sobre as pontes de Königsberg (1736) e ao ranqueamento de jogadores em campeonatos de xadrez por Landau,<sup>6,7</sup> foi mais recentemente, em meados de 2000, que a área de redes complexas se estabeleceu, principalmente por meio de estudos experimentais e analíticos fundamentados em física estatística, focando na caracterização e modelagem de grafos com propriedades topológicas livres de escala.<sup>8</sup> Desde então, essa área tem se desenvolvido significativamente, tanto em termos de caracterização de estruturas complexas quanto em diversas aplicações nas mais variadas áreas.<sup>9</sup>

Entretanto, uma porção significativa desses estudos concentrou-se em redes complexas não direcionadas e sem pesos associados às respectivas arestas, o que poderia ser esperado devido à maior simplicidade desse tipo de grafos, permitindo uma forma natural de obtenção de resultados que podem, então, ser estendidos para estruturas de interconexão incorporando propriedades adicionais.

Em princípio, um grafo básico, constituído apenas por nós e arestas, pode ser complementado de diversas formas, tanto em relação aos seus nós quanto às suas arestas. Por exemplo, podem-se associar rótulos e estados aos nós. No caso das arestas, as mesmas podem possuir orientação (direcionadas), rótulos, pesos com valores inteiros, reais, complexos, etc. Este segundo tipo de estruturas é, neste trabalho, referenciado como “redes complexas com semântica rica nas arestas”, sendo o principal objeto de estudo.

Redes enriquecidas possuem um potencial particularmente importante e promissor para a modelagem de sistemas complexos e aprendizado de máquina, pois permitem a incorporação de mais informações sobre os sistemas de interesse. Por exemplo, o relacionamento entre membros de uma rede social é frequentemente assimétrico. O termo assimétrico pode significar que, se um membro da rede segue o outro, o contrário não necessariamente é verdadeiro. Tais relações também podem ter um valor "negativo", representando uma discordância em relação a algum tópico ou pessoa.<sup>10,11</sup> Além disso, uma pessoa pode participar de diferentes grupos sociais e interagir com diferentes indivíduos em cada grupo de maneiras diferentes.<sup>12</sup>

No campo da sociometria, a multitude de relações assimétricas com retroalimenta-

ções positivas e negativas é representada por estruturas de dados conhecidas como grafos direcionados com sinal e arestas coloridas.<sup>13,14</sup> As cores representam diferentes tipos de interação social.<sup>15,16</sup> Na biologia, o mesmo tipo de estrutura de dados é usado para representar interações proteína-proteína,<sup>17</sup> onde as arestas negativas e positivas representam efeitos de inibição e ativação entre os elementos biológicos. Portanto, o desenvolvimento de técnicas visando a caracterização de grafos direcionados com ou sem sinal é de grande relevância, por exemplo, para a identificação de grupos criminosos,<sup>15</sup> a quantificação de polarização<sup>18</sup> e a identificação de comportamentos anômalos em redes sociais.<sup>19</sup>

No contexto acima caracterizado, o presente trabalho tem como objetivo uma abordagem conceitual e metodológica focada em redes complexas com arestas de semântica rica, em particular correspondendo à incorporação de pesos e direção. Além do desenvolvimento desses tópicos, também consideramos suas respectivas aplicações, que podem contribuir para a validação do trabalho e ilustrar o potencial do trabalho desenvolvido.

Embora a incorporação de pesos em redes complexas seja relativamente direta, podendo ser obtida pelo uso de matrizes simétricas de peso em vez de matrizes de adjacência, a consideração de arestas direcionadas representa um desafio muito maior, exigindo o uso de metodologias relativamente mais sofisticadas matematicamente e computacionalmente. Estruturas com assimetrias ainda mais acentuadas podem ser obtidas quando se consideram pesos em arestas direcionadas. Isso se deve principalmente ao fato de que não apenas as propriedades topológicas das redes direcionadas (também denominadas *digrafo*), mas principalmente as características de vários tipos de dinâmicas que ocorrem nessas estruturas, tornam-se muito mais complexas do que no caso de redes não-direcionadas, devido à perda de simetria Hermitiana das conexões.<sup>19,20</sup> De fato, a introdução de direcionalidade em uma rede complexa imediatamente implica em possíveis autovalores e autovetores complexos, que exigem abordagens mais abrangentes para sua caracterização e análise.

Na presente pesquisa, inicialmente abordamos como podemos usar o espectro de tais operadores deformados para caracterizar e comparar redes dirigidas<sup>1</sup> e modelos de redes complexas.<sup>21</sup> No momento atual da pesquisa estamos interessados em dois objetivos principais. O primeiro constitui-se em desenvolver novos algoritmos para encontrar “clusters” em redes dirigidas ou dados que podem ser mapeados em tais redes, como imagens. O segundo objetivo consiste em propor novos tipos de deformações, conseqüentemente novos operadores que possam abarcar casos mais genéricos tais como grafos dirigidos com sinais e cores. Para concluir o segundo objetivo estamos motivando nossa busca através de duas propriedades que aparecem em grafos direcionados: hierarquia e circularidade.

O conceito de hierarquia em uma rede pode ser definido formalmente através de um problema de ranqueamento. Esses problemas têm como objetivo associar a cada vértice um valor real, conhecido como score, para que, a partir deste, seja possível estabelecer uma ordem - pelo menos parcial - entre os vértices.<sup>22,23</sup> Esse princípio foi empregado, por

exemplo, na detecção de *trolls* em redes sociais.<sup>24</sup>

Recentemente, com o propósito de abordar o problema da hierarquia em redes, os autores em<sup>25</sup> propuseram uma nova técnica de ranqueamento conhecida como SpringRank. No entanto, mostramos que o SpringRank é apenas o método HodgeRank, mas abordado de uma maneira diferente. O HodgeRank é uma técnica proposta por (Jiang *et al.*, 2011) e é fundamentada no teorema de decomposição de Hodge-Helmoltz que discutiremos ao longo desta tese.

Embora esses conceitos possam parecer desconectados à primeira vista, o primeiro autovetor do operador Laplaciano deformado por um grupo de dilatação pode oferecer uma boa aproximação para a solução de HodgeRank.<sup>27</sup>

A propriedade de circularidade em digrafos está relacionada com a presença de ciclos; a tríade serve como exemplo de um ciclo que tem sido estudado de maneira extensiva para a definição<sup>13,28-30</sup> e a quantificação do equilíbrio em sistemas sociais.<sup>31</sup> Neste trabalho, optamos por associar a circularidade a uma medida de inconsistência em ciclos, que pode ser obtida a partir de um formalismo matemático proposto há décadas para o estudo de partículas quânticas carregadas em um espaço discreto na presença de potenciais vetores.<sup>32</sup> Devido à motivação histórica, tal formalismo é denominado Laplaciano Magnético. Este Laplaciano também pode ser interpretado como um caso específico de um operador deformado por um grupo unitário. Recentemente, descobriu-se que este formalismo, mediante deformação magnética, pode ser usado para aprimorar algoritmos de detecção de comunidades,<sup>33</sup> para caracterizar e comparar grandes redes direcionadas<sup>34</sup> e para definir algoritmos de processamento de sinal em grafos direcionados.<sup>27</sup>



## 2 DESENVOLVIMENTO

### 2.1 Conceitos Básicos

#### 2.1.1 Grupos de Simetria

A noção de simetria possui um papel crucial na física, particularmente na teoria da matéria condensada.<sup>35</sup> Ela proporciona insights fundamentais sobre como as propriedades de um sistema mudam ou permanecem inalteradas diante de diferentes transformações. As simetrias, por sua vez, são cruciais para determinar o comportamento e as propriedades dos sistemas físicos.

No campo do aprendizado de máquina, particularmente no domínio das redes neurais, a ideia de simetria também tem recebido uma atenção crescente. Na década de 1990,<sup>36</sup> o estudo de simetrias em redes neurais foi iniciado, focando em como a estrutura dessas redes poderia ser adaptada para refletir certas simetrias nos dados de entrada. No entanto, nas duas últimas décadas, essa noção evoluiu consideravelmente, resultando na concepção de redes neurais equivariantes.<sup>37-39</sup> Uma rede deste tipo possui a propriedade de que sua saída muda de maneira previsível quando os dados de entrada são transformados de acordo com uma simetria específica. Essa propriedade pode ser usada para melhorar a eficiência e a generalização das redes neurais, bem como para incorporar conhecimento *a priori* sobre as simetrias presentes nos dados.<sup>40</sup>

Um grupo de simetria, em matemática e física, é um conjunto de operações que transformam um objeto ou sistema, mantendo preservadas certas propriedades ou estruturas. Essas operações formam um grupo, uma estrutura algébrica que obedece a certas propriedades, como a operação associativa, a existência de uma identidade e a existência de um inverso para cada elemento.<sup>35</sup>

No contexto da teoria de grafos, um grupo de simetria pode representar as diferentes maneiras pelas quais um grafo pode ser transformado sem alterar sua estrutura global. As simetrias do grafo frequentemente têm um impacto significativo nas propriedades espectrais dos operadores associados, como o Laplaciano. Por exemplo, se o grafo possui uma matriz Laplaciana circulante,<sup>41</sup> que é invariante sob rotações, grandezas associadas ao espectro dessa matriz são invariantes em permutações de blocos. Utilizaremos esta propriedade para demonstrar algumas características do Laplaciano magnético em modelos de blocos estocásticos.

Neste trabalho, os grupos que nos interessam principalmente são  $U(1)$ ,  $\mathbb{R}_+$  e  $\mathbb{Z}_2$ . O grupo  $U(1)$  é o grupo de números complexos de módulo unitário, comumente associado a simetrias de rotação em sistemas físicos como as rotações de partículas carregadas em campos magnéticos. O grupo  $\mathbb{R}_+$  representa os números reais positivos, associado à

simetrias de escala. Um exemplo físico de sua aplicação é o conceito de fractais, onde a estrutura básica do objeto permanece a mesma independentemente da escala, ou na teoria das transições de fase, onde perto do ponto crítico, o sistema não tem uma escala de comprimento característica e as flutuações ocorrem em todas as escalas. Por fim,  $\mathbb{Z}_2$  é o grupo cíclico de ordem 2, o qual se relaciona com simetrias de reflexão, como no modelo de Ising, onde cada estado do spin pode ser invertido sem alterar o comportamento global do sistema.

Ao longo deste trabalho, vamos explorar como esses grupos de simetria e suas representações influenciam as propriedades dos operadores associados aos grafos, e particularmente dos Laplacianos deformados, permitindo-nos entender melhor a estrutura e o comportamento desses sistemas complexos.

### 2.1.2 Grafos: operadores e cálculo diferencial

A análise do comportamento de uma função (como o estudo de mínimos, máximos, pontos de inflexão, etc.) em uma variedade contínua pode ser feita através do uso de uma estrutura diferencial equipada em tais variedades. Essa estrutura diferencial é obtida através da definição de operadores como o gradiente. Adicionalmente, a análise das propriedades desses operadores diferenciais, como sua estrutura espectral, pode revelar características da variedade contínua. Portanto, surge a questão: é possível definir tais operadores e conseqüentemente tais estruturas para grafos? A resposta é sim.

Da mesma forma que no caso contínuo, para definir operadores diferenciais para um grafo, é necessário definir uma maneira de quantificar ângulos nesse espaço, o que é chamado de produto interno. Para  $f_V, g_V \in F_V$ , o produto interno no contexto de grafos pode ser definido como:

$$\langle f_V, g_V \rangle_V = \sum_{v \in V} f_V(v)g_V(v), \quad (2.1)$$

Assim, identificamos  $L_V = (F_V, \langle \cdot, \cdot \rangle_V)$  como sendo o espaço com produto interno para as funções nos vértices.

De maneira análoga, seja  $F_E = \{f_E : E \mapsto \mathbb{R} | f_E(u, v) = -f_E(v, u)\}$  o conjunto de funções assimétricas definidas nas arestas. Podemos definir um espaço com produto interno para as funções nas arestas,  $L_E = (F_E, \langle \cdot, \cdot \rangle_E)$  como

$$\langle f_E, g_E \rangle_E = \sum_{(u,v) \in E} w(u, v) f_E(u, v) g_E(u, v). \quad (2.2)$$

A razão de  $F_E$  ser restrito a funções assimétricas ficará mais explicitada nos próximos parágrafos. Com  $L_E$  e  $L_V$  definidos, podemos buscar análogos dos operadores gradiente e Laplaciano para o caso de grafos.

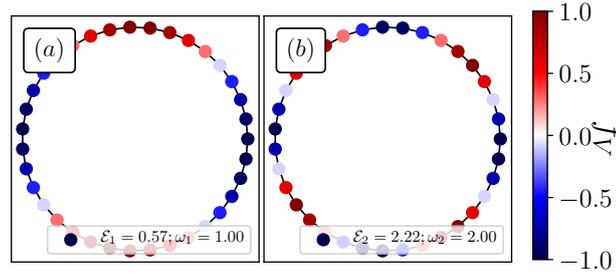


Figura 1 – Funcional de energia para um grafo do tipo anel. Energia para uma função sinal  $f_V = \sin(\frac{2\pi v\omega}{N})$  para dois valores de  $\omega$ . Note que a função que possui uma maior frequência possui um valor maior de energia.

Fonte: Elaborada pelo autor.

No cálculo diferencial, o operador gradiente é aquele que, quando aplicado a uma determinada função, retorna como informação o comportamento de máximo crescimento (e decréscimo) dessa função. De maneira análoga, o operador gradiente para grafos, quando aplicado a uma dada função  $f_V$ , indica as arestas onde tal função permanece constante, cresce ou decresce. Ou seja, o operador gradiente mapeia funções reais nos vértices em fluxos nas arestas<sup>2</sup> (matricialmente, mapeia vetores em matrizes assimétricas). Esse operador é dado por:

$$\begin{aligned} \text{grad} : L_V &\mapsto L_E \\ (\text{grad} f_V)(u, v) &= f_V(u) - f_V(v). \end{aligned} \quad (2.3)$$

Com o uso do operador grad e a definição do produto interno em  $L_E$ , é possível quantificar a suavidade de uma função em vértices específicos por meio da seguinte abordagem:

$$\begin{aligned} \mathcal{E}(f_V) &= \|\text{grad} f_V\|^2 \\ &= \langle \text{grad} f_V, \text{grad} f_V \rangle_E. \end{aligned} \quad (2.4)$$

Essa métrica pode ser interpretada como a energia de um grafo associado a uma determinada função  $f_V$ . Voltaremos a essa definição de energia para discutir o operador Laplaciano.

É conhecido que<sup>42</sup>(pg. 188), em qualquer espaço com produto interno (no nosso caso,  $L_V$ ), cada transformação linear (no nosso caso, grad) possui um operador dual e único denominado codiferencial, representado por  $\text{div} : L_E \mapsto L_V$  (no nosso caso, divergência). Este codiferencial obedece à seguinte relação:

$$\langle \text{grad} f_V, \text{grad} f_V \rangle_E = \langle f_V, \text{div} \circ \text{grad} f_V \rangle_V \quad (2.5)$$

O operador que obedece tal restrição é dado por

$$(\operatorname{div} f_E)(u) = \sum_{(u,v) \in E} w(u,v)(f_E(u,v) - f_E(v,u)). \quad (2.6)$$

É também interessante sua definição através de co-cadeias e operadores de cofronteira; diversas propriedades podem ser verificadas dos operadores acima definidos. Por exemplo, seja  $g_E \in F_E$  uma função que pode ser obtida via um potencial  $f_V$ ,  $\operatorname{grad} f_V = g_E$ , seja um  $f_E \in F_E$ . Então:

$$\langle \operatorname{grad} s, f_E \rangle_E = 0 \quad (2.7)$$

$$\langle s, \operatorname{div} f_E \rangle_V = 0, \quad (2.8)$$

como a equação acima precisa ser verificada para qualquer  $f_V \in F_V$ , a única possibilidade é que  $\operatorname{div} f_E$  seja o vetor nulo, remetendo assim às propriedades usuais do cálculo diferencial.

Equipados com esses operadores, estamos agora em condições de discutir a definição do operador Laplaciano.

É sabido que, em qualquer espaço dotado de um produto interno (no nosso caso,  $L_V$ ), toda transformação linear (no nosso caso,  $\operatorname{grad}$ ) possui um operador dual único, denominado codiferencial, representado por  $\operatorname{div} : L_E \mapsto L_V$  (no nosso caso, divergência). Este codiferencial obedece à seguinte relação:

$$\delta \mathcal{E}_f(g) := \lim_{\epsilon \rightarrow 0} \frac{\mathcal{E}(f + \epsilon g) - \mathcal{E}(f)}{\epsilon}. \quad (2.9)$$

A relação acima pode ser compreendida como a derivada direcional ao longo de  $g$ .

Verifica-se que  $\delta \mathcal{E}_f$  é uma transformação linear para qualquer  $g \in F_V$ . Portanto, pelo teorema de Riesz<sup>42</sup>(pg. 690), existe um único  $h \in F_V$  para o qual a equivalência abaixo é verificada:

$$\delta \mathcal{E}_f(g) = \langle h, g \rangle_V \quad (2.10)$$

Lembrando que a operação do gradiente está relacionada com a derivada direcional, é possível associar  $h = \nabla \mathcal{E}(f)$ , onde  $\nabla$  é um operador linear conhecido como gradiente.

Podemos nos questionar agora sobre como o gradiente do funcional de energia,  $\nabla \mathcal{E}(f)$ , pode ser expresso em termos dos operadores  $\operatorname{grad}$  e  $\operatorname{div}$ . Para isso, utilizamos o fato de que  $\operatorname{grad}$  é linear, então:

$$\|\operatorname{grad}(f + \epsilon g)\|^2 = \|\operatorname{grad} f\|^2 + \epsilon^2 \|\operatorname{grad} g\|^2 + 2\epsilon \langle \operatorname{grad} f, \operatorname{grad} g \rangle_E \quad (2.11)$$

Portanto,

$$\delta\mathcal{E}_f(g) = 2\langle \text{grad}g, \text{grad}f \rangle_E \quad (2.12)$$

Usando novamente a propriedade do codiferencial temos:

$$\delta\mathcal{E}_f(g) = 2\langle g, \text{div} \circ \text{grad}f \rangle_V \quad (2.13)$$

Conclui-se, então, que podemos associar o gradiente do funcional de energia em  $f$  ao resultado da aplicação do operador  $\text{div} \circ \text{grad}$  em  $f$ . Esse operador, que permite analisar o gradiente de energia,  $\nabla\mathcal{E}$ , é o operador Laplaciano, dado por:

$$\begin{aligned} L : \text{div} \circ \text{grad} : L_V &\mapsto L_V \\ (Lf_V)(u) &= \text{deg}_u f_V(u) - \sum_{v; u \sim v} w(u, v) f_V(v). \end{aligned} \quad (2.14)$$

Note que a formulação do operador Laplaciano depende da escolha do produto interno, essa escolha permite definir operadores Laplacianos de forma mais geral.

### 2.1.3 Problema da sincronização de grupos em grafos

Seja  $F$  um espaço vetorial com uma norma dada por  $\|\cdot\|$ , sobre a qual os elementos de um grupo  $\mathcal{G}$  atuam. Dado um grafo de medidas,  $G_\Gamma$ , o Problema de Sincronização de Grupos (Group Synchronization Problem) (GSP)<sup>43</sup> consiste em encontrar e/ou estimar um  $f \in F$  de tal forma que a frustração criada pelos potenciais nas arestas seja minimizada quando esses potenciais são aplicados a  $f$ :

$$\begin{aligned} f^* &= \arg \min_{f \in F} \eta(G_\Gamma, f) \\ &= \arg \min_{f \in F} \frac{\sum_{u, v \in V} w(u, v) \|f(u) - \gamma(u, v)f(v)\|}{\sum_{u \in V} d(u) \|f(u)\|^2}, \end{aligned} \quad (2.15)$$

onde  $\gamma(u, v)$  é o potencial na aresta  $(u, v)$  e  $d(u)$  é o grau do vértice  $u$ .

Embora à primeira vista possa parecer um conceito bastante abstrato, o estudo de problemas de sincronização de grupo encontrou aplicações práticas em uma variedade de contextos. Esses incluem aplicações biológicas,<sup>44</sup> jogos,<sup>44</sup> visão computacional<sup>45</sup> solução de quebra-cabeças,<sup>46</sup> a análise da evolução das espécies e até mesmo o estudo do Censo Escolar Brasileiro.<sup>5</sup> Portanto, este campo de estudo, apesar de sua abstração, é de relevância considerável em diversas áreas de pesquisa e aplicação prática.

Nas próximas seções, será mostrado como o GSP apresenta analogias com sistemas físicos e como isso se relaciona com propriedades interessantes, como ranqueamento de vértices e comunidades. Além disso, exploraremos os resultados que demonstram como essas analogias são úteis na resolução de problemas atuais em caracterização de redes e aprendizado de máquina em grafos. No entanto, antes de prosseguirmos, faremos uma breve conexão com o tema de cálculo em grafos previamente abordado.

### 2.1.4 GSP, cálculo em grafos e deformações

Em (Fanuel e Suykens, 2019), os autores definem novos operadores laplacianos para grafos direcionados, os quais permitem generalizar a função de ranqueamento obtida em um grafo direcionado. Essa nova classe de operadores é denominada pelos autores como *operadores laplacianos deformados*. O termo deformado se deve ao fato de que os produtos internos, laplacianos e demais elementos são modificados por elementos de um grupo.

No formalismo de laplacianos deformados, os produtos internos e as funções nos vértices e arestas precisam ser estendidas para o plano complexo. Dessa maneira, o produto interno para funções nos vértices passa a ser:

$$\langle f, g \rangle_V = \sum f(v)^* g(v) \quad (2.16)$$

e o produto interno para funções definidas sobre as arestas torna-se:

$$\langle f, g \rangle_E = \sum w_s(u, v) f(u, v)^* g(u, v), \quad (2.17)$$

onde  $w_s(u, v) = \frac{w(u, v) + w(v, u)}{2}$ .

Desse modo, é possível definir um operador laplaciano deformado,  $L_T : L_V \mapsto L_V$ , de tal forma que a energia de uma dada  $f \in L_V$  seja:

$$\langle f, L_T f \rangle_V = \sum (u, v) \|T(v, u)f(u) - T(u, v)f(v)\|^2 \quad (2.18)$$

onde  $T(v, u)$  é um operador linear Hermitiano com  $\mathbf{T}^\dagger = \mathbf{T}$ .

As relações de hermiticidade de  $\mathbf{T}$  permitem estabelecer a conexão de que os operadores laplacianos deformados estão relacionados com algum GSP. Uma opção plausível para  $T$  seria a identidade. Contudo, neste caso o operador laplaciano se converte no operador laplaciano combinatório usual.

A estratégia para desenvolver novas famílias de operadores laplacianos e métodos para análise de grafos, ou problemas definidos sobre eles, pode ser realizada através da definição de  $T$  como um operador de deformação. Esse operador deve apresentar propriedades benéficas que facilitam a solução e/ou análise dos problemas em consideração.

Por exemplo, uma deformação particular pode ser útil para ordenar elementos de um grafo, como times de futebol em um campeonato. Outra deformação pode ser aplicada para segmentar imagens ou tratar questões relacionadas ao Censo. Desta forma, a escolha apropriada do operador de deformação  $T$  expande a versatilidade e a capacidade de análise dos operadores laplacianos em diversos contextos de aplicação.

### 2.1.5 Deformação magnética ( $\mathbf{U}(1)$ )

No desenvolvimento desta tese, o foco principal de estudo foi a deformação conhecida como magnética. Recentemente, ocorreu um aumento expressivo na aplicação desta

técnica de deformação para abordar uma gama de problemas práticos e teóricos. Particularmente, ela tem sido aplicada em tarefas como detecção de comunidades,<sup>33</sup> comparação e caracterização de grafos,<sup>21,34,47,48</sup> e várias técnicas de aprendizado de máquina<sup>5,49,50</sup> e matemática.<sup>51,52</sup>

Um grafo direcionado é expresso por uma tupla  $G = (V, E, w)$ , onde  $V$  é o conjunto de vértices,  $E$  é o conjunto de arestas, com cada tupla ordenada  $e = (u, v) \in E$  atribuindo uma aresta direcionada do vértice  $u$  para  $v$ , e  $w : E \rightarrow \mathbb{R}$ . Existe a possibilidade de associar um grafo direcionado a um correspondente não direcionado  $G^{(s)} = (V, E^{(s)}, w^{(s)})$ , onde  $w^{(s)}(u, v) = \frac{w(u,v)+w(v,u)}{2}$ . Entretanto, a direcionalidade de  $G$  se perde em  $G^{(s)}$ .

Para preservar tanto a Hermiticidade quanto a informação sobre a direcionalidade, adota-se uma função  $\gamma$  definida como:  $\gamma : E \rightarrow \mathcal{G}$ , onde  $\mathcal{G}$  é um grupo tal que  $\gamma(u, v)^{-1} = \gamma(v, u)$ . Desta forma, o grafo simétrico equipado com  $\gamma_q$  consegue reter informações sobre as arestas direcionadas e, simultaneamente, manter a matriz de adjacência Hermitiana.

Um operador denominado Laplaciano magnético<sup>53</sup> -  $L_q$ , é então associado a  $(G^{(s)}, \gamma_q)$ . O termo "magnético" é devido à sua capacidade de descrever a fenomenologia de uma partícula quântica sujeita à ação de um campo magnético.<sup>54</sup> Devido a esse contexto físico, o parâmetro  $q$  é chamado de carga.

Procedendo com a construção do Laplaciano magnético, decompomos a matriz de adjacência em uma parte simétrica e assimétrica:  $\mathbf{W} = \frac{\mathbf{W} + \mathbf{W}^T}{2} + \frac{\mathbf{W} - \mathbf{W}^T}{2}$ . A matriz anti-simétrica, é conhecida como fluxo  $\mathbf{A} = \mathbf{W} - \mathbf{W}^T$ , que é usada para construção do potencial magnético em cada aresta dado por:

$$\gamma_q(u, v) = \exp(iA(u, v)q), \quad (2.19)$$

onde  $q \in [0, 2\pi]$ , observando-se que a deformação magnética é devida ao grupo unitário,  $U(1)$ .

O Laplaciano magnético, expresso em termos de matrizes, é dado por:

$$\mathbf{L}_q = \mathbf{D} - \mathbf{\Gamma}_q \odot \mathbf{W}^{(s)}, \quad (2.20)$$

onde  $\mathbf{D}$  é a matriz de graus;  $[\mathbf{\Gamma}_q]u, v = [\mathbf{\Gamma}_q^\dagger]v, u = \gamma_q(u, v)$  e  $[W^{(s)}]u, v = [W^{(s)}]v, u = w^{(s)}(u, v)$ .

Por construção, tanto  $\mathbf{D}$  quanto  $\mathbf{W}^{(s)}$  são simétricas, e  $\mathbf{\Gamma}_q$  é Hermitiana. Consequentemente,  $\mathbf{L}_q$  é Hermitiano.

Para tornar a discussão mais didática, consideremos uma analogia: a jornada de um caminhante aleatório que precisa percorrer um caminho do ponto inicial até um ponto final, enfrentando diversas opções de trajeto pelo caminho. Em cada ponto de parada (que corresponde aos nós do nosso grafo), o caminhante entra em um bar e, sob a influência da ingestão de álcool, tende a girar no sentido horário ou anti-horário Fig.2(a).

Acompanhando essa jornada está um observador externo - neste caso, um amigo sóbrio - que registra meticulosamente a quantidade de graus que o caminhante está girando a cada passo. Ele anota todas essas medidas em um caderno, sempre tendo como referência os pontos de parada anteriores (ou seja, os vértices do grafo) Fig.2(b). O que se obtém é um registro, ou “fluxo”, de fases ou orientações.

Por terem sido todas as medidas cuidadosamente registradas, a soma das diferenças de fases ao longo do trajeto não depende do caminho escolhido. Isso significa que, independentemente do trajeto que o caminhante optou por seguir, é possível reconstituir as mudanças de direção em cada bar (nó) sem qualquer ambiguidade. A única exceção a isso é a fase inicial, ou seja, a orientação inicial do caminhante, que não pode ser determinada apenas com essas informações Fig.3.

Esta independência do caminho percorrido é exatamente o que foi discutido na seção 2.1.3 sobre grupos de sincronização. Neste contexto, a frustração do GSP é zero, porque a informação coletada das diferenças permite reconstituir o trajeto, sendo a única ambiguidade a fase inicial, ou seja, a orientação inicial do caminhante, que não pode ser determinada apenas com essas informações.

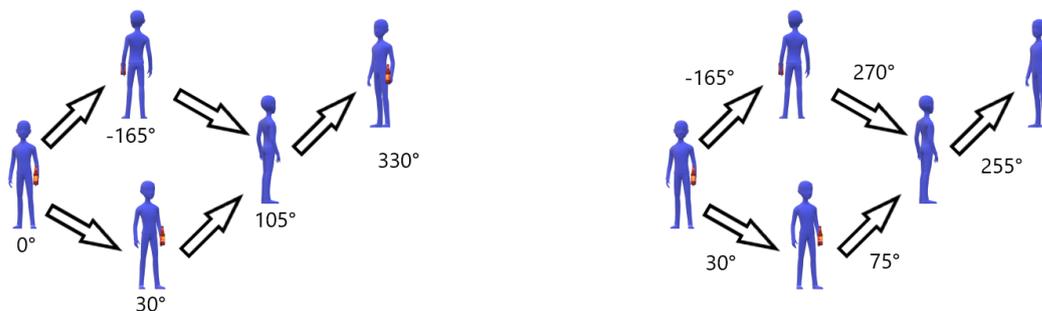


Figura 2 – Figura representando um caminhante aleatório confuso.. Na figura à esquerda, representamos os dois possíveis trajetos para o caminhante aleatório embriagado. Em cada etapa, o caminhante aleatório pode começar a girar no sentido anti-horário ou horário, e um observador externo registra os ângulos de rotação. Na figura à direita, representamos a diferença entre os ângulos de rotação para cada etapa.

Fonte: Elaborada pelo autor.

Vamos expandir nossa analogia para uma situação um pouco mais complexa. Suponha agora que, além do caminhante, o observador também comece a consumir bebida alcoólica sempre que passa pelo bar situado no caminho superior. Influenciado pelo álcool, o observador começa a anotar erroneamente as orientações, introduzindo ruído nas suas medidas.

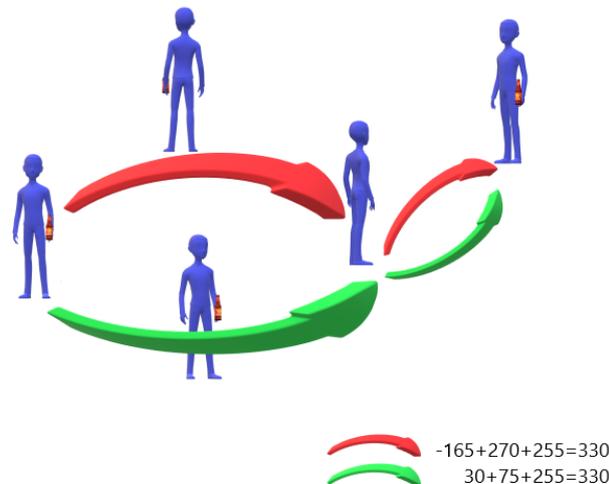


Figura 3 – Figura representando um GSP com independência de caminho. A figura acima ilustra a condição em que as fases do caminhante aleatório são registradas sem a presença de ruído. Neste cenário, a soma acumulada das diferenças de fase ao longo do trajeto independe do caminho escolhido. Isto é, é possível atribuir uma função potencial nos vértices.

Fonte: Elaborada pelo autor.

Nesse cenário, é como se o caminhante estivesse em uma “superposição” de rotações, o que lembra a situação de uma partícula carregada se movendo na presença de um potencial magnético, um fenômeno físico conhecido como efeito Aharonov-Bohm.<sup>32,55</sup> Com isso, a soma das diferenças de fase passa a depender do trajeto percorrido por ambos os amigos, e a frustração do GSP se torna maior que zero. Isso significa que a trajetória do caminhante não pode mais ser precisamente reconstituída com base nas notas do observador.

Em um ciclo fechado, os registros do observador sobre o comportamento do caminhante podem revelar características não triviais, como discontinuidades nas rotações do caminhante. Essa conexão com o tema de grafos e laplaciano magnético significa que, mesmo quando a informação coletada está corrompida por ruído, ainda é possível extrair *insights* importantes sobre o sistema em estudo. Isso é possível devido à estrutura subjacente do grafo e às propriedades dos operadores magnéticos, semelhante ao caso do efeito Aharonov-Bohm.

O formalismo da deformação magnética é uma ferramenta eficaz para a análise de ciclos em grafos direcionados e fluxos em grafos não direcionados. Entretanto, em certos contextos, como ao representar sistemas de agregação de preferências (ranqueamento), a deformação magnética pode não ser a abordagem mais adequada.

Por exemplo, no caso do Campeonato Brasileiro de Futebol,<sup>56</sup> os times podem ser

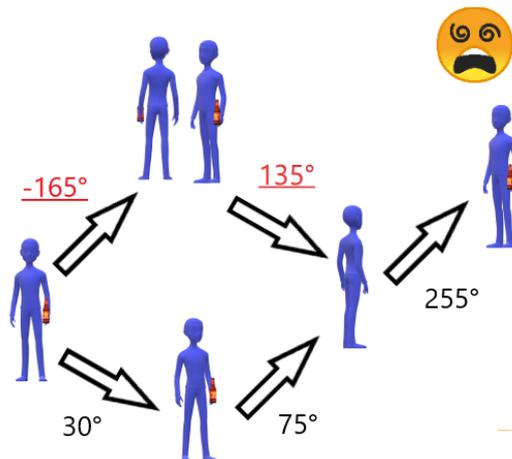


Figura 4 – Figura representando um GSP com a ruídos nas medidas. A ilustração acima representa um cenário no qual tanto o caminhante quanto o observador consomem álcool ao passar pelo bar no caminho superior. Sob a influência de bebida alcoólica, o observador começa a registrar as orientações de maneira imprecisa, introduzindo ruído nas medidas.

Fonte: Elaborada pelo autor.

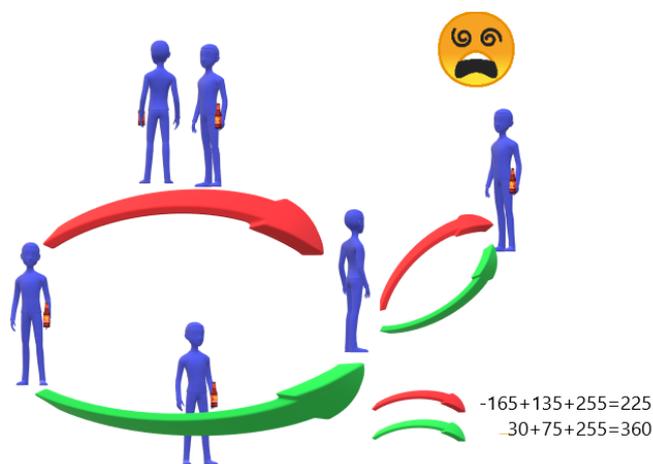


Figura 5 – Devido ao ruído na medição realizada pelo observador, o registro da caminhada agora depende do trajeto escolhido. Isto é, a soma dos giros (fases) não depende mais apenas dos vértices inicial e final, mas também do caminho percorrido por ambos. Neste cenário, a frustração do GSP é maior que zero.

Fonte: Elaborada pelo autor.

representados por vértices em um grafo direcionado, com os pontos de vitória e empate sendo pesos das arestas direcionadas. Neste cenário, optar por um método de ranqueamento

implica na decisão de como atribuir pesos aos vértices no grafo direcionado, com base nas conexões direcionadas que os vértices estabelecem entre si.

Para situações como essa, a abordagem da deformação por dilatação pode ser mais apropriada. Dessa forma, é crucial reconhecer o contexto e a natureza dos dados ao escolher a técnica de análise de grafos a ser aplicada.

### 2.1.5.1 Mergulho no espaço toroidal

O campo de estudo de mergulho de grafos consistem-se em encontrar métodos de mapear elementos de um dado grafo - vértices, arestas ou o próprio grafo em si - para uma dado elemento de uma dada variedade diferenciável de baixa dimensão.<sup>57</sup> Tal mergulho pode ser usado tanto para propósitos de visualização de dados como para aplicação de algoritmos de clusterização, predição de links e detecção de anomalias em imagens e redes sociais.<sup>58,59</sup>

Na literatura, diversos métodos de embebedimento de grafos foram propostos, cada um objetivando representar determinadas características dos objetos a serem mergulhados. Por exemplo, no contexto do formalismo Laplaciano combinatório o embebedimento do vértices utilizando o segundo autovetor associado ao segundo menor autovalor pode ser usado para resolver aproximadamente o problema de corte no caso de grafos não direcionados conectados.

No contexto de grafos direcionados, os autovetores do Laplaciano magnético também No contexto de grafos direcionados, os autovetores do Laplaciano magnético também estão relacionados a problemas de otimização e mergulho.<sup>60</sup> Especificamente, as fases do autovetor com menor autovalor, resolvem aproximadamente o problema de frustração. Tal problema consiste em encontrar um vetor  $\phi^* \in S^{|V|}$  que é solução para o seguinte problema de otimização

$$\Phi^* = \arg \min_{\Phi \in S^N} \left\{ \frac{\sum_{(u,v) \in E} w_{uv}^s |e^{2\pi i \phi_u} - \gamma_q(u,v) e^{2\pi i \phi_v}|}{2 \sum_{u \in V} d_u} \right\}. \quad (2.21)$$

### 2.1.6 Deformação por dilatação e ranqueamento ( $\mathbb{R}_+$ )

O método de HodgeRank<sup>26,56,61-63</sup> para ranqueamento se baseia na forma generalizada da decomposição de Hodge-Helmholtz,<sup>56,64</sup> amplamente aplicada nos campos do eletromagnetismo e da dinâmica de fluidos.

Um exemplo físico que ilustra a aplicação da decomposição de Hodge-Helmholtz é o estudo do fluxo de um fluido em torno de um objeto sólido, como um avião. Nesse caso, podemos analisar as propriedades do campo de velocidade do fluido utilizando a decomposição de Hodge-Helmholtz. O gradiente representa a parte do campo que está relacionada com a variação da velocidade ao longo do espaço. A componente harmônica

está associada a um comportamento suave e regular do campo, enquanto o rotacional captura a rotação do fluido em torno do objeto.

Essa decomposição é essencial para entender a interação entre o fluido e o objeto, permitindo uma análise mais precisa do escoamento e a identificação de fenômenos como a formação de vórtices. Além disso, a decomposição de Hodge-Helmholtz é uma ferramenta poderosa para o projeto e otimização da aerodinâmica, auxiliando na redução do arrasto e no aumento da eficiência de aeronaves.

Da mesma maneira, o método de HodgeRank, inspirado nessa decomposição matemática, permite decompor um grafo direcionado em componentes análogas (gradiente, harmônica e rotacional) e oferece uma abordagem rigorosa para o problema de ranqueamento de vértices em um grafo. Essa técnica proporciona uma mensuração mais precisa da qualidade da ordenação dos vértices e tem sido aplicada em diversos contextos, como análise de redes sociais, sistemas de recomendação e estudos de ecologia.

O método de HodgeRank tem sido redescoberto com o passar dos anos. No estudo de (Bacco, Larremore e Moore, 2018), os autores introduziram o termo “SpringRank” e propuseram um modelo inspirado na física que permite associar uma função de energia a um grafo direcionado. Em estudo mais recente, o método de HodgeRank foi novamente redescoberto sob a forma de níveis tróficos,<sup>65</sup> onde foi definido o Laplaciano trófico. No apêndice A, demonstramos que eles são todos equivalentes entre si.

Diferentemente do SpringRank e do Laplaciano trófico, o Laplaciano de dilatação apresenta uma variação em relação ao método de HodgeRank, apesar de ter este como caso limite. Este método, que representa uma alteração do Laplaciano combinatório, está relacionado ao grupo  $\mathbb{R}_+$ . Assim como o método do Laplaciano magnético, o Laplaciano de dilatação possui um parâmetro de ajuste. Este parâmetro permite controlar a qualidade do ranqueamento, podendo ajustar o foco para enfatizar os primeiros elementos (topo) ou os últimos elementos (base) na ordenação. O autovetor associado ao menor autovalor do Laplaciano de dilatação fornece diretamente uma pontuação de ranqueamento para seus nós. Esta abordagem pode ser vista como mais um aspecto a ser analisado ao abordar grafos dirigidos e/ou fluxos em grafos não dirigidos.

O Laplaciano de dilatação foi definido em<sup>27</sup> como:

$$(L_g f)(v_i) = \sum_{j \in V} (e^{g a_{ji}} f(v_i) - f(v_j)), \quad (2.22)$$

onde  $g > 0$ .

Em uma seção posterior, discutiremos como o Laplaciano de dilatação se complementa com o Laplaciano magnético, e como ambos podem ser utilizados para analisar grafos dirigidos e fluxos em grafos não dirigidos. Além disso, mostraremos uma maneira interessante de combinar as duas formulações. Mas antes, vamos transitar do tema de

grafos direcionados para o de grafos com sinais.

### 2.1.7 Grafos com sinais e sincronia de osciladores ( $\mathbb{Z}_2$ )

Os grafos com sinais<sup>66,67</sup> caracterizados por arestas com sinais positivos e negativos, são ferramentas poderosas para representar uma ampla gama de relações entre entidades. Para ilustrar, podemos visualizar uma rede social onde conexões amigáveis são representadas por arestas positivas e relações antagônicas por arestas negativas. Esses grafos desempenham um papel significativo em áreas como ciências sociais, biologia e ciência da computação, pois são capazes de captar interações complexas, incluindo tanto cooperação quanto competição entre indivíduos ou grupos.

Na linha do que foi discutido anteriormente, se o Laplaciano combinatório pode ser deformado para se tornar hermitiano em grafos direcionados, também podemos adaptar essa estrutura para lidar com grafos com sinais. O Laplaciano para esses grafos, conforme proposto por,<sup>10</sup> é definido da seguinte maneira em notação matricial:

$$\mathbf{L}_{ij} = \sum |A_{ij}| - A_{ij} \quad (2.23)$$

Nessa formulação, o Laplaciano adquire um grau definido pela soma dos valores absolutos da matriz de adjacência, tornando-se assim capaz de acomodar tanto relações positivas quanto negativas.

Como abordado em,<sup>66</sup> um gráfico com sinais pode ser visualizado como um sistema de osciladores acoplados, cada um podendo assumir um de dois estados, por exemplo, +1 ou -1. Aqui, as arestas positivas e negativas representam conexões sincronizadas e anti-sincronizadas, respectivamente. A dinâmica desses sistemas pode ser analisada por meio de técnicas de análise espectral aplicadas ao Laplaciano com sinais. Isso nos leva ao problema de sincronização  $\mathbf{Z}_2$ .

O estudo do primeiro autovetor do Laplaciano nos oferece uma aproximação para a solução do problema de sincronização  $\mathbf{Z}_2$ ,<sup>66</sup> demonstrando mais uma vez como a análise espectral pode revelar informações valiosas sobre a estrutura e a dinâmica de grafos complexos.

Em uma seção posterior, iremos explorar uma analogia interessante entre o problema de sincronização  $\mathbf{Z}_2$  e os grupos de renormalização na mecânica estatística, o que nos permite analisar transições de fase em sistemas complexos.

Com a introdução da noção de Laplaciano com sinais, estamos agora prontos para aprofundar a discussão sobre os Laplacianos de dilatação e magnético, bem como a forma como essas ferramentas podem fornecer *insights* únicos para a análise de grafos com características especiais. Nas próximas seções, abordaremos diversos métodos, analogias e *frameworks* propostos durante esta tese, com o objetivo tanto de aplicação prática em dados

reais como de investigação das propriedades intrínsecas desses Laplacianos deformados e dos grafos que representam.

## 2.2 Novas grandezas e métodos para grafos e fluxos

### 2.2.1 Além da simetrização

No estudo de grafos direcionados ou grafos com sinais, é prática comum recorrer a técnicas de simetrização ou à utilização do valor absoluto das conexões para construir um grafo não direcionado equivalente. Conforme observado nas seções anteriores, é possível sintetizar que um Laplaciano deformado por um grupo de simetria usa um grafo não direcionado e sem sinal como substrato, sobre o qual as operações de grupo são aplicadas para construir o Laplaciano deformado.

Em nossos estudos, propomos uma perspectiva inversa que nos levou à seguinte pergunta: Como podemos definir um operador Laplaciano generalizado a partir do qual podem ser extraídos grafos efetivos? Essa abordagem nos permite extrair informações relevantes dos grafos direcionados e com sinais sem perder a complexidade inerente a essas estruturas, permitindo decompor um grafo em toda uma família de grafos não direcionados, nos quais a solução por simetrização básica (ou seja,  $A + A.T$ ) e/ou absoluto ( $|A|$ ) são casos especiais.

Nessa seção iremos explorar essa proposta em mais detalhes, desenvolvendo uma metodologia para a definição de um operador Laplaciano generalizado e investigando as possíveis aplicações dessa abordagem.

#### 2.2.1.1 Generalizando o Grau de um Vértice

A agregação de um sinal em relação aos seus vizinhos é uma técnica tradicional para reduzir certos tipos de ruídos em imagens e grafos, como ilustrado na Fig.6. Podemos conceber o grau de um vértice de um grafo como uma soma ponderada de uma função constante ( $f(v)$ ) nos vizinhos desse vértice. Este entendimento do grau e sua relação com a filtragem nos levam à uma definição mais geral de uma função grau.

Antes de prosseguirmos com a definição de tal generalização, baseada no problema de sincronização descrito em 2.1.3, vamos considerar  $T \in \mathcal{G}$  com uma representação em  $\mathcal{F}$ . Definimos o tensor de promoção,  $P$ , como

$$P_{uv}^{tz} = T_{vu}^* T_{tz}. \quad (2.24)$$

Com o tensor de promoção agora definido, passamos à função de grau generalizada.

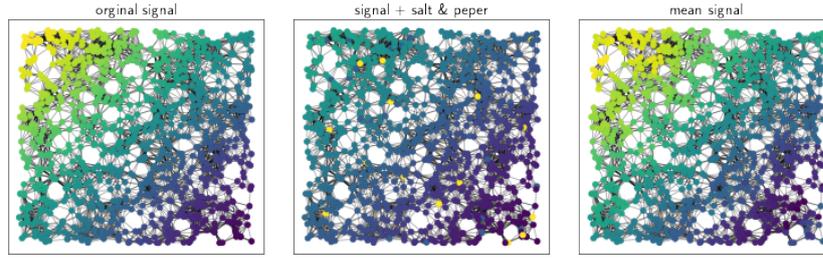


Figura 6 – Na figura da esquerda, é mostrado um sinal em um grafo geométrico. Na figura central, o mesmo grafo é apresentado com o sinal contaminado por um ruído ‘salt & peper’. Na figura da direita, é mostrado o resultado da filtragem utilizando a agregação de discrepâncias entre o sinal, ou seja,  $f(u) = f(u) + \frac{1}{D(u)} \sum_{v \in Nei(u)} (f(v) - f(u))$ .

Fonte: Elaborada pelo autor.

**Definição 1 (Função de Grau Generalizada)** *Seja  $H : V \times V \mapsto \mathcal{V}$ , seja  $f : V \mapsto \mathcal{F}$  uma função sinal definida nos vértices.*

*Denominamos como função grau generalizada um  $D : \mathcal{H} \times V \mapsto \mathcal{V}$  tal que*

$$\begin{aligned}
 D_{\mathcal{G}}(H, f, u) &= \\
 & \sum_{v \in Nei(u)} w(u, v) (P_{vu}^{uv} H(u, v) f(v) - P_{uv}^{uv} H(u, u) f(u)) \\
 &= \sum_{v \in Nei(u)} w(u, v) (\omega_H(u, v) f(v) - w_{H, \mathcal{G}}(u, v) f(u))
 \end{aligned} \tag{2.25}$$

Note que se  $\mathcal{V}$  são os reais,  $T = 1$  e  $h(u, v) = 1 - \delta_{uv}$  então a função acima nada mais é que o conhecido grau do vértice  $u$  utilizado na literatura.

Em notação matricial, seja  $\mathbf{F}$  um vetor representando um sinal nas arestas, então podemos associar a seguinte matriz:

$$\begin{aligned}
 \mathbf{D}_{\mathcal{G}}(H) &= \mathbf{W} \odot (\tilde{\mathbf{P}} \odot \mathbf{H} - \mathbf{D}^{(P)}) \\
 &= \mathbf{W} \odot \mathbf{H}_d,
 \end{aligned} \tag{2.26}$$

onde:

$$\mathbf{D}_{uv}^{(P)} = \begin{cases} \sum_{s \in Nei(u)} P_{us}^{us} h(u, u), & \text{se } u = v \\ 0 & \text{caso contrário} \end{cases} \tag{2.27}$$

Um caso especial é quando  $\mathbf{H} = \mathbf{A}$ . Se  $T_{uv} = 1$ , a definição de grau é recuperada,  $\mathbf{D}_{\mathcal{G}} = \mathbf{W} \odot \tilde{\mathbf{P}}$ .

A definição de grau generalizado nos permite expressar o Laplaciano deformado como:

**Definição 2 (Laplaciano por deformação de Grupo)**

$$(L_{\mathcal{G}}f)(u) = D_{\mathcal{G}}(A, u)f(u) - \sum_v w_{\mathcal{G}}(u, v)f(v) \quad (2.28)$$

$A$  é adjacência de  $G$ .

A definição de um grau generalizado dada pela Eq.(2.25) foi essencial para expressar o Laplaciano deformado Eq.(2.28) de maneira análoga ao combinatorial Eq.(2.14).

O grau generalizado em relação a um vértice  $u$  pode ser compreendido como a contabilização das discrepâncias de um sinal ( $f$ ) em um vértice  $u$  em relação aos seus vizinhos. Essa discrepância é representada pelo termo  $\omega_H(u, v)f(v) - w_{H, \mathcal{G}}(u, v)f(u)$ . Agora, vamos analisar o caso em que  $\mathbf{H} = \mathbf{A} + \mathbf{I}_{|V|}$ . Nesse caso específico, vamos omitir o subíndice  $H$  de  $\omega_H$  e  $w_{H, \mathcal{G}}$ .

2.2.1.2 Discrepância nas Arestas e Matrizes de Pesos Efetivas

A discussão sobre grau generalizado pode parecer extensa, ou mesmo desnecessária, à primeira vista, mas, a partir de agora e na seção de resultados (referenciada posteriormente), demonstraremos como podemos nos beneficiar dessas definições para traçar analogias e definir novos métodos. Começemos com a definição de discrepância nas arestas.

**Definição 3 (Discrepância na Aresta)** Definimos a discrepância em  $\xi : E \mapsto \mathbb{R}_+$  para um dado  $f \in \mathcal{V}$  como

$$\xi_f(u, v) = \| \omega(u, v)f(v) - w_{\mathcal{G}}(u, v)f(u) \|_{\mathcal{V}} \quad (2.29)$$

A partir da definição de discrepância acima, podemos definir uma medida de frustração generalizada da seguinte forma:

**Definição 4 (Frustração Generalizada)**

$$\eta_f = \frac{1}{2} \sum_{(u,v) \in E} \frac{w(u, v)\xi_f(u, v)^2}{\sum_{u \in V} D_{\mathcal{G}}(A, u) \| f(u) \|_F^2} \quad (2.30)$$

Como mencionado anteriormente, se  $F_{uv}^{vu} = \varepsilon$ , a função de grau em  $u$  com  $H = A$  é idêntica ao grau do vértice  $u$  no grafo não direcionado. Também teremos que  $\xi_f(u, v) = f(v) - w_{\mathcal{G}}f(u)$ . Se  $\| f(u) \|_F^2 = 1$ , então a frustração generalizada é equivalente à frustração usada no problema de sincronização de grupo discutida na seção 2.1.3, dado por:

$$\frac{1}{2} \sum_{(u,v) \in E} \frac{w(u, v) \| f(v) - w_{\mathcal{G}}f(u) \|_F^2}{\sum_{u \in V} \text{Vol}(G)} \quad (2.31)$$

As duas definições anteriores nos motivam a definir uma função de peso efetivo associada a um grafo.

**Definição 5 (Peso Efetivo)** Para um  $\beta \in \mathbb{R}_+$

$$w_e^{(f,\beta)}(u, v) = w(u, v)e^{-\beta\xi_f(u,v)^2} \quad (2.32)$$

Por simplicidade, assumindo que  $\|f(u)\|_F^2 = 1$ , a frustração generalizada pode ser escrita em termos de um potencial logarítmico nas arestas

$$\eta_f = -\frac{\beta}{\text{Vol}_G(G)} \sum_{(u,v) \in E} w \ln \left( \frac{w_e^{(f)}}{w} \right) (u, v) \quad (2.33)$$

Uma implicação imediata da definição acima é que, para cada grafo não direcionado sem sinal equipado com um potencial de grupo,  $G = (V, E, w, T)$ , temos a ele associada uma família de grafos,  $G_e^{(f,\beta)} = (V, E, w_e^{(f,\beta)})$ . Com isso, novas caracterizações e medidas de  $G$  podem ser extraídas levando-se em conta  $G_e^{(f,\beta)}$ . No caso de que  $P_{uv}^{vu} = 1$ , tais grafos são não direcionados.

Concluimos aqui a discussão sobre grupos de sincronização, laplacianos deformados e pesos efetivos. Na seção de resultados, mostrando como uma análise de dados em grafos, levando-se em conta essas formulações de laplacianos deformados desenvolvidas em trabalhos anteriores<sup>27,66</sup> e nesta tese, contribuem para a compreensão de dados em grafos.

A seguir, discutiremos grandezas e métodos definidos com base nessas novas formulações para grafos direcionados, com sinais e/ou fluxos em grafos não direcionados.

## 2.2.2 Espectro de um Grafo no Contexto de Deformações de Grupo e Grandezas Associadas

No emblemático trabalho “Can One Hear the Shape of a Drum?”<sup>68</sup> Kac propõe um questionamento intrigante: conhecendo o espectro (som) de um determinado operador que define as oscilações de uma membrana (tambor), seria possível identificar univocamente a forma dessa membrana? A discussão aqui é análoga: focaremos no espectro dos operadores laplacianos e matrizes de pesos generalizadas e como podemos utilizá-los para caracterizar, identificar e comparar redes complexas?

### 2.2.2.1 Análogos termodinâmicos e mecânico estatísticos

O operador descrito na Eq.(2.14) é um mapa linear em um espaço vetorial de dimensão finita. Desta maneira, existem  $f_\lambda \in F_V$  tal que  $(Lf_\lambda)(u) = \lambda f_\lambda(u)$ , onde  $\lambda$  pertence a  $\mathbb{R}$  (grafos não direcionados) ou  $\mathbb{C}$  (direcionados) sendo que os valores de  $\lambda$  possíveis podem ser obtidos através da solução das raízes do polinômio característico dado por  $\chi(\lambda) = \det(\lambda\mathbf{I} - \mathbf{L})$ .

No que se refere à caracterização de redes sociais e biológicas, uma infinidade de trabalhos recentes abordou essa caracterização usando métodos espectrais.<sup>69,70</sup> Muitos deles utilizaram conceitos de mecânica estatística para definir novas funções e medidas, permitindo a inclusão de um parâmetro de escala,<sup>70,71</sup> tal como a temperatura. No entanto, devido aos operadores Laplaciano e adjacência combinatórios no caso direcionado não serem hermitianos, essa analogia carece de sentido.

Frente a esse desafio, propomos e analisamos essas analogias usando o formalismo de operadores deformados, com foco em operadores magnéticos.<sup>1,21</sup> Estes são hermitianos, mesmo na presença de direção, expandindo assim o escopo de aplicação das analogias termodinâmicas para redes direcionadas. Na seção seguinte, discutiremos mais detalhadamente como as técnicas espectrais podem ser empregadas para analisar e caracterizar redes, incluindo as novas medidas e conceitos apresentados neste trabalho.

### 2.2.2.2 Entropia

As analogias que construímos em nossos trabalhos anteriores<sup>1,72</sup> puderam ser desenvolvidas através da definição da função de partição associada a um dado grafo. Essa função de partição é essencial, pois ela permite caracterizar a estrutura e as propriedades de um sistema em termos termodinâmicos, e pode ser expressa pela seguinte equação:

$$Z(\beta, q) = \sum_{i=1}^{|V|} e^{-\beta \lambda_i(q)} \quad (2.34)$$

Aqui,  $\beta \in \mathbb{R}_+$  representa o inverso da temperatura, sendo o parâmetro de escala, e  $\lambda_i(q)$  é o  $i$ -ésimo autovalor associado ao operador Laplaciano magnético normalizado para um valor de carga  $q$ . Note que essa função é similar à função de partição da mecânica estatística, em que se pondera os estados do sistema pela sua energia através de uma distribuição exponencial.

### 2.2.2.3 Calor Específico

Com base na função de partição definida na Eq.2.34, podemos analisar o comportamento de diversas grandezas termodinâmicas. Por exemplo, consideramos a função definida a seguir, que é determinada em termos da variância do espectro:

$$c_\lambda(q, \beta) = \beta^2 (\mathbb{E}(\lambda_q^2) - \mathbb{E}(\lambda_q)^2), \quad (2.35)$$

onde

$$\mathbb{E}(\lambda) = \frac{\sum_{i=1}^{|V|} \lambda_i e^{-\beta \lambda_i}}{Z}. \quad (2.36)$$

A Eq.2.35 é análoga à definição de calor específico na mecânica estatística, razão pela qual optamos por manter essa nomenclatura.

Na teoria da matéria condensada, o comportamento do calor específico em relação a parâmetros externos é uma característica importante para identificar o tipo de material. Similarmente, no contexto dos grafos, analisamos como o calor específico se comporta em relação à temperatura e à carga. Essa medida, assim como a entropia, nos permite entender melhor as propriedades estruturais dos grafos e como elas se alteram em resposta a diferentes parâmetros, potencializando nossa capacidade de caracterização e análise de redes.

Essas analogias termodinâmicas ganham ainda mais relevância quando consideramos as propriedades de simetria do Laplaciano magnético. Como discutido anteriormente, essas simetrias, especialmente as que surgem de ciclos no grafo, têm um impacto significativo nas propriedades espectrais do operador como é mostrado na seção 3.1. Conseqüentemente, elas também afetam grandezas como a entropia e o calor específico, oferecendo uma maneira intrigante de conectar a estrutura e as simetrias do grafo com as suas propriedades termodinâmicas efetivas.

### 2.2.3 Análise de Gases Matriciais e Teoria de Matrizes Aleatórias para Grafos Direcionados

Na física, os autovalores do operador Hamiltoniano têm um papel crucial ao determinar o comportamento de um sistema quântico. Por exemplo, o conjunto de diferenças entre esses autovalores representa o conjunto de valores de energia que um fóton pode ter para ser absorvido por esse sistema. No entanto, para certos tipos de sistemas - como núcleos atômicos pesados - a identificação desse Hamiltoniano é complexa. Motivado por essa dificuldade, Wigner e Dyson propuseram estudar um núcleo atômico analisando o que seria conhecido como gás matricial.<sup>73</sup>

Os gases matriciais podem ser analisados com a teoria de matrizes aleatórias (RMT - Random Matrix Theory). Esta teoria alcançou aplicações em diversas áreas, como economia,<sup>74</sup> sociologia, e matemática pura.<sup>75</sup> Em suma, dado um possível modelo de uma matriz aleatória, a RMT tem como objetivo estudar o comportamento estatístico dos autovalores e autovetores desse modelo. A conjectura é que esse comportamento estatístico nos permite caracterizar o modelo que gera tais matrizes. Um exemplo disso é a demonstração de propriedades universais no que se refere ao comportamento do espectro.

Pode-se questionar por que o estudo de tais propriedades universais é interessante. Podemos citar como exemplo que os preços de fechamento de ativos no mercado de ações

são comumente utilizados para construir matrizes de correlação. Estas, por sua vez, podem ser usadas em algoritmos de otimização de portfólios de investimento. No entanto, em<sup>76</sup> os autores usaram a distribuição de espaçamentos de autovalores dessas matrizes de correlação para demonstrar que, em determinados períodos, essas matrizes pertencem a certas classes de matrizes aleatórias que caracterizam sistemas não correlacionados. Portanto, os autores concluem que há uma presença de ruído que invalida certos algoritmos de construção de portfólios de investimento. Posteriormente, é demonstrado como é possível utilizar resultados de RMT para remover tais ruídos, o que tende a aprimorar a performance de algoritmos de construção de portfólios.<sup>74,77</sup>

No presente trabalho, utilizamos a RMT para analisar grafos direcionados, com ênfase no modelo de grafo aleatório direcionado Erds-Rényi,  $G(n, p)$ . Este modelo é definido por dois parâmetros:  $n$  (número de vértices) e  $p$  (probabilidade de conexão entre dois vértices). Fixando o valor de  $n$  e  $p$ , dizemos que um determinado número  $N_{exp}$  de realizações de  $G(n, p)$  forma um conjunto (ensemble) de grafos, cada um com suas respectivas matrizes de adjacência e Laplaciana. Assim, podemos associar com cada tupla de parâmetros, e com as respectivas matrizes (Laplaciana ou adjacência) aleatórias, um ensemble de autovalores e autovetores.

Seja, por exemplo,  $G = (V, E)$  um grafo de uma realização de  $G(n, p)$ . Associamos a este grafo uma matriz aleatória cujas entradas são dadas por:

$$\mathcal{W}_{uv} = \begin{cases} \mathcal{N}(0, 1/2) & \text{se } (u, v) \in E, \\ \mathcal{N}(0, 1) & \text{se } u = v, \\ 0 & \text{caso contrário} \end{cases}, \quad (2.37)$$

Observe que a matriz acima pode ser obtida pela multiplicação ponto a ponto da matriz de adjacência do grafo  $G$  por uma realização de uma matriz aleatória gaussiana ortogonal  $GOE$  (se  $G$  é não direcionado) ou de Ginibre (se  $G$  é direcionado).

Em relação à caracterização dos autovalores do ensemble de matrizes aleatórias simétricas, uma função de caracterização bem conhecida é a função de espaçamento dos autovalores.<sup>78</sup> Tal função é dada por:

$$r_k = \frac{\min(\lambda_k - \lambda_{k-1}, \lambda_{k+1} - \lambda_k)}{\max(\lambda_k - \lambda_{k-1}, \lambda_{k+1} - \lambda_k)} \quad (2.38)$$

onde  $k$  é um inteiro com  $0 < k < |V|$  e  $\lambda_k$  representa o  $k$ -ésimo autovalor do espectro  $\Lambda$  de  $G$ .

Dado a Eq.(2.38), associa-se com cada tupla de parâmetros  $(n, p)$  uma distribuição de probabilidade (pdf) da função de espaçamento para o ensemble de grafos  $\rho_r$ . Surpreendentemente, quando  $p = 1$ , a distribuição de espaçamento dos autovalores do grafo também representa o espaçamento entre números primos consecutivos.<sup>73</sup>

No caso em que  $G$  é direcionado e lidamos com o operador de adjacência usual, a função Eq.(2.38) não é válida uma vez que  $\Lambda \subset \mathbb{C}$ . Recentemente, (Sá, Ribeiro e Prosen, ) propuseram um análogo a tal equação para o caso de matrizes aleatórias não simétricas e não-hermitianas. Esta função é conhecida como espaçamento dos primeiros vizinhos e é dada por:

$$r_{\mathbb{C},k} = \frac{\lambda_k^{NN} - \lambda_k}{\lambda_k^{NNN} - \lambda_k} \quad (2.39)$$

Onde o índice  $NN(NNN)$  se refere ao primeiro(segundo) vizinho de  $\lambda_k$ . Note que, devido à impossibilidade de se definir um operador de ordenamento total em  $\mathbb{C}$ , não é possível ter a Eq.(2.38) como limite.

No desenvolvimento da presente tese utilizamos a Eq.(2.39) para analisar o comportamento de modelos de grafos aleatórios.<sup>21</sup>

Propomos também que outra maneira de se analisar ensembles de grafos direcionados é a utilização do formalismo de deformações que retornam operadores Hermitianos, mesmo quando as matrizes de adjacência não são simétricas. Salientamos que tal análise também pode auxiliar na compreensão de como os parâmetros envolvidos na construção desses operadores deformam  $\Lambda$ .

Adicionalmente, a análise espectral de matrizes aleatórias também envolve a avaliação do comportamento estatístico dos autovetores associados a essas matrizes. Um desses comportamentos refere-se à evolução de descritores estatísticos da entropia do autovetor com o aumento da conectividade. Essa entropia é um valor real positivo calculado usando a seguinte função para uma dada realização:

$$S_\lambda = -\mathbf{f}_\lambda^\dagger \mathbf{f}_\lambda \ln \mathbf{f}_\lambda^\dagger \mathbf{f}_\lambda, \quad (2.40)$$

Onde  $\mathbf{f}_\lambda^\dagger$  é o transporte conjugado do autovetor associado ao autovalor  $\lambda$ . Comumente, se avalia a média dessa função para várias realizações  $\langle \sum_{\lambda \in \Lambda} \frac{S_\lambda}{N} \rangle$  em função de  $p$ .

O leitor que acompanhou até aqui pode ter observado a diversidade de operadores e quantidades propostos para explorar o espaço espectral de redes direcionadas. Modelos de redes direcionadas podem ser estudados utilizando os espaços associados a operadores combinatórios, magnéticos, de dilatação e simetrização. Entretanto, cada operador pode gerar um espectro diferente, até mesmo em campos distintos (Complexo e Real), proporcionando diferentes caracterizações dos modelos de rede. Por isso, torna-se crucial quantificar quão diferentes são esses espaços, operadores e quantidades.

No caso específico do operador magnético, a dificuldade de estudar diretamente as matrizes de adjacência deformadas surge porque os valores agora são invariantes por uma rotação de  $2\pi n$ . Para contornar essa questão e entender melhor os valores das entradas das matrizes de adjacência magnética, recorreremos ao formalismo da estatística circular. No entanto, apesar deste formalismo fornecer informações sobre como os pesos magnéticos se comportam, ele ainda não é suficiente para mensurar quão distante um espaço espectral magnético está de outro. Para essa quantificação, propomos a utilização de ferramentas provenientes da teoria do transporte ótimo e técnicas de aprendizado de variedades aplicadas ao espaço espectral.

Embora a primeira vista, o estudo da estatística circular e a quantificação de distâncias entre espaços utilizando o transporte ótimo possam parecer pouco práticos, eles fornecem métodos valiosos para explorar conjuntos de dados compostos de redes direcionadas. Como exemplo dessa aplicação, estudamos o espectro da matriz de adjacência magnética relacionada com as redes de interação gênica. Nossos resultados indicam que as relações entre alguns genes podem ser meramente efeitos de ruído.

Além disso, é importante frisar a relevância dos estudos nessa área, especialmente à luz de pesquisas recentes que revelaram comportamentos ruins em redes neurais magnéticas de grafos dirigidos com heterofilia.<sup>80</sup> Tais observações destacam a importância de analisar a estrutura espectral dessas redes para entender suas propriedades, estendendo as discussões anteriores acerca da RMT e suas aplicações para analisar tais sistemas complexos.

## 2.3 Algumas aplicações para análise de dados e aprendizado de máquina

### 2.3.1 Dados tabulares e clusterização de colunas

Em 1895, Edmund Landau<sup>6</sup> propôs em uma revista de xadrez um método para distribuir prêmios entre jogadores, considerando uma tabela de vitórias. O problema levantado por Landau pode ser posto da seguinte forma: dado um conjunto  $V$  de jogadores e  $E$  um conjunto de tuplas, onde para cada  $u, v \in E$ , temos que se  $(u, v) \in E$ , então o jogador  $u$  disputou uma partida com o jogador  $v$ . A essa tupla associa uma função de pontuação  $w : E \mapsto 0, 0.5, 1$  de maneira que:

$$w(u, v) = \begin{cases} 0 & \text{se } u = v, \text{ ou se } u \text{ não disputou com } v, \\ 1 & \text{se } u \text{ ganhou de } v, \\ 0.5 & \text{em caso de empate,} \\ 0 & \text{se } u \text{ perdeu a partida para } v. \end{cases}$$

A estrutura formada por  $(V, E, w)$  é um grafo, tal que se todas as partidas terminarem em empate, o grafo é não direcionado. Aqui,  $V$  representa o conjunto de vértices,  $E$  o conjunto de arestas e  $w$  a função peso.

A possibilidade mais simples para definir o melhor jogador de xadrez é definir uma função que soma todas as pontuações

$$d : V \mapsto \mathbb{R}$$

$$d(u) = \sum_{(u,v) \in E} w(u,v), \quad (2.41)$$

a função acima é a previamente mencionada função grau do vértice  $u$  e é membro do conjunto de funções definidas nos vértices,  $d \in F_V = f_V : V \mapsto \mathbb{R}$ . Embora a função  $d$  possa ser usada para distribuir prêmios entre os jogadores, é importante notar que um jogador mediano poderia obter um score alto apenas jogando contra adversários iniciantes, e conseqüentemente, ganhar um prêmio maior que um jogador que tenha disputado e empatado com jogadores mais experientes.

Nesse contexto, Landau apresenta uma questão relevante: como encontrar uma função  $f \in F_V$  que permita uma distribuição de prêmios mais justa? Isso equivale a um problema similar ao de quantificar a influência de um usuário em uma dada rede social.<sup>81</sup> Portanto, o trabalho de Landau é pioneiro na análise quantitativa de redes sociais, tomando a rede social formada por jogadores de xadrez como um exemplo.

A solução proposta por Landau busca encontrar uma função  $f$  que satisfaz a seguinte equação para todo  $u \in V$ :

$$\lambda_{\max} f(u) = \sum_{v \in V} w(u,v) f(v), \quad (2.42)$$

onde  $\lambda_{\max} \in \mathbb{R}$ . Dessa forma, Landau sugere como solução um uma função de ranking  $\mathbf{f}$  ( $\mathbf{f}_u = f(u)$ ), que é o autovetor associado ao maior autovalor da matriz  $\mathbf{W}$ , isto é:

$$\mathbf{f}^T \mathbf{W} = \lambda_{\max} \mathbf{f}^T. \quad (2.43)$$

É notável que a abordagem de Landau aproveita as propriedades do teorema de Perron-Frobenius antes mesmo de sua publicação. Por essa razão, Landau revisita o problema de ranqueamento em (Landau, 1915), após a publicação do teorema de Perron.<sup>81</sup>

Com o advento de novos desafios na análise de dados tabulares, as limitações do método proposto por Landau, inicialmente voltado para o dados tabulares de jogos de xadrez, tornaram-se aparentes. Seja na determinação do desempenho socioeconômico de indivíduos ou na identificação de fatores de saúde que impactam conjuntos de doenças.<sup>83</sup>

Para abordar esses novos desafios, estudos recentes<sup>84-87</sup> propõem abordagens de mapeamento de dados tabulares em grafos visando uma melhor análise exploratória dos

dados. Em alguns estudos, como,<sup>84,85,87</sup> essas relações são expressas por arestas não direcionadas, cujos pesos são determinados com base em medidas de informação mútua ou valores de correlação. No entanto, a construção desses grafos pode resultar em vários componentes desconexos, um problema se as relações entre *todos* os pares de vértices forem importantes.

Dentro desse cenário, introduzimos<sup>5</sup> um arcabouço para o mapeamento e análise exploratória de dados tabulares. Reconhecemos que além de mapear dados tabulares em um grafo, é essencial identificar e lidar com grupos de colunas que apresentam interferências mútuas. Essa proposta foi inspirada na à necessidade de remover ruído ou plano de fundo de imagens utilizando o espectro do Laplaciano combinatório relacionado com sua imagem.

E é nesse ponto que a relevância do Laplaciano magnético se destaca. O uso desta abordagem é particularmente útil para desvendar as interações em ciclos de relações, tornando a análise exploratória mais clara.

Portanto, o coração do nosso arcabouço está na identificação de quando as colunas (ou recursos) formam ciclos de relações entre si, situações que podem dificultar a análise exploratória. Para isso, nós implementamos uma abordagem que integra blocos estocásticos aninhados, a esparsificação de arestas e a Explicação Automatizada de Inteligência (XAI), visando desvendar a estrutura subjacente do grafo associado a um conjunto de dados tabular.

O diagrama Fig.7 representa nosso método proposto. Dividimos o método em quatro etapas. O método inicia através da construção de grafo direcionado com pesos a partir dos dados tabulares. Nessa construção, cada atributo (ou coluna) é associado a pelo menos um vértice do grafo. As arestas representam as relações entre duas colunas. Consideramos  $C$  como sendo o conjunto de todas as colunas. Uma coluna  $c \in C$  é selecionada aleatoriamente e as colunas restantes são usadas como atributos para treinar um algoritmo de *gradient boosting machine* (GBM) a fim de prever os valores da coluna  $c$ . Chamamos  $\bar{V}_c$  as colunas mapeadas de  $c$ .

Depois do treinamento, guardamos a contribuição do vértice  $v$  na tarefa de prever os vértices relacionados a  $c$ . Repetimos este procedimento para cada vértice em  $V$ , resultando em um grafo direcionado e ponderado completo.

Para determinar a contribuição de  $v$  na previsão de  $c$ , colocamos algumas restrições. Exigimos que o grau de entrada do vértice  $v_c$  reflita a precisão do GBM treinado, ou seja,  $k_{in}(v_c) = \sum_{u \in \bar{V}_c} w(u, v_c) = Acc(v_c)$ . Se uma coluna não tem uma relação significativa com as demais colunas ou não pode ser explicada por elas, o grau de entrada será baixo, diminuindo a contribuição do vértice para a estrutura geral do grafo.

Definimos  $\epsilon(u \rightarrow v_c)$  como uma função que quantifica a contribuição de uma coluna associada a  $u$  na previsão dos valores da coluna  $p$  usando o GBM. Assim, escolhemos a

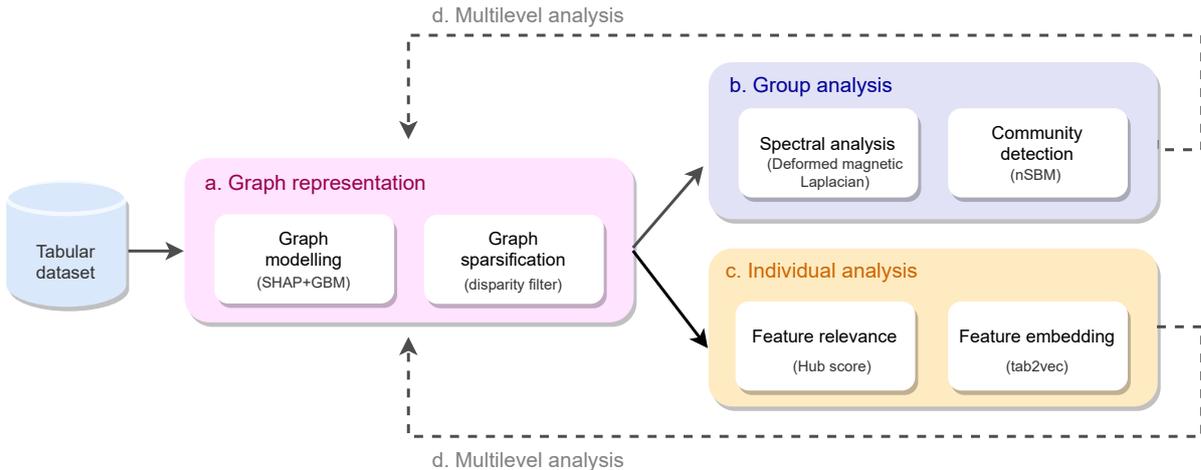


Figura 7 – Diagrama sequencial da metodologia proposta para análise exploratória de dados tabulares. O dataset tabular é primeiramente transformado em um grafo direcionado e ponderado. Posteriormente, um processo de esparsificação do grafo é aplicado. Este grafo é então empregado para conduzir (b) uma análise de agrupamento utilizando técnicas de análise espectral e identificação de comunidades. Além disso, possibilita a execução de (c) análises individuais (ranqueamento) e comparativas (mergulho das colunas via Laplaciano magnético).

Fonte: RESENDE; TOKUDA; COSTA.<sup>5</sup>

função de peso de uma aresta  $(u, v)$  como:

$$w(u, v) = Acc(v) \frac{\epsilon(u \rightarrow v)}{\sum_{z \in V} \epsilon(z \rightarrow v)}. \quad (2.44)$$

Para aplicar a Eq. 2.44, precisamos escolher uma maneira de quantificar a contribuição de uma variável  $v$  devido à presença de um atributo  $u$ . Existem vários métodos na literatura para isso.<sup>88</sup> Em nosso trabalho, decidimos usar as Explicações Aditivas de Shapley (SHAP, na sigla em inglês).<sup>89,90</sup> O método SHAP, inspirado na teoria dos jogos cooperativos, quantifica a contribuição marginal de um recurso para uma única tarefa de previsão.

Como o valor SHAP é calculado para cada elemento do conjunto de dados, temos um grafo diferente definido pela Eq. 2.44 para cada instância. Por exemplo, se os dados tabulares são de uma pesquisa, o grafo pode ser utilizado para estudar as respostas individuais. No entanto, mesmo permitindo a criação de um grafo para cada instância nos dados, neste trabalho nos concentramos em um único grafo para descrever todo o conjunto de dados. Neste caso, o peso da aresta  $(u, v)$  é definido como a média dos valores absolutos do SHAP, ou seja,

$$w(u, v) = Acc(v) \frac{\mathbb{E}[|\text{SHAP}_i(u \rightarrow v)|]}{\sum_{z \in V} \mathbb{E}[|\text{SHAP}_i(z \rightarrow v)|]}. \quad (2.45)$$

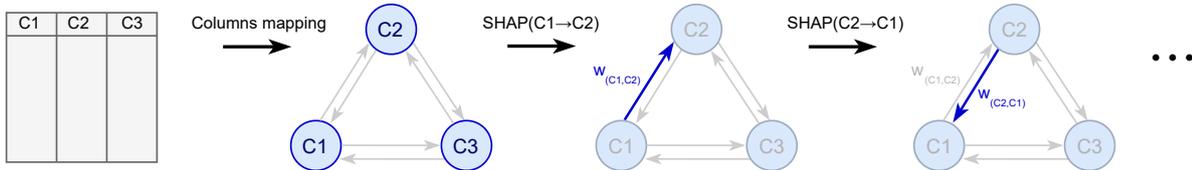


Figura 8 – Representação da construção do grafo direcionado de interpretabilidade. Inicialmente, as colunas dos dados tabulares são mapeadas para um grafo direcionado completo. Em seguida, o peso  $w_{(C1,C2)}$  do arco  $(C1, C2)$  é calculado a partir do valor SHAP de  $C1$  na tarefa de previsão de  $C2$ . O mesmo procedimento é repetido para atribuir o peso de cada arco no grafo.

Fonte: RESENDE; TOKUDA; COSTA.<sup>5</sup>

É compreensível que, à luz dos benefícios apresentados pelo formalismo do Laplaciano magnético, surja a tentação de aplicá-lo imediatamente. No entanto, é preciso ter cautela. Em nosso caso, onde estamos lidando com um grafo completo, a aplicação direta desse formalismo pode levar a um alto custo computacional devido ao cálculo das propriedades em questão. Além disso, como será discutido na seção de resultados o excesso de informação (relação, arestas) ofusca o resultado de métodos de visualização e da própria solução do problema de sincronização de grupo.

Uma abordagem simples para reduzir o número de arestas e aumentar a interpretabilidade das técnicas de visualização de grafo consiste na aplicação de um limiar aos pesos das arestas, mantendo apenas as conexões mais fortes. No entanto, é difícil escolher o valor usado para o parâmetro de limiar<sup>91</sup> que pode ocasionar muitos componentes desconectados.

Na última década, um grande número de métodos de filtragem de grafos (também conhecidos como esparsificação de grafos) foi desenvolvido para mitigar os problemas presentes na abordagem de filtragem de arestas baseada em limiar ingênuo.<sup>92–94</sup> Neste trabalho, adotamos o critério de filtro de disparidade desenvolvido por<sup>92</sup> para filtrar as arestas.

Seja  $s(u) = \sum_{v \in V | (u,v) \in E} w(u, v)$  o grau de saída de um atributo associada ao nó  $u$  no grafo de interpretabilidade. Definido desta forma,  $s(u)$  está relacionado à contribuição do atributo  $u$  para explicar todas as outros atributos. Assim,  $p(u, v) = w(u, v)/s(u)$  quantifica o quanto a explicação dada pelo atributo  $u$  na tarefa de prever o atributo  $v$  contribui para a quantidade total de interpretabilidade do atributo  $u$ . Então, com  $k_{out}(u)$  sendo o grau

de saída do nó  $u$ , podemos associar com cada aresta  $(u, v)$  a seguinte quantidade:

$$w_\alpha(u, v) = 1 - (k_{out}(u) - 1) \int_0^{p(u,v)} (1 - x)^{k_{out}(u)-2} dx. \quad (2.46)$$

Arestas com  $w_\alpha$  acima de um determinado limiar  $\alpha \in [0, 1]$  são filtradas. Portanto, esse método permite filtrar as arestas e, ao mesmo tempo, manter a estrutura principal do grafo, como apontado em.<sup>92</sup>

Com o grafo filtrado, podemos finalmente aplicar o formalismo do Laplaciano magnético. Especificamente, nos concentramos na técnica de mergulho pelas fases dos autovetores desse Laplaciano, conforme descrito na seção 2.1.5.1.

### 2.3.2 Clusterização de vértices e imagens

A solução aproximada para a Eq.2.21 permite associar a cada vértice do grafo uma posição em um espaço  $S^1$ . Por isso, em (Fanuel *et al.*, 2018) os autores discutem sobre como usar tal mergulho para performar algoritmos de clusterização ou visualização, o que eles denominam de *magnetic eigenmaps* (automapas magnéticas).

Contudo, os automapas magnéticos não podem ser aplicados no caso em que as arestas tem sinal ou cores. Por isso, estamos estudando como generalizar tais mergulhos para o caso de dígrafos com sinal e cores.

No presente trabalho, antes de buscar tal generalização iniciamos estudando como as propriedades de carga e modularidade afetam o mergulho definido pelos automapas magnéticos, o que não foi discutido em (Fanuel *et al.*, 2018). Para isso, estudamos tanto redes direcionadas artificiais como reais. Visando uma maior aplicação da teoria de deformação nos questionamos também em como usar o mergulho magnético para segmentar imagens. Pois no caso da teoria espectral de grafos não direcionados o vetor de Fielder é extensivamente usado para tal propósito.

Obviamente, para definir uma segmentação via mergulho magnético é necessário antes definir métodos de mapear imagens em um grafo direcionado. Para tal, propusemos e estudamos dois métodos para realizar tal mapeamento.

**Kernel:** Um método comumente utilizado para mapear imagens em grafos não direcionados, consiste em utilizar uma função kernel com domínio limitado a uma certa vizinhança do pixel.<sup>7</sup> O valor dessa função para dois pixels pertencentes a uma vizinhança define o peso da aresta relacionada. Isto é, para todo pixel  $u$  e para cada pixel na vizinhança de  $u$   $v \in \mathcal{N}ei(u)$ , podendo ser associado a aresta o valor de peso dado pelo kernel gaussiano,

$$w(u, v) = e^{-\frac{\|p(u)-p(v)\|_P^2}{\sigma_s} - \frac{|I(u)-I(v)|}{\sigma_I}}, \quad (2.47)$$

onde  $\sigma_s, \sigma_I > 0$ ;  $p$  é uma função de posição e  $I$  é a função de intensidade do pixel.

No presente trabalho incorporamos como direção da aresta  $(u, v)$  o sentido de crescimento da função intensidade do pixel. Isto é, para um  $\Delta I_{min} \geq 0$  Uma aresta  $(u, v)$  é adicionada usando os seguintes critérios:

- Se  $I(v) - I(u) > I_{min}$  a aresta vai de  $u$  para  $v$ .
- Se  $I(v) - I(u) < -I_{min}$  a aresta vai de  $v$  para  $u$ .
- Se  $|I(v) - I(u)| \leq I_{min}$ , a aresta é não direcionada.

**Gradiente:** (Casaca, Paiva e Nonato, 2011) propuseram usar o produto interno do gradiente da imagem (em escala de cinza)- calculado usando o filtro de Sobel - e o vetor de deslocamento entre os pixels. Utilizando de tal produto interno os autores definem a seguinte função peso entre os pixels  $u$  e  $v$

$$w(u, v) = \frac{1}{1 + \eta g^2(u, v)}, \quad (2.48)$$

onde  $\eta \geq 0$  e  $g$  é dado por

$$g(u, v) = \max \left\{ \nabla C(v) \cdot \vec{d}_{uv}, \nabla C(u) \cdot \vec{d}_{uv}, 0 \right\} \quad (2.49)$$

Note que a função  $g$  é simétrica,  $g(u, v) = g(v, u)$ , O que é essencial para garantir que os autovalores da matriz associada a  $w$  seja real e a segmentação pelo vetor de Fielder seja possível. Contudo, na formulação de deformação podemos analisar o embebimento gerado pela solução do problema apresentado pela Eq.2.21. Portanto, torna-se possível remover uma das restrições da Eq.2.49, obtendo

$$h(u, v) = \max \left\{ \nabla C(x) \cdot \vec{d}_{uv}, 0 \right\}. \quad (2.50)$$

### 2.3.3 Uma perspectiva sobre transições de fase em sistemas físicos via GSPs

Até este ponto, nossa exploração concentrou-se primordialmente na aplicação de problemas de sincronização de grupo e laplacianos deformados para auxiliar nossa compreensão e manipulação de dados organizados em redes complexas. Essa abordagem foi inspirada tanto por sistemas físicos, como o fenômeno de Aharonov-Bohm, quanto por sistemas computacionais, incluindo o processamento de sinais em biologia. No entanto, durante o desenvolvimento deste trabalho, surgiu uma pergunta intrigante: Será que poderíamos inverter a abordagem? Poderíamos empregar problemas de sincronização de grupo em grafos para aprimorar nossa compreensão de sistemas físicos?

Essa questão não foi levantada de maneira meramente retórica. Nesta seção, buscamos estabelecer conexões entre problemas de sincronização de grupo em grafos e sistemas físicos que sofrem transições de fase. Especificamente, nosso objetivo foi fornecer um novo

olhar sobre as transições de fase em sistemas discretos emergentes,<sup>96</sup> estudando os GSPs associados a eles.

Nas últimas décadas, transições de fases topológicas têm desafiado os físicos. A ausência de um parâmetro de ordem em sistemas onde ocorre tal transição é uma de suas características mais peculiares e provavelmente mais desafiadoras. Nesse sentido, as transições de fase topológica contrastam com as transições de fase por quebra de simetria. Por exemplo, para o modelo XY,<sup>97</sup> o teorema de Mermin-Wagner<sup>98-100</sup> garante que nenhuma transição de fase ocorra por quebra espontânea de simetria  $O(2)$ , contudo uma transição de fase por mudança de um invariante topológico pode ocorrer, e ocorre abaixo de uma certa temperatura crítica.

Nos últimos anos, uma abordagem para caracterizar transições de fase tem sido desenvolvida,<sup>101-103</sup> utilizando conceitos da ciência de redes. Esta nova abordagem baseia-se na construção de um grafo em que as arestas codificam alguma noção de correlação entre os elementos do sistema. Com o grafo construído, as quantidades usuais de redes complexas podem ser calculadas e, como mostram os resultados, essas quantidades podem ser usadas para identificar transições de fase. Este é um resultado notável devido ao fato de que essa abordagem é muito mais simples do que os métodos de aprendizado profundo. Além disso, não é um algoritmo de caixa preta. Neste trabalho, estamos mais intimamente relacionados com os objetivos deste tipo de abordagem. Ou seja, usar ferramentas comumente desenvolvidas na ciência de dados matemática, na esperança de obter novas percepções e interpretações para estudar sistemas mecânicos estatísticos.

Aqui, mostramos como os procedimentos de granulação grosseira utilizados em métodos de renormalização com o objetivo de caracterizar tais sistemas permite associar a eles um fluxo de GSPs. Também mostramos, por meio de simulação, que a quantidade que captura quão boa é a solução de relaxamento espectral para os GSPs pode ser usada como indicador para transições de fase. Essa discussão nos leva à próxima seção, que abordará como um processo de granulação grosseira e o fluxo de GSPs.

### 2.3.3.1 Granulação grosseira e conexão com GSPs via grafos de medidas

Consideremos  $S$  como um sistema discreto finito composto por  $V$  elementos com parâmetros externos fixos. A estrutura de dados que descreve a localização e a interação entre os elementos é um grafo,  $G_S$ , possivelmente geométrico. As interações entre os elementos são fornecidas pelos termos cruzados do Hamiltoniano associado, com estas interações capturadas pelas arestas de  $G_S$ .

Para ser mais didático, suponha o modelo de Ising 1D cujo Hamiltoniano para  $N$  spins é

$$H = J \left( \sum_{i=1}^{|V|} s_i s_{i+1} + s_1 s_{|V|} \right) + h \sum_{i=1}^{|V|} s_i, \quad (2.51)$$

onde  $s_i \in [-1, 1]$  é o spin do  $i$ -ésimo elemento;  $J$  e  $h \in \mathbb{R}$  são, respectivamente, o acoplamento e o campo magnético. Comumente, para modelos descritos por Eq.2.51, o procedimento de granulação grosseira usado em procedimentos de renormalização no espaço real é apenas a partição da vizinhança,

$$\begin{aligned} \mathcal{P} : V &\mapsto \mathcal{V}_1 \\ \{s_1, s_2, \dots, s_n\} &\mapsto \{\{s_1, s_2\}, \dots, \{s_{n-1}, s_n\}\}. \end{aligned} \quad (2.52)$$

No grupo de renormalização do espaço real (RG), cada subconjunto de spins particionados é mapeado em um novo spin efetivo. Por exemplo, para o modelo Ising 2d usando a granulação grosseira de blocos de spins, o novo spin efetivo pode ser escolhido usando uma regra de votação majoritária. O ponto principal do RG é encontrar pontos fixos do sistema usando aplicações sucessivas da granulação grosseira. Por exemplo, quando o procedimento de RG é aplicado em alguma configuração de spins extraída na temperatura crítica ( $T = T_c$ ), as aplicações sucessivas de RG mantêm a configuração invariante. Isso é esperado devido à propriedade de invariância de escala no ponto crítico.<sup>104</sup>

Suponha um sistema composto por graus de liberdade que podem ser descritos por elementos de um grupo,  $\mathcal{G}$ . Para cada par de elementos,  $u, v \in V$ , é possível associar uma medida em pares,  $\gamma(u, v) = \gamma(u, v)^{-1} = \gamma(v, u) \in \mathcal{G}$ , que quantifica o alinhamento entre os elementos. Por exemplo, para sistemas do tipo Ising, temos  $\gamma(u, v) = s_u s_v$ . Aqui, chamamos de grafo de medidas um grafo  $(V_m, E_m, w, \gamma)$  onde  $w : E_m \mapsto \mathbb{R}_+$  é a função de peso e  $\gamma : E_m \mapsto \mathcal{G}$  é uma função de potencial de aresta. A função de peso pode estar associada ao grau de confiança da medida em pares relacionada. Como veremos, o potencial de aresta pode ter um comportamento análogo quando comparado com potenciais físicos.

Aqui propomos usar os mapeamentos de granulação grosseira em cada passo para construir um grafo de medidas de ruído em pares. As medidas em pares são construídas amostrando um spin em cada bloco antes de mapear este bloco em um spin efetivo. Portanto, em cada passo de granulação grosseira, temos um GSP dado por Eq.2.15. Por causa disso, para cada sistema em cada configuração específica de parâmetros externos, associamos o seguinte fluxo com os sistemas considerados

$$\{V, \eta^{(0)}, f^{(0)}\} \xrightarrow{\mathcal{P}} \{V^{(1)}, \eta^{(1)}, f^{(1)}\} \dots \quad (2.53)$$

Usando simulações de Monte Carlo, investigamos como esses fluxos se comportam no ponto crítico.

O operador laplaciano deformado é hermitiano e positivo-definido. Portanto, associamos a cada sistema em cada passo do procedimento de granulação grosseira um conjunto de autovalores reais positivos,  $\Lambda^{(n)}$ , e um conjunto de autovetores normalizados,  $v_j^{(n)}$ . Além disso, este operador define um fluxo espectral para os sistemas considerados,

$$\{\Lambda^{(0)}, \{v_0^{(0)}, v_1^{(0)}, \dots\}\} \xrightarrow{\mathcal{P}} \{\Lambda^{(1)}, \{v_0^{(1)}, v_1^{(1)}, \dots\}\}. \quad (2.54)$$



### 3 RESULTADOS

#### 3.1 Pétalas e simetrias de rotação

Um dos aspectos que focamos no desenvolvimento desta tese foi a análise do espectro do Laplaciano magnético e sua relação com as estruturas de comunidade nas redes. As estruturas de comunidade, também conhecidas como clusters, blocos ou módulos.

Comunidades afetam o comportamento funcional de uma rede complexa tais como a propagação de uma informação falsa em bolhas de redes sociais. Assim, entender como as estruturas de comunidade afetam o espectro do Laplaciano magnético representa um outro passo para entender como dinâmicas se comportam em redes dirigidas.

Começamos estudando a influência da estruturas de blocos em redes direcionadas no espectro do Laplaciano magnético e seu impacto no calor específico, representado por  $c_\lambda$ . Nossa análise considerou redes com conexões dentro e entre as comunidades, representadas por  $\mathbf{W}_{\text{in}}$  e  $\mathbf{W}_{\text{out}}$ , respectivamente, sem fazer qualquer distinção entre os diferentes blocos. Desta maneira, a matriz de adjacência para redes consideradas é dada pela matriz abaixo:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{\text{in}} & \mathbf{W}_{\text{out}} & \mathbf{0}_{N_c} & \cdots & \mathbf{0}_{N_c} \\ \mathbf{0}_{N_c} & \mathbf{W}_{\text{in}} & \mathbf{W}_{\text{out}} & \cdots & \mathbf{0}_{N_c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{\text{out}} & \mathbf{0}_{N_c} & \mathbf{0}_{N_c} & \cdots & \mathbf{W}_{\text{in}} \end{bmatrix} \quad (3.1)$$

onde  $\mathbf{0}_{N_c}$  é uma matriz nula de tamanho  $N_c \times N_c$ .

O Laplaciano magnético no caso acima é:

$$\mathbf{H}_q = \begin{bmatrix} \mathbf{H}_{\text{in}} & \mathbf{H}_{\text{out}} & \mathbf{0}_{N_c} & \cdots & \mathbf{H}_{\text{out}}^\dagger \\ \mathbf{H}_{\text{out}}^\dagger & \mathbf{H}_{\text{in}} & \mathbf{H}_{\text{out}} & \cdots & \mathbf{0}_{N_c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{\text{out}} & \mathbf{0}_{N_c} & \mathbf{0}_{N_c} & \cdots & \mathbf{H}_{\text{in}} \end{bmatrix}, \quad (3.2)$$

A matriz  $\mathbf{H}_q$  é uma matriz de Toeplitz.<sup>105</sup> Portanto, os autovalores podem ser determinados considerando a propriedade de que todas as colunas da matriz original podem ser obtidas através de permutações cíclicas da primeira coluna.

É bem estabelecido<sup>41,105</sup> que os autovetores de uma matriz cíclica pode ser obtida através de

$$|\psi_u\rangle = \begin{bmatrix} |\phi\rangle \\ \rho_u |\phi\rangle \\ \vdots \\ \rho_u^{N_f-1} |\phi\rangle \end{bmatrix}, \quad (3.3)$$

onde  $u \in 0, \dots, N_f - 1$  e  $\rho_u = \rho_{N_f - u}^* = \exp(\frac{2\pi i u}{N_f})$ .

Em<sup>1</sup> demonstramos que, graças às propriedades acima, a seguinte equação para a  $u$ -ésima matriz  $N_f$  blocos é:

$$\tilde{\mathbf{H}}_u = \mathbf{H}_{\text{in}} + e^{\frac{2\pi i u}{N_f}} \mathbf{H}_{\text{out}} + e^{-\frac{2\pi i u}{N_f}} \mathbf{H}_{\text{out}}^\dagger. \quad (3.4)$$

No caso em que as conexões são uniformes dentro e entre os blocos, as matrizes intra e entre blocos são dadas por:

$$\mathbf{H}_{\text{in}} = \frac{\mathbf{I}_{N_c}(1 + d) - \mathbf{1}_{N_c}}{d}, \quad (3.5)$$

e

$$\mathbf{H}_{\text{out}} = -\frac{\exp(2\pi i q)}{2d} \mathbf{1}_{N_c} \quad (3.6)$$

respectivamente.

Como mostrado em,<sup>1</sup> a equação acima leva à seguinte equação de autovalores:

$$\begin{aligned} \lambda_{u,v} = 1 - \frac{\cos(2\pi(\frac{u}{N_f} - q))}{d} \\ + \frac{2}{d} \left( 1 + \cos(2\pi(\frac{u}{N_f} - q)) \right) f(v, N_c, m_c) + \Delta, \end{aligned} \quad (3.7)$$

onde  $f(v, N_c, m_c) = \sum_{l=1}^{m_c-1} \cos(\frac{2\pi v l}{N_c})$ , tal que

$$f(v, N_c, m_c) = \begin{cases} m_c & \text{if } v = 0, \\ \frac{\sin(\frac{\pi v m_c}{N_c})}{\sin(\frac{\pi v}{N_c})} \cos(\frac{\pi v}{N_c} (m_c - 1)) & \text{otherwise} \end{cases}. \quad (3.8)$$

A equação Eq.3.7 indica uma simetria de rotação relacionada ao parâmetro de carga na rede direcionada modular. Essas simetrias também refletem o comportamento da estrutura em forma de pétala do calor específico mostrado na Figura 9. Embora esse resultado possa parecer um caso muito especial, como mostramos em<sup>1</sup> a estrutura de pétalas se faz presente mesmo quando as conexões não são uniformes. Portanto, a estrutura de pétalas é uma característica do modelo de blocos abordado.

### 3.2 Medidas e comparação de grafos dirigidos em web escala

Um dos desafios mais significativos no desenvolvimento desta tese foi o cálculo das propriedades espectrais de grafos dirigidos grandes, que incluem centenas de milhares de vértices. A complexidade computacional deste cálculo, mesmo para matrizes esparsas, é da ordem de  $O(|V|^3)$ . Esta magnitude tornaria muitas medições espectrais de grafos inviáveis. Nos métodos e análises que desenvolvemos, é necessário calcular essas propriedades para

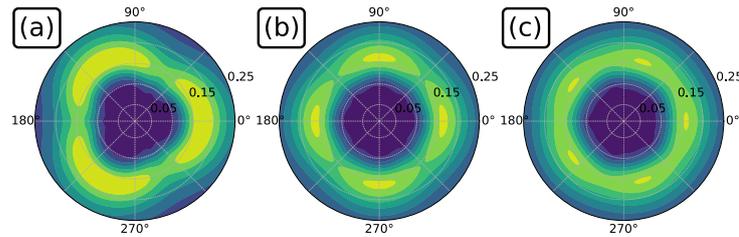


Figura 9 – Estrutura de pétalas do calor específico para redes direcionadas modulares. Calor específico em cores em termos da carga  $2\pi q$  (coordenadas polares) e temperatura (coordenada radial).

Fonte: RESENDE; COSTA.<sup>1</sup>

diversos valores de cargas. Portanto, o tempo necessário para calcular essas propriedades se tornaria impraticável, mesmo para a geração de uma figura de pétalas para um grafo com apenas algumas dezenas de milhares de vértices, mesmo utilizando algum cluster.

Para contornar essa limitação, buscamos alternativas que possibilitassem um cálculo eficiente. Estudos recentes propuseram abordagens utilizando a Quadratura de Lanczos Estocástica (Stochastic Lanczos Quadrature - SLQ) para grafos não dirigidos.<sup>70</sup> Este método tem permitido o cálculo e a comparação de redes não dirigidas em escala web, trazendo avanços consideráveis para o campo.

No entanto, nesta tese, decidimos ir um passo além e trazer outra conexão com a física para o estudo dos grafos dirigidos. Inspiramo-nos em um método utilizado em matéria condensada, o Método Polinomial do Kernel (Kernel Polynomial Method - KPM).<sup>2,106</sup> Graças à natureza hermitiana do Laplaciano magnético, pudemos utilizar esse método, apesar dos elementos das matrizes serem complexos.

O KPM se mostrou particularmente eficiente para calcular propriedades espectrais do Laplaciano magnético. Esta abordagem apresentou excelentes resultados, como demonstrado na figura 10 e no trabalho<sup>1</sup> obtivemos excelente acurácia na estimativa de medidas, como o calor específico, com poucos termos na aproximação.

O método KPM foi implementado em um pacote Python utilizando TensorFlow<sup>107</sup> e CuPy.<sup>108</sup> Na época, as únicas implementações de KPM open-source eram construídas em CPU e exclusivamente para sistemas quânticos, o que resultava em uma complexidade de uso e baixa performance. Por essa razão, avançamos no campo do software livre, criando uma solução open-source, o eMaTe Fig. 11, que permitiu a obtenção dos resultados apresentados. Este software robusto foi capaz de computar propriedades para grafos dirigidos e não dirigidos com milhões, ou até mesmo dezenas de milhões, de vértices. Desenvolvido com a utilização de TensorFlow e CUDA, o eMaTe foi concebido para se integrar a um ecossistema chamado stdog (Structure and Dynamics on Graphs). O objetivo era expandir ainda mais este projeto, mas não conseguimos reunir um número suficiente

de pessoas para trabalhar nele.

O eMaTe é um pacote Python implementado em TensorFlow e CuPy, com o objetivo principal de fornecer métodos eficientes capazes de estimar densidades espectrais e funções de traço de matrizes esparsas, sejam elas hermitianas (via transformação magnética) ou não. O desenvolvimento deste pacote foi fundamental para a publicação do trabalho disponível em.<sup>1</sup>

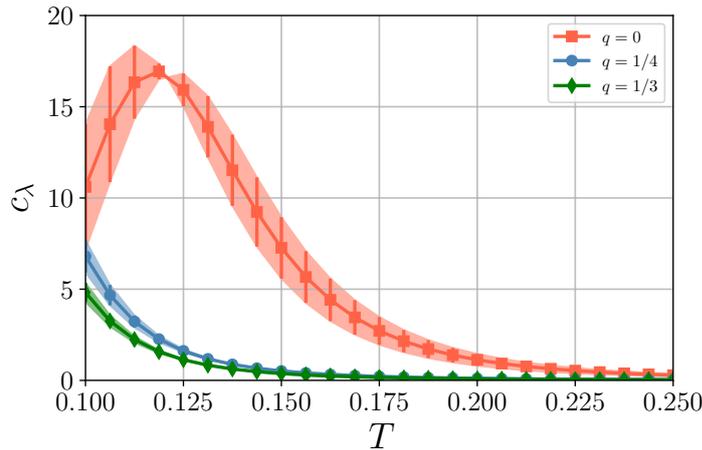


Figura 10 – Calor específico aproximado via KPM para uma rede com  $|V| = 3000$ ,  $N_f = 3$ ,  $p_c = 0.25$  and  $p_d = 0.5$  veja<sup>1</sup> para mais detalhes. As barras de erros representam o desvio obtido entre o valor exato (obtido numericamente via método iterativo) e o valor aproximado obtido via KPM e integração numérica.

Fonte: RESENDE; COSTA.<sup>34</sup>



Figura 11 – Logo do eMaTe. eMaTe é um pacote python implementado em tensorflow e cupy cujo objetivo principal é fornecer métodos úteis capazes de estimar densidades espectrais e funções traço de matrizes esparsas sejam elas hermitianas (via transformação magnética) ou não. O Desenvolvimento deste pacote foi essencial para publicação do trabalho disponível em.<sup>1</sup> <https://github.com/st-dogpkg/emate>.

Fonte: Elaborada pelo autor.

Para ilustrar a utilidade da abordagem KPM juntamente com Laplacianos magnéticos, apresentamos em<sup>1</sup> um método que, a partir do cálculo aproximado do espectro do

Laplaciano para vários valores de carga, permite a utilização das densidades espectrais para mensurar distâncias entre grafos dirigidos. Esse método mostra como é possível usar medidas espectrais para fazer comparações significativas entre grafos dirigidos, uma tarefa tipicamente desafiadora devido à natureza direcional das conexões.

Na figura 12, demonstramos o resultado de uma tarefa onde, dada a estrutura de um grafo de dois modelos específicos - Barabasi-Albert e Erdos-Renyi -, buscamos recuperar os parâmetros generativos do mesmo. Nesta tarefa, os parâmetros do modelo foram estimados minimizando a distância de Wasserstein<sup>109</sup> entre as densidades espectrais dos grafos modelados e o grafo alvo.

Surpreendentemente, utilizando apenas 100 termos na expansão do KPM e 20 vetores aleatórios, o mínimo da distância foi atingido corretamente. Este resultado mostra o potencial de aplicação dessa combinação de métodos. A precisão alcançada, mesmo com um número limitado de termos na expansão e vetores aleatórios, destaca a eficácia do KPM quando combinado com o uso de Laplacianos magnéticos para a análise de propriedades espectrais em grafos dirigidos.

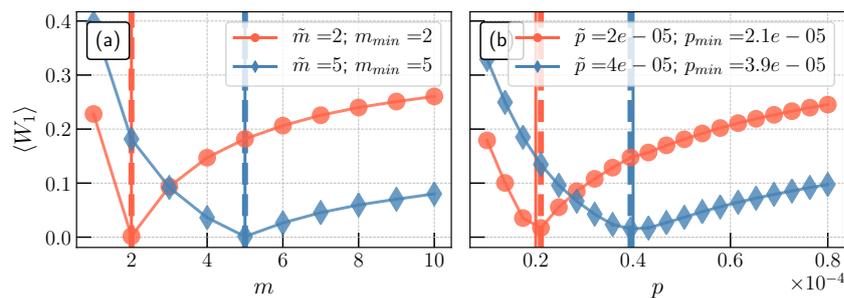


Figura 12 – Distâncias 1-Wasserstein Médias para Redes BA e ER Utilizando a Estimativa Espectral via KPM. As curvas em (a) e (b) representam a média das distâncias 1-Wasserstein, respectivamente para os modelos de redes BA (Barabási-Albert) e ER (Erdos-Rényi), considerando os parâmetros adotados na geração dessas redes. Esses resultados foram obtidos em redes com cem mil vértices. A estimativa espectral foi realizada utilizando o Método Polinomial do Kernel (KPM), com 100 termos na expansão do KPM e 20 vetores aleatórios para aplicação do “truque de Hutchinson”.<sup>2</sup> Calor específico aproximado via KPM para uma rede com

Fonte: RESENDE; COSTA.<sup>34</sup>

### 3.3 RMT para operadores deformados

Na seção de desenvolvimentos, discorreremos sobre os variados operadores que podem estar associados a grafos. Desta forma, cada grafo pode ser vinculado a múltiplos espaços espectrais possíveis associados ao operador Laplaciano Combinatório ou aos deformados.

Isso ressalta a necessidade de quantificar quão distintos esses espaços, operadores e quantidades podem ser.

Como observamos na Fig.13, a transformação magnética mapeia um ensemble de Ginibre para um GUE. É essencial enfatizar que isso não significa que os valores de entrada da matriz de adjacência magnética são *amostrados* de uma distribuição gaussiana do tipo  $\mathcal{N}(0, 1) + i\mathcal{N}(0, 1)$ . O fato de observarmos uma distribuição de espaçamento do GUE em Fig.13 é simplesmente uma constatação do caráter de universalidade de uma matriz complexa em que os valores de entrada são i.i.d. Por isso, torna-se interessante descobrir qual tipo de variável aleatória reside na matriz de adjacência magnética quando esta é oriunda de uma matriz de adjacência Gaussiana. Acreditamos que uma melhor compreensão de como tais transformações ocorrem pode auxiliar na exploração de aplicações de transformações magnéticas.

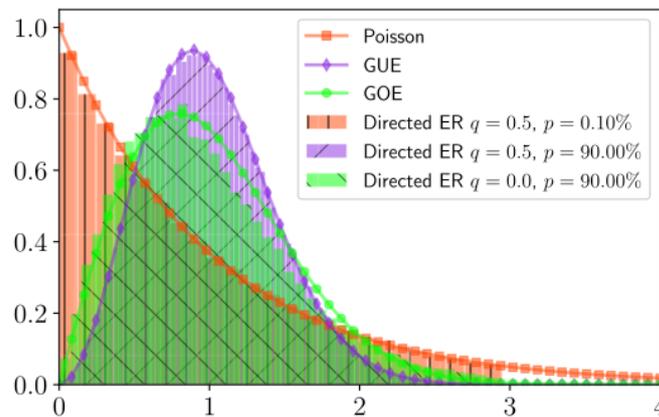


Figura 13 – Distribuição da razão de espaçamento de espectros de matrizes de adjacência magnética extraídas de grafos ER para diversos valores de probabilidade de conexão ( $p$ ) e conectividade ( $q$ ). Quando  $q = 0$ , a transformação magnética mapeia o ensemble de Ginibre para um ensemble GOE. Com  $q \neq 0$  o ensemble é mapeado em um com distribuição  $P(S)$  idêntica ao GUE. No limite em que a rede é totalmente desconectada,  $p = 0$ , obtêm-se a distribuição de Poisson, como esperado.

Fonte: Elaborada pelo autor.

Para determinar o resultado do mapeamento de um ensemble Gaussiano pela transformação magnética, precisamos compreender o efeito desta transformação em dois tipos de variáveis aleatórias. Primeiramente, as variáveis trigonométricas  $Y_{\mathbb{R}} \sim \cos \mathcal{N}(0, 2)$  e  $Y_{\mathbb{C}} \sim \sin \mathcal{N}(0, 2)$ . O segundo tipo é a variável aleatória resultante do produto de uma variável aleatória normal por uma trigonométrica, sendo  $X_{\mathbb{R}} \sim \mathcal{N}(0, 1)Y_{\mathbb{R}}$  e  $X_{\mathbb{C}} \sim \mathcal{N}(0, 1)Y_{\mathbb{C}}$ .

Para determinar a distribuição de probabilidade associada a cada função trigo-

nométrica, utilizamos o fato de que essas funções são periódicas. Portanto, é possível demonstrar que a distribuição de probabilidade associada a  $Y_{\mathbb{R}}$  apresenta a seguinte forma:

$$f_Z(z) \approx \frac{e^{-\frac{z^2}{4\sigma_x^2}} K_0\left(\frac{z^2}{4\sigma_x^2}\right)}{\sqrt{2\sigma_x^2\pi^{3/2}}} \quad (3.9)$$

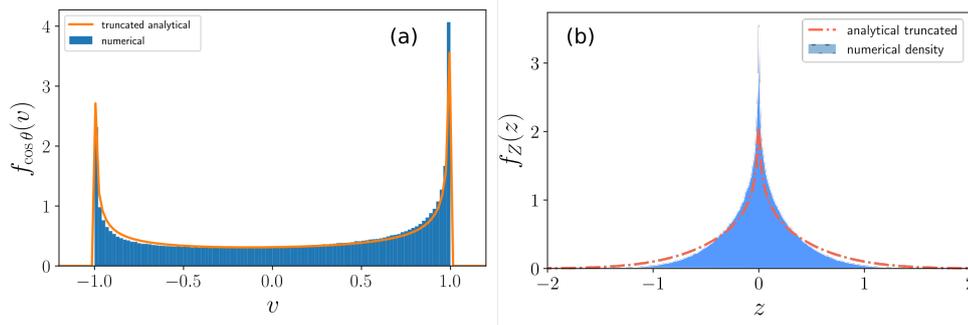


Figura 14 – Em (a) é apresentado a função densidade de probabilidade da variável aleatória  $\cos\mathcal{N}(0, 2)$  a qual está relacionada com o termo de fase gerado pela transformação magnética no operador de adjacência combinatório. Tal distribuição pode ser obtida via formalismo de estatística circulante. Em (b) é apresentado a densidade da variável aleatória  $\mathcal{N}(0, 1) \cos\mathcal{N}(0, 1)$  que representa a componente real dos valores da matriz de adjacência magnética. A expressão analítica para tal distribuição foi obtida via teorema de convolução para variáveis aleatórias.

Fonte: Elaborada pelo autor.

Os resultados numéricos e analíticos da expressão são apresentados na Fig.16(b). Note que as distribuições de valores seguem um comportamento de decaimento rápido, o que confirma que as entradas não são do tipo Gaussiana.

A Teoria das Matrizes Aleatórias (RMT), sob uma perspectiva magnética, demonstra um potencial notavelmente promissor para a análise de grafos direcionados. Esta perspectiva se revela especialmente relevante quando consideramos sua aplicação em redes gênicas. Nessas redes, os nós simbolizam genes e as arestas denotam interações entre eles. Estas interações, que podem ser de ativação ou inibição, são intrinsecamente direcionadas. A aplicação de técnicas de matriz aleatória a essas redes é tradicionalmente desafiadora, devido, em grande parte, à ausência de simetria hermitiana nas matrizes de adjacência dos respectivos grafos.

Exploramos uma dessas aplicações, onde a RMT magnética é empregada para estudar conjuntos de dados de grafos direcionados. Como as matrizes envolvidas são hermitianas, podemos utilizar técnicas de desdobramento de forma segura. Essas técnicas estão disponíveis em ferramentas como o `empyrcalrmt`.

Na Figura 15, é apresentado o resultado de um procedimento de desdobramento aplicado à matriz de adjacência magnética de uma rede gênica específica. A densidade dos dados observada sugere que as interações podem ser puramente aleatórias, dado que o resultado se alinha com um formato GUE.

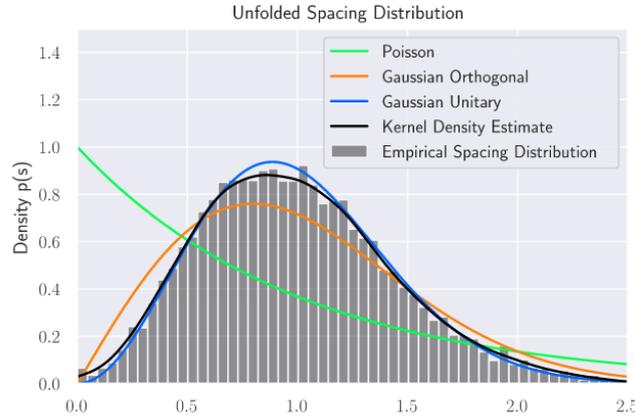


Figura 15 – Distribuição dos autovalores da matriz de adjacência magnética em uma rede gênica com a aplicação da técnica de desdobramento. Esta figura ilustra a distribuição dos autovalores da matriz de adjacência magnética após a aplicação da técnica de desdobramento em uma rede gênica. As curvas teóricas para as distribuições GOE (Gaussian Orthogonal Ensemble), GUE (Gaussian Unitary Ensemble) e Poisson são apresentadas para comparação. A coincidência do formato da distribuição com a curva GUE é uma forte indicação que as interações na rede serem puramente aleatórias.

Fonte: Elaborada pelo autor.

### 3.4 Transições de fase e GSPs

Nos resultados das seções anteriores, utilizamos a física para caracterizar problemas computacionais. Em particular, exploramos como a Teoria das Matrizes Aleatórias (RMT) e medidas termodinâmicas podem ser aplicadas na caracterização de grafos direcionados, como aqueles que representam redes biológicas. Nesse contexto, os resultados seguiram em uma via de mão única, do mundo da física para o domínio dos problemas da ciência de dados em grafos. No entanto, neste trabalho, também invertemos essa direção, conforme discutido na seção 2.3.3. Demonstramos como um problema computacional pode ser trazido de volta ao domínio da física, explorando como as transições de fase em sistemas discretos podem ser caracterizadas por meio dos GSPs. Nossa abordagem envolve a simulação de um modelo de Ising 2D em uma grade periódica de dimensão  $48 \times 48$ , utilizando o algoritmo de Metrópolis-Hastings. Para assegurar a termalização, realizamos  $10^4$  passos e extraímos  $10^3$  configurações de spins. Em seguida, aplicamos o procedimento de RG em um bloco de

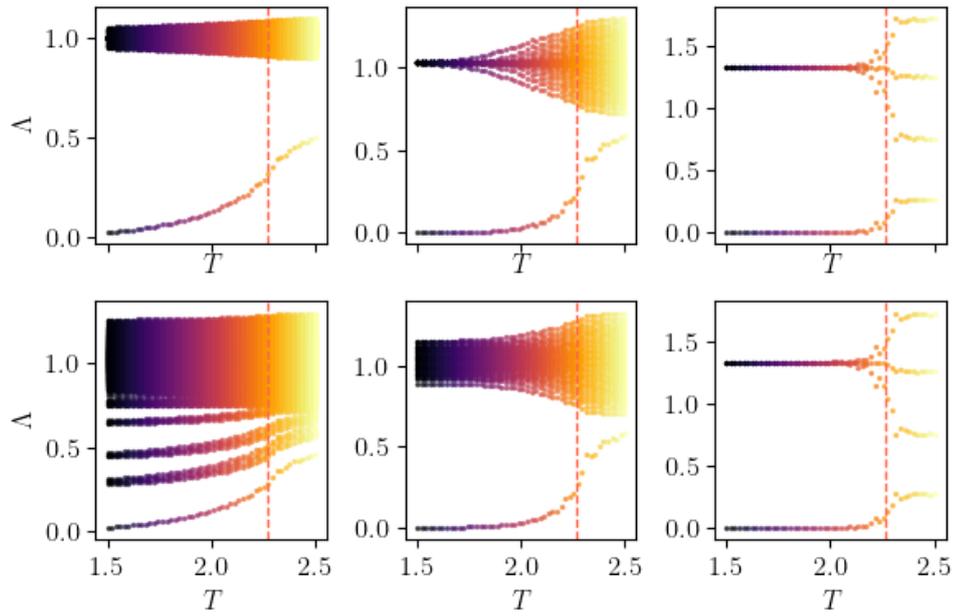


Figura 16 – Fluxo do espectro do Laplaciano deformado  $\mathbb{Z}_2$  para um modelo de Ising em uma grade periódica 2D de tamanho  $48 \times 48$  em função da temperatura. A primeira linha representa os espectros dos GSPs usando um grafo completo. A segunda linha representa os espectros usando um modelo de Watts-Strogatz 2D. Cada coluna representa uma etapa de renormalização para blocos de  $3 \times 3$ . As linhas verticais tracejadas representam a temperatura crítica associada ao modelo de Ising 2D no limite de  $N$ ,  $T_c \approx 2.269$ . As configurações de spins foram obtidas usando o algoritmo de Metrópolis-Hastings.

Fonte: Elaborada pelo autor.

dimensões  $3 \times 3$ . Geramos dois tipos de grafos para essa análise: um grafo completo e um modelo de rede 2D de Watts-Strogatz com 4 vizinhos.

Escolhemos o modelo de Ising em duas dimensões como sistema para prova de conceito, que podemos associar a Laplacianos deformados em  $\mathbb{Z}_2$ . O fluxo espectral para o grafo completo e para o Watts-Strogatz (esparso) é apresentado na figura 16. Para ambos os tipos de grafos, observamos uma característica interessante. Mesmo utilizando o modelo de rede esparsa para realizar as medições, ainda somos capazes de obter uma caracterização satisfatória da transição de fase. Isso é significativo, pois indica que não é necessário realizar um grande número de medições em pares para caracterizar a transição de fase.

Aqui, ao invés de utilizar conceitos e técnicas físicas para solucionar problemas de ciência da computação, estamos usando uma problemática computacional para lançar luz sobre um fenômeno físico.

### 3.5 Grafos, dados tabulares e remoção de atributos via Laplaciano Magnético

Na seção de metodologia, apresentamos um arcabouço concebido no decorrer desta tese, cuja finalidade é associar um grafo direcionado a um conjunto de dados tabular. Neste cenário, os vértices representam as colunas do mencionado conjunto de dado. Como ilustração prática desse modelo, escolhemos um conjunto de dados tabular derivado de um questionário como caso de estudo. Uma versão pre-print encontra-se disponível em.<sup>5</sup>

Dentro deste contexto, observa-se que as vivências durante a adolescência podem exercer uma influência significativa sobre a vida adulta. Este é um tópico que tem sido objeto de diversas pesquisas.<sup>110,111</sup> Notavelmente, temos a PeNSE (Pesquisa Nacional de Saúde do Escolar),<sup>112</sup> uma iniciativa conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em colaboração com o Ministério da Saúde e o Ministério da Educação. O objetivo desta pesquisa é compreender melhor os riscos à saúde e os padrões comportamentais entre os adolescentes brasileiros.

Com edições realizadas nos anos de 2009, 2012 e 2015, a pesquisa tem como público-alvo estudantes do nono ano, geralmente com quatorze anos, matriculados em escolas públicas e privadas de todo o Brasil. A escolha desse grupo etário alinha-se às diretrizes éticas internacionais para a aplicação de questionários socioeconômicos direcionados a adolescentes. Em nosso estudo, focamos nos dados da pesquisa de 2015, que contou com a participação de quase 130.000 estudantes em todo o território brasileiro. O questionário da pesquisa abrange uma variedade de áreas, incluindo: educação dos pais, profissão, posses; aspectos de saúde como saúde sexual, oral e mental; hábitos alimentares e fatores de risco; dinâmicas familiares e violência doméstica; e infraestrutura educacional. Um estudo prévio<sup>83</sup> explorou este dataset com o intuito de identificar correlações entre indicadores cruciais e perfis sociodemográficos. Como exemplo, os pesquisadores detectaram correlações entre hábitos alimentares saudáveis (avaliados pela frequência das refeições e pelo tipo de alimentos consumidos) e variáveis como idade, gênero e status socioeconômico. Contudo, tal estudo limitou-se a análises de regressão linear desses fatores.

Os resultados obtidos com a aplicação de nossa metodologia foram elucidativos. Como mostrado na Fig.17, obtemos uma boa concordância geral entre as comunidades obtidas por nosso arcabouço e a classificação qualitativa original das perguntas na pesquisa. Notavelmente, essa correspondência foi particularmente forte para os grupos Alimentação e Imagem Corporal. Isso sugere que nossa abordagem pode fornecer perspectivas valiosas sobre as categorias subjacentes dentro de um conjunto de dados complexo e permite avaliar de maneira não-supervisionada a classificação de perguntas em um questionário.

Entretanto, nossa análise também revelou algumas discordâncias interessantes. Por exemplo, constatamos que algumas perguntas classificadas originalmente na classe Segurança seriam melhor enquadradas na categoria Consumo de Drogas. Isso não apenas

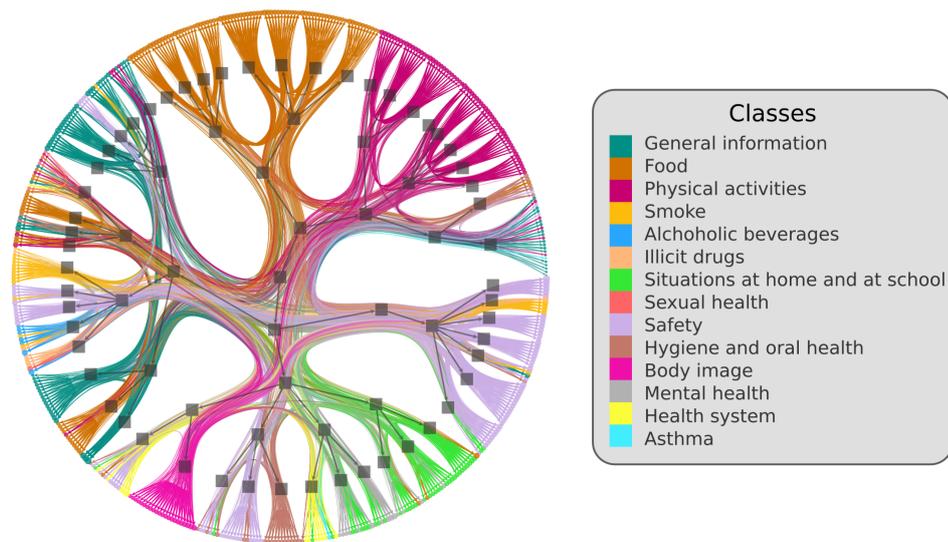


Figura 17 – Visualização hierárquica das comunidades de perguntas do questionário PeNSE 2015 obtida por nosso framework. Os vértices no círculo exterior representam as perguntas do questionário e as arestas direcionadas indicam as relações entre elas. A cor de cada vértice representa a classificação original da pergunta. A estrutura hierárquica sobreposta, obtida por meio da aplicação do Nested Stochastic Block Model,<sup>3,4</sup> revela as comunidades no grafo. Destaca-se uma boa concordância entre as comunidades identificadas e a classificação original das questões, sugerindo a eficácia do modelo proposto. Detalhes adicionais podem ser encontrados no artigo,<sup>5</sup> que expande além do escopo desta tese.

Fonte: RESENDE; TOKUDA; COSTA.<sup>5</sup>

desafia a classificação existente, mas também sugere a presença de conexões potencialmente ocultas entre diferentes fatores de risco.

Finalmente, um ponto interessante revelado pelo estudo foi a possível correlação entre os alimentos relacionados a comer com os pais e as perguntas da classe Situações em Casa. Essa observação sugere que a relação harmoniosa com os pais pode influenciar positivamente os hábitos alimentares dos adolescentes, um aspecto fundamental para o desenvolvimento saudável nesta fase da vida.

Os resultados obtidos na análise da PeNSE<sup>5</sup> podem parecer, à primeira vista, um tanto desconectados do foco central desta tese - a aplicação de grafos direcionados com ênfase em Laplaciano Magnético. Surge então a questão: por que desenvolvemos esse arcabouço e por que direcionamos nossa atenção para a análise de dados tabulares?

O cerne da questão é que a motivação inicial para este trabalho não se centrava em como analisar um conjunto de dados tabulares, mas sim em como extrair significado de um conjunto de relações de preferências entre diferentes conjuntos de dados. O que realmente implica dizer que as características A, B e C são cruciais para prever uma característica X?

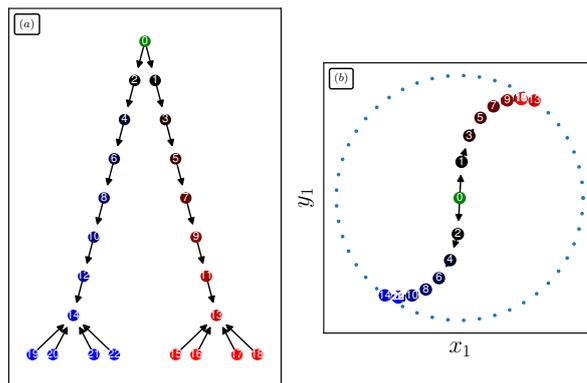


Figura 18 – Em (a), apresentamos um grafo direcionado estruturado com nódulos coloridos para facilitar a inspeção visual, com posições definidas manualmente. Em (b), o resultado do mergulho é exibido, com as fases do primeiro autovetor do Laplaciano Magnético ( $q = 1/3$ ) servindo como coordenadas polares e a coordenada radial determinada pelo ranking do Laplaciano de dilatação ( $g = 3/10$ ). Realizamos uma transformação para posicionar o vértice de maior rank no centro. Notavelmente, os nós de alta hierarquia estão mais próximos do centro do círculo e os braços do digrafo são mergulhados de forma a refletir a simetria do grafo original. Este exemplo simples serve para ilustrar como a aplicação de um mergulho usando as soluções dos problemas de frustração (GSP), considerando tanto a hierarquia quanto os ciclos em um grafo, pode ser um recurso útil na análise de estruturas direcionadas.

Fonte: Elaborada pelo autor.

E se também pudermos utilizar X, C e B para prever A? Como exposto anteriormente, nós modelamos essas relações através de um grafo direcionado, permitindo identificar grupos de colunas (ou questões) que possuem relevância mútua na predição entre si, mas que apresentam pouca influência ao prever colunas pertencentes a outros grupos. Em essência, estamos lidando com a dinâmica de fluxos e ciclos em nossos grafos direcionados.

Na seção 2.1.5.1, abordamos a aplicação do mergulho no espaço toroidal - determinado pelas fases dos autovetores do Laplaciano Magnético - como uma estratégia para entender melhor esses ciclos, uma abordagem discutida no artigo de Fanuel *et al.*, que emprega tais fases para resolver um problema análogo ao clássico Graph-CUT. No decorrer do desenvolvimento desta tese a seguinte questão surgiu: como poderíamos empregar esse mergulho no espaço toroidal para aprimorar nossa compreensão de um conjunto de atributos?

A fim de exemplificar como essa estratégia poderia ser aplicada para melhorar a nossa compreensão de um conjunto de atributos, consideramos a Fig18. Na sub-figura (a), apresentamos um grafo direcionado estruturado, com as cores dos nós usadas apenas para facilitar a inspeção visual e as posições ajustadas manualmente. Já na sub-figura (b), ilustramos os resultados do mapeamento para o espaço que combina a solução para o

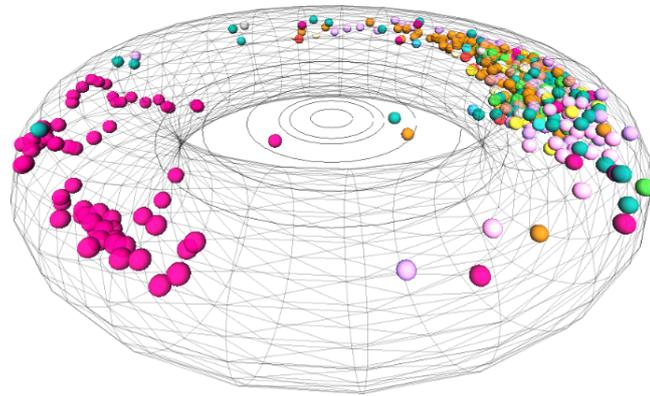


Figura 19 – Mergulho no espaço toroidal com informação do ranqueamento (radial) do grafo de interpretabilidade do PeNSE. As features com maior rank (conforme definido em<sup>5</sup>) estão mais próximas ao centro axial. À esquerda, observa-se um grupo bem definido de vértices rosa que representa questões da classe *Atividades Físicas*.

Fonte: RESENDE; TOKUDA; COSTA.<sup>5</sup>

problema de frustração magnética e dilatação. Utilizamos as fases do segundo autovetor do Laplaciano Magnético como coordenadas polares e a coordenada radial é determinada pelo ranqueamento do Laplaciano de dilatação, transformado para que o vértice com maior rank fique em 0.

Neste cenário, o mergulho Fig.18, embora seja um exemplo simplificado, revela com clareza a hierarquia e as relações simétricas inerentes ao grafo direcionado. Com os nós de alta hierarquia mais próximos ao centro e a preservação da simetria do grafo, “braços”, este exemplo oferece uma representação visual intuitiva do que esperávamos ao aplicar tal técnica ao nosso grafo de interpretabilidade do PeNSE e foi o que foi feito no trabalho Resende, Tokuda e Costa e é apresentado aqui na figura 19.

Comumente, para facilitar a visualização e análise de dados tabulares de alta dimensão, utilizam-se técnicas de redução de dimensionalidade, como t-SNE e UMAP.<sup>113,114</sup> Quando aplicamos esse processo ao questionário PeNSE, cada linha dos dados tabulares, que corresponde às respostas de um estudante, é representada como um vetor multidimensional. As técnicas mencionadas então projetam esses vetores para um espaço de dimensão reduzida, preservando, tanto quanto possível, as relações de proximidade entre eles. Para aumentar a clareza da visualização, pode-se ponderar as características com base em uma medida específica, como demonstrado em.<sup>115</sup> No contexto do nosso trabalho, optamos por uma medida de centralidade derivada do grafo de interpretabilidade - o hub score - para determinar essa ponderação.

As duas sub-figuras em 20 (parte superior) ilustram essas mergulhos feitas por

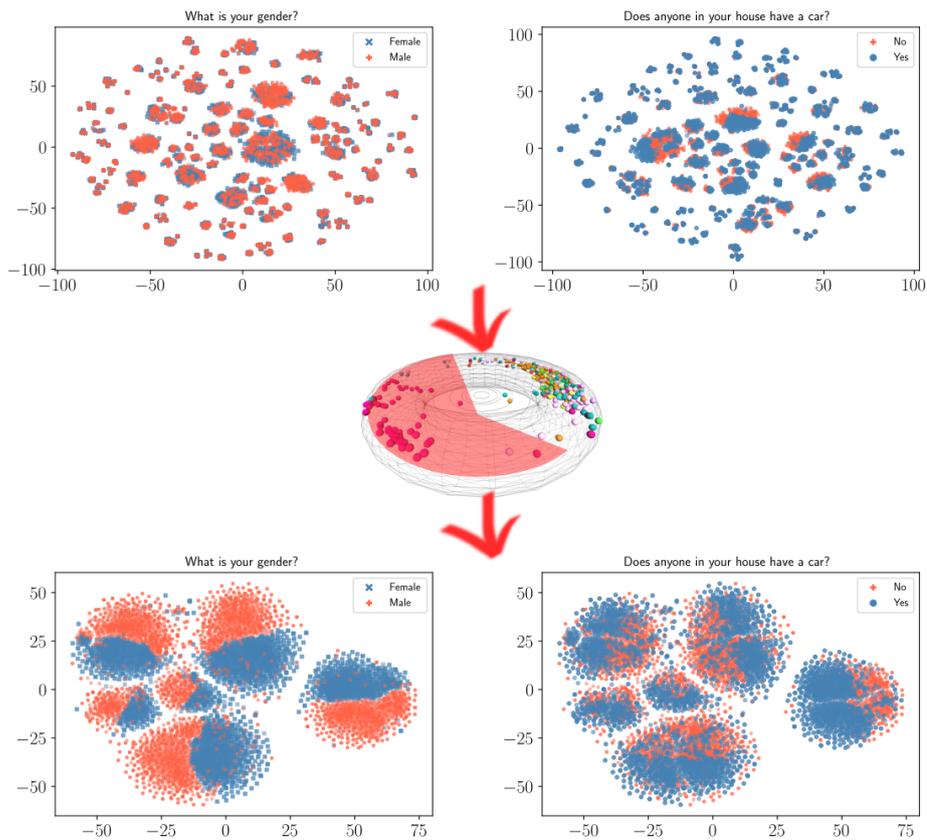


Figura 20 – Esta figura ilustra os mergulhos resultantes da pesquisa PeNSE após a remoção de features via Laplaciano Magnético ( $q = 1/10$ ), aprimorando a análise exploratória. As respostas são codificadas por cores de acordo com duas questões específicas: gênero e a posse de carro por alguém na família. As imagens superiores representam os mergulhos obtidos por meio do UMAP ponderado pelo hub score do grafo de interpretabilidade, utilizando todas as questões. As duas imagens inferiores, por outro lado, retratam os mergulhos após a remoção de um subconjunto de questões (excluindo aquelas marcadas em vermelho no meio da figura

Fonte: RESENDE; TOKUDA; COSTA.<sup>5</sup>

meio do UMAP. Em cada sub-figura, cada ponto no espaço de mergulho representa as respostas de um estudante, com a cor do ponto indicando uma resposta específica à pergunta indicada na imagem. No entanto, observamos que a ausência de agrupamentos claramente definidos sugere que a estrutura das respostas não é facilmente discernível, mesmo ao utilizar a ponderação proposta.

Como uma estratégia para superar esse obstáculo, usamos o Laplaciano magnético, que fornece informações sobre os ciclos em um grafo direcionado, para orientar a remoção de algumas perguntas. Baseado no espaço toroidal obtido do grafo de interpretabilidade, figura 19, dividimos nossos atributos (questões do PeNSE) em dois grupos. Ao reaplicar a técnica de mergulho com UMAP ponderado, conseguimos obter agrupamentos mais

claramente definidos, como pode ser observado na Figura 20 (parte inferior), especialmente em relação à pergunta sobre o gênero.

### 3.6 Além da Simetrização

O estudo do Laplaciano magnético, e, por extensão, das grandezas termodinâmicas associadas, foi um dos focos principais desta tese. Entretanto, um recente trabalho discutiu como o Graph Magnetic Neural Network, não obtêm resultados satisfatórios de classificação em alguns conjuntos de dados de grafos dirigidos. Aliado com esse resultado recente, o questionamento que era frequentemente levantado durante o desenvolvimento desta tese e a cada vez que o Laplaciano magnético era aplicado era: A caracterização de redes dirigidas via Laplaciano magnético e grandezas termodinâmicas associadas sempre fornecerá a informação necessária criada pela direção?

#### 3.6.1 Decomposição de HodgeRank e o impacto nas deformações magnéticas e de dilatação.

Para abordar esse questionamento, propomos uma maneira mais formal de avaliar essa questão do que apenas medir quando uma medida de acurácia aumenta ou diminui. Nossa abordagem para isso é aplicar a decomposição associada ao método de Hodge-Rank. Este método tem uma conexão com a deformação por dilatação como dito na seção metodologia e apresentado no trabalho de Fanuel e Suykens.

Tal decomposição permite que o fluxo definido pelas arestas dirigidas ( $\mathbf{A} - \mathbf{A}^T$ ) seja decomposto em fluxos gradiente, livre de rotação e harmônico.

Em cada um desses fluxos, cada aresta é associada a um valor real, positivo ou negativo. Portanto, é possível inverter o fluxo em cada aresta para representar um grafo dirigido. Na imagem 21 mostramos o resultado dessa decomposição para um grafo simples.

A imagem é organizada da seguinte maneira: no canto superior esquerdo, apresentamos um grafo direcionado. As imagens b, c, d e e na mesma linha ilustram os resultados da decomposição de HodgeRank do grafo direcionado em suas componentes curl-free, divergence-free, harmônica e rotacional, respectivamente. Na segunda linha, demonstramos o calor específico magnético para o grafo original e para cada uma de suas componentes. A última linha exibe o score obtido pelo laplaciano de dilatação para o grafo original e para cada uma das suas decomposições.

Destaca-se que, em algumas componentes, o resultado do Laplaciano magnético é informativo, enquanto o do Laplaciano de dilatação não é, e vice-versa. Isso sugere a necessidade de abordagens diferentes para obter uma compreensão completa da estrutura e das propriedades de um grafo direcionado.

Essa constatação leva a um dilema. Por um lado, podemos investir esforço na

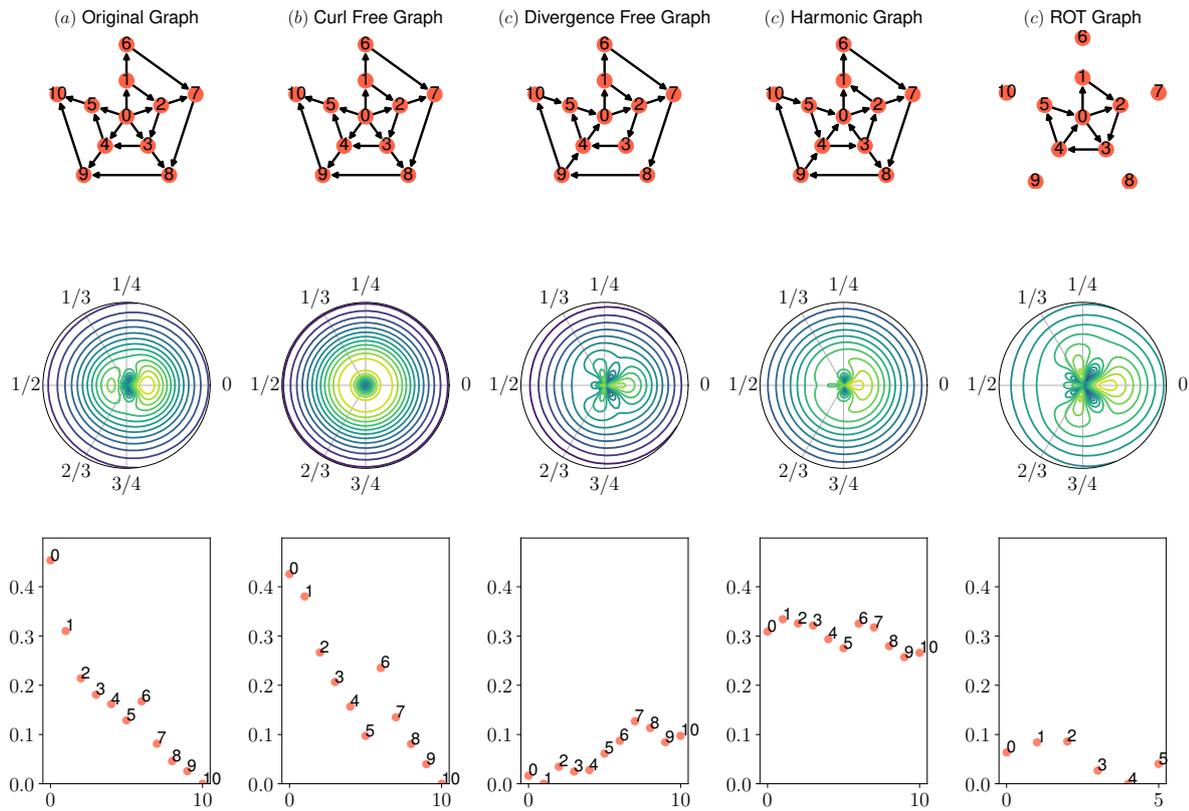


Figura 21 – Decomposição de Helmholtz-Hodge de um grafo direcionado e análise dos seus respectivos laplacianos magnético e de dilatação para cada componente. Na imagem localizada no canto superior esquerdo, é apresentado um grafo direcionado. Nas sequências ilustradas em b, c, d, e, na mesma linha superior, podemos observar os resultados da decomposição de HodgeRank do referido grafo direcionado, representados nas componentes curl-free, divergence-free, harmônica e rotacional, respectivamente. Na segunda linha, está demonstrado o calor específico magnético, tanto para o grafo original quanto para cada uma de suas componentes. Finalmente, na última linha, é exibido o score obtido pelo laplaciano de dilatação, novamente tanto para o grafo original quanto para cada decomposição do mesmo. É fácil notar que, em algumas componentes, o resultado do laplaciano magnético é informativo enquanto o do laplaciano de dilatação não é, e vice-versa

Fonte: Elaborada pelo autor.

definição de novas grandezas para o Magnético e a Dilatação, mas ainda assim precisaríamos decidir qual usar e também tornaria a necessidade de adaptar e implementar todas essas novas medidas e métodos. Esse dilema foi um dos motivadores para propomos nesta tese a definição de matrizes efetivas, como discutido na seção 2.2.1.

Essas matrizes, além de incluir parâmetros de escala e deformação, permitem combinar ambas as deformações para definir matrizes de adjacência reais e simétricas

para grafos dirigidos. Tal abordagem abre um leque de medidas e técnicas para grafos não dirigidos para grafos dirigidos e/ou com sinais.

A seguir, apresentaremos resultados que corroboram que esses grafos efetivos fornecem informação útil, demonstrando que não constituem meras abstrações matemáticas.

### 3.6.2 Aplicando propriedades de grafos não dirigidos em dirigidos.

Avançando em nossa discussão, a imagem 22 ilustra a criação de três grafos efetivos para três valores de carga diferentes e um parâmetro de dilatação fixo. Em (a), é apresentada uma matriz de adjacência de um grafo direcionado amostrado de um modelo de blocos. A simetrização dessa matriz,  $(\mathbf{A} + \mathbf{A}^T)$ , é apresentada em (b). Por fim, em (c), (d) e (e), são apresentadas diferentes matrizes de adjacência efetivas obtidas pela equação 2.32. Os valores de frustração ( $\eta$ ) e a carga para deformação magnética ( $q$ ) são indicados nos títulos de cada uma dessas imagens. Para todas as matrizes efetivas, o parâmetro de dilatação foi definido como  $0.1/(|V| - 1)$ , onde  $|V| = 500$ .

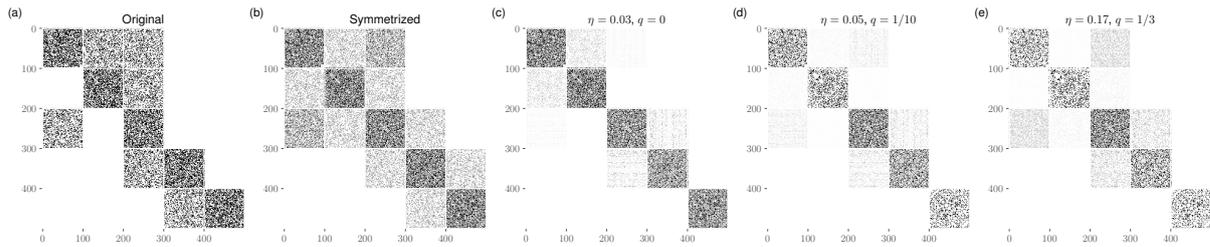


Figura 22 – Visualização de uma matriz de adjacência de um grafo direcionado e suas representações na forma simétrica e efetivas. Na figura acima, em (a), é apresentada uma matriz de adjacência de um grafo direcionado amostrado de um modelo de blocos. Em (b), é apresentada a matriz de adjacência obtida através da simetrização  $(\mathbf{A} + \mathbf{A}^T)$ . Em (c), (d) e (e), são apresentadas diferentes matrizes de adjacência efetivas obtidas pela equação 2.32. Nos títulos de (c), (d) e (e), é apresentado o valor de frustração ( $\eta$ ) e o valor da carga para deformação magnética ( $q$ ). Para todos os grafos efetivos, o parâmetro de dilatação foi definido como  $0.1/(|V| - 1)$ , onde  $|V| = 500$ .

Fonte: Elaborada pelo autor.

Conforme mencionado anteriormente, temos que os grafos efetivos possibilitam a aplicação de todas as medidas de grafos não direcionados em grafos direcionados. No entanto, é essencial investigar se essas medidas permanecem úteis quando aplicadas. Para isso, utilizamos o mesmo grafo direcionado usado para gerar a figura 22 e calculamos a *betweenness centrality* tanto para o grafo simetrizado quanto para o efetivo. A *betweenness* é uma medida amplamente utilizada devido à sua interpretação como uma medida do fluxo de informações em uma rede complexa.<sup>116</sup>

Essa medida, para um vértice  $v$ , pode ser calculada pela seguinte equação:

$$B(v) = \sum_{i,j} \frac{\sigma_{ij}(v)}{\sigma_{ij}}, \quad (3.10)$$

onde  $\sigma_{ij}$  é o número de menores caminhos conectando os vértices  $i$  e  $j$ , e  $\sigma_{ij}(v)$  é o número de menores caminhos que passam por  $v$  e conectam os vértices  $i$  e  $j$ .

Deve-se observar que a *betweenness* é mal definida em grafos direcionados, uma vez que a existência de um caminho não é necessariamente garantida, mesmo que o grafo seja fracamente conectado. Contudo, o mesmo não pode ser dito para o grafos efetivos definidos nesta tese.

Para analisar as relações entre os *betweenness* para o grafo simetrizado e efetivo, optamos por mergulhar o grafo direcionado original de acordo com a mesma abordagem discutida para o resultado apresentado em 18. Portanto, vamos formalizar a definição de um mergulho usando as fases dos autovetores do laplaciano magnético e a solução para o problema de dilatação (valor absoluto do primeiro autovetor), da seguinte forma:

$$\begin{aligned} \mathcal{C}_{q,g}^{(l)} : V &\mapsto S^1 \times \mathbb{R}^+ \\ \mathcal{C}_{q,g}^{(l)}(u) &= ((\angle v_q^{(l)})(u), |v_g^{(1)}(u)|) \\ &= (\theta_l(u), s(u)), \end{aligned} \quad (3.11)$$

onde  $v_q^{(l)}$  e  $v_g^{(0)}$  são o  $l$ -ésimo autovetor do Laplaciano magnético com uma carga  $g$  e o primeiro autovetor do laplaciano de dilatação com um parâmetro  $g$ , respectivamente.

A representação deste mapeamento em um gráfico 2D, como o apresentado na figura 18, é definida pela seguinte equação para um vértice  $u$ :

$$\mathcal{C}(u) = \left( 1 - \frac{s(u) - s_{\min}}{s_{\max} - s_{\min}} \right) (\cos(\theta_2(u)), \sin(\theta_2(u))), \quad (3.12)$$

onde  $s_{\max} = \max_{u \in V} s(u)$  e  $s_{\min} = \min_{u \in V} s(u)$ .

Na Figura 23, mergulhamos o grafo direcionado usado para produzir o resultado exibido na Figura 22. As cores utilizadas correspondem à medida de *betweenness* para a versão simetrizada e efetiva com  $q = 1/10$ . Optamos pelo mapa de cores viridis para esta representação. Como se pode observar, a medida de *betweenness* efetiva conseguiu capturar de forma eficaz a noção de hierarquia do grafo direcionado.

### 3.6.3 Grupo de Renormalização Deformado

O conceito de Grupo de Renormalização (RG) é um pilar fundamental na teoria de universalidade e transições de fase, conforme discutido na seção 3.4. Esse conceito

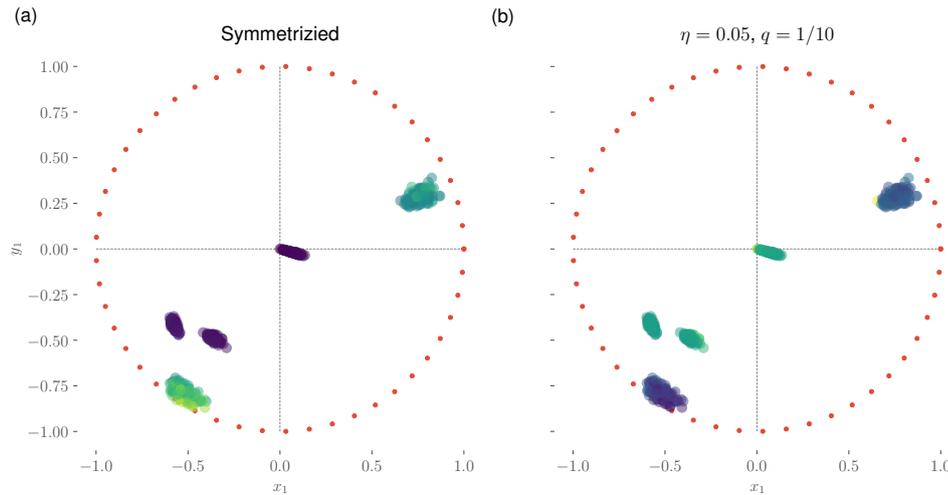


Figura 23 – Visualização de um grafo e sua representação no espaço que combina a solução para o problema de frustração magnética e dilatação. Cores representam betweenness obtido no grafo simetrizado e efetivo. A figura ilustra, em (a) e (b), o mergulho definido pela equação 3.12 do grafo direcionado utilizado para gerar os resultados da figura 22. As cores representam a betweenness para a versão simetrizada e efetiva ( $q = 1/10$ ) do grafo mencionado. Conforme pode ser observado, a medida efetiva de betweenness conseguiu captar de maneira efetiva a hierarquia no grafo direcionado.

Fonte: Elaborada pelo autor.

ganhou significativa relevância no estudo de redes complexas, pois os RGs funcionam como ferramentas para examinar simetrias e escalas em sistemas complexos. Entretanto, sua aplicação em redes complexas tem sido desafiadora devido à heterogeneidade dessas redes. Vale lembrar que os primeiros métodos de RG foram desenvolvidos para sistemas estatísticos discretos em grades, como Ising e Kondo.

Recentemente, foi definido o Grupo de Renormalização Laplaciano (RGL),<sup>117</sup> que introduziu uma nova perspectiva baseada na difusão em uma rede complexa. O RGL é capaz de identificar escalas espaço-temporais apropriadas em redes heterogêneas, representando um avanço em relação a métodos anteriores, como os baseados em vizinhança. As técnicas tradicionais de RG, incluindo o RGL, mostraram-se altamente eficazes na análise de simetrias e escalas organizacionais em grafos não direcionados. Recentemente, demonstrou-se como essa técnica pode aprimorar a qualidade de Redes Neurais de Grafos.<sup>118</sup> Uma generalização para hipergrafos (*high-order networks*) foi proposta em,<sup>119</sup> enquanto o trabalho de um grupo brasileiro<sup>120</sup> formalizou a abordagem do RGL utilizando o formalismo da teoria dos campos, amplamente empregado em teoria quântica de campos. No entanto, todos esses métodos de RG não são aplicáveis em contextos mais gerais, como grafos direcionados e/ou com sinais.

Logo, existe uma demanda para o desenvolvimento de uma extensão das técnicas

do grupo de renormalização que possa considerar as características únicas dos grafos direcionados. Tal extensão abriria caminho para uma análise mais completa de uma ampla gama de sistemas complexos e *datasets*, já que muitas redes do mundo real são, de fato, direcionadas. Aqui, discutiremos brevemente como podemos explorar RG utilizando o formalismo de grafos efetivos e, além disso, mostraremos que essa nova abordagem também pode ser usada para reforçar que tal formalismo de grafos efetivos contém mais informações do que apenas a simples simetrização.

Antes de prosseguirmos para a apresentação de nossa versão do Grupo de Renormalização, aqui denominada RGEG, é relevante abordar um procedimento de granulação grosseira que discutimos anteriormente em 2018, conforme apresentado em.<sup>53</sup> Este procedimento, que tinha como objetivo principal estabelecer uma conexão entre o espaço toroidal (seção 2.1.5.1) e a grandeza do calor específico do Laplaciano magnético (equação 2.35). A inspiração para este procedimento veio do fato que, para certos sistemas físicos discretos, dos quais o Ising 2D é um exemplo, o calor específico é uma propriedade crucial. No ponto de transição de fase, ocorre uma divergência deste.

Para estudar a relação entre uma grandeza  $\mathcal{M}$  (que pode ser o calor específico) em um grafo  $G^{(0)} = (V, E, w)$  e o espaço toroidal, primeiramente, mapeamos  $G$  em  $\mathbb{T}^2$ , de tal modo que cada vértice  $u \in V$  seja definido por um ponto  $\mathbf{r}_u^{(0)} \in \mathbb{T}^2$ . Assim, o resultado do mapeamento é um conjunto de pontos  $r_u^{(0)}$ , conforme ilustrado na Figura 24(e).

Em seguida, definimos um procedimento de granulação similar ao modelo de blocos.<sup>?,?</sup> Após calcular a grandeza  $\mathcal{M}$  em  $G^{(0)}$ ,  $\mathcal{M}^{(0)}$ , e mapear o grafo,  $r_u^{(0)}$ , agrupamos os vértices em super-vértices de tal modo que a soma das distâncias (utilizando a métrica euclidiana) entre seus elementos constituintes seja minimizada. Cada super-vértice representa um vértice de um novo grafo  $G^{(1)} = (\frac{|V|}{2}, E^{(1)}, w^{(1)})$ , conforme exemplificado na Figura 24(c), onde  $E^{(1)}$  e  $w^{(1)}$  podem ser definidos de forma a preservar uma propriedade específica da rede original. Escolhemos definir as arestas entre os super-vértices  $l$  e  $m$  no passo  $t + 1$  da granulação da seguinte forma:

$$w(l, m, t + 1) = \begin{cases} 1, & \text{if } \sum_{\substack{u \in l \\ v \in m}} w(u, v, t) \geq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (3.13)$$

Em seguida, definimos um procedimento de granulação similar ao modelo de blocos.<sup>?,?</sup> Após calcular a grandeza  $\mathcal{M}$  em  $G^{(0)}$ ,  $\mathcal{M}^{(0)}$ , e mapear o grafo,  $r_u^{(0)}$ , agrupamos os vértices em super-vértices de tal modo que a soma das distâncias (utilizando a métrica euclidiana) entre seus elementos constituintes seja minimizada. Cada super-vértice representa um vértice de um novo grafo  $G^{(1)} = (\frac{|V|}{2}, E^{(1)}, w^{(1)})$ , conforme exemplificado na Figura 24(c), onde  $E^{(1)}$  e  $w^{(1)}$  podem ser definidos de forma a preservar uma propriedade

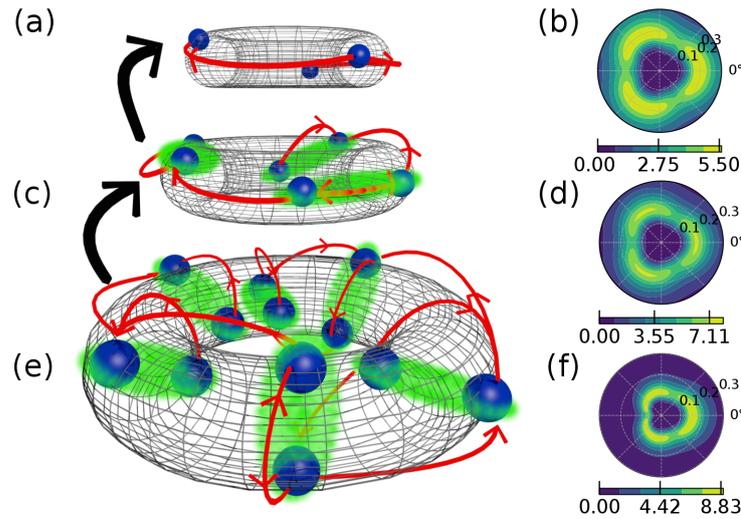


Figura 24 – Processo de granulação grosseira definida via espaço toroidal e evolução do calor específico. Em (a), (c) e (e), está representado o procedimento de granulação grosseira definido no espaço toroidal, descrito na seção 2.1.5.1, para uma rede dirigida com três módulos. Esta rede foi gerada com o mesmo modelo utilizado para obter os resultados apresentados na figura 9. Os calores específicos para cada etapa do procedimento de granulação grosseira são apresentados em (f), (d) e (b). Observe que a estrutura de pétalas é preservada pelo procedimento de granulação grosseira.

Fonte: RESENDE; COSTA.<sup>53</sup>

específica da rede original. Escolhemos definir as arestas entre os super-vértices  $l$  e  $m$  no passo  $t + 1$  da granulação da seguinte forma:

$$\left. \begin{array}{l} G^{(0)} \\ \mathcal{M}^{(0)} \\ \{r_u^{(0)}\} \end{array} \right\} \mapsto \left. \begin{array}{l} G^{(1)} \\ \mathcal{M}^{(1)} \\ \{r_u^{(1)}\} \end{array} \right\} \mapsto \dots \mapsto \left. \begin{array}{l} G^{(N)} \\ \mathcal{M}^{(N)} \\ \{r_u^{(N)}\} \end{array} \right\}, \quad (3.14)$$

Decidimos implementar o procedimento de granulação, detalhado anteriormente, no modelo de blocos que originou o resultado das pétalas na figura ???. Essa escolha se baseia na simplicidade do modelo e em nosso conhecimento da expressão analítica em certos casos, permitindo uma comparação eficaz com a abordagem do RGEG que iremos abordar posteriormente.

Como ilustrado na Figura 24, os subgráficos (f), (d) e (b) apresentam o mapa bidimensional do calor específico para cada grafo, obtido em cada etapa do procedimento de granulação grosseira implementado no espaço  $\mathbb{T}^2$ . É importante salientar que a estrutura de pétalas se preserva durante todo o procedimento de granulação, resultado que se deve à simplicidade do grafo selecionado. Naturalmente, ao nos limitarmos a apenas duas

dimensões, é improvável que obtenhamos um processo de granulação robusto para redes altamente heterogêneas. No entanto, este modelo simples serve como uma referência útil para a discussão subsequente sobre o RGEg, cujo procedimento é descrito no quadro abaixo.

Um grafo direcionado,  $G$ , um valor de carga,  $q$ , o parâmetro de dilatação  $g$ , parâmetro de escala  $\beta$  e uma estratégia  $\mathcal{S}$  para solução do problema de frustração associado, um método de esparsificação (remoção de arestas)  $\mathcal{E}$  e finalmente um  $\mathcal{G}$  são necessários.

- O Laplaciano deformado é construído com os parâmetros  $g$  e  $q$ ;
- A frustração é obtida através da equação 2.31 e estratégia  $\mathcal{S}$ ;
- O peso efetivo é calculado para cada aresta usando a equação 2.32, formando um grafo não direcionado  $G_U$ ;
- O grafo  $G_U$  é esparsificado através de  $\mathcal{E}$ ;
- $|V|//2$  pares de vértices selecionados por  $\mathcal{G}$  aplicado em  $G_U$  são contraídos em  $|V|//2$  super-vértices.
- Um grafo direcionado  $\tilde{G}$  com  $|V|//2$  vértices é criado. Uma aresta entre dois vértices é estabelecida se existia pelo menos uma aresta entre os vértices originais.

O RGEg, como descrito no quadro acima, é extremamente flexível. No entanto, para realizar um estudo de caso comparativo com a granulação utilizada no resultado da figura 24, selecionamos  $\mathcal{E}$ ,  $\mathcal{S}$  e  $\mathcal{G}$  da maneira mais simples possível.

Para  $\mathcal{E}$ , optamos pelo filtro de disparidade que nós utilizamos em<sup>5</sup> e que foi implementado na biblioteca de código aberto `edgeseraser`.<sup>121</sup> Para  $\mathcal{S}$ , o processo é descrito no quadro abaixo.

- Construímos um Laplaciano magnético normalizado com carga  $q$ ;
- Extraímos as fases do primeiro autovetor do Laplaciano magnético e as armazenamos em  $\theta_1$ ;
- Construímos um Laplaciano de dilatação com o parâmetro  $g$ ;
- Obtemos uma solução aproximada para o problema de frustração da dilatação, representada pelo vetor formado pela norma de cada componente do primeiro autovetor do Laplaciano de dilatação,  $\mathbf{s}$ ;
- Construímos o vetor solução cuja componente  $u$  é dada por:  $(e^{iq\theta_1} \mathbf{s})(u)$ .

Finalmente, para  $\mathcal{G}$ , escolhemos a pseudo-inversa do Laplaciano combinatório, conforme discutido em<sup>120</sup> para redes complexas não dirigidas. Assim, os  $|V|//2$  pares de vértices a serem contraídos representam uma noção de correlação, ou seja, os super-vértices serão formados pelos pares de vértices mais correlacionados.

Com a implementação do RGEN para o grafo direcionado, realizamos uma amostragem do modelo de blocos, mantendo os mesmos parâmetros utilizados na geração do grafo para a obtenção dos resultados na figura 24. Isto é, um grafo com estrutura de ciclos com  $N_f = 3$  blocos, onde cada bloco não direcionado possui uma probabilidade de conexão  $p_c = 50\%$  entre os vértices dos blocos e conexões direcionadas entre blocos com probabilidade  $p_d = 70\%$ , conforme escolhido em.<sup>53</sup>

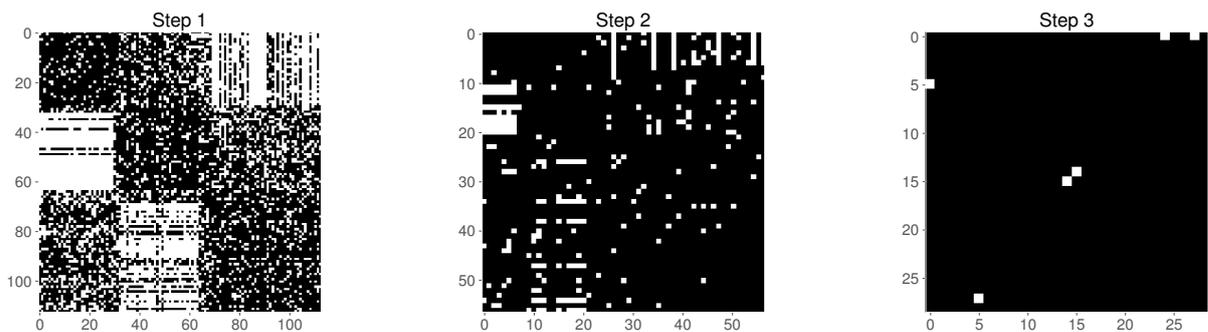


Figura 25 – Evolução da matriz de adjacência obtida pelo processo de RG para o Grafo Simetrizado. O processo de RG eliminou toda a informação de direção. Três passos de renormalização aplicados na versão simetrizada de um grafo direcionado com três blocos em ciclo, conforme discutido em.<sup>1</sup> Note o processo de homogeneização indicando que a abordagem de RG não se aplica adequadamente ao grafo direcionado usando a versão simétrica.

Fonte: Elaborada pelo autor.

A figura 25 apresenta o resultado do RG considerando apenas o grafo simetrizado,

isto é, é equivalente a setar  $q = g = 0$ . É importante destacar que, já no terceiro passo, toda a informação sobre a direção foi perdida, uma vez que o grafo se homogeneizou em uma rede totalmente conectada. Portanto, isso mostra que a abordagem de RG não é útil se utilizada a versão simétrica do grafo, isto é,  $\mathbf{W} + \mathbf{W}^T$ .

A figura 26 apresenta o resultado do RGEg, ao utilizar os parâmetros  $q = 1/10$  e  $g = 0$ . É possível observar que o processo de RGEg fortalece a estrutura de blocos e a direção dos ciclos do modelo. No quarto passo, temos um grafo direcionado que poderia ter sido obtido com os parâmetros  $N_f = 3$ ,  $p_d = 100\%$  e  $p_c = 100\%$ . O que corroborará com o resultado obtido pela granulação grosseira que é apresentado na imagem 24. Portanto, estes resultados são um indicativo promissor de que podemos explorar o RGEg em outras aplicações futuras. Em particular, seria interessante verificar se, assim como o RG em grafos não direcionados pode aprimorar os resultados de GCN, como mostrado por,<sup>118</sup> o mesmo ocorre com o RGEg proposto aqui.

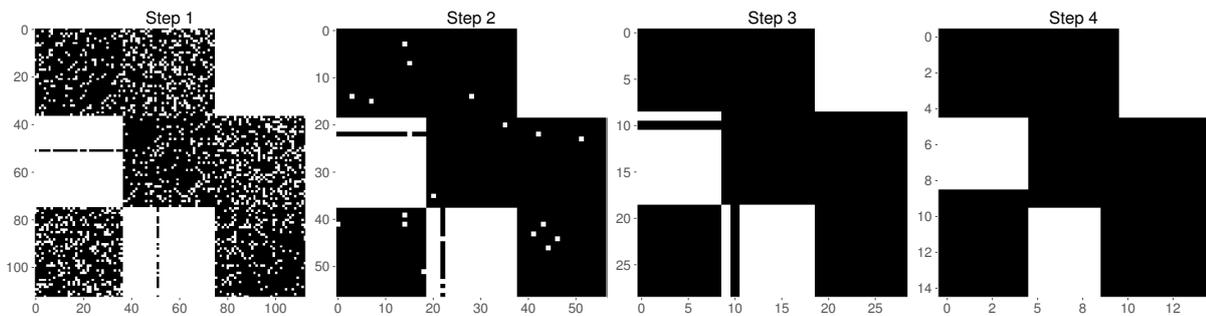


Figura 26 – Evolução da matriz de adjacência obtida pelo processo de RGEg para o Grafo Efetivo. Quatro passos de RGEg aplicados utilizando  $q = 1/10$  e  $g = 0$  em um grafo direcionado com três blocos em ciclo, conforme discutido em.<sup>1</sup> O processo de RGEg reforça a estrutura de ciclos e preserva a direção.

Fonte: Elaborada pelo autor.

## 4 CONCLUSÃO

Esta tese, se aprofundou na complexa interação entre grafos, grupos e termodinâmica. Enquanto motivada pelo estudo de grafos dirigidos e/ou com pesos, a respectiva abordagem através das operações de grupo emergiu como o principal pilar de nossa investigação, lançando uma nova luz sobre as propriedades dos grafos com arestas com propriedades enriquecidas, assim como pela sua aplicabilidade na caracterização de conjuntos de dados.

Realizamos avanços significativos na caracterização, comparação e identificação de simetrias termodinâmicas, especialmente ao desvendar as propriedades do espectro do Laplaciano magnético em redes direcionadas modulares. Conseguimos obter expressões analíticas através do formalismo de matrizes circulantes, revelando a estrutura de pétalas no calor específico. O Método Polinomial do Kernel (KPM) provou ser uma ferramenta indispensável para superar desafios computacionais, permitindo a análise de grandes grafos direcionados. Adicionalmente, a Teoria das Matrizes Aleatórias (RMT) se mostrou particularmente útil na análise de grafos magnéticos e conjuntos de dados de redes gênicas. Em particular, observa-se que conseguimos derivar expressões analíticas para os elementos das matrizes de adjacência magnética, utilizando o formalismo de estatística circulante.

A discussão do método de Hodge-Rank, apresentado na seção "Além da Simetrização", decompõe o fluxo definido pelas arestas dirigidas em fluxos gradiente, livre de rotação e harmônico, proporcionando uma visão mais aprofundada das características e propriedades de redes dirigidas e sua representação por meio de grafos efetivos.

A extensão do Grupo de Renormalização (RG) para grafos direcionados e o estabelecimento do Grupo de Renormalização de Grafos Efetivos (RGEF) forneceram novos métodos para examinar simetrias e escalas em sistemas complexos que incluem direções. Este é um campo que tem sido um desafio devido à heterogeneidade das redes complexas.

Como uma contribuição adicional, desenvolvemos o arcabouço Tab2Graph, que possibilita a associação de um grafo direcionado a um conjunto de dados tabulares, conforme exemplificado pelos dados do Censo Escolar PeNSE de 2015. Nesse contexto, o conceito de espaço de frustração, através do mergulho toroidal (GSP magnético), foi fundamental para aprimorar a análise exploratória dos dados do PeNSE, proporcionando uma compreensão mais aprofundada do questionário.

Exploramos também como problemas de sincronização de grupo podem ser aplicados para investigar fenômenos físicos. Estabelecemos uma conexão entre esses problemas, intimamente relacionados aos Laplacianos deformados por grupos, e a física estatística, particularmente em como o fluxo espectral pode efetivamente caracterizar as transições de

fase, ilustrado pelo modelo de Ising 2D.

Um aprendizado de particular importância que esta tese busca transmitir é que antes de se optar por uma simetrização ou mesmo por um Laplaciano deformado (uma abordagem moderna), é essencial entender o significado semântico da direção, do sinal, entre outros aspectos presentes nas arestas do problema em questão.

Reconhecemos que nossa pesquisa possui limitações, como a falta de exploração da Teoria das Matrizes Aleatórias para dilatação e sinal, e a ausência de investigação sobre como a heterofilia se manifesta nas matrizes de adjacência e Laplaciana deformadas. Para futuras pesquisas, sugerimos a exploração de novas deformações de grupo, o desenvolvimento de novas métricas de rede baseadas em grafos efetivos, e a utilização do grafo de interpretabilidade como entrada para técnicas de aprendizado de máquina.

Em conclusão, esta tese proporcionou uma exploração substancial do domínio onde os grafos, os grupos e a termodinâmica se intersectam, em particular sob a perspectiva de arestas dirigidas e/ou com pesos. Os desenvolvimentos e abordagens aqui descritas contribuíram para o avanço do conhecimento na área, abrindo caminho para futuras pesquisas e aplicações no aprendizado de máquina em grafos e na análise de redes complexas, com um enfoque especial na semântica do que direção e sinal representam em cada problema.

## REFERÊNCIAS

- 1 RESENDE, B. M. F.; COSTA, L. da F. Characterization and comparison of large directed networks through the spectra of the magnetic Laplacian. **Chaos: an interdisciplinary journal of nonlinear science**, v. 30, n. 7, p. 073141, July. 2020. DOI: 10.1063/5.0006891.
- 2 HUTCHINSON, M. A Stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. **Communications in Statistics - simulation and computation**, Taylor & Francis, v. 18, n. 3, p. 10591076, Jan. 1989.
- 3 PEIXOTO, T. P. Hierarchical block structures and high-resolution model selection in large networks. **Physical Review X**, American Physical Society, v. 4, n. 1, p. 011047, Mar. 2014. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevX.4.011047>. Acesso em: 16 jun. 2023.
- 4 MORELLI, L.; GIANANTI, V.; CITTARO, D. Nested stochastic block models applied to the analysis of single cell data. **bioRxiv**, Apr. 2021. Disponível em: <https://www.biorxiv.org/content/10.1101/2020.06.28.176180v3>. Acesso em: 15 jun. 2023.
- 5 RESENDE, B. M. F. de; TOKUDA, E. K.; COSTA, L. da F. **Unraveling the graph structure of tabular data through Bayesian and spectral analysis**. Disponível em: <http://arxiv.org/abs/2110.01421>. Acesso em: 16 jun. 2023.
- 6 LANDAU, E. Zur relativen wertbemessung der turnierresultate. **Deutsches Wochenschach**, v. 11, p. 366–369, 1895.
- 7 LANDAU, E. über Preisverteilung bei Spielturnieren. **Zeitschrift für Mathematik und Physik**, v. 64, 1915.
- 8 BROIDO, A. D.; CLAUSET, A. Scale-free networks are rare. **Nature Communications**, Nature Publishing Group, v. 10, n. 11, p. 1017, Mar. 2019. ISSN2041-1723.
- 9 COSTA, L. D. F. *et al.* Characterization of complex networks: a survey of measurements. **Advances in Physics**, v. 56, n. 1, p. 167242, Jan. 2007.
- 10 LESKOVEC, J.; HUTTENLOCHER, D.; KLEINBERG, J. Signed networks in social media. *In: INTERNATIONAL CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS - CHI '10*. 28., 2010, Atlanta. **Proceedings** [...]. Atlanta: ACM Press, 2010. p. 1361.
- 11 ZHENG, Q.; ZHENG, Q. Social networks with Rich Edge semantics. Boca Raton: Taylor & Francis, CRC Press, 2017. (Chapman & hall/crc data mining and knowledge discovery, 30).
- 12 PADGETT, J. F.; ANSELL, C. K. Robust action and the Rise of the Medici, 1400-1434. **American Journal of Sociology**, v. 98, n. 6, p. 1259–1319, May 1993. DOI: 10.1086/230190.

- 13 HARARY, F. On the notion of balance of a signed graph. **Michigan Mathematical Journal**, v. 2, n. 2, p. 143–146, 1953. Disponível em: <http://projecteuclid.org:80/Dienst/getRecord?id=euclid.mmj/1028989917/>. Acesso em: 16 jun. 2023.
- 14 CRNOVRSANIN, T. *et al.* Visualization techniques for categorical analysis of social networks with multiple edge sets. *Social Networks*, v. 37, p. 56–64, May 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0378873313001044>. Acesso em: 16 jun. 2023.
- 15 ZHENG, Q.; SKILLICORN, D. B. Analysis of criminal social networks with typed and directed edges. *In: IEEE INTERNATIONAL CONFERENCE ON INTELLIGENCE AND SECURITY INFORMATICS (ISI)*. 2015, Baltimore. **Proceedings** [...] Baltimore: IEEE, 2015. p. 1–6.
- 16 ZHENG, Q. Social networks with multiple relationship semantics. *In: IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING 2015 - ASONAM '15.*, 2015, Paris. **Proceedings** [...] Paris: ACM Press, 2015. p. 894–901.
- 17 AITTOKALLIO, T. Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, v. 7, n. 3, p. 243–255, May 2006. ISSN 1467-5463, 1477-4054. Disponível em: <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bb1022>. Acesso em: 16 jun. 2023.
- 18 BONCHI, F. *et al.* Discovering polarized communities in signed networks. *In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT*. 28., 2019, Beijing. **Proceedings** [...] Beijing, China: ACM, 2019. p. 961–970. DOI: 10.1145/3357384.3357977.
- 19 LI, Y. *et al.* On spectral analysis of directed signed graphs. **International Journal of Data Science and Analytics**, v. 6, n. 2, p. 147–162, Sept. 2018. ISSN 2364-415X, 2364-4168. Disponível em: <http://link.springer.com/10.1007/s41060-018-0143-9>. Acesso em: 16 jun. 2023.
- 20 JAVARI, A. *et al.* ROSE: role-based signed network embedding. *In: THE WEB CONFERENCE 2020*. 2020, Taipei. **Proceedings** [...] Taipei, Taiwan: ACM, 2020. p. 2782–2788. DOI: 10.1145/3366423.3380038.
- 21 PERON, T. *et al.* Spacing ratio characterization of the spectra of directed random networks. **Physical Review E**, v. 102, n. 6, p. 062305, Dec. 2020. ISSN 2470-0045, 2470-0053. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevE.102.062305>. Acesso em: 16 jun. 2023.
- 22 SHAHRIARI, M.; JALILI, M. Ranking nodes in signed social networks. **Social Network Analysis and Mining**, v. 4, n. 1, p. 172, Dec. 2014. ISSN 1869-5450, 1869-5469. Disponível em: <http://link.springer.com/10.1007/s13278-014-0172-x>. Acesso em: 16 jun. 2023.
- 23 GEMICI, K.; VASHEVKO, A. Visualizing hierarchical social networks. **Socius: sociological research for a dynamic world**, v. 4, p. 237802311877298, Jan. 2018. ISSN 2378-0231, 2378-0231. Disponível em: <http://journals.sagepub.com/doi/10.1177/>

2378023118772982. Acesso em: 16 jun. 2023.

24 WU, Z.; AGGARWAL, C. C.; SUN, J. The troll-trust model for ranking in signed networks. *In: INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING - WSDM '16*. 9.; 2016, San Francisco. **Proceedings** [...] San Francisco: ACM Press, 2016. p. 447–456. DOI: 10.1145/2835776.2835816.

25 BACCO, C. D.; LARREMORE, D. B.; MOORE, C. A physical model for efficient ranking in networks. **Science Advances**, v. 4, n. 7, p. eaar8260, July 2018. ISSN 2375-2548. Disponível em: <https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aar8260>. Acesso em: 16 jun. 2023.

26 JIANG, X. *et al.* Statistical ranking and combinatorial Hodge theory. **Mathematical Programming**, v. 127, n. 1, p. 203–244, Mar. 2011. ISSN 0025-5610, 1436-4646. Disponível em: <http://link.springer.com/10.1007/s10107-010-0419-x>. Acesso em: 16 jun. 2023.

27 FANUEL, M.; SUYKENS, J. Deformed Laplacians and spectral ranking in directed networks. **Applied and Computational Harmonic Analysis**, v. 47, n. 2, p. 397–422, Sept. 2019. ISSN 10635203. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1063520317301008>. Acesso em: 16 jun. 2023.

28 AREF, S. *et al.* Multilevel structural evaluation of signed directed social networks based on balance theory. **Scientific Reports**, v. 10, n. 1, p. 15228, Dec. 2020. ISSN 2045-2322. Disponível em: <http://www.nature.com/articles/s41598-020-71838-6>. Acesso em: 16 jun. 2023.

29 FAUST, K. A Puzzle concerning triads in social networks: graph constraints and the triad census. **Social Networks**, v. 32, n. 3, p. 221–233, July 2010. ISSN 03788733. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0378873310000158>. Acesso em: 16 jun. 2023.

30 O'MALLEY, A. J.; ONNELA, J.-P. Introduction to social network analysis. *In: VAN GINNEKEN, E.; BUSSE, R. (ed.). Health care systems and policies*. Boston: Springer, 2017. p. 1–44. DOI: 10.1007/978-1-4939-6704-9\_15-1..

31 CHEN, Y. *et al.* Bridge: enhanced signed directed network embedding. *In: INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT*. 27., 2018, New York. **Proceedings** [...] New York: Association for Computing Machinery, 2018. p. 773–782. DOI: 10.1145/3269206.3271738.

32 SHUBIN, M. A. Discrete magnetic Laplacian. **Communications in Mathematical Physics**, v. 164, n. 2, p. 259–275, Aug. 1994. ISSN 0010-3616, 1432-0916. Disponível em: <http://link.springer.com/10.1007/BF02101702>. Acesso em: 16 jun. 2023.

33 FANUEL, M.; ALAÍZ, C. M.; SUYKENS, J. A. K. Magnetic eigenmaps for community detection in directed networks. **Physical Review E**, v. 95, n. 2, p. 022302, Feb. 2017. DOI: 10.1103/PhysRevE.95.022302. Acesso em: 16 jun. 2023.

- 34 RESENDE, B. M. F. de; COSTA, L. da F. Characterization and comparison of large directed networks through the spectra of the magnetic Laplacian. **Chaos: an interdisciplinary journal of nonlinear science**, v. 30, n. 7, p. 073141, July 2020.
- 35 WOLFRAM, T. **Applications of group theory to atoms, molecules, and solids**. Cambridge: Cambridge University Press, 2014. ISBN 978-1-107-02852-4.
- 36 WOOD, J.; SHAW-TAYLOR, J. A unifying framework for invariant pattern recognition. **Pattern Recognition Letters**, v. 17, n. 14, p. 14151422, Dec. 1996. ISSN 0167-8655.
- 37 GERKEN, J. E. *et al.* **Geometric deep learning and equivariant neural networks**. 2021. Disponível em: <http://arxiv.org/abs/2105.13926>. Acesso em: 16 jun. 2023.
- 38 VELIVCKOVIC, P. Everything is connected: graph neural networks. **Current Opinion in Structural Biology**, v. 79, p. 102538, Apr. 2023. ISSN 0959440X.
- 39 AGRAWAL, D. *et al.* Group-equivariant autoencoder for identifying spontaneously broken symmetries. **Physical Review E**, v. 107, n. 5, p. 054104, May 2023.
- 40 SMIDT, T. E.; GEIGER, M.; MILLER, B. K. Finding symmetry breaking order parameters with Euclidean neural networks. **Physical Review Research**, v. 3, n. 1, p. L012002, Jan. 2021. ISSN 2643-1564.
- 41 KRA, I.; SIMANCA, S. R. On circulant matrices. **Notices of the American Mathematical Society**, v. 59, n. 03, p. 368, Mar. 2012. ISSN 0002-9920, 1088-9477.
- 42 THE Princeton companion to mathematics. Princeton: Princeton University Press, 2008. ISBN 978-0-691-11880-2.
- 43 SINGER, A. angular synchronization by eigenvectors and semidefinite programming. 2009. Disponível em: <http://arxiv.org/abs/0905.3174>. Acesso em: 16 jun. 2023.
- 44 D'ASPREMONT, A.; CUCURINGU, M.; TYAGI, H. Ranking and synchronization from pairwise measurements via SVD. 2020. Disponível em: <http://arxiv.org/abs/1906.02746>. Acesso em: 16 jun. 2023.
- 45 BIRDAL, T. *et al.* Synchronizing probability measures on rotations via optimal transport. 2020. Disponível em: <http://arxiv.org/abs/2004.00663>. Acesso em: 16 jun. 2023.
- 46 HUROYAN, V.; LERMAN, G.; WU, H.-T. Solving Jigsaw puzzles by the graph connection Laplacian. **SIAM Journal on Imaging Sciences**, v. 13, n. 4, p. 17171753, Jan. 2020. ISSN 1936-4954.
- 47 PERON, T. *et al.* Spacing ratio characterization of the spectra of directed random networks. **Physical Review E**, v. 102, n. 6, p. 062305, Dec. 2020. ISSN 2470-0045, 2470-0053.
- 48 BÖTTCHER, L.; PORTER, M. A. **Complex networks with complex weights**. 2022. Disponível em: <http://arxiv.org/abs/2212.06257>. Acesso em: 16 jun. 2023.

- 49 ZHANG, X. *et al.* **MagNet**: a magnetic neural network for directed graphs. 2021. Disponível em: <http://arxiv.org/abs/2102.11391>. Acesso em: 16 jun. 2023.
- 50 GEISLER, S. *et al.* **Transformers meet directed graphs**. 2023. Disponível em: <http://arxiv.org/abs/2302.00049>. Acesso em: 16 jun. 2023.
- 51 BERKOLAIKO, G. Nodal count of graph eigenfunctions via magnetic perturbation. **Analysis PDE**, v. 6, n. 5, p. 12131233, Nov. 2013. ISSN 1948-206X.
- 52 VERDIÈRE, Y. C. de. Magnetic interpretation of the nodal defect on graphs. **Analysis PDE**, v. 6, n. 5, p. 12351242, Nov. 2013. ISSN 1948-206X, 2157-5045.
- 53 MESSIAS, B.; COSTA, L. da F. **Characterization and space embedding of directed graphs and social networks through magnetic Laplacians**. 2018. Disponível em: <http://arxiv.org/abs/1812.02160>. Acesso em: 15 jun. 2023.
- 54 SHUBIN, M. A. Discrete magnetic Laplacian. **Communications in Mathematical Physics**, v. 164, n. 2, p. 259275, Aug. 1994. ISSN 0010-3616, 1432-0916.
- 55 VERDIÈRE, Y. C. de. Magnetic interpretation of the nodal defect on graphs. **Analysis PDE**, v. 6, n. 5, p. 12351242, Nov. 2013. ISSN 1948-206X, 2157-5045.
- 56 AYRES, J. E. de A. Teoria de HodgeRank aplicada aos dados históricos do campeonato brasileiro de futebol. 2016. Tese (Doutorado) — Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2016.
- 57 CAI, H.; ZHENG, V. W.; CHANG, K. C.-C. A comprehensive survey of graph embedding: problems, techniques and applications. 2018. Disponível em: <http://arxiv.org/abs/1709.07604>. Acesso em: 16 jun. 2023.
- 58 MIZ, V. *et al.* Anomaly detection in the dynamics of web and social networks using associative memory. *In: THE WORLD WIDE WEB CONFERENCE ON - WWW '19*. 2019., San Francisco. Proceedings [...] San Francisco: ACM Press, 2019. p. 1290–1299.
- 59 MUNOZ, R. Laplacian eigenmaps manifold learning and anomaly detection methods for spectral images. 2010. Dissertation (Master of Science) - Rochester Institute of Technology, Rochester, 2010.
- 60 FANUEL, M. *et al.* Magnetic eigenmaps for the visualization of directed networks. **Applied and Computational Harmonic Analysis**, v. 44, n. 1, p. 189–199, Jan. 2018.
- 61 YUICHI, K. *et al.* **Hierarchical and circular flow structure of the interfirm transaction network in Japan**. 2019. Disponível em: <https://ideas.repec.org/p/eti/dpaper/19063.html>. Acesso em: 15 jun. 2023.
- 62 YU, Y. *et al.* Dags with no curl: an efficient dag structure learning approach. 2021. Disponível em: <http://arxiv.org/abs/2106.07197>. Acesso em: 15 jun. 2023.

63 GEBHART, T.; FU, X.; FUNK, R. J. **Go with the flow? a large-scale analysis of health care delivery networks in the United States using Hodge theory**. Disponível em: <https://arxiv.org/pdf/2110.09637.pdf>. Acesso em: 15 jun. 2023.

64 HARUNA, T.; FUJIKI, Y. Hodge decomposition of information flow on small-world networks. **Frontiers in Neural Circuits**, v. 10, Sept. 2016. ISSN 1662-5110. Disponível em: <http://journal.frontiersin.org/Article/10.3389/fncir.2016.00077/abstract>. Acesso em: 16 jun. 2023.

65 MACKAY, R. S.; JOHNSON, S.; SANSOM, B. How directed is a directed network? **Royal Society Open Science**, v. 7, n. 9, p. 201138, Sept. 2020. ISSN 2054-5703, 2054-5703. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsos.201138>. Acesso em: 16 jun. 2023.

66 CUCURINGU, M. Synchronization over  $Z_2$  and community detection in signed multiplex networks with constraints. **Journal of Complex Networks**, v. 3, n. 3, p. 469506, Sept. 2015. ISSN 2051-1310, 2051-1329.

67 ZHENG, Q.; SKILLICORN, D. ZHENG, Q.; SKILLICORN, D. **Spectral embedding of signed networks**. In: INTERNATIONAL CONFERENCE ON DATA MINING. 2015. p. 55–63. Disponível em: <https://epubs.siam.org/doi/epdf/10.1137/1.9781611974010.7.doi/10.1137/1.9781611974010>. Acesso em: 15 jun. 2023.

68 KAC, M. Can one hear the shape of a drum? **The American Mathematical Monthly**, v. 73, n. 4, p. 1, Apr. 1966. ISSN 00029890.

69 RESENDE, B. M. F. de. RESENDE, B. M. F. de. **Variações do teorema central do limite para matrizes aleatórias: de núcleos atômicos a filtragem de matrizes de correlação**. 2021. Disponível em: <https://opencodecom.net/post/2021-12-14-variacoes-do-teorema-central-do-limite-para-matrizes-aleatorias-de-nucleos-atomicos-a-filtragem-de-matrizes-de-correlaca/>. Acesso em: 15 jun. 2023.

70 DEHMER, M.; MOWSHOWITZ, A. A history of graph entropy measures. *Information Sciences*, v. 181, n. 1, p. 57–78, Jan. 2011. ISSN 00200255. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0020025510004147>. Acesso em: 16 jun. 2023.

71 TSITSULIN, A.; MUNKHOEVA, M.; PEROZZI, B. Just SLaQ when you approximate: accurate spectral distances for web-scale graphs. In: WEB CONFERENCE 2020. Taipei. **Proceedings** [...] Taipei, Taiwan: ACM, 2020. p. 2697–2703. DOI: 10.1145/3366423.3380026.

72 NICOLINI, C.; VLASOV, V.; BIFONE, A. Thermodynamics of network model fitting with spectral entropies. **Physical Review E**, v. 98, n. 2, p. 022322, Aug. 2018. ISSN 2470-0045, 2470-0053. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevE.98.022322>. Acesso em: 16 jun. 2023.

73 MESSIAS, B.; COSTA, L. da F. **Characterization and space embedding of directed graphs and social networks through magnetic Laplacians**. 2018. Disponível em: <http://arxiv.org/abs/1812.02160>. Acesso em: 16 jun. 2023.

74 LIVAN, G.; NOVAES, M.; VIVO, P. Introduction to Random matrices. Cham: Springer International Publishing, 2018. (SpringerBriefs in mathematical physics, v. 26). ISBN 978-3-319-70883-6 978-3-319-70885-0.

75 LALOUX, L. *et al.* Random matrix theory and financial correlations. **International Journal of Theoretical and Applied Finance**, v. 03, n. 03, p. 391–397, July 2000. ISSN 0219-0249, 1793-6322. DOI: 10.1142/S0219024900000255.

76 FROM Prime Numbers to Nuclear Physics and Beyond. Disponível em: <https://www.ias.edu/ideas/2013/primes-random-matrices>. Acesso em: 15 jun. 2023.

77 LALOUX, L. *et al.* Noise dressing of financial correlation matrices. **Physical Review Letters**, v. 83, n. 7, p. 1467–1470, Aug. 1999. DOI: 10.1103/PhysRevLett.83.1467.

78 EDELMAN, A.; WANG, Y. Random matrix theory and its innovative applications. In: MELNIK, R.; KOTSIREAS, I. S. (ed.). **Advances in applied mathematics, modeling, and computational science**. Boston, MA: Springer, 2013. p. 91–116. DOI: 10.1007/978-1-4614-5389-5\_5.

79 ATAS, Y. Y. *et al.* Distribution of the ratio of consecutive level  $s$  in Random matrix Ensembles. **Physical Review Letters**, v. 110, n. 8, p. 084101, Feb. 2013. DOI: 10.1103/PhysRevLett.110.084101.

80 SÁ, L.; RIBEIRO, P.; PROSEN, T. Complex spacing ratios: a signature of dissipative quantum chaos. **Physical Review X**, v. 10, n. 2, p. 021019, Apr. 2020. DOI: 10.1103/PhysRevX.10.021019.

81 MASKEY, S. *et al.* **A fractional graph Laplacian approach to oversmoothing**. 2023. Disponível em: <http://arxiv.org/abs/2305.13084>. Acesso em: 15 jun. 2023.

82 VIGNA, S. **Spectral ranking**. 2019. Disponível em: <http://arxiv.org/abs/0912.0238>. Acesso em: 15 jun. 2023.

83 LANDAU, E. Über Preisverteilung Bei Spieltunnieren. **Zeitschrift für Mathematik und Physik**, v. 64, 1915.

84 LEVY, R. B. *et al.* Consumo e comportamento alimentar entre adolescentes brasileiros: pesquisa nacional de saúde do escolar (PeNSE), 2009. **Ciência & Saúde Coletiva**, v. 15, p. 3085–3097, 2010.

85 LANGFELDER, P.; HORVATH, S. Wgcna: an r package for weighted correlation network analysis. **BMC Bioinformatics**, v. 9, n. 1, p. 559, Dec. 2008. ISSN 1471-2105.

86 NIEMIRA, M. *et al.* Molecular signature of subtypes of non-small-cell lung cancer by large-scale transcriptional profiling: identification of key modules and genes by weighted gene co-expression network analysis (wgcn). **Cancers**, v. 12, n. 1, p. 37, Dec. 2019. ISSN 2072-6694.

87 OKU, A. Y. A. *et al.* Potential confounders in the analysis of Brazilian adolescent's

health: a combination of machine learning and graph theory. **International Journal of Environmental Research and Public Health**, v. 17, n. 1, p. 90, Dec. 2019. Disponível em: <https://www.mdpi.com/1660-4601/17/1/90>. Acesso em: 16 jun. 2023.

88 ADAMS, J. L. *et al.* **Sirius**: a mutual information tool for exploratory visualization of mixed data. 2021. Disponível em: <http://arxiv.org/abs/2106.05260>. Acesso em: 16 jun. 2023.

89 MOLNAR, C. **Interpretable machine learning**. 2019. 185 p. E-book.

90 SAMEK, W. Learning with explainable trees. **Nature Machine Intelligence**, v. 2, n. 1, p. 16–17, Jan. 2020. ISSN 2522-5839. Disponível em: <http://www.nature.com/articles/s42256-019-0142-0>. Acesso em: 16 jun. 2023.

91 LUNDBERG, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. **Nature Machine Intelligence**, v. 2, n. 1, p. 56–67, Jan. 2020. Disponível em: <http://www.nature.com/articles/s42256-019-0138-9>. Acesso em: 16 jun. 2023.

92 COSCIA, M. The Atlas for the aspiring network scientist. 2021. Disponível em: <http://arxiv.org/abs/2101.00863>. Acesso em: 16 jun. 2023.

93 SERRANO, M. A.; BOGUNA, M.; VESPIGNANI, A. Extracting the multiscale backbone of complex weighted networks. **Proceedings of the National Academy of Sciences**, v. 106, n. 16, p. 6483–6488, Apr. 2009. Disponível em: <http://www.pnas.org/cgi/doi/10.1073/pnas.0808904106>. Acesso em: 16 jun. 2023.

94 MARCACCIOLI, R.; LIVAN, G. A Pólya urn approach to information filtering in complex networks. **Nature Communications**, v. 10, n. 1, p. 745, Dec. 2019. Disponível em: <http://www.nature.com/articles/s41467-019-08667-3>. Acesso em: 16 jun. 2023.

95 BATSON, J. *et al.* Spectral sparsification of graphs: theory and algorithms. **Communications of the ACM**, v. 56, n. 8, p. 87–94, Aug. 2013. Disponível em: <https://dl.acm.org/doi/10.1145/2492007.2492029>. Acesso em: 16 jun. 2023.

96 CASACA, W.; PAIVA, A.; NONATO, L. G. Spectral segmentation using cartoon-texture decomposition and inner product-based metric. *In: SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES*. 24., 2011, Alagoas. **Proceedings [...]** Alagoas: IEEE, 2011. p. 266–273. Disponível em: <http://ieeexplore.ieee.org/document/6134741/>. Acesso em: 16 jun. 2023.

97 ANDERSON, P. W. More is different. **Science**, v. 177, n. 4047, p. 393396, Aug. 1972. ISSN 0036-8075, 1095-9203.

98 KOSTERLITZ, J. M. The critical properties of the two-dimensional xy model. **Journal of Physics C: solid state physics**, v. 7, n. 6, p. 1046–1060, Mar. 1974. Disponível em: <https://iopscience.iop.org/article/10.1088/0022-3719/7/6/005>. Acesso em: 16 jun. 2023.

99 MERMIN, N. D.; WAGNER, H. Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. **Physical Review Letters**, v. 17, n. 22, p. 1133–1136, Nov. 1966. DOI: 10.1103/PhysRevLett.17.1133.

- 100 MERKL, F.; WAGNER, H. Recurrent Random walks and the absence of continuous symmetry breaking on graphs. **Journal of Statistical Physics**, v. 75, n. 1-2, p. 153–165, Apr. 1994. ISSN 0022-4715, 1572-9613. Disponível em: <http://link.springer.com/10.1007/BF02186284>. Acesso em: 16 jun. 2023.
- 101 ILLING, B. *et al.* Mermin–Wagner fluctuations in 2D amorphous solids. **Proceedings of the National Academy of Sciences**, v. 114, n. 8, p. 1856–1861, Feb. 2017. DOI: 10.1073/pnas.1612964114.
- 102 VALDEZ, M. A. *et al.* Quantifying complexity in quantum phase transitions via mutual information complex networks. **Physical Review Letters**, v. 119, n. 22, p. 225301, Nov. 2017. DOI: 10.1103/PhysRevLett.119.225301.
- 103 SUNDAR, B. *et al.* Complex-network description of thermal quantum states in the ising spin chain. **Physical Review A**, v. 97, n. 5, p. 052320, May 2018. DOI: 10.1103/PhysRevA.97.052320.
- 104 ZAMAN, S.; LEE, W.-C. Real-space visualization of quantum phase transitions by network topology. **Physical Review E**, v. 100, n. 1, p. 012304, July 2019. DOI: 10.1103/PhysRevE.100.012304.
- 105 ANDERSON, P. W. More is different. **Science**, v. 177, n. 4047, p. 393–396, Aug. 1972. DOI: 10.1126/science.177.4047.393.
- 106 GRAY, R. M. Toeplitz and circulant matrices: a review. **Foundations and Trends in Communications and Information Theory**, v. 2, n. 3, p. 155239, 2005. ISSN 1567-2190, 1567-2328.
- 107 WEIßE, A. *et al.* The kernel polynomial method. **Reviews of Modern Physics**, v. 78, n. 1, p. 275306, Mar. 2006. ISSN 0034-6861, 1539-0756.
- 108 ABADI, M. *et al.* Tensorflow: large-scale machine learning on heterogeneous systems. 2015. Disponível em: <https://www.tensorflow.org/>. Acesso em: 15 jun. 2023.
- 109 OKUTA, R. *et al.* CuPy: a NumPy-compatible library for NVIDIA GPU calculations. 2017. Disponível em: [http://learningsys.org/nips17/assets/papers/paper\\_16.pdf](http://learningsys.org/nips17/assets/papers/paper_16.pdf). Acesso em: 15 jun. 2023.
- 110 PEYRÉ, G.; CUTURI, M. Computational optimal transport: with applications to data science. **Foundations and Trends® in Machine Learning**, v. 11, n. 56, p. 355607, 2019. ISSN 1935-8237, 1935-8245.
- 111 GRUNBAUM, J. A. *et al.* Youth risk behavior surveillance - United States, 2003 (Abridged). **Journal of School Health**, v. 74, n. 8, p. 307–324, Oct. 2004. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.1746-1561.2004.tb06620.x>. Acesso em: 15 jun. 2023.
- 112 CURRIE, C. Inequalities in young people’s health: HBSC international report from the 2005/2006 Survey. Geneva: World Health Organization, 2008. (1,5).

113 OLIVEIRA, M. M. de *et al.* Characteristics of the national adolescent school-based health survey-PeNSE, Brazil. **Epidemiologia e Serviços de Saúde**, v. 26, p. 605–616, 2017.

114 KOBAK, D.; BERENS, P. The art of using t-sne for single-cell transcriptomics. **Nature Communications**, v. 10, n. 11, p. 5416, Nov. 2019. ISSN 2041-1723.

115 MCINNES, L.; HEALY, J.; MELVILLE, J. Umap: uniform manifold approximation and projection for dimension reduction. 2020. Disponível em: <http://arxiv.org/abs/1802.03426>. Acesso em: 15 jun. 2023.

116 GRISCI, B. I.; KRAUSE, M. J.; DORN, M. Relevance aggregation for neural networks interpretability and knowledge discovery on tabular data. **Information Sciences**, v. 559, p. 111129, jun. 2021. ISSN 0020-0255.

117 NEWMAN, M. E. A measure of betweenness centrality based on random walks. **Social Networks**, v. 27, n. 1, p. 39–54, 2005.

118 VILLEGAS, P. *et al.* Laplacian renormalization group for heterogeneous networks. **Nature Physics**, v. 19, n. 3, p. 445450, Mar. 2023. ISSN 1745-2473, 1745-2481.

119 CASO, F. *et al.* **Renormalized graph neural networks**. 2023. Disponível em: <http://arxiv.org/abs/2306.00707>. Acesso em: 16 jun. 2023.

120 CHENG, A.; SUN, P.; TIAN, Y. Simplex path integral and renormalization group for high-order interactions. 2023. Disponível em: <http://arxiv.org/abs/2305.01895>. Acesso em: 16 jun. 2023.

121 LOURES, M. de C.; PIOVESANA, A. A.; BRUM, J. A. Laplacian coarse graining in complex networks. 2023. Disponível em: <http://arxiv.org/abs/2302.07093>. Acesso em: 16 jun. 2023.

122 MESSIAS, B. Python, Edges Eraser. 2022. Disponível em: <https://github.com/devmessias/edgeseraser>. Acesso em: 15 jun. 2023.

123 GONG, X.; HIGHAM, D. J.; ZYGALAKIS, K. Directed network Laplacians and Random graph models. **Royal Society Open Science**, v. 8, n. 10, p. 211144, Oct. 2021. ISSN 2054-5703.

## APÊNDICES



## APÊNDICE A – HODGERANK, SPRINGRANK E LAPLACIANO TRÓFICO

A abordagem HodgeRank foi redescoberta ao longo dos anos por meio de diferentes abordagens. Por exemplo, Bacco, Larremore e Moore derivaram um método de classificação conhecido como *SpringRank* usando um modelo inspirado em osciladores harmônicos. Especificamente, eles propuseram uma função de classificação de posição,  $s : V \rightarrow \mathbb{R}_+$ , que permite que uma energia seja associada ao grafo. Essa função de energia é dada pela seguinte equação

$$H(s) = \frac{1}{2} \sum_{u,v \in V} A(u,v)(s(u) - s(v) - 1)^2. \quad (\text{A.1})$$

Portanto, devido à convexidade, eles propõem encontrar o  $s$  tal que a seguinte equação é satisfeita

$$\left( \mathbf{D}^{in} + \mathbf{D}^{out} - (\mathbf{A} + \mathbf{A}.T) \right) \mathbf{s}^* = (\mathbf{D}^{out} - \mathbf{D}^{in}) \mathbf{1}. \quad (\text{A.2})$$

Com a equação acima e as suposições propostas por Bacco, Larremore e Moore, podemos encontrar as conexões com o método SprinRank e a decomposição de Hodge-Helmholtz e, conseqüentemente, com o método HodgeRank. Primeiro, observe que a parte à esquerda é apenas o Laplaciano combinatório associado a um grafo não ponderado. Além disso, a parte à direita é a divergência do Laplaciano combinatório aplicado ao fluxo  $A - A^T$ . Assim, a Eq.A.2 pode ser reescrita como

$$\begin{aligned} (\Delta_0 s^*)(u) &= \sum_{v \in V} (w(v,u) - w(u,v)) \\ &= -(\text{div} F)(u). \end{aligned} \quad (\text{A.3})$$

A expressão acima é a mesma apresentada no Teorema 3 por Jiang *et al.* na ausência de preferências ponderadas. Conseqüentemente, a solução do HodgeRank e SpringRank são exatamente as mesmas. Portanto, a interpretação também é semelhante. O potencial Helmholtz obtido pela decomposição de Hodge permite associar um fluxo de gradiente,  $F_g$ , a um grafo direcionado. Esse fluxo tem o mesmo significado obtido pela Eq.A.1. Ou seja, o fluxo é devido a um gradiente de uma função potencial,  $\phi$ . Em essência, Eq.A.2 representa a versão não ponderada do método Hodge-rank.

Recentemente, MacKay, Johnson e Sansom derivaram um método conhecido como níveis tróficos definindo um Laplaciano Trófico,  $\Lambda$ . Este método tem sido usado em aplicações desde análise biológica até sistemas financeiros<sup>65</sup> e modelos geradores adicionais foram propostos.<sup>122</sup> No entanto, é notável que o Laplaciano Trófico é o mesmo que o

Laplaciano combinatório para  $G_s$ . Portanto, a expressão para  $\mathbf{\Lambda}$  é exatamente a mesma apresentada à esquerda de Eq.A.2. Podemos perceber que a única diferença entre os métodos SpringRank e Trophic é que este último busca um  $h : V \rightarrow \mathbb{R}_+$  tal que a seguinte equação seja respeitada

$$\left(\mathbf{D}^{in} + \mathbf{D}^{out} - (\mathbf{A} + \mathbf{A}.T)\right) \mathbf{h}^* = (\mathbf{D}^{in} - \mathbf{D}^{out})\mathbf{1}. \quad (\text{A.4})$$

Que é a decomposição de Hodge-Helmholtz com um fluxo invertido,  $A^T - A$ .

$$\begin{aligned} (\Delta_0 s^*)(u) &= \sum_{v \in V} (w(v, u) - w(u, v)) \\ &= -(\text{div}(-F))(u). \end{aligned} \quad (\text{A.5})$$

As relações entre os operadores e funções Laplacianos discutidos neste apêndice estão resumidos na Tabela 1.

Tabela 1 – Relações entre os conceitos nos métodos HodgeRank, SpringRank e Trophic Laplacian

<b>HodgeRank</b>	<b>SpringRank</b>	<b>Trophic Laplacian</b>
Helmholtz Potential ( $\phi$ )	Node Positional rank ( $s$ )	trophic levels ( $h$ )
Flow divergence ( $\nabla F$ )	$D_{out} - D_{in}$	node imbalances ( $v$ )
Combinatorial Laplacian ( $\Delta_0$ )	$D_{out} + D_{in} - (A + A^T)$	Trophic Laplacian ( $\Lambda$ )
Residual	$\frac{1}{2} \sum_{u, v \in V} A(u, v)(s(u) - s(v) - 1)^2$	$\frac{\sum_{m, n} w_{mn}(h_n - h_m - 1)^2}{w_{mn}}$

Fonte: Elaborada pelo autor.