



Universidade Federal
do Rio de Janeiro

Escola Politécnica

ANÁLISE DA REGULARIDADE ESTRUTURAL DE REDES COMPLEXAS ATRAVÉS DE PARTICIONAMENTO EM ESTRELAS

Matheus Guedes de Andrade

Projeto de Graduação apresentado ao Curso de Engenharia de Computação e Informação da Escola Politécnica da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

Orientador: Daniel Ratton Figueiredo

Rio de Janeiro
Março de 2018

ANÁLISE DA REGULARIDADE ESTRUTURAL DE REDES COMPLEXAS
ATRAVÉS DE PARTICIONAMENTO EM ESTRELAS

Matheus Guedes de Andrade

PROJETO SUBMETIDO AO CORPO DOCENTE DO CURSO DE ENGENHARIA DE COMPUTAÇÃO E INFORMAÇÃO DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO DE COMPUTAÇÃO E INFORMAÇÃO.

Examinadores:

Prof. Daniel Ratton Figueiredo, Ph.D.

Prof. Luidi Simonetti, D.Sc.

Prof. Franklin Marquezino, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2018

Guedes de Andrade, Matheus

Análise da Regularidade Estrutural de Redes Complexas Através de Particionamento em Estrelas/Matheus Guedes de Andrade. – Rio de Janeiro: UFRJ/POLI – COPPE, 2018.

XI, 46 p.: il.; 29, 7cm.

Orientador: Daniel Ratton Figueiredo

Projeto (graduação) – UFRJ/ Escola Politécnica/ Curso de Engenharia de Computação e Informação, 2018.

Referências Bibliográficas: p. 44 – 46.

1. Complexidade Estrutural. 2. Entropia de Redes.
3. Redes Complexas. 4. Teoria de Grafos. I. Ratton Figueiredo, Daniel. II. Universidade Federal do Rio de Janeiro, Escola Politécnica/ Curso de Engenharia de Computação e Informação. III. Título.

*Dedico este trabalho,
essencialmente, a meus avós.*

Agradecimentos

Agradeço a minha mãe por me ensinar que o amor é a maior lição da vida. Obrigado por sempre acreditar em mim e por me criar para ser um homem de caráter. Nada do que fiz até hoje seria possível sem sua incansável dedicação e suas palavras de carinho e conforto nos momentos difíceis.

Agradeço a meu pai por me incentivar a realizar meus sonhos, por me ensinar que devemos tratar os momentos ruins como oportunidades de aprendizado e por mostrar que a humildade e a bondade são as qualidades mais importantes da existência humana.

Agradeço a meu querido padrasto por cativar em mim o amor pela matemática, por ser ouvinte atento de todas as teorias malucas que surgem em minha cabeça e por participar de forma primordial em minha educação e criação.

A meus irmãos, Clara e Carlos, agradeço por serem inesgotáveis fontes de alegria no meu cotidiano.

A meus primos, de sangue e postigos, agradeço por me acompanharem nos caminhos da vida.

Agradeço ao professor Daniel Rattton por apresentar a Teoria dos Grafos de forma tão fantástica aos alunos de ECI, cativando em mim a imensa vontade de estudá-los, e por aceitar orientar-me na empreitada deste trabalho de conclusão.

Agradeço a toda minha família, aos amigos que fiz e a todos meus professores. As lições que me ensinaram fazem parte do que sou hoje, e me ajudam a esboçar o homem que quero ser amanhã.

Por fim, agradeço a você que, por qualquer que seja o motivo, está dedicando seu precioso tempo a ler este trabalho.

“For a successful technology, reality must take precedence over public relations, for nature cannot be fooled.”

— RICHARD P. FEYNMAN

Resumo do Projeto de Graduação apresentado à Escola Politécnica/COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

ANÁLISE DA REGULARIDADE ESTRUTURAL DE REDES COMPLEXAS ATRAVÉS DE PARTICIONAMENTO EM ESTRELAS

Matheus Guedes de Andrade

Março/2018

Orientador: Daniel Ratton Figueiredo

Curso: Engenharia de Computação e Informação

A análise de Redes Complexas é fundamental para o entendimento de fenômenos observados por pesquisadores em diversas áreas da ciência. Nessa perspectiva, o relacionamento latente entre as características topológicas das redes com as propriedades dos fenômenos observados surge como fator chave. Dentre os diversos aspectos topológicos que aparecem como alicerce para o entendimento desses fenômenos, a regularidade topológica da rede é, por vezes, foco de observação. No âmbito citado, o uso de uma adaptação da entropia de von Neumann, inicialmente definida para misturas de estados quânticos, é comumente utilizada para descrever a regularidade de uma topologia, oferecendo uma base de comparação entre redes distintas. Não obstante, o cálculo dessa métrica é custoso em termos computacionais, dado que depende da obtenção de autovalores de matrizes. Em efeito, quando os objetos de estudo são redes de tamanho elevado, calcular diretamente essa entropia é, em geral, impraticável. Este trabalho propõe uma alternativa computacionalmente viável para obter uma aproximação da entropia de von Neumann através de particionamentos da rede. De forma direta, o método descrito consiste em decompor o grafo que representa a rede original em subgrafos estrela e utilizar o particionamento resultante para obter possíveis limitantes superiores e inferiores para a métrica em questão. A contribuição deste trabalho consiste na formalização matemática que embasa a obtenção destes limitantes, na descrição de algoritmos para o específico particionamento de grafos em estrelas e na análise experimental dos limitantes subsequentes, nos quesitos de efetividade de aproximação e de tratabilidade computacional para redes grandes. **Palavras-Chave:** Complexidade Estrutural, Entropia de Redes, Redes Complexas, Teoria de Grafos.

Abstract of the Undergraduate Project presented to Poli/COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Computer and Information Engineer.

COMPLEX NETWORKS STRUCTURAL REGULARITY ANALISYS THROUGH STAR PARTITIONS

Matheus Guedes de Andrade

March/2018

Advisor: Daniel Ratton Figueiredo

Course: Computer and Information Engineering

Complex Networks analisys are fundamental to understanding phenomenological behavior observed by researches in multiple areas of science. In this perspective, relationships between topological aspects of networks with properties of observed phenomena are key concepts for effective behavioral descriptions. Among the core topological characteristics used to study these relationships, regularity metrics are some of the observation focus. Within this context, an adaptation of the von Neumann entropy discribed for quantum state mixtures is used to quantify regularity. Nonetheless, the computational cost of calculatating this entropy directly makes sometimes impractical the task of analysing big networks. The work discribed in this document proposes an alternative approach to compute an approximantion for the von Neumann entropy through network partition schemes. Directly, the described method consists in decomposign the network's underlying graph into star subgraphs, where the resulting partition is used to compute both lower and upper bounds for the entropy in question. In essence, this document addresses the mathematical basis for obtaining the expressed bounds, algortihms used to decompose graphs into star partitions and an experimental analysis of the obtained results when applying the method to sampled real networks and stastical models, regarding computing execution time and approximation distance.

Keywords: Structural Complexity, Network Entropy, Complex Networks, Graph Theory.

Sumário

Lista de Figuras	x
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivação	2
1.2 Objetivo	2
2 Conceitos Relacionados	4
2.1 Redes e Grafos	4
2.1.1 Nomenclatura e definições básicas	4
2.1.2 Classes de grafos	6
2.1.3 Representação através de matrizes	7
2.1.4 O Espectro do Laplaciano	8
2.2 Propriedades básicas do traço de matrizes	9
2.3 Modelos aleatórios de Redes	10
2.3.1 O modelo de Erdős-Rényi	10
2.3.2 O modelo de Barabási-Albert	11
2.4 A entropia de distribuições de probabilidade	11
2.4.1 Quantificando a incerteza de distribuições	12
2.4.2 A entropia de von Neumann	13
2.4.3 Complexidade da entropia pelo cálculo de autovalores	14
2.5 Teoria de Majoração	14
2.6 Particionamento em subgrafos	15
2.6.1 Particionamentos em classes específicas	16
2.6.2 Estrelas e cobertura mínima de vértices	17
3 Um algoritmo aproximativo para a entropia de von Neumann	18
3.1 Ideias preliminares	18
3.1.1 Entropia de particionamentos	18
3.1.2 Subgrafos de espectro simples e complexidade de particiona- mentos	19

3.2	Descrevendo limitantes superiores	20
3.2.1	Entropia dos graus normalizados	20
3.2.2	Cardinalidade da união dos espectros de estrelas	21
3.2.3	Partições maximais e a redução da entropia	22
3.2.4	Entropia de constelação como limitante superior	23
3.3	Conjectura do particionamento mínimo	24
3.4	Descrevendo limitantes inferiores	24
3.5	Algoritmos de particionamento em estrelas	25
3.5.1	Estrelas maximais pela ordenação de vértices	25
3.5.2	Uma heurística para a busca de estrelas maximais	26
3.5.3	Complexidade assintótica	27
4	Avaliação Experimental	30
4.1	Implementação do processo experimental	30
4.2	Descrição do processo de avaliação	30
4.2.1	Medidas de entropia calculadas	31
4.2.2	Medições de entropia	31
4.2.3	Medições do tempo de execução	31
4.2.4	Instâncias de redes reais	32
4.3	Eficiência dos limitantes para a aproximação	32
4.3.1	Redes aleatórias com referencial	33
4.3.2	Redes reais	34
4.3.3	Análise de distribuições empíricas de estrelas	36
4.4	Eficiência do cálculo computacional	38
4.4.1	Média empírica nos modelos aleatórios	38
4.4.2	Tempo médio para redes reais amostradas	38
5	Conclusões e trabalhos futuros	41
5.1	Conclusões finais	41
5.2	Próximos passos	42
5.2.1	Algoritmo de particionamento mínimo	42
5.2.2	Particionamentos em classes de grafo distintas	42
5.2.3	Conjecturas sobre a entropia mínima	43
5.2.4	Entropia característica de redes reais	43
	Referências Bibliográficas	44

Lista de Figuras

2.1	A comum representação visual de um grafo. Os círculos representam vértices, as linhas representam arestas e os números uma rotulação para os nós.	5
2.2	Exemplos de classes de grafos instanciadas com 4 vértices. Da esquerda para direita: grafo caminho P_4 , grafo estrela S_4 , grafo bipartido completo $K_{2,2}$, grafo completo K_4	7
3.1	Um grafo (meio) com entropia de von Neumann $H_n = 3.05$ decomposto em duas partições distintas: a partição de cardinalidade mínima (esquerda), com entropia de estrelas $H_s = 3.25$, e a partição obtida pela heurística do algoritmo 2 (direita), com $H_s = 3.31$	27
4.1	Medições empíricas do modelo BA, com $m = 5$. Curvas de mesma cor representam os limites superiores e inferiores obtidos dos particionamentos.	33
4.2	Medições empíricas do modelo $G(N, p)$ no regime $\frac{\log p}{p}$. Curvas de mesma cor representam os limites superiores e inferiores obtidos dos particionamentos	34
4.3	CCDF's empíricas para as redes de colaboração de Matéria Condensada amostradas em 1999 e 2005 (cond-mat, cond-mat-2005), da esquerda para direita, respectivamente.	36
4.4	CCDF's empíricas para a rede de amizades no YouTube (com-youtube).	37
4.5	CCDF's empíricas para a rede roadNet.	38
4.6	Média empírica do tempo médio de execução das métricas, em segundos, para o modelo BA com $m = 5$	39
4.7	Média empírica do tempo médio de execução das métricas, em segundos, para o modelo $G(N, p)$ no regime $\frac{\log N}{N}$	39

Lista de Tabelas

4.1	Descrição das redes pelo número de vértices, arestas e seu valor de entropia de graus normalizados.	35
4.2	Limitantes superiores e inferiores calculados para redes reais para os algoritmos de particionamento 1 e 2, respectivamente.	35
4.3	Tempo médio, em segundos, obtido por 10 amostras do cálculo de aproximação dos particionamentos dos algoritmos 1 e 2 para redes reais, respectivamente.	40

Capítulo 1

Introdução

De todo esforço realizado por cientistas atuais em entender o mundo que nos cerca, analisar, descrever e entender sistemas complexos aparece como foco de pesquisa em diversas instituições. Dentre esses sistemas, as ditas Redes Complexas tem papel fundamental para descrever comportamentos sistêmicos que podem emergir do relacionamento conjunto entre pares de objetos. Por seu embasamento no uso de Grafos como entidades matemáticas que modelam relacionamentos deste tipo, é central analisar redes através de sua estrutura, buscando, na topologia, possíveis explicações para fenômenos por nós observados.

Nesse âmbito, o estudo das propriedades topológicas remete, por vezes, na concepção de métricas que permitam a comparação de sistemas e de suas partes constituintes. Essas medições propiciam, por meio das características estruturais, generalizações para os métodos e ferramentas de análise do comportamento de redes reais que emergem em cenários profundamente distintos e que são compostas por elementos díspares. Tem-se, como exemplo, o PageRank [1], a popular métrica de classificação de relevância de páginas web utilizada pela Google em sua ferramenta de busca que, em complemento à toda sua importância para a pesquisa na web, foi utilizada em [2] para modelar o impacto da extinção de espécies em cadeias alimentares.

Devido a considerável magnitude da escala desses sistemas contemplando, por vezes, trilhões de objetos (i.e a web), é imprescindível definir métricas capazes de serem computadas na prática, que enalteçam alguma perspectiva da topologia da rede e que, efetivamente, possam ser utilizadas nas mais diversas aplicações. Dentre estas perspectivas, pode-se questionar o quão complexo é, de fato, a topologia de uma rede, cabendo definir um conceito de complexidade adequado à análise de estruturas, estipulando um ponto de partida para responder tal questão.

1.1 Motivação

Por sua multidisciplinaridade, diversas áreas da ciência contribuem ativamente para avanços na pesquisa de Redes Complexas. Dentre estas, a física aparece como contribuinte fundamental, trazendo consigo diversos métodos analíticos desenvolvidos durante séculos para o campo, bem como formidáveis fontes de inspiração.

Dentre as leis e conceitos físicos de suma importância para a compreensão de nosso universo, são protagonistas a Entropia, que é, fundamentalmente, uma medida de desordem, e a segunda lei da Termodinâmica, que, até então, define a unicidade de sentido da passagem do tempo. Cunhado por Carnot no estudo de máquinas térmicas e sendo o alicerce da mecânica estatística estudada por Boltzmann, Maxwell e Gibbs, o conceito de entropia é generalizado para a análise de distribuições de probabilidade através do célebre trabalho de Shannon [3] o qual, por toda sua relevância no estudo de diversos sistemas, acaba por iniciar o surgimento da Teoria da Informação.

Ainda na física, ao estudar sistemas quânticos e conceber diversos fundamentos matemáticos chave para a descrição da mecânica quântica, John von Neumann define em [4] um conceito de entropia para misturas de estados quânticos. Esta definição estipula a reversibilidade de processos desta natureza e, assim como a entropia de Boltzmann, Maxwell e Gibbs, vai de encontro a definição obtida por Shannon para distribuições de probabilidade.

No âmbito do estudo de redes, a definição formal de entropia ainda é um problema em aberto. Há, em contrapartida, esforços realizados na utilização da entropia quântica de von Neumann para a análise de redes complexas, como expresso em [5], [6] e [7]. Esses esforços são sustentados pelo caráter matricial comum entre a formalização de von Neumann dos sistemas de misturas, com as matrizes de adjacência dos grafos, que formaliza o estudo das redes.

A entropia de von Neumann definida para redes é calculada através do espectro do Laplaciano dos grafos e enaltece a regularidade topológica da disposição estrutural dos relacionamentos de objetos. O custo computacional do cálculo de autovalores para matrizes de dimensão elevada pode, por vezes, inviabilizar o uso dessa medida de entropia para diversas aplicações em redes complexas, abrindo espaço para a busca de métricas alternativas que ofereçam um compromisso entre expressividade e desempenho computacional.

1.2 Objetivo

Este trabalho tem como objetivo definir um processo para a obtenção de métricas de entropia derivadas da entropia de von Neumann e capazes de serem efetivamente

calculadas para redes que contemplam um grande número de objetos, bem como a formalização da análise matemática que é base para tal. É também pertinente ao escopo usar as técnicas definidas em um estudo empírico de seu comportamento, ao aplicá-las em instâncias de redes reais obtidas de bases de dados existentes e em instâncias obtidas por modelos aleatórios generativos.

O processo definido se baseia na análise de particionamentos de grafos em subgrafos estrelas sendo a métrica de entropia proposta referida como entropia de particionamento em estrelas. A ideia por trás do método é usar as propriedades espectrais dos subgrafos estrela que permitem o cálculo de seus autovalores de forma simples.

O foco principal consiste em analisar particionamentos para buscar limitantes superiores e inferiores para a entropia de von Neumann, oferecendo uma base de comparação para redes distintas. Por fim, traz-se um algoritmo para a obtenção desses limitantes que possa ser efetivamente aplicado em grandes redes e que permita atingir um nível de compromisso entre a aproximação da métrica de von Neumann e o tempo de processamento.

Capítulo 2

Conceitos Relacionados

Esse capítulo tem como propósito definir o arcabouço teórico do projeto, contemplando os conceitos e propriedades matemáticas de grafos e redes, a nomenclatura utilizada, o conceito de entropia que norteia a métrica a ser proposta, as propriedades dos modelos generativos de redes de interesse e o particionamento de grafos.

2.1 Redes e Grafos

Das mais diversas teorias matemáticas usadas na computação, a teoria dos Grafos é de fundamental importância para a solução de problemas. Como expresso em [8], grafos podem ser definidos como *estruturas discretas que codificam o relacionamento entre pares de objetos* e são extensamente utilizados para representar redes. Devido a sua capacidade de abstração, grafos podem representar redes dos mais diversos tipos objetos e relacionamentos. Pode-se, por exemplo, representar a rede de páginas web existentes na internet e conectadas por *hyperlinks*, a rede de distribuição de energia formada por postes e cabos e a rede de computadores, roteadores e os diversos sistemas que compõem a Internet.

2.1.1 Nomenclatura e definições básicas

Quanto a terminologia, nomeiam-se os objetos de uma rede como *nós* ou *vértices* e seus relacionamentos como *arestas* ou *links*. As arestas podem ser direcionadas ou não e possuir um valor de peso associado, o que está relacionado com a simetria dos relacionamentos abstraídos. Assim, constrói-se um grafo ao definir seu conjunto de vértices V , de arestas $E \in [V \times V] \setminus \{u, u\}$ para todo $u \in V$ e uma função peso $w(u, v)$ que mapeia arestas em valores. Essa classificação define grafos direcionados com ou sem peso e não direcionados com ou sem peso. Grafos sem peso podem ser visualizados como tendo peso unitário para todas suas arestas, de forma que w seja comumente omitida ao tratar este tipo de grafo. Cabe notar que os objetos

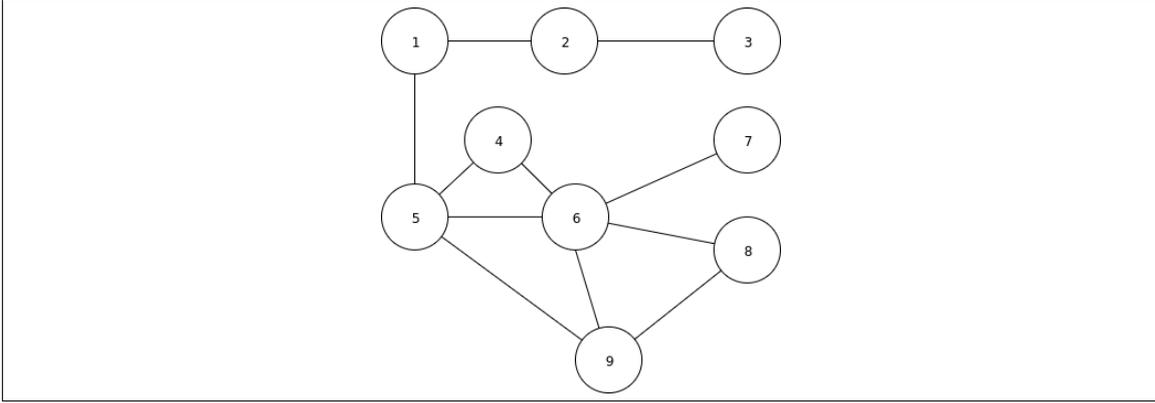


Figura 2.1: A comum representação visual de um grafo. Os círculos representam vértices, as linhas representam arestas e os números uma rotulação para os nós.

de estudo desse trabalho são definidos como *grafos sem peso e não-direcionados*, e todas as relações explicitadas neste documento são definidas para estes.

A aresta $e = (u, v)$ é dita *incidente* aos vértices u, v e o número de arestas $d(u)$ incidentes em u é definido como o *grau* de u em G . Quando a aresta (u, v) existe, diz-se que u e v são *vizinhos* e, sendo $|V| = N, |E| = M$, seguem as relações abaixo.

$$\sum_{u=1}^N d(u) = 2M \quad (2.1)$$

$$d(v) \leq N \quad (2.2)$$

$$M \leq \frac{N(N-1)}{2} \quad (2.3)$$

Um exemplo de um grafo com $|N| = 9$ e $|E| = 11$ é mostrado na figura 2.1.

Tem-se que um grafo $H = (V', E')$ é *subgrafo* de $G = (V, E)$ quando $V' \subset V$ e $E' \subset E$. H é dito *subgrafo gerador* de G se $V' = V$. Usa-se, ao longo do texto, a notação $H \subset G$ para denotar que H é subgrafo de G .

Um *caminho simples* $P = [(u, x_1), (x_1, x_2), \dots, (x_n, v)]$ entre u e v em G é definido como uma sequência de arestas que não incidem em vértices repetidos, de modo que $u \neq x_i \neq v$ para todo x_i . Quando tal caminho existe, diz-se que é possível alcançar v a partir de u seguindo as arestas de P . Grafos são denominados *conexos* quando existe um caminho simples entre quaisquer $u, v \in V$. Em contrapartida, um *ciclo* C em G é semelhante a P , mas com extremidades iguais, i.e $u = v$.

Pela *distância* entre dois vértices u e v entende-se o tamanho do menor caminho em G que conecta ambos os vértices.

Um subgrafo $H = (V', E') \subset G$ é dito uma *componente conexa* de G quando $V' \subset V, E' \subset E$ são subconjuntos maximais perante a propriedade¹ de conexidade

¹Exprime-se a maximalidade de um conjunto C perante uma condição P quando é impossível aumentar a cardinalidade de C sem desrespeitar P .

dos vértices de V' . Caso G seja conexo, há uma única componente conexa $H = G$; caso contrário, G pode ser escrito como uma união de componentes conexas distintas.

Um *isomorfismo* $\phi(V)$ de G é uma função de rotulação de vértices e um grafo $H = (V', E')$ é dito *isomorfo* de $G = (V, E)$ quando existe $\phi(V)$ tal que o subgrafo induzido por $\phi(V')$ faz $E' = E$.

Definidos estes conceitos básicos, é necessário explicitar a notação matemática envolvida e algumas propriedades importantes de suas representações.

2.1.2 Classes de grafos

Alguns conjuntos de grafos compartilham propriedades semelhantes, sendo classificados especificamente.

Grafos são ditos *completos* quando todas as possíveis arestas entre os vértices de V pertencem a E (i.e $E = [V \times V] \setminus \{u, u\}$, para todo $u \in V$), e são comumente representados por K_n , onde n é a cardinalidade de V . K_n possui $\frac{n(n-1)}{2}$ arestas e $d(u) = n - 1$ para todo vértice u . Um grafo é dito *d-regular* quando todos os seus vértices possuem grau igual a d . Grafos completos são grafos $(|V| - 1)$ -regulares.

Quando apresentam topologia centralizada em um único vértice, grafos são denominados *estrelas*. A estrela de n vértices S_n possui $n - 1$ arestas, o vértice central é denominado *centro* de S_n e o grau dos vértices respeita a seguinte relação:

$$d(u) = \begin{cases} (n - 1), & \text{se } u \text{ é o centro.} \\ 1, & \text{caso contrário} \end{cases}$$

Um *grafo caminho* P_n é definido por uma sequência de vértices ligados dois a dois, possuem os vértices de ponta com grau 1, e os vértices intermediários com grau 2, tendo $n - 1$ arestas. Grafos *ciclos* C_n podem ser formados a partir de grafos caminho ao se conectar os vértices das pontas, totalizando n arestas. Em C_n , todos os vértices possuem grau 2.

Grafos *k-partidos* $G_{V_1, \dots, V_k} = (V, E)$ possuem $V = \bigcup_{i=1}^k V_i$ e $(v, u) \notin E$ dado $v, u \in V_i$ para todo i . Um grafo é dito *bipartido* quando $k = 2$ e estrelas são também grafos desta classe, o que pode ser visto fazendo colocando em um conjunto o vértice central e em outro todos os seus vizinhos.

Grafos que não possuem ciclos são denominados *árvores*, de forma que S_n e P_n se encaixam nesta categoria. Todas as árvores de n vértices, assumindo a ausência de vértices desconetados, possuem $n - 1$ arestas e, quando conexas, possuem exatamente um único caminho simples entre qualquer par de vértices. Há diversas outras classes de grafos definidas na literatura, e.g grafos pirulito, mas que não cabem ao escopo desse documento. Exemplos de instancias para algumas das classes citadas seguem na figura 2.2.

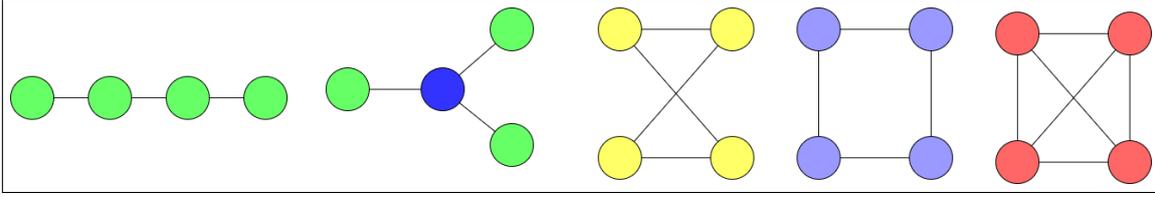


Figura 2.2: Exemplos de classes de grafos instanciadas com 4 vértices. Da esquerda para direita: grafo caminho P_4 , grafo estrela S_4 , grafo bipartido completo $K_{2,2}$, grafo completo K_4 .

2.1.3 Representação através de matrizes

A representação mais comumente utilizada na literatura de grafos consiste em matrizes de adjacência de vértices. Da definição usual, admite-se um grafo $G = (V, E)$, onde V e E possuem, respectivamente, N e M elementos. A matriz de adjacência $A(G)$ é quadrada e possui dimensão N , com entradas usualmente definidas pela relação abaixo, como descrito por **CORMEN** em [9].

$$a_{ij} = \begin{cases} 1, & (u, v) \in E \\ 0, & \text{caso contrário} \end{cases}$$

No caso de grafos não-direcionados, que efetivamente são o escopo desse trabalho, $A(G) = A(G)^T$, i.e simétrica, e $\sum_{j=1}^N a_{uj} = d(u)$.

Importantes propriedades dos grafos são definidas pelo espectro da matriz de adjacências, porém, algumas propriedades chave são estudadas através da *matriz Laplaciano* do grafo, comumente chamada de $L(G)$. Assumindo uma matriz diagonal $D(G)$ com $D(G)_{uu} = d(u), u \in V$ e dimensão N , a equação 2.4 introduz o Laplaciano de um grafo G qualquer de N vértices, como definido por **CHUNG** em [10].

$$L(G) = D(G) - A(G) \tag{2.4}$$

Tem-se que a matriz $L(G)$ é simétrica, que $\sum_{j=1}^N D(G)_{ij} = 0$ para todo i , e que $L(G)$ é semi-positiva definida, tendo todos seus autovalores reais e não-negativos.

De forma a exemplificar essa representação matricial, pode-se observar $A(G), D(G)$ e $L(G)$ para o grafo completo K_4 mostrado na figura 2.2. As matrizes seguem abaixo.

$$A(K_4) = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad D(K_4) = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \quad L(K_4) = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

2.1.4 O Espectro do Laplaciano

Os resultados desse trabalho estão fundamentados nas propriedades espectrais do Laplaciano, por isso, é importante ponderar sobre o espectro de $L(G)$ através do uso dessa matriz como uma transformação linear aplicada a um vetor.

Sendo $G_{(u,v)}$ o grafo simples que contém somente a aresta (u, v) e assumindo $u < v$, seu Laplaciano $L(u, v)$ é definido em [11] por

$$L(u, v)_{ij} = \begin{cases} 1, & ij = vv \text{ e } ij = uu \\ -1, & ij = uv \text{ e } ij = vu \\ 0, & \text{caso contrário} \end{cases}$$

Admitindo $x \in \mathbb{R}^N$ um vetor com coordenada $x_i \in \mathbb{R}$, pode-se então escrever

$$L(u, v)x = [0, 0, \dots, x_u - x_v, \dots, x_v - x_u, \dots, 0] \quad (2.5)$$

$$x^T L(u, v)x = x_u^2 + x_v^2 - 2x_u x_v = (x_u - x_v)^2 \quad (2.6)$$

De forma efetiva, os valores $x_u - x_v$ e $x_v - x_u$ retratados na equação 2.5 correspondem, respectivamente, à u -ésima e à v -ésima coordenadas do vetor $L(u, v)x$. A equação 2.6 mostra que somente as entradas de x referentes aos vértices que incidem à aresta (u, v) são contempladas pela transformação expressa. Generalizando para um grafo qualquer, podemos estender 2.6 utilizando $L(G) = \sum_{(u,v) \in E} L(u, v)$ e chegando em:

$$x^T L(G)x = \sum_{(u,v) \in E} (x_u - x_v)^2. \quad (2.7)$$

A equação 2.7 é conhecida como expressão quadrática do Laplaciano e permite, efetivamente, encontrar seus autovalores. Usando a equação característica de matrizes, chega-se na expressão 2.10.

$$L(G)x = \lambda x \quad (2.8)$$

$$x^T L(G)x = \lambda x^T x \quad (2.9)$$

$$\sum_{(u,v) \in E} (x_u - x_v)^2 = \lambda x^T x \quad (2.10)$$

As implicações de 2.10 estão expressas em [10]. Assumindo uma ordenação crescente para os autovalores ($\lambda_i < \lambda_j$ quando $i < j$) e que a multiplicidade do autovalor

λ_i é dada por π_{λ_i} , seguem alguns dos resultados importantes:

- $\pi_0 \geq 1$ para todo grafo G .
- O número de componentes conexas de G é dado por π_0 , sendo G conexo se, e somente se $\lambda_2 = 0$;
- $\sum_{i=1}^N \lambda_i = \sum_{v \in V} d(v) = \text{tr}[L(G)]$, pelo teorema do traço de matrizes, que será formalmente referenciado na seção 2.2.

O espectro de um grafo qualquer pode ser obtido através de procedimentos numéricos para a resolução da equação característica, tendo o custo de convergência do método utilizado. Há, porém, resultados bem definidos para classes específicas de grafos que, pela estrutura de suas matrizes de adjacência, oferecem fórmulas fechadas para seus autovalores. O uso de fórmulas fechadas permitem reduzir o custo do cálculo do espectro, substituindo complexos métodos numéricos pela instancição de fórmulas algébricas mais simples. De [10], denotando novamente π_λ ² como a multiplicidade do autovalor λ , explicitam-se as seguintes relações:

- O espectro de K_n tem $\pi_n = n - 1$, $\pi_0 = 1$;
- O espectro de S_n tem $\pi_n = 1$, $\pi_1 = n - 2$, $\pi_0 = 1$.

Tem-se resultados similares para grafos ciclo e grafos caminho, bem como outras classes comuns à literatura de grafos.

2.2 Propriedades básicas do traço de matrizes

Como pode ser visto em [12], o traço de uma matriz, representado por $\text{tr}(\cdot)$, é definido como a soma dos elementos de sua diagonal principal. Supondo $A = \sum_{i=1}^K A_i$ uma matriz quadrada de ordem N e λ_i seu i -ésimo, tem-se as seguintes relações:

$$\text{tr}(A) = \sum_{i=1}^N a_{ii} = \sum_{i=1}^N \lambda_i \quad (2.11)$$

$$\text{tr}(A) = \sum_{i=1}^K \text{tr}(A_i) \quad (2.12)$$

O segundo lado da equação 2.11 é conhecido como *teorema do traço* e relaciona os valores da diagonal principal de A com seu espectro e permite que qualquer função linear dos autovalores de A possa ser calculada através das entradas de sua diagonal principal. O traço é uma *operação linear* da diagonal de A , o que é expressado pela equação 2.12.

²Cabe notar que, usando essa notação, tem-se $\sum_{\lambda \in \Lambda} \pi_\lambda = N$.

2.3 Modelos aleatórios de Redes

A busca por padrões e características específicas da topologia de redes reais que sejam comumente observadas em diversos âmbitos é de fundamental importância para o estudo de Redes Complexas. Essa busca culmina, em diversas ocasiões, no aparecimento de modelos aleatórios capazes de gerar instâncias que herdem algumas destas características e que possam servir como objetos de estudo para a avaliação de métricas e algoritmos.

Os estudos desses modelos permitem que sejam feitas generalizações para propriedades matemáticas, bem como a definição de teoremas, propiciando tratabilidade formal para a análise das redes. Dentre estes modelos, dois tem cunho central perante a literatura: o modelo aleatório de Erdős-Rényi e o modelo de anexação preferencial de Barabási-Albert.

2.3.1 O modelo de Erdős-Rényi

Como inicialmente descrito em [13], o modelo em questão assume um grafo aleatório de N vértices onde cada aresta existe com probabilidade fixa p , sendo comumente conhecido como modelo $G(N, p)$. Definido os parâmetros descritos, o grafo pode ser gerado através de um processo de amostragem de uma variável aleatória com distribuição de Bernoulli para cada possível aresta $e \in V \times V$. O conjunto de arestas contempla aquelas que tiveram amostragem igual a 1, existindo com probabilidade p .

Pelo uso da distribuição de Bernoulli, diversas propriedades matemáticas podem ser simplesmente extraídas como consequência. Algumas destas seguem:

- no regime que $p = 0.5$, todos os grafos de N vértices possuem a mesma probabilidade de aparecerem como instâncias, dado que todas as sequências de combinações entre 0 e 1 são equiprováveis;
- o grau dos vértices segue uma distribuição binomial;
- o grau médio da rede é pN e o número médio de arestas é $\binom{N(N-1)/2}{2}p$.

Além destas propriedades, diversos teoremas em [13] são provados na condição *Asymptotically almost surely*, i.e *a.a.s*. Sendo uma cláusula X definida, ela é dita *a.a.s* quando se verifica que

$$\lim_{N \rightarrow \infty} P(X) = 1$$

onde $P(X)$ representa a probabilidade da cláusula ser verdadeira e N é o tamanho da rede. Esses teoremas analisam regimes de valores de p relacionados com N que, por exemplo, garantem *a.a.s* que o grafo gerado será conexo ou que possuirá uma

componente conexa composta por uma fração constante dos vértices. Esses teoremas permitem que o modelo de Erdős-Rényi, quando instanciado para valores grandes de N , tenha utilidade no estudo das características estruturais por eles garantidas.

Dentre os regimes de p de interesse, cita-se a função $p = \frac{\log(N)}{N}$, que, como mostrado em [13], oferece um *sharp treshold* para a conexidade a.a.s no $G(N, p)$. Essa afirmação indica que qualquer função que decresça não mais rápido que $\frac{\log(N)}{N}$ com N implicará em instâncias conexas, dado o crescimento indefinido do número de nós.

2.3.2 O modelo de Barabási-Albert

O modelo de anexação preferencial de Barabási-Albert, a partir de agora referido como modelo BA e definido em [14], surge no estudo de propriedades de redes reais. O modelo gera redes a partir de um processo de crescimento, onde nós são continuamente adicionados junto com um número fixo de arestas incidentes a eles. Uma representação simples do algoritmo de construção descrito em [14] segue:

1. Um clique inicial de tamanho m_0 é instanciado.
2. Em cada iteração, um vértice v é adicionado ao grafo com $m < m_0$ arestas incidentes a ele e escolhidas ao acaso, com probabilidade $p_{u,v} = \frac{d(v)}{\sum_{i=1}^n d(i)}$, onde n é o número de vértices existentes no grafo antes da adição de v .
3. O algoritmo termina quando $|V| = N$, i.e a rede atinge o tamanho escolhido.

Através de uma equação de continuidade, ALBERT e BARABÁSI provaram que a distribuição de graus de redes obtidas pelo algoritmo segue uma lei de potência, expressa por $P(K = k) \propto k^{-\gamma}$, tornando estas redes livres de escala. É também estipulado pelos autores que a distribuição de lei de potência é obtida mesmo com um crescimento indefinido da rede. Ainda em [14], diversas características topológicas das redes obtidas desse modelo são especificadas, como exemplo o tamanho médio de caminhos, o coeficiente de clusterização e a dispersão de espectro

A propriedade livre de escala tem impacto importante perante redes reais de interesse na literatura. O modelo BA fornece, com a anexação preferencial, uma possível explicação para o aparecimento desta propriedade por um processo dinâmico.

2.4 A entropia de distribuições de probabilidade

Como expresso por COVER e THOMAS em [15], o conceito de entropia é fundamental para o entendimento de problemas de cunho probabilístico. Inicialmente estudado na física, e aprofundado por Shannon, ele descreve uma noção de incerteza

para uma variável aleatória perante uma distribuição de probabilidades. Sendo X uma variável aleatória discreta e $p(X)$ uma distribuição de massa probabilidades de X , a entropia da distribuição é definida em 2.13.

$$H_y(X) = - \sum_{i=1}^N p(x_i) \log_y p(x_i) = \mathbb{E}\left[\frac{1}{\log_y p(x)}\right] \quad (2.13)$$

A base da entropia y é definida através da base do logaritmo utilizado no cálculo da definição e, neste trabalho, usa-se a entropia de base 2, sendo, a partir de agora, suprimida a base dos logaritmos nas equações. Segundo [15], a entropia possui as seguintes propriedades:

- $H(X)$ é concava em p ;
- $H(X) \leq \log_2 N$, com igualdade quando p é uniforme com suporte de N elementos;
- $H(X) > 0$, caso X seja aleatória.

O conceito de entropia e suas características são extensamente utilizados em aplicações de Aprendizado por Máquina, Sistemas de Comunicação, Compressão de Arquivos e diversos outros campos, como definido em [15]. A concavidade da função definida em 2.13 permite que métodos de otimização sejam efetivamente utilizados e garante a convergência de alguns algoritmos relacionados a entropia.

2.4.1 Quantificando a incerteza de distribuições

É central notar que a entropia é uma função da própria densidade de probabilidade e abstrai a variável aleatória associada, *i.e variáveis aleatórias distintas seguindo a mesma distribuição tem o mesmo valor de entropia*.

Na equação 2.13, a função $\log(\frac{1}{p(x)})$ quantifica a incerteza, ou ainda a surpresa, de que uma variável aleatória assumirá um valor x_i associado a uma probabilidade determinada $p(x_i)$, pois decresce o quanto mais provável for o valor referido, *i.e* quanto maior é $p(x)$. Efetivamente, no limite onde $p(x) = 1$, $H(X) = 0$ e a variável regida nestas condições é determinística. No limite que $p(x) = 0$, a parcela $p(x) \log(\frac{1}{p(x)})$ também tende para zero, de forma que valores que apareçam com probabilidade nula não interfiram no cálculo de $H(X)$.

Em suma, $H(X)$ é uma média da incerteza capturada pela distribuição de probabilidades de X , e pode-se pensar que está relacionada com a dispersão da densidade de probabilidade perante seu suporte. Das propriedades citadas anteriormente, vê-se que entropia máxima é obtida ao se considerar uma distribuição uniforme e é

minima quando a densidade de probabilidade colapsa para o regime determinístico, como denotado em [15].

2.4.2 A entropia de von Neumann

O uso de entropia nos sistemas físicos encontra a definição de Shannon utilizada na teoria da informação. Como descrito por PETZ 16, von Neumann desenvolveu este conceito para matrizes de densidade que descrevem conjuntos de misturas de estados quânticos. Como expresso na publicação de VON NEUMANN 4, sendo ρ uma matriz de densidade e λ_i seu i -ésimo autovalor, a entropia de von Neumann é expressa por

$$S = -\text{tr}(\rho \log \rho) = -\sum_{i=1}^N \lambda_i \log(\lambda_i). \quad (2.14)$$

A definição inicial faz uso do logaritmo matricial, de forma que, quando ρ é escrita através de seus autovetores, chega-se ao lado direito da equação.

Em [5], PASSERINI e SEVERINI estendem este conceito para o campo de Redes Complexas, através do Laplaciano de grafos. Os autores especificam que a métrica de von Neumann está relacionada com a regularidade da disposição de arestas e é crescente conforme $|E|$ cresce.

Das análises realizadas em [5], vê-se uma interessante amostragem da entropia de von Neumann para os grafos de 6 vértices, onde, em um processo de crescimento de $|E|$ fixado V , são encontradas as sequências de arestas a serem adicionadas em G que maximizam e minimizam a entropia. Foi observado que as variações mínimas de entropia obtidas são referentes à adição de arestas de forma concentrada e local, aumentando gradualmente número de subgrafos completos em G . Em contra partida, a variação é máxima quando arestas são adicionadas de forma esparsa, formando componentes conexas e provocando, por exemplo, o surgimento de distâncias elevadas.

Além destes resultados, é também descrito em [5] que cliques minimizam a entropia de von Neumann para a cardinalidade de seu conjunto de arestas, mas, concomitantemente, maximizam seu valor para grafos com cardinalidade de vértices N . Essa dualidade deve-se pois cliques são os grafos d -regulares com maior d quando fixado um número de arestas da forma $\frac{n(n-1)}{2}$, i.e maior regularidade, e os grafos de maior número de arestas dado um número fixo de vértices.

O Laplaciano, como citado anteriormente, não pode ser diretamente observado como uma matriz de densidade, posto que $\text{tr}(L(G)) > 1$. A fim de se construir o paralelo com a equação 2.14, é necessário normalizar o Laplaciano o que, segundo

[5], pode ser realizado através da normalização pela soma de graus, expressa em 2.15.

$$\rho_{L(G)} = \frac{L(G)}{\text{tr}(L(G))}. \quad (2.15)$$

Dentre os trabalhos relacionados à entropia de von Neumann para grafos e suas ramificações, cita-se também uma comparação da entropia de von Neumann com uma possível definição da entropia de Shannon para grafos em [6], o uso de uma alternativa aproximada para a análise de redes em [7] e uma aproximação para a entropia de von Neumann de grafos direcionados em [17].

Por fim, cabe notar que a ideia central para se obter a entropia de von Neumann para um grafo é transformar o espectro de seu Laplaciano em uma distribuição discreta de probabilidades. Diferentemente da abordagem de observar os autovalores como variável aleatória, seus valores normalizados é que são vistos como respectivas probabilidades. Nos trabalhos citados, não são feitas relações entre a densidade obtida e possíveis variáveis de suporte, porém, isto não traz nenhum impacto negativo quanto a viabilidade da técnica, dado que a entropia depende, essencialmente, da distribuição de probabilidades, e não dos valores de uma variável aleatória.

2.4.3 Complexidade da entropia pelo cálculo de autovalores

Como pode-se ver em [18], a complexidade de se obter todos os autovalores de uma matriz quadrada de ordem n é da ordem de $O(n^3)$. No regime em que $n > 10^6$, o cálculo dos autovalores é, por vezes, impraticável para certas operações. No caso da entropia de von Neumann, a obtenção do espectro por métodos numéricos de solução para equações características obedecem, em geral, esta complexidade.

2.5 Teoria de Majoração

Um conceito semelhante a definição de entropia de uma distribuição aleatória pode ser observado ao tentar classificar o quão espalhado é um vetor quando comparado a um outro. O estudo desse comportamento deu origem a teoria de Majoração, que, como retratada por MARSHALL *et al.* [19], permitiu o desenvolvimento de diversos teoremas que definem desigualdades e limitantes. Como pode ser visto em [19], sendo dois vetores $x, y \in \mathbb{R}^N$ e assumindo que $x_{(i)}$ é a i -ésima menor entrada de x , diz-se que $x \succ y$ quando as relações 2.16 e 2.17 são verdadeiras.

$$\sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)} \quad 1 \leq k < N \quad (2.16)$$

$$\sum_{i=1}^N x_{(i)} = \sum_{i=1}^N y_{(i)} \quad (2.17)$$

Diversas propriedades seguem quando $x \succ y$, e que são exploradas na referência citada. Dentre estas, cabe, no escopo deste trabalho, retratar aquela referida ao valor de funções convexas quando aplicadas a sequência de valores majorados e a um importante resultado obtido para o espectro de matrizes hermitianas.

Sendo $x, y \in \mathbb{R}^N$ e f uma função convexa qualquer, $x \succ y$ implica em:

$$f\left(\sum_{i=1}^N x_i\right) > f\left(\sum_{i=1}^N y_i\right). \quad (2.18)$$

De fato, uma das condições equivalentes da majoração é que todas as funções convexas no espaço dos vetores em questão respeitam a equação 2.18, como observado em [19].

O teorema de Horn-Schur define o relacionamento entre as entradas da diagonal de uma matriz hermitiana M com seus autovalores. Uma matriz hermitiana é definida no domínio complexo $\mathbb{I}^n \times \mathbb{I}^n$, sendo a generalização complexa de uma matriz real simétrica, i.e as entradas de M respeitam $M_{ij} = \overline{M_{ji}^H}$, onde M^H a matriz conjugada transposta de M . Sendo λ_M o vetor de autovalores de M e m_{ii} o vetor de valores de sua diagonal principal, o teorema diz que $\lambda_M \succ m_{ii}$. Como as matrizes hermitianas generalizam matrizes reais simétricas, o teorema é válido para estas e, em essência, rege $L(G)$.

2.6 Particionamento em subgrafos

Alguns problemas combinatórios em grafos podem ser identificados pela a busca por subgrafos particulares, que detenham certas características. Pela definição de grafos, podemos exprimir $G = (u, v)$ como uma união de seus subgrafos, particionando seu conjunto de arestas. Assumindo um conjunto $[H_1, \dots, H_K]$ de K subgrafos de G , sendo todos disjuntos em arestas entre si, temos um particionamento de G fazendo:

$$G = \bigcup_{i=1}^K H_i. \quad (2.19)$$

A condição de disjunção em arestas faz com que cada aresta de G apareça uma

única vez dentre os subgrafos H_i . Dessa forma, pode-se escrever as matrizes de adjacência e o Laplaciano de G através da soma das respectivas matrizes dos subgrafos. Essas matrizes serão quadradas de ordem N e admitem as equações abaixo.

$$A(G) = \sum_{i=1}^K A(H_i) \quad (2.20)$$

$$L(G) = \sum_{i=1}^K L(H_i) \quad (2.21)$$

De forma a ater-se a definição feita na seção 2.1.3, fazem-se os subgrafos $H_i = (V_i, E_i)$ serem geradores, com os vértices não incidentes as arestas de E_i isolados em H_i . Com essa representação, as matrizes dos subgrafos possuem a mesma dimensão das matrizes de G , como indicado anteriormente.

Unindo as propriedades do traço de matrizes definidas na seção 2.2 com a equação 2.21, tem-se:

$$\text{tr}(L(G)) = \sum_{i=1}^N \lambda_i = \sum_i^K \text{tr}(H_i) = \sum_{j=1}^K \sum_{i=1}^N \lambda_{H_{ji}}. \quad (2.22)$$

A equação 2.22 impõe uma condição de conservação para o espectro de G com qualquer particionamento da forma 2.19, o que é, de certa forma, intuitivo, dado que todas as arestas de G aparecem no particionamento uma única vez. Essa equação implicaria, através de argumentos de linearidade, que qualquer função linear do espectro de $L(G)$ poderia ser efetivamente calculada usando o espectro conjunto dos subgrafos do particionamento.

2.6.1 Particionamentos em classes específicas

Na busca particionamentos de G , pode-se ater àqueles cujos elementos, em totalidade, são isomorfos a classes específicas de grafos. Estes particionamentos acrescentam condições à conservação de arestas expressa em 2.19, as quais podem aumentar a dificuldade de obtenção da subdivisão de forma significativa. Pode-se, por exemplo, buscar por particionamentos onde todos os subgrafos resultantes sejam caminhos, ciclos, estrelas ou cliques. Em complemento, algumas restrições podem, inclusive, impossibilitar a existência de particionamentos não triviais. Um simples exemplo é observado ao se tentar particionar uma árvore qualquer em ciclos, já que a própria definição de árvore implica na ausência de subgrafos ciclo.

No âmbito deste trabalho, as partições completas em estrelas norteiam o desenvolvimento do algoritmo aproximativo proposto para a métrica de von Neumann

e suas propriedades são extensamente exploradas no próximo capítulo. Cabe, por fim, notar que esse tipo específico de particionamento será também referido como *constelação*.

2.6.2 Estrelas e cobertura mínima de vértices

Um dos problemas NP-Completo definidos em [20] consiste em decidir se um grafo possui uma cobertura de vértices de tamanho K . Uma cobertura de vértices consiste em encontrar um subconjunto de vértices $V' \subset V$ tais que todas as arestas $e \in E$ sejam incidentes em ao menos um dos elementos de V' . O problema de otimização NP-Difícil associado consiste na busca pelo menor K que satisfaz o problema de cobertura. Em um paralelo com o particionamento em subgrafos, o problema de encontrar o menor conjunto de estrelas $S = [S_1, \dots, S_k]$, tais que $S_i \subset G$ e $E(S_i) \cap E(S_j) = \emptyset$ dado $i \neq j$, é reduzido ao problema de cobertura mínima de vértices.

Prova. Seja P um conjunto de particionamento em estrelas mínimo e C um conjunto de cobertura mínima. Pela definição, os centros das estrelas não se repetem em P e todas as arestas de G são incidentes nos vértices centrais de P . Supondo que $|P| < |C|$ pode-se enumerar o conjunto de centros de P que cobre todas as arestas de G e é menor que C , o que leva em uma contradição sobre a minimalidade de $|C|$. Em contrapartida, supondo $|P| > |C|$, pode-se enumerar um particionamento $|P'|$ de G com estrelas maximais centradas nos elementos de C . $|P'| = |C|$ porque, caso fosse menor, uma contradição semelhante a expressa anteriormente seria encontrada, e é claramente não maior, já que os $|C|$ são utilizados. Por fim, a suposição $|P| > |C|$ levaria a $|P| > |P'|$, e, conseqüentemente, em uma contradição da minimalidade de P . Chega-se, assim, à conclusão de que $|P| = |C|$, provando a redução. \square

Capítulo 3

Um algoritmo aproximativo para a entropia de von Neumann

Como expresso na seção 2.4.2, calcular a entropia de von Neumann para grandes redes pode ser uma tarefa impraticável, devido ao tempo de solução de equações características. Este capítulo tem como objetivo descrever um algoritmo para obter uma aproximação para métrica de von Neumann através de limites inferiores e superiores obtidos através de particionamentos de grafos em estrelas maximais.

3.1 Ideias preliminares

O teorema do traço impõe uma interessante relação dos autovalores de $L(G)$ e do espectro de um particionamento de G . Pela seção 2.2, vemos que, se a métrica de von Neumann fosse linear, poderíamos calcular exatamente o valor da entropia particionando G em subgrafos, resolvendo as respectivas equações características e aplicando a definição da métrica para a união dos espectros. Porém, este não é o caso, e assim, o que se pode fazer é buscar por aproximações da métrica através de particionamentos.

Os particionamentos oferecem a oportunidade de reduzir o problema de computar os autovalores de $L(G)$ em subproblemas de autovalores de $L(H_i)$, na esperança de que estes subproblemas possam ter soluções de cálculo mais simples. Porém, é intrínseco do processo um compromisso entre a busca por uma decomposição em subgrafos, i.e a complexidade do algoritmo que gera partições respeitando a relação 2.19, e a facilidade de solução dos subproblemas provenientes da decomposição.

3.1.1 Entropia de particionamentos

Admitindo um particionamento P de G que obedece a equação 2.19 e sendo $\lambda = [\lambda_1, \dots, \lambda_T]$ a união dos autovalores de cada subgrafo, define-se a entropia de

particionamento de G induzida por P na equação 3.1.

$$H_P(G) = \sum_{i=1}^T \lambda_i \log\left(\frac{\text{tr}(L(G))}{\lambda_i}\right) \quad (3.1)$$

Vê-se os autovalores provenientes dos subgrafos H_i normalizados pela soma de graus, i.e $\text{tr}(L_G)$, representando os valores de uma distribuição de massa de probabilidades, capaz de ser aplicada na definição de von Neumann. É claro que essa relação é dependente do particionamento escolhido e que, variados os subgrafos, seu valor é alterado.

O processo descrito pode ser observado como uma transformação de G para um grafo de dimensão superior. Sendo K o número de subgrafos usados na decomposição, a equação 3.1 é numericamente equivalente a entropia de von Neumann para um grafo com $\sum_{i=1}^K N_i$ vértices, sendo N_i a cardinalidade do conjunto de vértices de cada subgrafo, e K componentes conexas, onde cada uma destas componentes é representada por um H_i .

É certo que as propriedades das matrizes bloco diagonais e desse tipo de transformação relacionam o valor da entropia de von Neumann usual com aquele calculado para o particionamento. É simples verificar que os autovalores dessa matriz bloco diagonal são iguais à união dos espectros de $L(H_i)$, dado que os blocos não interferem nos autovalores um dos outros.

3.1.2 Subgrafos de espectro simples e complexidade de particionamentos

A simplicidade de encontrar o espectro dos subgrafos envolvidos no particionamento é parte do compromisso da efetividade do método. Da seção 2.1.4, tem-se algumas das classes de grafos específicas que possuem fórmulas fechadas para seus autovalores. Traz-se, assim, a ideia de buscar por partições isomorfas a estas classes, permitindo a substituição da usual solução de equações características pela instanciação de fórmulas específicas.

Observando as fórmulas fechadas expressas para espectro de $L(G)$ na seção 2.1.4, é possível propor particionamentos de grafos em cliques, estrelas, caminhos e ciclos. Opta-se, neste trabalho, por analisar a viabilidade do uso de estrelas e cliques. Encontrar por grafos completos como subgrafos de um grafo qualquer é sensivelmente mais complexo do que procurar por uma decomposição em estrelas. Como denotado em [20], encontrar um clique de tamanho k em um grafo é um problema NP-Completo, e, conseqüentemente, buscar por um particionamento não trivial ¹ de

¹O conjunto de todas as arestas de G é um particionamento em subgrafos K_2 .

G em cliques é NP-Difícil.

Em contrapartida à busca por cliques, um subgrafo estrela de G pode ser simplesmente obtido por escolher um vértice e listar todas arestas incidentes a ele. Esse tipo de construção permite a concepção de um algoritmo para decomposição de G com complexidade tratável, e que será explorado na seção 3.5. Porém, é de se esperar que a cardinalidade do particionamento obtido influencie no valor de entropia aproximado, cabendo notar o problema de encontrar o particionamento mínimo em estrelas de G é NP-Difícil, como descrito na seção 2.6.2.

Ainda que um particionamento mínimo em estrelas não possa, a priori, ser computado em tempo polinomial para um grafo qualquer, pode-se encontrar particionamentos úteis para o cálculo da entropia. Por este motivo, escolhe-se a entropia de particionamento em estrelas, que, a partir deste ponto, será referida como $H_s(P)$.

3.2 Descrevendo limitantes superiores

A análise do particionamento como limitante depende, profundamente, do modo como uma cobertura completa de arestas em estrelas impacta a entropia de particionamento. Para obter uma perspectiva significativa, aplica-se na equação 3.1 o padrão do espectro completo de uma constelação. Os autovalores $\lambda_i = 1$ fazem $\lambda_i \log \lambda_i = 0$, de forma que os únicos autovalores efetivos são os relacionados ao número de vértices da estrela em questão. Denotando C_i o número de vértices da i -ésima estrela de um particionamento P , a entropia de estrelas, termo que se referirá a entropia de particionamento em estrelas de agora em diante, é definida na equação 3.2.

$$H_s(P) = \log 2M - \frac{1}{2M} \sum_{i=1}^K C_i \log(C_i) \quad (3.2)$$

Indaga-se, em condições qualitativas, quais os efeitos da cardinalidade da constelação no valor de H_s , bem como seu relacionamento com a métrica de von Neumann. De forma a encontrar estas relações, é necessário descrever as propriedades chave de uma partição em estrelas, e usá-las para análise necessária.

3.2.1 Entropia dos graus normalizados

Antes de prosseguir com a análise da entropia de particionamento, será definida uma métrica auxiliar para algumas das provas que asseguram as propriedades de H_s como limitante superior para H_n . Sendo $D = [d(u_1), \dots, d(u_n)]$, $u_i \in v$ a sequência de graus dos vértices de G , introduz-se na equação 3.3 a *entropia dos graus normalizados*,

H_d .

$$H_d(G) = \log 2M - \frac{1}{2M} \sum_{i=1}^N d(u_i) \log(d(u_i)) \quad (3.3)$$

Do teorema de Horn-Schur descrito na seção 2.5, pode-se derivar um lema que relaciona H_d e H_n .

Lema 1. $H_d(G) \geq H_n(G)$, sendo a entropia de graus normalizados um limitante superior para a entropia de von Neumann.

Demonstração. A diferença entre as entropias é expressa por:

$$\Delta_{dn}(G) = \frac{1}{2M} \left(\sum_{i=1}^N \lambda_i \log \lambda_i - \sum_{i=1}^N d(u_i) \log(d(u_i)) \right) \quad (3.4)$$

Pelo teorema de Horn-Schur, como $L(G)$ é simétrica, seus autovalores majoram as entradas de diagonal principal, as quais, precisamente, representam os graus dos vértices de G . Diretamente da definição de majoração, tem-se que $\sum_{i=1}^N \lambda_i \log \lambda_i > \sum_{i=1}^N d(u_i) \log(d(u_i))$, já que $x \log(x)$ é côncava, levando em $\Delta_{dn}(G) > 0$ e $H_d > H_n$. \square

3.2.2 Cardinalidade da união dos espectros de estrelas

A formula específica dos autovalores de estrelas traz uma interessante consequência para o particionamento de estrelas, definida no teorema 1.

Teorema 1. *Seja P qualquer particionamento completo das arestas de E em estrelas, o número R de elementos não nulos da concatenação dos espectros das estrelas de P é constante, independente da cardinalidade $|P|$ do particionamento, e igual a $|E| = M$, i.e $R = M$.*

Prova. Seja P um particionamento em estrelas qualquer de um grafo $G = (V, E)$ com $|P| = K$ e $C = [C_1, \dots, C_T]$ o conjunto obtido da concatenação dos autovalores das estrelas de P com $|C| = T$ ordenado em ordem decrescente. Pela fórmula dos autovalores de estrelas, C possui K elementos maiores que 1, i.e autovalores respectivos ao tamanho da estrela, e K autovalores iguais a 0, i.e pois cada estrela tem exatamente 1 autovalor 0 em seu espectro. Sendo R o número de elementos de C maiores que 0, $R - K$ é o número exato elementos iguais a 1. Novamente, pela fórmula dos autovalores de estrelas, sabe-se que S_n possui $n - 2$ autovalores de valor 1, assim, $R - K = \sum_{i=1}^K (C_i - 2) = (\sum_{i=1}^K C_i) - 2K$. Aplicando o teorema do traço, tem-se:

$$\sum_{i=1}^T C_i = \sum_{i=1}^R C_i = \sum_{i=1}^K C_i + \sum_{i=1}^K (C_i - 2) = 2M \quad (3.5)$$

$$2 \sum_{i=1}^K C_i = 2M + \sum_{i=1}^K 2 = 2M + 2K \quad (3.6)$$

$$\sum_{i=1}^K C_i = K + M \quad (3.7)$$

$$R - K = \sum_{i=1}^K C_i - 2K = K + M - 2K = M - K \quad (3.8)$$

$$R = M \quad (3.9)$$

□

Efetivamente, o número total T de autovalores de um particionamento P é, por definição, $T = R + K$. Aplicando a equação 3.9, tem-se que $T = M + K$, mostrando, também, que o número total de autovalores cresce linearmente com K .

3.2.3 Partições maximais e a redução da entropia

Um particionamento de estrelas $C_G = [C_1, \dots, C_K]$ é dito *maximal* quando nenhum par de estrelas $C_i, C_j \in C_G$ compartilha os vértices centrais, e, quando uma estrela C_i é definida somente por uma aresta $e = (u, v)$, nem v nem u existem como centros em C_G . Diretamente desta definição segue o teorema 2.

Teorema 2. *O particionamento de G em estrelas de cardinalidade mínima é maximal.*

Prova. Seja P_{min} um particionamento de cardinalidade mínima de G . Supondo P_{min} não maximal, é possível obter um outro particionamento P' unindo quaisquer que fossem as estrelas de P_{min} que compartilhassem centros, fazendo $|P'| < |P_{min}|$, o que é uma contradição.

□

Dado que particionamentos mínimos são também maximais, cabe estipular se o caráter maximal anda em harmonia com particionamentos de entropia baixa. De forma a aclarar esta questão, segue o teorema 3.

Teorema 3. *Seja $C = [C_1, \dots, C_K]$ a sequência de autovalores diferentes de 1 de um particionamento P não maximal de G e $C_m = [C_{m1}, \dots, C_{ml}]$ a sequência para o particionamento maximal P' obtido de P . Tem-se que a entropia do particionamento de P' é menor que a entropia de particionamento de P .*

Demonstração. Supondo que $|C| = |C_m| + 1$, de forma que C_m é obtido ao se juntar duas estrelas $i, j \in P$ de centros iguais. Tem-se que $C_{mi} = C_i + C_j - 1$ e que a diferença de entropia dos particionamentos é expressa por:

$$\Delta H_{PP'} = H_C - H_{C_m} = \frac{1}{2M}(C_{mi} \log(C_{mi}) - C_i \log(C_i) + C_j \log(C_j)) \quad (3.10)$$

$$\Delta H_{PP'} = \frac{1}{2M}((C_i + C_j - 1) \log(C_{mi}) - C_i \log(C_i) + C_j \log(C_j)) \quad (3.11)$$

$$\Delta H_{PP'} = \frac{1}{2M}(C_i \log \frac{C_{mi}}{C_i} + C_j \log \frac{C_{mi}}{C_j} - \log C_{mi}) \quad (3.12)$$

Como, na equação 3.12, $C_i \log \frac{C_{im}}{C_i} + C_j \log \frac{C_{jm}}{C_j} > C_i + C_j > \log(C_i + C_j) > \log(C_{mi}) > 0$, tem-se que $\Delta H_{PP'} > 0$, de forma que $H_P > H_{P'}$. Por fim, um particionamento não maximal com cardinalidade $|P| > |P'|$ pode ser reduzido a P' por sucessivas uniões de estrelas de mesmo centro, reduzindo continuamente a entropia e fazendo $H_P > H_{P'}$ no caso geral. \square

3.2.4 Entropia de constelação como limitante superior

A relação qualitativa entre a entropia de particionamento em estrelas e a entropia de von Neumann é expressa pela conjectura 1.

Conjectura 1. *Seja P um particionamento em estrelas maximais qualquer de $G = (V, E)$ com $|V| = N, |E| = M$, tem-se que $H_s(P) \geq H_n(G)$.*

Argumentação. Como expresso na seção 3.1.1, a entropia de um particionamento em estrelas P , $|P| = K$ pode ser vista como a entropia de von Neumann calculada para um grafo $|G'|$ com K componentes conexas. Como expresso na seção 3.2.2, existem exatamente K autovalores do espectro de G' que são maiores que 1, e são centros de estrelas. Calculando a fórmula da diferença entre as entropias, tem-se:

$$\Delta_{G'G} = \log(2M) - \frac{1}{2M} \sum_{i=1}^K \lambda_i \log(\lambda_i) - \log(2M) + \frac{1}{2M} \sum_{i=1}^N C_i \log(C_i) \quad (3.13)$$

$$\Delta_{G'G} = \frac{1}{2M} \left(\sum_{i=1}^N \lambda_i \log(\lambda_i) - \sum_{i=1}^K C_i \log(C_i) \right) \quad (3.14)$$

Sabe-se que $\sum_{i=1}^N \lambda_i - \sum_{i=1}^K C_i = M - K$. Assumindo que G é conexo sem perda de generalidade, tem-se que, como P é maximal, $K \leq (N - 1)$, com igualdade somente quando G é um clique de tamanho N . Portanto, sabe-se que existe um $J < (N - 1)$ tal que $\sum_{i=1}^J \lambda_i > \sum_{i=1}^K C_i$, fazendo com que a sequência C_i dos autovalores das K estrelas seja majorada de forma fraca pelos autovalores de $L(G)$. Assim, diz-se que

$\sum_{i=1}^N \lambda_i \log(\lambda_i) > \sum_{i=1}^K C_i \log(C_i)$, fazendo $\Delta_{G'G} > 0$ devido à convexidade da função $x \log(x)$. □

3.3 Conjectura do particionamento mínimo

De resultados observados em grafos distintos, vê-se que particionamentos máximos de menor cardinalidade aparecem com menor entropia. Não obstante, não se encontrou exemplos cuja constelação de cardinalidade mínima obtida não fosse a constelação de entropia mínima do grafo. Esses indícios permitem argumentar que é provável que a busca por particionamentos de cardinalidade mínima seja, de certa forma, equivalente à busca pelo particionamento de entropia mínima. Consequentemente, o particionamento que melhor aproxima a entropia de von Neumann, caso a conjectura 1 prove-se verdadeira em todos os casos, é o particionamento de cardinalidade mínima. Dessas observações, conjectura-se:

Conjectura 2. *O particionamento de entropia mínima de G é o particionamento de cardinalidade mínima e, sendo dois particionamentos P, P' de G em estrelas máximos de menor entropia possível, $|P| \leq |P'|$ implica em $H_s(P) \leq H_s(P')$.*

Argumentação. Sabe-se que a soma S_k dos autovalores diferentes de 1 em um particionamento de cardinalidade k é constante e inferior a soma obtida de particionamentos com $k + 1$ elementos em uma unidade. Um particionamento de tamanho K com entropia mínima representa a decomposição mais concentrada (i.e regular) de S_k em k valores. A diferença unitária entre as somas totais não seria capaz de fornecer uma decomposição mais concentrada, de forma que a entropia seria, então, crescente perante a cardinalidade dos particionamentos de entropia mínima com número de elementos fixados. □

3.4 Descrevendo limitantes inferiores

Após a seção 3.2, tem-se indícios de que a entropia de estrelas é um limitante superior para a entropia de von Neumann, dado qualquer particionamento. Seria interessante, porém, obter também um limite inferior dado uma subdivisão em estrelas, já que assim, ao resolver o problema de encontrar um particionamento em estrelas, ambos limites seriam obtidos. Dessa motivação surge a ideia de remover os elementos iguais a 1 do espectro de particionamento gerado, e aplicar a definição da entropia sobre uma nova distribuição renormalizada. Da seção 3.2.2, sabe-se que $\sum_{i=1}^k C_i = M + K$, propõe-se, o limitante inferior expresso na equação 3.15.

$$l_{inf} = - \sum_{i=1}^K \frac{C_i}{M+K} \log\left(\frac{C_i}{M+K}\right) = \log(M+K) - \sum_{i=1}^K C_i \log\left(\frac{C_i}{M+K}\right). \quad (3.15)$$

Conjectura-se, assim, a seguinte propriedade:

Conjectura 3. *Sendo $C = [C_1, \dots, C_K]$ a sequência dos K autovalores obtidos de um particionamento maximal P qualquer de um grafo $G = (V, E)$ com $|E| = M$, a função 3.15 é um limitante inferior para a entropia de von Neumann $H_n(G)$.*

Argumentação. A expressão 3.15 pode ser vista como a entropia de von Neumann obtida para um grafo de K vértices e $\lceil \frac{M+K}{2} \rceil \leq M$ arestas, dada a equação 3.7. Como a entropia de von Neumann é crescente perante adição de arestas, tem-se $H_n(G) \geq l_{inf}$. □

3.5 Algoritmos de particionamento em estrelas

De forma geral, um particionamento em grafos estrela pode conter subgrafos de tamanho variado, respeitando duas restrições: a disjunção de arestas entre estrelas distintas e a própria restrição sobre os graus dos vértices nas estrelas, proveniente da definição de S_n . Todavia, como proveniente dos resultados da seção 3.2, acrescentaremos mais uma restrição para os particionamentos de interesse: quer-se encontrar por particionamentos de estrelas maximais.

3.5.1 Estrelas maximais pela ordenação de vértices

Um simples algoritmo capaz de gerar um particionamento em estrelas maximais qualquer pode ser descrito através de um processo ordenado de sucessivas remoções de vértices e de suas arestas incidentes.

Dado o conjunto V e uma ordenação $O = [O_1, \dots, O_n]$, basta iterar sobre elementos de V seguindo O , criando na i -ésima iteração a estrela com centro O_i . Todas as arestas incidentes a O_i que ainda estão em E serão adicionadas a esta estrela e esse respectivo subgrafo será removido de G . A ordenação O define quais vértices aparecem como centro de estrelas e qual conjunto de arestas é contemplado por cada uma, já que nós processados primeiro possuem maior prioridade em receber suas arestas incidentes. O processo descrito é formalizado no algoritmo 1.

É correto afirmar que todas as estrelas obtidas pela aplicação do algoritmo 1 são maximais no particionamento. Cada nó de G só é processado uma única vez, e, após o vértice v_i ser definido como centro de estrela, todas as arestas incidentes a v_i já

Algorithm 1 Partição em Estrelas por remoções sucessivas

```
1: procedure CONSTELLATION( $G$ ,  $vertexOrder$ )
2:    $vertexQueue \leftarrow vertexOrder(G.V)$ 
3:    $edgeSet \leftarrow \emptyset$ 
4:    $constellation \leftarrow \emptyset$ 
5:   while  $|edgeSet| < |E|$  do
6:      $v \leftarrow vertexQueue.pop()$ 
7:      $vertexStar \leftarrow InitStar(v)$ 
8:     if  $degree(v) \neq 0$  then
9:       for  $u \in neighbors(v)$  do
10:        if  $(u, v) \notin edgeSet$  then
11:           $vertexStar.V \leftarrow vertexStar.V + \{u\}$ 
12:           $vertexStar.E \leftarrow vertexStar.E + \{(u, v)\}$ 
13:           $edgeSet \leftarrow edgeSet + \{(u, v)\}$ 
14:        if  $|vertexStar.V| \geq 2$  then
15:           $constellation \leftarrow constellation + \{vertexStar\}$ 
16:   return  $constellation$ 
```

estão presentes no particionamento. Como nenhum nó pode aparecer mais de uma vez como centro de estrela, a condição de maximalidade estipulada é respeitada.

A ordem inicial dos vértices é de importância central para a cardinalidade do particionamento gerado. Isto faz com que a busca por particionamentos de cardinalidades distintas demande a execução do algoritmo para diferentes ordenações de vértices. Além disso, nada pode ser garantido nas cardinalidades resultantes da troca de ordenação inicial, de forma que, caso cardinalidades baixas sejam desejadas, é necessário que as ordenações iniciais sejam norteadas por alguma função informativa dos vértices.

3.5.2 Uma heurística para a busca de estrelas maximais

Da análise do algoritmo 1 surge espaço para a proposição de uma heurística que permita encontrar particionamentos de cardinalidade baixa. Ainda observando a estratégia de remoções sucessivas, é intuitivo pensar que, caso os vértices de maior grau sejam priorizados, i.e sejam iterados primeiro, um significativo número de arestas será coberto em cada iteração. Por conseguinte, caso as iterações sejam efetivas em reduzir o número de arestas faltantes, é de se esperar que um pequeno número de iterações seja necessário, repercutindo em um particionamento de cardinalidade baixa.

De forma paralela, pode-se ver a ordenação dos vértices por grau como uma tentativa gulosa de maximizar o tamanho das estrelas criadas. De fato, nessa estratégia, o vértice de maior grau e todas suas arestas incidentes seriam a primeira estrela a ser gerada, a qual é o maior subgrafo estrela de G existente. Porém, as

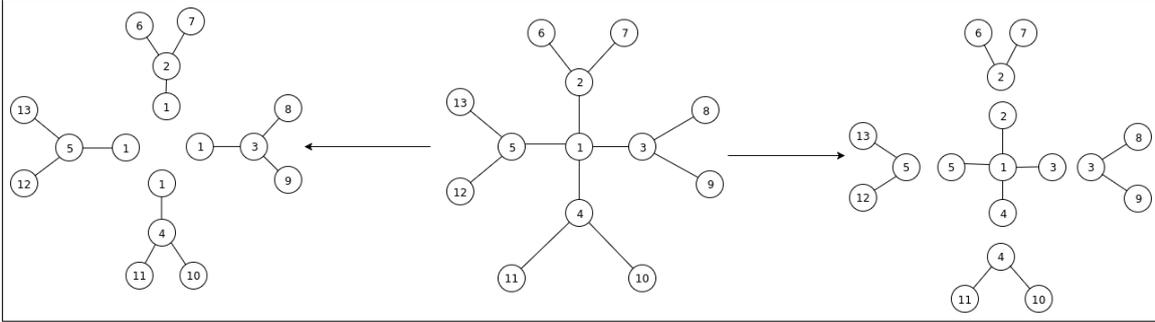


Figura 3.1: Um grafo (meio) com entropia de von Neumann $H_n = 3.05$ decomposto em duas partições distintas: a partição de cardinalidade mínima (esquerda), com entropia de estrelas $H_s = 3.25$, e a partição obtida pela heurística do algoritmo 2 (direita), com $H_s = 3.31$.

sucessivas remoções reduzem continuamente os graus de cada vértice, de forma que, para $i > 1$, a garantia de que a maior estrela possível para a i -ésima iteração será encontrada é quebrada. Não obstante, é natural pensar que, caso os vértices na i -ésima iteração sejam ordenados pelo número de arestas que ainda existem após i remoções, i.e seu *grau instantâneo*, a estrela gerada nessa iteração será o maior subgrafo estrela existente em $G - C^i$, onde C^i é o conjunto de particionamento gerado até então.

Em suma, a heurística proposta consiste na busca do particionamento que maximiza localmente o tamanho de estrelas. Ao ser comparada com a simples ordenação de vértices, é correto afirmar que existem ocasiões onde a cardinalidade obtida por uma ordenação qualquer será menor do que a aquela obtida pela heurística. Pode-se mostrar também que o particionamento mínimo não necessariamente é aquele que maximiza localmente as estrelas de cada iteração com um simples contra-exemplo mostrado na figura 3.1. Entretanto, a heurística traz uma abordagem que permite aplicar o particionamento em grandes redes, respeitando o compromisso entre cardinalidade do particionamento e tempo de execução.

A formalização do processo descrito segue, em pseudo-código, o algoritmo 2.

A implementação faz uso de uma fila de prioridades, onde os vértices tem como prioridade inicial seus respectivos graus. Durante a execução do algoritmo, as estrelas adicionadas no particionamento decrementam a prioridade dos vértices não centrais em uma unidade. A atualização da prioridade reorganiza a fila de acordo com o grau dos vértices no grafo $G - C_i$, onde C_i é o conjunto de estrelas da iteração i .

3.5.3 Complexidade assintótica

A complexidade assintótica do algoritmo 1 depende, fundamentalmente, do método de marcação das arestas que já foram processadas e da complexidade de verificação

Algorithm 2 Partição em estrelas por graus máximos

```
1: procedure CONSTELLATION( $G$ )
2:    $vertexPriorityQueue \leftarrow (V_G, d(V))$ 
3:    $edgeSet \leftarrow \emptyset$ 
4:    $constellation \leftarrow \emptyset$ 
5:   while  $|edgeSet| < |E|$  do
6:      $v \leftarrow vertexPriorityQueue.pop()$ 
7:      $vertexStar \leftarrow InitStar(v)$ 
8:     if  $degree(v) \neq 0$  then
9:       for  $u \in neighbors$  do
10:        if  $(u, v) \notin edgeSet$  then
11:           $vertexStar.V \leftarrow vertexStar.V + \{u\}$ 
12:           $vertexStar.E \leftarrow vertexStar.E + \{(u, v)\}$ 
13:           $edgeSet \leftarrow edgeSet + \{(u, v)\}$ 
14:           $vertexPriorityQueue.update(u, u.priority - 1)$ 
15:        if  $|vertexStar.v| \geq 2$  then
16:           $constellation \leftarrow constellation + \{vertexStar\}$ 
17:   return  $constellation$ 
```

destas marcações. Pode-se usar uma estrutura em *hash*, com complexidade de busca $O(1)$ no caso médio e $O(n)$ no pior caso. Como cada aresta é checada duas vezes, a complexidade assintótica torna-se $O(|E|)$ no caso médio e $O(|E|^2)$ no pior caso.

Vê-se que o algoritmo 2 é uma modificação do algoritmo 1, removendo a ordenação inicial de vértices e usando uma fila de prioridade, de forma a manter os vértices ordenados por graus instantâneos eficientemente. A fila de prioridades escolhida é descrita como uma fila de prioridade de altura limitada, definida em [21]. A escolha desse tipo de fila de prioridade se dá pois o valor de graus instantâneos p_i dos vértices de G obedece $0 \leq p_i \leq N - 1$.

A complexidade deste algoritmo é dependente do processo de atualização de prioridade dos graus. As filas de prioridade de altura limitada funcionam em dois níveis: o nível mais externo da estrutura é indexado com os valores concretos das prioridades representadas, e o nível interno guarda os nós que possuem o respectivo valor de prioridade. É guardado um ponteiro para a classe não vazia de maior prioridade, fazendo a busca da prioridade máxima $O(1)$. Usando tabelas *hash* como estruturas interiores permite obter tempo médio de adição e remoção constantes, mas lineares no pior caso. Na i -ésima iteração do algoritmo, tem-se um máximo de $N - 1 - i$ vértices a sofrerem atualização. O pior caso aparece em todas as iterações quando se tem grafo completo como entrada. Analisando esta situação e fazendo T o número total de atualizações, tem-se que:

$$\begin{aligned} T &= \sum_{i=1}^N (N - 1 - i) \\ T &= N^2 - N - \frac{N^2 + N}{2} \\ T &= \frac{N^2 - 3N}{2}. \end{aligned} \tag{3.16}$$

Da equação 3.16 vê-se que, assumindo o caso médio de remoção e adição da tabelas *hash*, o algoritmo tem complexidade $O(N^2)$.

Capítulo 4

Avaliação Experimental

Este capítulo tem como objetivo avaliar os resultados obtidos pela a execução do algoritmo proposto como alternativa para o cálculo da entropia de von Neumann, tanto em sua efetividade como métrica, quanto sua tratabilidade computacional.

4.1 Implementação do processo experimental

Os algoritmos de particionamento descritos foram implementados utilizando a linguagem *Python*, versão 3.6, com o auxílio do framework de computação numérica NumPy [22] e a biblioteca de análise de redes *graph-tool* [23]. A biblioteca *graph-tool* funciona como uma interface para Python com a biblioteca *boost::Graph*, implementada em C++ e fornece diversas ferramentas para instanciar grafos através de modelos probabilísticos e de arquivos externos. O framework NumPy foi usado para avaliar as funções matemáticas em questão, bem como calcular autovalores através de métodos numéricos implementados no framework. Os gráficos desta seção foram gerados através da Matplotlib [24], uma biblioteca para a linguagem Python voltada para a produção de figuras e imagens com qualidade de publicação.

4.2 Descrição do processo de avaliação

De forma a obter uma visão geral das propriedades do algoritmo proposto, utiliza-se, na avaliação dos resultados, redes obtidas pela instanciação dos modelos de Erdős-Reyní ($G(N, p)$) e de Barabasi-Albert (BA) descritos na seção 2.3, bem como redes amostradas na literatura. Os modelos de rede permitem traçar curvas de médias empíricas, analisando o comportamento das métricas de interesse com a variação do tamanho da rede. Em complemento, as redes amostradas na literatura oferecem medições pontuais das métricas envolvidas, e permitem uma verificação sobre a viabilidade do uso das técnicas em questão para redes reais.

4.2.1 Medidas de entropia calculadas

Na análise da medida proposta, as seguintes métricas foram mensuradas:

- A entropia de estrelas H_{s1} usando o algoritmo de particionamento 1 com ordenação inicial de grau e seu respectivo limitante inferior;
- A entropia de estrelas H_{s2} usando a heurística de maximização do grau instantâneo, descrita no algoritmo de particionamento 2 e seu respectivo limitante inferior;
- A entropia de graus normalizados H_d ;
- A entropia de von Neumann H_n , calculada somente para as redes obtidas de modelos aleatórios em regimes de tamanho viáveis.

A escolha desses valores permite observar quantitativamente o comportamento da aproximação quando variadas as estratégias de partição em estrelas, bem como observar a efetividade de um particionamento ao fornecer limitantes superiores e inferiores, em comparação a entropia de graus normalizados.

4.2.2 Medições de entropia

Segue-se, para os modelos aleatórios, que o processo de medição busca médias empíricas. Supondo $G = M(\theta)$ o grafo resultante de um dos modelos probabilísticos utilizados sobre os parâmetros θ e seja $f(G)$ uma determinada função que deseja ser medida para o grafo, a média amostral para $f(G)$ perante θ é obtida pela equação 4.1.

$$\frac{\sum_{i=1}^T f(M(\theta))}{T} \quad (4.1)$$

Como os modelos são probabilísticos, as redes geradas por cada instanciação de um grupo de parâmetros tendem a ser diferentes, e, por conseguinte, é de se esperar que valores distintos de entropia sejam obtidos. A busca pela média empírica permite suavizar as flutuações de entropia calculadas e permitir uma análise de redes de tamanhos díspares com significado.

4.2.3 Medições do tempo de execução

Como, para sistemas computacionais genéricos, a execução de processos e programas dependem das condições atuais de execução, como a disponibilidade de memória e o uso do disco, bem como do número de tarefas sendo executadas simultaneamente

pelo processador, o tempo real de execução de um processo também sofre flutuações. Conseqüentemente, médias empíricas foram utilizadas para obter medições comparativas entre o tempo de execução do algoritmo de entropia de particionamento em estrelas com o cálculo da entropia de von Neumann. Para uma rede G qualquer, o tempo médio de execução de uma métrica é expresso na equação 4.2.

$$\tau_M = \frac{\sum_{i=1}^t \tau(f(G))}{t} \quad (4.2)$$

No caso das redes amostradas da literatura, a métrica definida em 4.2 é precisamente utilizada, dado que as avaliações são pontuais para cada grafo. Não obstante, na análise das redes provenientes em modelos, a média dos tempos médios é utilizada, aplicando a equação 4.2 em 4.1.

4.2.4 Instâncias de redes reais

De forma a utilizar o algoritmo proposto para obter aproximações da entropia de von Neumann, as instâncias de redes reais escolhidas possuem um regime de número de vértices no qual o tempo necessário para a obtenção de autovalores de $L(G)$ torna a métrica de von Neumann impraticável de ser computada. Uma breve descrição destas redes segue nesta seção.

- Rede de colaboração de publicações em Astrofísica (astro-ph), com referência em [25].
- Rede de co-autoria em publicações da área de matéria condensada (cond-mat), amostrada nos anos de 1999, 2003 e 2005 e inicialmente descrita por NEWMAN em [25].
- Rede social do YouTube (com-youtube), cujos relacionamentos representam relações de amizade dentro da plataforma por seus usuários, descrita em [26].
- Rede de estradas da Califórnia (roadNet), cujos relacionamentos identificam pontos de encontro entre ruas, definida em [27].

4.3 Eficiência dos limitantes para a aproximação

A análise do comportamento dos limitantes segue duas fases distintas. Inicialmente, os modelos aleatórios são analisados em regimes onde H_n pôde ser computado, fornecendo valores referências diretos para comparação. Na segunda etapa, H_n deixa de ser computado e os algoritmos são redes reais de cardinalidades elevadas.

4.3.1 Redes aleatórias com referencial

As imagens 4.1 e 4.2 mostram os resultados obtidos do cálculo das médias empíricas para redes de ambos os modelos, com N variando de 10 a 500. Em ambas as imagens, as curvas de cores iguais mostram os limitantes inferiores e superiores da mesma métrica, cabendo notar que as curvas de *ordenação por grau* e de *graus máximos instantâneos* correspondem, respectivamente, aos particionamentos resultantes dos algoritmos 1 e 2.

É visível que os valores obtidos como limitantes superiores pelas estrelas estão consideravelmente mais distantes de H_n do que H_d . Além disso, vê-se claramente que, para o modelo BA , a diferença na estratégia de particionamento pouco afeta o valor de entropia obtido e que o valor de H_n é muito próximo de H_d . Em contrapartida, no caso do $G(N, p)$, vê-se um espaçamento efetivo entre H_n e H_d e entre os limites inferiores dos particionamentos distintos, porém a diferença entre H_{S1} e H_{S2} é somente um pouco mais acentuada que a do modelo BA .

A semelhança extrema entre H_{S1} e H_{S2} observada para o modelo BA pode ter origem na topologia livre de escala da rede, de forma que a heurística proposta não forneça melhoria expressiva na cardinalidade dos particionamentos obtidos. Pode-se, inclusive, cogitar que os particionamentos gerados sejam muito semelhantes, tendendo a igualdade, o que, na perspectiva da heurística, poderia ser explicado ou por uma diferença considerável do grau dos vértices de maior grau com os de grau intermediário, ou por um isolamento dos vértices de alto grau entre si, i.e baixo número de vizinhos em comum. Pela distribuição dos graus em lei de potência e pela anexação preferencial, a primeira explicação aparece mais plausível.

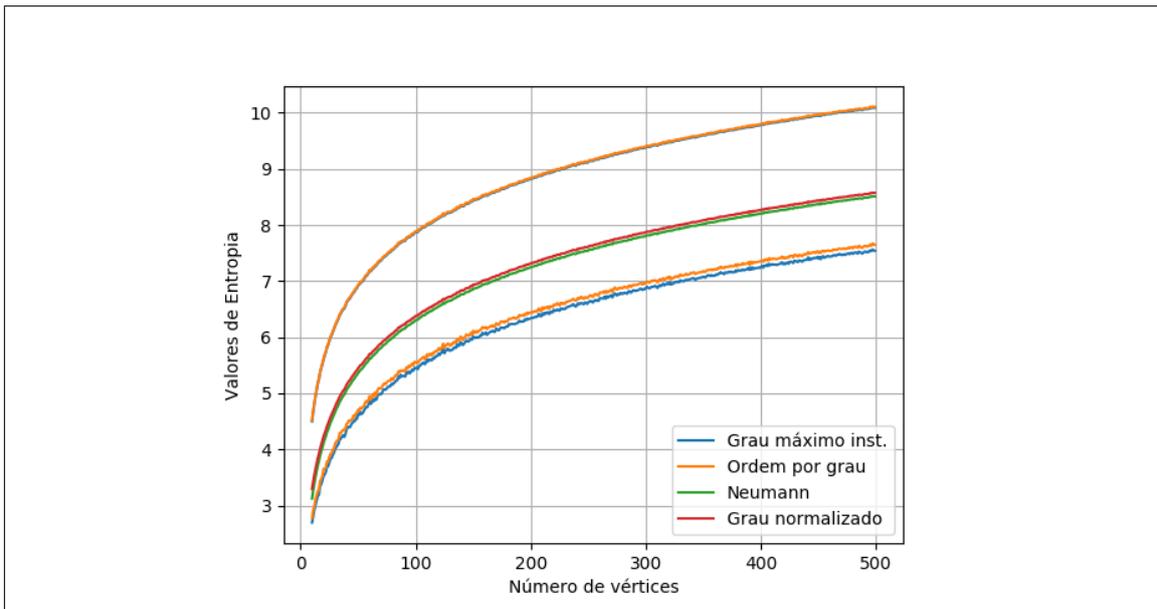


Figura 4.1: Medições empíricas do modelo BA , com $m = 5$. Curvas de mesma cor representam os limites superiores e inferiores obtidos dos particionamentos.

Não obstante, para o modelo $G(N, p)$, a possível monotonia da topologia perante o regime proposto poderia explicar a diferença sutil entre os limitantes superiores de estrelas. Como, no modelo, todas as arestas existem com mesma probabilidade, é de se esperar que a topologia do $G(N, p)$ não possua muita variabilidade, o que é embasado pela distribuição binomial dos graus, que possui cauda curta. Em complemento, é de se esperar que o regime $p = \frac{\log(N)}{N}$ leve a uma topologia esparsa, já que $\lim_{N \rightarrow \infty} p = 0$, o que, na perspectiva da heurística, não forneça as características topológicas que ofereçam diferenças expressivas no tamanho das estrelas encontradas.

Ainda que, para ambos os modelos analisados, a heurística não forneça uma redução muito significativa, ela gera, de fato, para o regime de parâmetros analisado, particionamentos com entropia não menor do que o fornecido pela ordem decrescente dos graus.

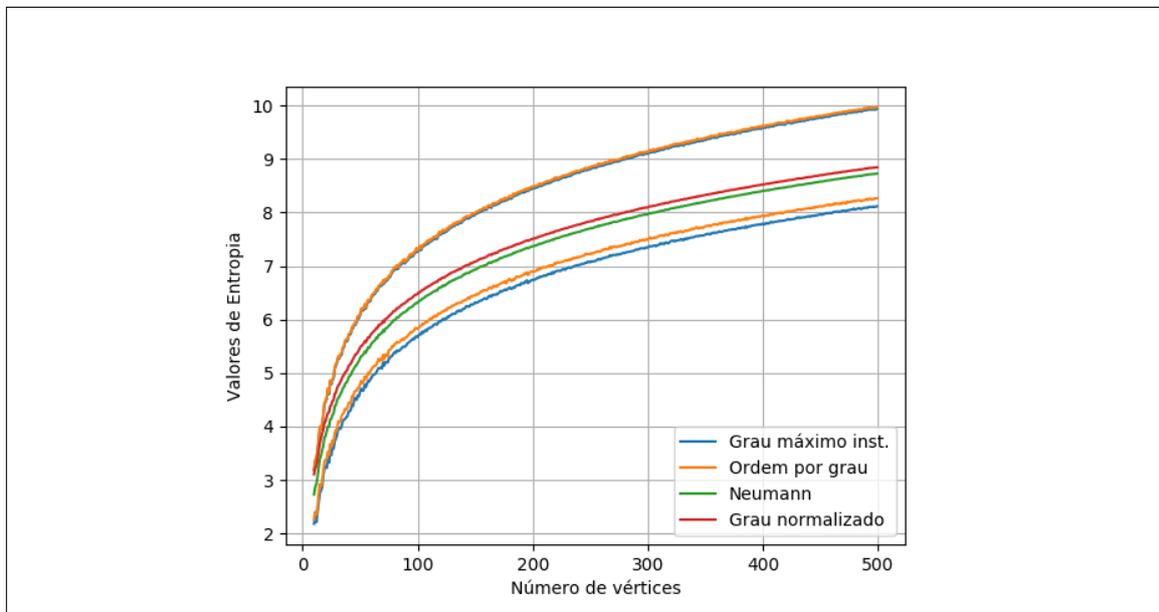


Figura 4.2: Medições empíricas do modelo $G(N, p)$ no regime $\frac{\log p}{p}$. Curvas de mesma cor representam os limites superiores e inferiores obtidos dos particionamentos

Em complemento ao distanciamento dos limitantes entre si, a entropia de estrelas e os limitantes inferiores no $G(N, p)$ aparecem consideravelmente menos distante de H_n do que no modelo BA . Por conseguinte, pode-se dizer que, dados os parâmetros utilizados, a aproximação do valor de entropia por estrelas é mais efetiva para $G(N, p)$ do que para o BA .

4.3.2 Redes reais

Uma breve análise das redes reais usadas consta na tabela 4.1 e os resultados obtidos ao se computar as métricas definidas na seção 4.2.1 seguem na tabela 4.2. Como

descrito na seção 2.4, a entropia de von Neumann é crescente de acordo com o número de arestas do grafo e, pelos resultados obtidos, vê-se o mesmo comportamento para as medidas propostas, o que era esperado.

Rede	Vértices	Arestas	H_D
cond-mat	16726	47594	13.42
astro-ph	16706	121251	13.06
cond-mat-2003	31163	120029	14.16
cond-mat-2005	40421	175693	14.47
com-youtube	1134890	2987624	17.64
roadNet	1965206	5533214	20.80

Tabela 4.1: Descrição das redes pelo número de vértices, arestas e seu valor de entropia de graus normalizados.

Rede	$ S_2 $	H_{S_2}, l_{inf}	$ S_1 $	H_{S_1}, l_{inf}
cond-mat	9461	(14.63,12.61)	9598	(14.63,12.64)
astro-ph	10259	(15.34,12.31)	10404	(15.35,12.33)
cond-mat-2003	17815	(15.71,13.31)	18094	(15.72,13.33)
cond-mat-2005	23150	(16.14,13.58)	23536	(16.15,13.60)
com-youtube	279118	(19.14,15.47)	301923	(19.15,15.54)
roadNet	1099063	(21.08,19.99)	1342209	(21.15,20.28)

Tabela 4.2: Limitantes superiores e inferiores calculados para redes reais para os algoritmos de particionamento 1 e 2, respectivamente.

Ao explorar os resultados dispostos na tabela 4.2, interessantes aspectos das entropias calculadas aparecem. É notório observar que, para as redes analisadas, o uso da heurística de graus maximais forneceu entropia não menor do que a outra estratégia utilizada e particionamentos com cardinalidades inferiores em todos os casos, o que era esperado, como descrito na seção 3.2. Em oposição à pequena redução de entropia obtida, com diminuição efetiva somente na segunda casa decimal, a redução de cardinalidade pelo uso da heurística foi considerável perante a perspectiva da diferença bruta entre os valores.

As tabelas dispostas corroboram os resultados obtidos pela análise dos modelos teóricos, para ambos os limites calculados. Vê-se que a redução da cardinalidade do particionamento impacta positivamente no limite superior, porém aumenta a distância da métrica perante os limitantes inferiores. A entropia, porém, tem um caráter de crescimento não linear, de forma que a redução de valores mais altos pode, de certa forma, representar mais do que a redução de valores baixos.

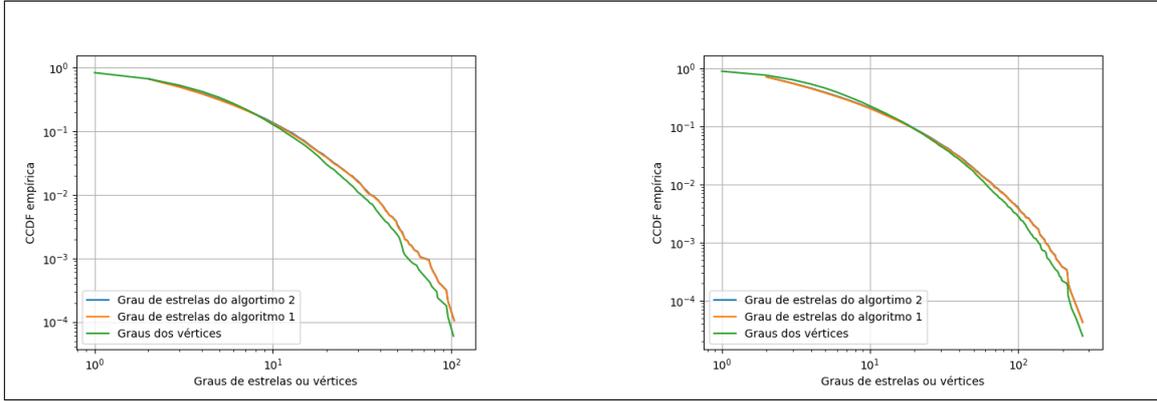


Figura 4.3: CCDF's empíricas para as redes de colaboração de Matéria Condensada amostradas em 1999 e 2005 (cond-mat, cond-mat-2005), da esquerda para direita, respectivamente.

4.3.3 Análise de distribuições empíricas de estrelas

Das divergências entre os limitantes superiores de ambos os particionamentos, o valor obtido para a *roadNet* tem maior valor bruto, trazendo um interessante ponto de análise para a conjuntura geral da entropia de particionamento.

De modo a observar as possíveis causas dessas diferenças, calculou-se as curvas da distribuição cumulativa complementar de probabilidade (CCDF) dos valores dos graus dos vértices e da cardinalidade das estrelas obtidas nos particionamentos para 4 das redes descritas em 4.2.4. As CCDF's foram dispostas usando a escala logarítmica para facilitar a visualização devido a disparidade nas ordens de grandeza dos graus envolvidos.

Ao analisar as curvas das figuras 4.3 e 4.4, nota-se uma profunda semelhança entre as três distribuições obtidas quando as redes são observadas em separado. Além disso, as redes cond-mat e cond-mat-2005 mantêm extrema similaridade entre a forma das curvas. Em complemento, é aparente que há uma aproximação para uma relação linear entre elas, dado que a concavidade é reduzida para a rede cond-mat-2005. A proximidade com uma função linear é ainda mais evidente para as curvas da rede com-youtube, que é ordens de grandeza maior que as de matéria condensada.

A CCDF empírica das redes livres de escalas é expressa, de forma singular, por uma relação linear com derivada negativa entre os graus e as respectivas probabilidades, dado o uso da escala logarítmica em ambos os eixos. Nesse âmbito, as redes de matéria condensada capturam a evolução dos relacionamentos com o passar do tempo, e mostram uma evidente aproximação a uma relação linear, que pode indicar uma tendência ao domínio livre de escala. Esse indício é fortemente evidenciado para rede de amigos no YouTube.

Em oposição ao comportamento concentrado das distribuições das três redes, a

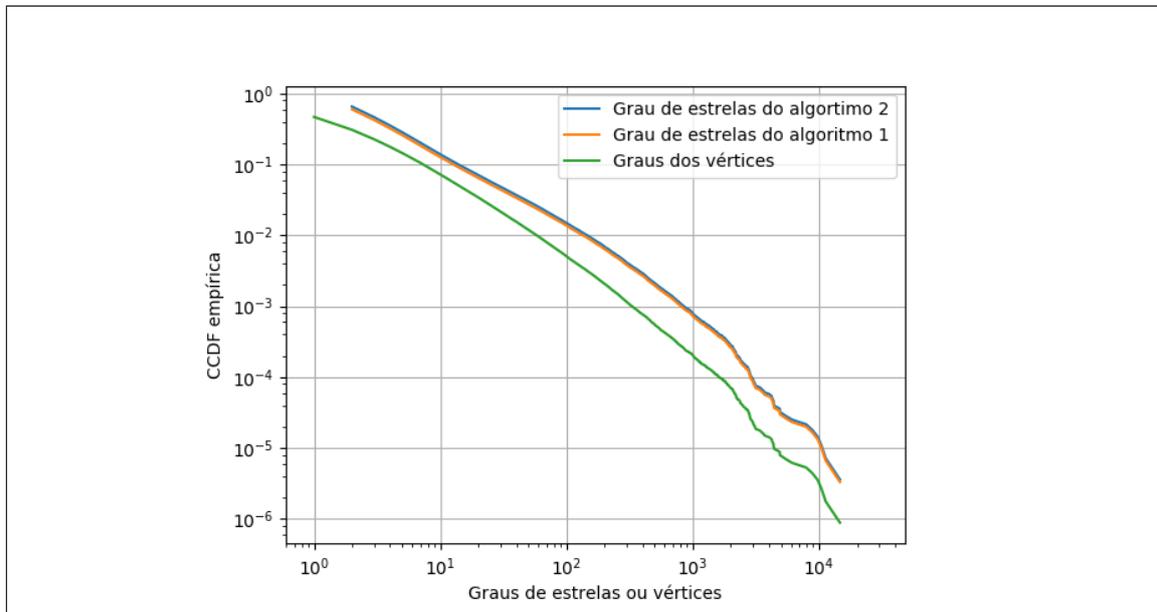


Figura 4.4: CCDF's empíricas para a rede de amizades no YouTube (com-youtube).

rede roadNet apresenta uma diferença significativa em suas distribuições. As curvas mostradas na figura 4.5 estão deslocadas significativamente entre si. Em complemento, há um maior decaimento das funções expressas, indicando que a probabilidade de graus altos decresce mais rapidamente do que aquela indicada pelas três redes analisadas anteriormente.

Por fim, essa diferença de padrões explica a disparidade relativa à efetividade da heurística na redução do limite superior quando aplicada as topologias analisadas. A iminência da propriedade livre de escala aparece como fator fundamental de explicação, e traz indícios de que, para redes com distribuição de grau neste regime, a formação de estrelas pela heurística se reduz, de forma qualitativa, àquela referente a ordenação por grau, o que foi também evidente ao se observar a análise das instâncias do modelo BA.

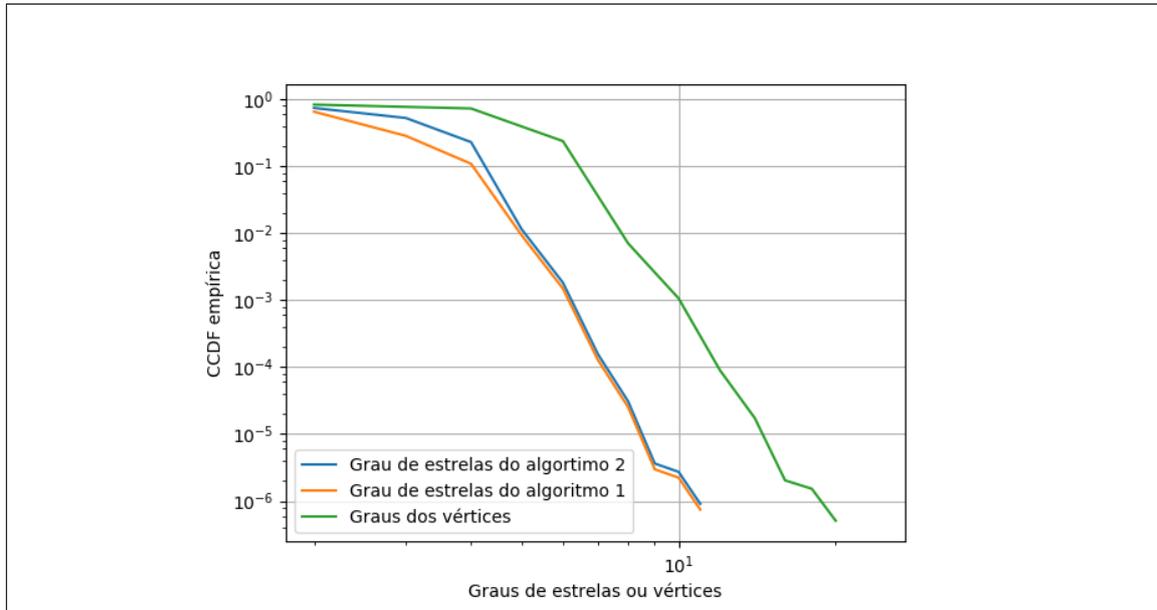


Figura 4.5: CCDF's empíricas para a rede roadNet.

4.4 Eficiência do cálculo computacional

A análise da eficiência computacional é focada na mensuração dos modelos no regime que o referencial da entropia de von Neumann é viavelmente computado, permitindo a verificação do resultado esperado em relação à redução do tempo de execução.

4.4.1 Média empírica nos modelos aleatórios

Os gráficos 4.6 e 4.7 mostram claramente a profunda diferença que emerge no tempo de execução medido em segundos dos algoritmos quando as redes atingem um número de vértices da ordem de 10^2 .

Comparando o algoritmo proposto com o cálculo da entropia de von Neumann, vê-se que o objetivo de design de fornecer maior tratabilidade perante grandes redes foi cumprido. Quanto ao uso da heurística do algoritmo 2, é aparente que existe uma redução no tempo de computação do particionamento, quando comparado à ordenação decrescente de vértices por grau, aplicada ao algoritmo 1. Essa redução fornece indício de que, mesmo que a redução de entropia obtida não seja muito satisfatória, o uso da heurística funciona melhor para algumas topologias específicas, quando observado o tempo de computação.

4.4.2 Tempo médio para redes reais amostradas

De forma a terminar a análise quantitativa do desempenho computacional, avaliou-se o tempo médio obtido ao se calcular a aproximação pelos limites propostos para a rede de estradas (roadNet) e para a rede de amizades do YouTube (com-youtube).

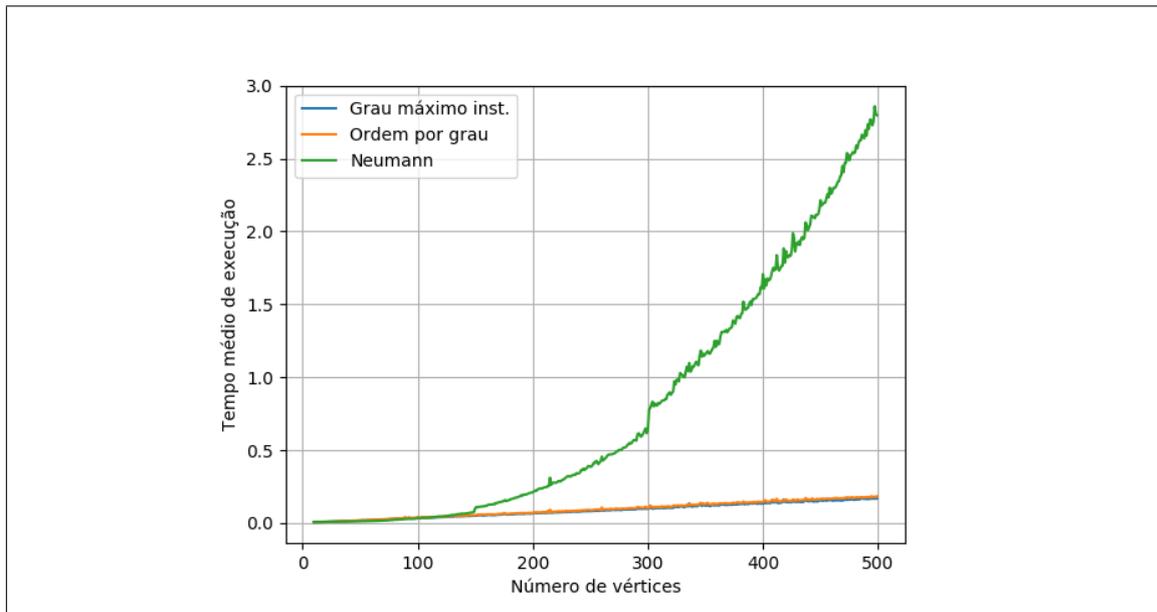


Figura 4.6: Média empírica do tempo médio de execução das métricas, em segundos, para o modelo BA com $m = 5$.

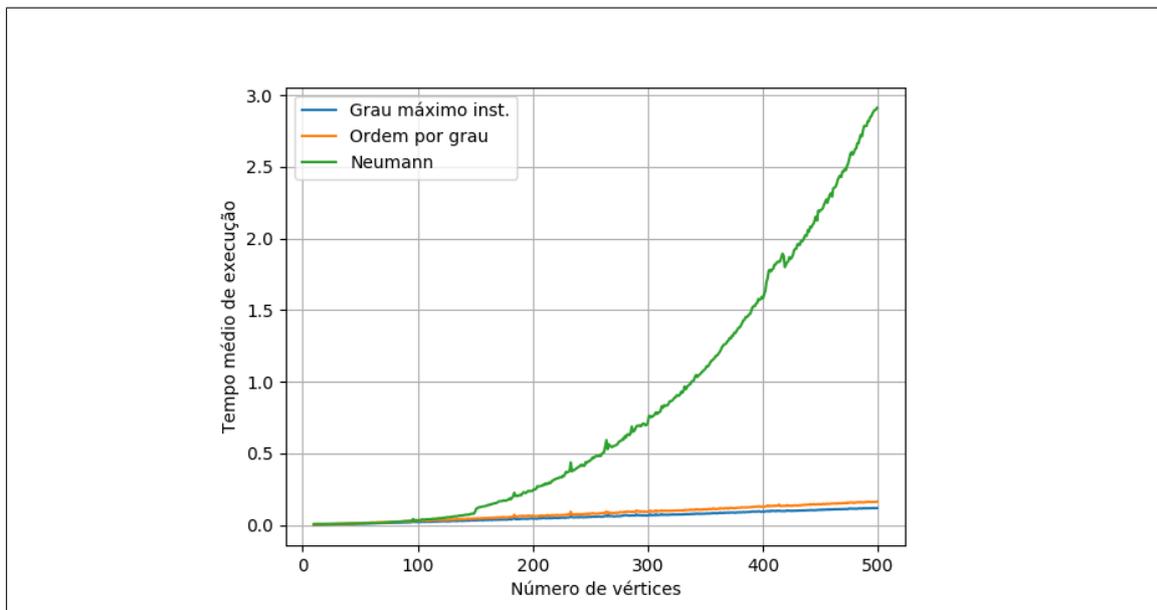


Figura 4.7: Média empírica do tempo médio de execução das métricas, em segundos, para o modelo $G(N, p)$ no regime $\frac{\log N}{N}$.

Rede	$\tau(H_{S_1})$	$\tau(H_{S_2})$
com-youtube	445.291	374.305
roadNet	1130.686	1178.646

Tabela 4.3: Tempo médio, em segundos, obtido por 10 amostras do cálculo de aproximação dos particionamentos dos algoritmos 1 e 2 para redes reais, respectivamente.

Os resultados expressos na tabela 4.3 concluem, por fim, a viabilidade do uso dos particionamentos quanto à avaliação de grandes redes. Além disso, os valores expostos mostram que a heurística, além de obter cardinalidades menores de particionamentos como referenciado na tabela 4.2, pode, em algumas instâncias, reduzir o tempo de computação, mesmo pagando o preço de atualizar os graus na fila de prioridades.

Capítulo 5

Conclusões e trabalhos futuros

Após a descrição dos procedimentos realizados, das medições experimentais e da teoria definida, este capítulo tem como propósito traçar as conclusões provindas deste trabalho, bem como estipular possíveis trabalhos futuros.

5.1 Conclusões finais

Inicialmente, foi proposto que a métrica de von Neumann era intensamente custosa para ser calculada para grandes redes e que a entropia de estrelas forneceria uma alternativa aproximativa para exploração de topologias extensas. Nesta perspectiva, a análise da aproximação proposta demandou, de forma clara, dois caminhos de observação, os quais tangem a eficiência computacional e a distância de aproximação entre os limitantes fornecidos.

Quanto ao primeiro quesito, a entropia de particionamento em estrelas cumpre seu propósito, oferecendo tratabilidade às grandes redes. Quanto a faixa de comparação fornecida pelas constelações, observa-se, porém, que os limites distanciam-se da entropia de von Neumann em valor absoluto. Os resultados experimentais obtidos mostram também indícios empíricos de que as conjecturas propostas podem ser validadas. Não obstante, o comportamento similar perante a forma das curvas obtidas pode indicar a manutenção das propriedades topológicas captadas pela métrica derradeira.

A melhor proximidade em valor absoluto da entropia de graus normalizados quando comparada à entropia de estrelas não pode ser contestada. Porém, cabe-se o questionamento sobre como a sequência de graus do grafo pode, efetivamente, enaltecer as propriedades topológicas da rede as quais os autovalores da entropia de von Neumann são efetivos em caracterizar. Esta é, certamente, uma tarefa não trivial, bem como a busca para o entendimento profundo sobre como a estrutura de particionamentos pode, ou não, obter expressões mais significativas que o valor dos graus.

Cabe também ponderar sobre a verdadeira natureza do método de particionamento. O limitante superior obtido trata-se da entropia de von Neumann do grafo transformado para um espaço de vértices de dimensão superior com diversas componentes conexas. Pode-se, assim, visualizar uma função de transformação que leva um grafo G em um grafo G' onde os relacionamentos são mais esparsos e, de certa forma, negligenciados, dado que a transformação remove a informação dos vértices vizinhos conjuntos que existe no grafo inicial.

Por fim, espera-se que as soluções trazidas e os resultados obtidos forneçam um ponto de partida para um maior entendimento sobre a entropia de particionamentos e sobre a entropia de von Neumann para grafos, bem como que seja frutífero o estudo de suas propriedades para as diversas aplicações fundamentais endereçadas no estudo de redes complexas.

5.2 Próximos passos

O uso da técnica de particionamento proposta abre espaço para futuros desenvolvimentos dos resultados expressos, tanto na parte teórica como na parte experimental, de forma que cabe elicitare alguns dos possíveis aprofundamentos a serem realizados.

5.2.1 Algoritmo de particionamento mínimo

A fórmula de entropia expressa para o particionamento em estrelas oferece uma função que pode ser minimizada de forma recursiva, abrindo caminho para a proposição de um algoritmo de programação dinâmica que ataque diretamente a minimização da entropia, ao invés da busca por partições de baixa cardinalidade.

5.2.2 Particionamentos em classes de grafo distintas

Como grafos caminho possuem fórmulas fechadas para seus autovalores e são facilmente encontrados em grafos, pode-se propor o conceito de entropia de particionamento para caminhos, usando bases similares àquelas referidas ao particionamento de estrelas. Esse trabalho forneceria uma base de comparação importante na análise da técnica de particionamento de forma geral e poderia abrir portas para o desenvolvimento de outros esquemas de partição. Com a análise de caminhos e estrelas, pode-se, também, tentar entender como o particionamento de árvores em geral pode ser utilizado, abrindo caminho para possíveis conexões com problemas de arborecência.

5.2.3 Conjecturas sobre a entropia mínima

Dos resultados teóricos expressos, a conclusão das conjecturas levantadas permitiriam terminar o embasamento teórico do método. Em específico, a conjectura da entropia mínima estipula, de forma essencial, o paralelo entre o problema de minimização da entropia com o problema de particionamento mínimo. Essa resolução reduziria o problema de minimização da entropia de estrelas num problema NP-Difícil, sendo de profundo interesse das mais diversas áreas da computação.

5.2.4 Entropia característica de redes reais

Da análise realizada no processamento de redes reais, os padrões de entropia observados oferecem indícios de que redes de tipos diferente podem oferecer curvas características de entropia. A verificação e validação dessas curvas pode fornecer poderosas ferramentas para melhor entender o surgimento e a evolução de redes complexas reais, bem como a proposição de modelos aleatórios de crescimento baseados na variação de entropia que gerassem redes com propriedades topológicas semelhantes as redes reais regidas pelo mesmo padrão entrópico.

Referências Bibliográficas

- [1] PAGE, L., BRIN, S., MOTWANI, R., et al. *The PageRank citation ranking: Bringing order to the web*. Relatório técnico, Stanford InfoLab, 1999.
- [2] ALLESINA, S., PASCUAL, M. “Googling food webs: can an eigenvector measure species’ importance for coextinctions?” *PLoS computational biology*, v. 5, n. 9, pp. e1000494, 2009.
- [3] SHANNON, C. E. “A mathematical theory of communication”, *The Bell System Technical Journal*, v. 27, n. 3, pp. 379–423, July 1948. ISSN: 0005-8580. doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [4] VON NEUMANN, J. *Mathematical Foundations of Quantum Mechanics*. Investigations in physics. Princeton University Press, 1955. ISBN: 9780691028934.
- [5] PASSERINI, F., SEVERINI, S. “The von Neumann entropy of networks”, *ArXiv e-prints*, dez. 2008. doi: 10.4018/978-1-60960-171-3.ch005.
- [6] ANAND, K., BIANCONI, G., SEVERINI, S. “Shannon and von Neumann entropy of random networks with heterogeneous expected degree”, 83(3): 036109, mar. 2011. doi: 10.1103/PhysRevE.83.036109.
- [7] AURYE, C., WILSON, R. C., HANCOCK, E. R. “Network analysis using entropy component analysis”, *Journal of Complex Networks*.
- [8] KLEINBERG, J., TARDOS, E. *Algorithm Design*. Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc., 2005. ISBN: 0321295358.
- [9] CORMEN, T. H. *Introduction to algorithms*. MIT press, 2009.
- [10] CHUNG, F. R. K. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [11] MOHAR, B. “The Laplacian spectrum of graphs”. In: *Graph Theory, Combinatorics, and Applications*, pp. 871–898. Wiley, 1991.

- [12] STRANG, G. *Linear Algebra and Its Applications*. Thomson, Brooks/Cole, 2006. ISBN: 9780534422004.
- [13] ERDŐS, P., RÉNYI, A. “On the Evolution of Random Graphs”. In: *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pp. 17–61, 1960.
- [14] ALBERT, R., BARABÁSI, A.-L. “Statistical mechanics of complex networks”, *Reviews of Modern Physics*, v. 74, pp. 47–97, jan. 2002. doi: 10.1103/RevModPhys.74.47.
- [15] COVER, T. M., THOMAS, J. A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. ISBN: 0471241954.
- [16] PETZ, D. “Entropy, von Neumann and the von Neumann Entropy”. In: Rédei, M., Stöltzner, M. (Eds.), *John von Neumann and the Foundations of Quantum Physics*, pp. 83–96, Dordrecht, Springer Netherlands, 2001. ISBN: 978-94-017-2012-0. doi: 10.1007/978-94-017-2012-0_7.
- [17] YE, C., WILSON, R. C., COMIN, C. H., et al. “Approximate von Neumann entropy for directed graphs”, *Physical Review E*, v. 89, n. 5, pp. 052804, 2014.
- [18] PAN, V. Y., CHEN, Z. Q. “The Complexity of the Matrix Eigenproblem”. In: *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, STOC '99, pp. 507–516, New York, NY, USA, 1999. ACM. ISBN: 1-58113-067-8. doi: 10.1145/301250.301389. Disponível em: <http://doi.acm.org/10.1145/301250.301389>.
- [19] MARSHALL, A. W., OLKIN, I., ARNOLD, B. C. “Inequalities: theory of majorization and its applications”, *New York*, 1979.
- [20] KARP, R. “Reducibility among combinatorial problems”. In: Miller, R., Thatcher, J. (Eds.), *Complexity of Computer Computations*, Plenum Press, pp. 85–103, 1972.
- [21] SKIENA, S. S. *The Algorithm Design Manual*. Springer Publishing Company, Incorporated, 2008. ISBN: 1848000693, 9781848000698.
- [22] NUMPY. “NumPy: a numerical extension for the computer language Python”. Disponível em: <http://www.numpy.org/>.
- [23] GRAPH TOOL. “graph-tool - Efficient network analysis”. Disponível em: <https://graph-tool.skewed.de/>.

- [24] MATPLOTLIB. “MatPlotLib, a plotting library for publication quality figures”.
. Disponível em: <https://matplotlib.org/>.
- [25] NEWMAN, M. E. “The structure of scientific collaboration networks”, *Proceedings of the national academy of sciences*, v. 98, n. 2, pp. 404–409, 2001.
- [26] YANG, J., LESKOVEC, J. “Defining and Evaluating Network Communities based on Ground-truth”, *ArXiv e-prints*, maio 2012.
- [27] LESKOVEC, J., LANG, K. J., DASGUPTA, A., et al. “Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters”, *ArXiv e-prints*, out. 2008.