

Entropia de grafos

CRISTIANE MARIA SATO

ORIENTADOR: YOSHIHARU KOHAYAKAWA

Durante a elaboração deste trabalho,
a autora recebeu apoio financeiro da FAPESP através do processo 05/60504-6.

Sumário

Introdução	4
1 Preliminares e notação	6
1.1 Conjuntos e funções	6
1.2 Teoria dos grafos	7
1.3 Probabilidade	8
I Entropia de grafos	10
2 Definição e propriedades básicas	11
2.1 Codificação e entropia de grafos	11
2.2 Uma caracterização alternativa	13
2.3 O politopo dos conjuntos estáveis	16
2.4 Propriedades básicas	17
2.5 O lema da contração	19
2.6 O lema da substituição	21
3 Entropia de hipergrafos	27
3.1 Definições	27
3.2 Propriedades	27
3.3 A entropia de alguns hipergrafos especiais	30
4 Cantos convexos	32
4.1 Entropia de cantos convexos	32
4.2 Pares geradores e antibloqueadores	33
4.3 O politopo fracionário dos conjuntos estáveis	35
5 Grafos perfeitos	37
5.1 Grafos perfeitos e cantos convexos	37
5.2 Grafos perfeitos e entropia de grafos	40

Sumário	3
6 Casais perfeitos e normais	43
6.1 Casais perfeitos	43
6.2 Casais normais	45
7 Modularidade	49
7.1 Pares submodulares	49
7.2 Pares supermodulares	50
7.3 Pares modulares	52
II Aplicações	53
8 Ordenação	54
8.1 Preliminares e notação	54
8.2 Ordenação a partir de informação parcial	55
8.3 Uma visão geral	56
8.4 Grafos de comparabilidade	56
8.5 Decomposição laminar	57
8.6 Limitantes	59
8.7 Encontrando uma boa comparação	63
8.8 Computando respostas	68
9 Funções de hashing	71
9.1 O problema	71
9.2 Limitando superiormente	71
9.3 Limitando inferiormente	72

Introdução

Breve histórico

O conceito de entropia de grafos tem suas raízes na teoria da informação, aparecendo pela primeira vez como solução de um problema de codificação proposto por Körner [15] em 1973. Considere uma fonte que emite símbolos de acordo com uma distribuição de probabilidade. Concatenando os símbolos, obtemos palavras. Körner queria medir o quão boa podia ser uma codificação de palavras de tamanho fixo emitidas pela fonte, de acordo com uma certa medida de desempenho.

Uma característica especial é que o conjunto de símbolos é ambíguo, isto é, os símbolos podem ou não ser distinguíveis. O mesmo vale para as palavras. Isso permite que várias palavras indistinguíveis sejam codificadas da mesma maneira. O desafio então é usar esse fato de uma forma inteligente para diminuir o tamanho da codificação.

A definição de entropia de grafos é justamente a solução para o problema de Körner, ou seja, é uma medida de desempenho da melhor codificação possível. No entanto, não é fácil trabalhar com essa definição. O próprio Körner, para mostrar que ela é válida, provou sua equivalência com uma função de minimização relacionada a entropia de variáveis aleatórias. A entropia de uma variável aleatória é usualmente interpretada como uma medida da quantidade de informação contida na variável aleatória em questão.

Uma importante propriedade de entropia de grafos é a subaditividade, isto é, com relação a uma distribuição de probabilidade fixada, a entropia da união de dois grafos nunca ultrapassa a soma das entropias desses grafos. A busca por condições em que a soma da entropia de um grafo e a de seu complemento é exatamente a entropia do grafo completo mostrou-se um caminho frutífero. Os estudos nessa direção foram iniciados por Körner e Longo [17]. Em 1988, Körner e Marton [18] provaram que uma condição suficiente é que, para qualquer distribuição de probabilidade, os grafos em questão sejam um grafo bipartido e seu complemento.

Em 1990, Csiszár, Körner, Lovász, Marton e Simonyi [2] mostraram uma nova caracterização de entropia de grafos. Essa caracterização, além de sua simplicidade, relaciona a entropia de um grafo com o politopo dos conjuntos estáveis desse grafo, sobre o qual são conhecidas diversas propriedades interessantes. Usando essa caracterização, Csiszár, Körner, Lovász, Marton e Simonyi mostraram que a soma da entropia de um grafo e a de seu complemento é igual à entropia do grafo completo para toda distribuição de probabilidade se e somente se o grafo é perfeito.

Os resultados de Csiszár, Körner, Lovász, Marton e Simonyi foram um grande avanço no

estudo da entropia de grafos. Uma das conseqüências de seus resultados é que é possível calcular em tempo polinomial a entropia de um grafo perfeito. Isso foi muito importante para algumas aplicações de entropia de grafos.

Körner, Simonyi e Tuza [21] apresentaram também condições necessárias e suficientes para que a soma das entropias de grafos cuja união é um grafo completo seja igual à entropia do grafo completo para toda distribuição de probabilidade.

Dentre as aplicações mais conhecidas, destacamos o uso de entropia de grafos para o problema de ordenação a partir de informação parcial (Kahn e Kim [11]); para a determinação de cotas do tipo Fredman-Komlós para funções de espalhamento (*hashing*) perfeitas e sistemas separadores (Körner [16] e Körner e Marton [19]); e em complexidade computacional (Radhakrishnan [24, 25]).

Organização do texto

Este texto é dividido em duas partes.

Na primeira parte, apresentamos a definição de entropia de grafos, caracterizações e algumas propriedades básicas. Apresentamos duas generalizações de entropia de grafos: a primeira para hipergrafos e a segunda para cantos convexos (*convex corners*). Mostramos também algumas condições suficientes para a aditividade, isto é, para que a soma da entropia de dois grafos seja igual a entropia de sua união. Nessa linha, apresentamos uma caracterização de grafos perfeitos usando entropia de grafos.

Na segunda parte, apresentamos duas aplicações de entropia de grafos. A primeira é uma aplicação ao problema de ordenação a partir de informação parcial. Na segunda, usamos entropia de grafos para a determinação de cotas do tipo Fredman-Komlós para funções de espalhamento (*hashing*) perfeitas.

Observamos que algumas das demonstrações mais fáceis omitidas nos artigos e incluídas neste texto foram elaboradas pela aluna. Boa parte das demonstrações originais foram ligeiramente modificadas, com o objetivo de facilitar a leitura.

Finalmente, a aluna acredita que a vasta gama de assuntos envolvidos neste estudo torna-o muito interessante e desafiador, mas igualmente gratificante. Esperamos que este texto possa transmitir um pouco do entusiasmo com que elaboramos este trabalho.

Capítulo 1

Preliminares e notação

Neste capítulo introduzimos a terminologia e a notação adotadas neste texto. Assumimos do leitor alguma familiaridade com teoria de grafos e combinatória poliédrica.

1.1 Conjuntos e funções

Em todo texto, usamos V e U para referirmos a conjuntos finitos. Denotamos por $\binom{V}{2}$ o conjunto $\{\{u, v\} : u \in V, v \in V, u \neq v\}$ dos pares não-ordenados de elementos de V .

Para cada inteiro positivo n , definimos $[n] := \{1, \dots, n\}$.

O conjunto dos números reais é denotado por \mathbb{R} . Os símbolos \mathbb{R}^V (respectivamente, \mathbb{R}_+^V) denota o conjunto de todos os vetores indexados por V e com coordenadas reais (respectivamente, reais não-negativas).

Seja $U \subseteq V$. Definimos o *vetor característico de U* como o vetor $\chi^U \in \mathbb{R}_+^V$ tal que

$$\chi_v^U = \begin{cases} 1, & \text{se } v \in U; \\ 0, & \text{caso contrário.} \end{cases}$$

Abreviamos $\log_2 x$ como $\lg x$. Denotamos o logaritmo natural de x por $\ln x$.

Uma função $f: R \rightarrow \mathbb{R}$, onde $R \subseteq \mathbb{R}$ é dita *convexa* se

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \tag{1.1.1}$$

para quaisquer $x, y \in R$ e qualquer $0 \leq \lambda \leq 1$. Dizemos que f é *côncava* se $-f$ é convexa. Uma função f é dita *estritamente convexa* se a relação (1.1.1) é estrita para quaisquer $x, y \in R$ e qualquer $0 < \lambda < 1$. Dizemos que f é *estritamente côncava* se $-f$ é estritamente convexa.

A seguinte desigualdade é bastante conhecida e será muito usada ao longo do texto:

Lema 1.1.1 (Desigualdade de Jensen) *Seja $f: R \rightarrow \mathbb{R}$ uma função convexa e sejam $x_1, \dots, x_k \in R$. Então*

$$f\left(\sum_{i=1}^k \lambda_i x_i\right) \leq \sum_{i=1}^k \lambda_i f(x_i), \tag{1.1.2}$$

sempre que $\sum_{i=1}^k \lambda_i = 1$ e $0 \leq \lambda_i \leq 1$ para todo i .

Sejam $x, y \in \mathbb{R}^V$. Usamos a notação $x = y$ para indicar que $x_v = y_v$ para todo $v \in V$. Usaremos a mesma notação para $x < y$ e $x > y$ e também para $x \leq y$ e $x \geq y$.

Denotamos o vetor nulo por $\mathbf{0}$ e o vetor com todas as coordenadas iguais a 1 por $\mathbf{1}$.

Sejam $a, b \in \mathbb{R}_+^V$. Definimos $\lg a \in \mathbb{R}^V$ como

$$(\lg a)_v = \lg a_v.$$

Definimos $a/b \in \mathbb{R}_+^V$ como

$$(a/b)_v = a_v/b_v.$$

Sejam $x^1, \dots, x^k \in \mathbb{R}^n$. Sejam $\lambda_1, \dots, \lambda_k$ reais não-negativos tais que $\sum_{i=1}^k \lambda_i = 1$. Dizemos que

$$\sum_{i=1}^k \lambda_i x^i$$

é uma *combinação convexa* de x^1, \dots, x^k .

Um conjunto $A \subseteq \mathbb{R}^n$ é *convexo* se

$$\lambda x + (1 - \lambda)y \in A$$

para quaisquer $x, y \in A$ e qualquer $0 \leq \lambda \leq 1$. Isto é, A é fechado por combinações convexas.

O *fecho convexo* de um conjunto $A \subseteq \mathbb{R}^n$ é o conjunto formado por todas as combinações convexas dos vetores de A . Denotamos o fecho convexo de A por $\text{conv}(A)$.

O seguinte resultado é bem conhecido:

Lema 1.1.2 (Média geométrica e média aritmética) *Sejam $x_1, \dots, x_k \in \mathbb{R}$. Então*

$$\left(\prod_{i=1}^k x_i \right)^{1/k} \leq \frac{1}{k} \sum_{i=1}^k x_i. \quad (1.1.3)$$

1.2 Teoria dos grafos

Um *grafo* é um par $G = (V, E)$, onde V é um conjunto finito e $E \subseteq \binom{V}{2}$. Dizemos que G é um *grafo sobre V* , e que V é o *conjunto de vértices* e E é o *conjunto de arestas* de G . Chamamos os elementos de V de *vértices* e os de E , de *arestas*.

Dado um grafo G , denotamos por $V(G)$ o conjunto de vértices de G e por $E(G)$ o conjunto de arestas de G .

Uma aresta $\{u, v\}$ será abreviada como uv .

Seja uv uma aresta. Dizemos que uv *liga* os vértices u e v , e que u e v são *pontas de uv* . Dizemos também que u e v são *adjacentes* ou *ligados*.

O *complemento* de um grafo G é o grafo $\overline{G} := (V, \binom{V}{2} \setminus E(G))$.

Um grafo G é dito *completo* se $E(G) = \binom{V}{2}$ e *vazio* se $E(G) = \emptyset$. Denotamos por K_V (respectivamente, $\overline{K_V}$) o grafo completo (respectivamente, vazio) sobre V . Denotamos por K_n (respectivamente, $\overline{K_n}$) qualquer grafo completo (respectivamente, vazio) com n vértices.

Denotamos por K_{n_1, \dots, n_k} qualquer grafo G para o qual existe uma partição $\{V_1, \dots, V_k\}$ de $V(G)$ que satisfaça as seguintes condições:

- $|V_i| = n_i$ para todo $1 \leq i \leq k$;
- V_i é um conjunto estável de G para todo $1 \leq i \leq k$;
- se $v \in V_i$, então v é adjacente a todo $u \in V \setminus V_i$.

Sejam G e F grafos. Dizemos que F é um *subgrafo de G* se $V(F) \subseteq V(G)$ e $E(F) \subseteq E(G)$. Se $V(F) = V(G)$, então F é um *subgrafo gerador de G* . Se $V(F) \cup E(F) \subsetneq V(G) \cup E(G)$, dizemos que F é um *subgrafo próprio de G* . Se $E(F)$ consiste de todas as arestas de G que têm as duas pontas em $V(F)$, então F é um *subgrafo induzido de G* ou, mais precisamente, F é o *subgrafo de G induzido por $V(F)$* . O subgrafo de G induzido por $U \subseteq V(G)$ é denotado por $G[U]$.

Seja G um grafo e $U \subseteq V(G)$. Dizemos que U é uma *clique de G* se $G[U]$ é completo. Se $G[U]$ é vazio, dizemos que U é um *conjunto estável de G* . Denotamos por $\omega(G)$ o tamanho da maior clique de G .

Denotamos por $\mathcal{S}(G)$ a família de conjuntos estáveis de G e por $\mathcal{S}_{\max}(G)$ a família de conjuntos estáveis maximais de G .

Os *componentes* de um grafo G são os subgrafos induzidos pelas classes de equivalência de $V(G)$ da relação de equivalência \sim dada por: para cada $u, v \in V(G)$, temos $u \sim v$ se e somente se $uv \in E(G)$.

Uma função $c: V(G) \rightarrow C$ é uma *coloração dos vértices de G* se $c(v) \neq c(u)$ sempre que v é adjacente a u . Os elementos de C são chamados de *cores* e $|C|$ é o *número de cores*. Dizemos que v *recebeu* a cor $c(v)$ ou ainda que $c(v)$ é a cor *atribuída a v* . Note que um conjunto de vértices que receberam a mesma cor é um conjunto estável de G . Uma *k -coloração dos vértices de G* é uma coloração dos vértices de G com k cores. Uma coloração de vértices é dita *mínima* se o número de cores é o menor possível.

O *número cromático* $\chi(G)$ de um grafo G é o número de cores em uma coloração mínima. É evidente que

$$\omega(G) \leq \chi(G).$$

Seja G um grafo e $U \subseteq V(G)$. Denotamos por $G - U$ o grafo $G[V \setminus U]$. Abreviamos $G - \{u\}$ como $G - u$. Seja $E' \subseteq E(G)$. Denotamos por $G - E'$ o grafo $(V(G), E(G) \setminus E')$.

Sejam G e F grafos. A união de G e F é definida como

$$G \cup F := (V(G) \cup V(F), E(G) \cup E(F)).$$

1.3 Probabilidade

Um *espaço de probabilidade finito* consiste de um conjunto finito Ω e de uma função $\mathbb{P}: \Omega \rightarrow [0, 1]$ tal que $\sum_{x \in \Omega} \mathbb{P}[x] = 1$. Um *evento* é um subconjunto de Ω . A probabilidade de um evento A é definida como

$$\mathbb{P}[A] := \sum_{x \in A} \mathbb{P}[x].$$

Seja A, B eventos de Ω . Definimos a *probabilidade conjunta entre A e B* como

$$\mathbb{P}[A, B] := \mathbb{P}[A \cap B].$$

Se $\mathbb{P}[B] > 0$, definimos a *probabilidade condicional de A dado B* como

$$\mathbb{P}[A | B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Um vetor p é uma *distribuição de probabilidade sobre V* , se $p \in \mathbb{R}_+^V$ e $\sum_{v \in V} p_v = 1$. Dizemos que uma distribuição de probabilidade p sobre V é *uniforme* se $p_v = 1/|V|$ para todo $v \in V$. Dizemos que uma distribuição de probabilidade $p \in \mathbb{R}_+^V$ é *positiva* se $p > 0$.

Uma *variável aleatória* é uma função $X: \Omega \rightarrow V$. Usamos a expressão $X = v$ para denotar o evento $\{x \in \Omega: X(x) = v\}$. A *distribuição de probabilidade de uma variável aleatória X* é um vetor em \mathbb{R}_+^V , denotado por $\text{dist}(X)$, tal que

$$\text{dist}(X)_v := \mathbb{P}[X = v]$$

para todo $v \in V$.

Parte I

Entropia de grafos

Capítulo 2

Definição e propriedades básicas

Neste capítulo apresentamos a definição de entropia de grafos e algumas de suas propriedades básicas. Primeiro fornecemos a definição dada originalmente por Körner em 1973. Em seguida, mostramos duas caracterizações com as quais é mais fácil trabalhar. Por fim, provamos algumas propriedades básicas.

2.1 Codificação e entropia de grafos

A entropia de grafos surgiu naturalmente de um problema proposto por Körner [15] em 1973. Primeiro, damos uma descrição informal do problema, com o intuito de proporcionar uma visão geral. Em seguida, definimos entropia de grafos formalmente.

Suponha que tenhamos uma fonte que emite *símbolos* um após o outro, de acordo com uma certa distribuição de probabilidade. Uma característica especial de nossa fonte é que nem todos os símbolos emitidos são distinguíveis dois-a-dois.

Concatenando símbolos emitidos pela fonte, formamos *palavras*. Dizemos que duas palavras de mesmo comprimento são *distinguíveis* se possuem símbolos distinguíveis em pelo menos uma de suas posições.

Estamos interessados em codificar todas as palavras de um certo comprimento fixo. Isto é, queremos associar um *codeword* a cada palavra de modo que palavras distinguíveis sejam mapeadas a codewords diferentes. É permitido não codificar uma fração insignificante das palavras, isto é, uma fração de palavras com baixíssima probabilidade de emissão.

Uma codificação ingênua poderia simplesmente associar um codeword diferente a cada palavra. Mas uma codificação mais esperta se aproveitaria do fato de que é permitido codificar palavras indistinguíveis a um mesmo codeword para diminuir o número de codewords necessários. Nosso problema central é, de alguma forma, medir o desempenho de uma codificação e calcular qual seria o melhor desempenho possível.

Agora descreveremos o problema mais formalmente. Seja V um conjunto finito e p uma distribuição de probabilidade sobre V . Chamamos os elementos de V de *símbolos*. Suponha que a fonte emite símbolos de V . Em um dado instante, a probabilidade de um símbolo $v \in V$ ser emitido é p_v . Como já foi dito, nem todos os símbolos emitidos pela fonte são

distinguíveis dois-a-dois. Podemos considerar distinguibilidade como uma relação binária, simétrica e arbitrária (mas conhecida e fixa) que nos diz, para cada par de símbolos, se estes são distinguíveis ou não. A relação de distinguibilidade entre os símbolos pode ser descrita através de um grafo sobre V , no qual dois vértices são adjacentes se são distinguíveis. Tal grafo é chamado de *grafo dos símbolos de V* .

Fixe t um inteiro não-negativo. Seja U um conjunto finito. Denotamos por U^t o conjunto de todas as t -uplas (u_1, \dots, u_t) , onde $u_i \in U$ para todo i . Uma *palavra* de comprimento t (emitida pela fonte) é uma t -upla $(v_1, \dots, v_t) \in V^t$ de símbolos emitidos consecutivamente pela fonte. Duas palavras $x = (x_1, \dots, x_t)$ e $y = (y_1, \dots, y_t)$ são *distinguíveis* se x_i e y_i são distinguíveis para algum i .

Considere um grafo cujo conjunto de vértices é o conjunto de todas as palavras de comprimento t , onde vértices são adjacentes se são distinguíveis. Tal grafo é chamado de *grafo das palavras de V^t* . A seguinte construção mostra como obter o grafo das palavras de V^t a partir do grafo dos símbolos de V .

Seja G um grafo. A *t -ésima potência co-normal G^t* de G é o grafo com sobre $V(G)^t$ com conjunto de arestas

$$E(G^t) := \{\{x, y\} : \{x_i, y_i\} \in E(G) \text{ para algum } 1 \leq i \leq t\}.$$

Note que o grafo das palavras de V^t é a t -ésima potência co-normal do grafo dos símbolos de V .

Defina a probabilidade de uma palavra $u = (u_1, \dots, u_t)$ como $p(u) := \prod_{i=1}^t p(u_i)$. A probabilidade de um subconjunto $U \subseteq V^t$ é definida como $p(U) := \sum_{u \in U} p(u)$.

Seja $U \subseteq V(G^t)$. Uma *codificação das palavras de U* é uma função que associa a cada vértice de U um codeword de modo que vértices adjacentes são associados a codewords diferentes. Fixe $0 < \varepsilon < 1$. Lembrando que é permitido que uma fração de palavras de baixíssima probabilidade deixe de ser codificada, definimos uma *codificação das palavras de comprimento t* como uma codificação das palavras de um conjunto $U \subseteq V(G^t)$ tal que $p(U) > 1 - \varepsilon$.

O desempenho de uma codificação é medida pela razão

$$\frac{\lg M}{t},$$

onde M é o número de codewords diferentes que a codificação utiliza. Essa razão indica o número de bits necessários pela codificação para descrever cada símbolo de uma palavra. Assim, quanto menor a razão, melhor é o desempenho da codificação. Estamos interessados em medir o quão boa pode ser uma codificação para palavras muito longas. A entropia de grafos será a resposta para essa questão.

Observe que um conjunto estável em G^t é um conjunto de palavras duas-a-duas não-distinguíveis e que, portanto, podem ser mapeadas para um mesmo codeword. Assim, o número de codewords necessários para uma codificar as palavras de $U \subseteq V^t$ é o número de conjuntos estáveis de G^t necessários para cobrir U . Isto é, o número de codewords necessários para codificar U é o número cromático $\chi(G^t[U])$. Portanto, o desempenho da melhor codificação de U é

$$\frac{\lg \chi(G^t[U])}{t}.$$

Finalmente, podemos apresentar a definição de entropia de grafos dada originalmente por Körner [15]. Seja G um grafo e p uma distribuição de probabilidade sobre $V(G)$. A *entropia de G com relação a p* é definida como

$$H(G, p) := \lim_{t \rightarrow \infty} \min \left\{ \frac{1}{t} \lg \chi(G^t[U]) : U \subseteq V(G^t), p(U) > 1 - \varepsilon \right\}.$$

Para mostrar que essa é uma fórmula válida, é necessário provar que o limite existe e é independente de $\varepsilon \in (0, 1)$. Körner fez isso mostrando que a expressão acima é equivalente a uma fórmula computável que será apresentada na seção seguinte.

Uma idéia intuitiva para a entropia de grafos é a seguinte: suponha que G é o grafo de símbolos de um conjunto finito V e que p é uma distribuição de probabilidade sobre V . Então, o número médio de bits necessários em uma codificação ótima para as palavras em V^t é $tH(G, p)$.

2.2 Uma caracterização alternativa

Nesta seção apresentamos uma caracterização de entropia de grafos dada por Körner [15]. Para isso, revisamos alguns conceitos básicos de entropia de variáveis aleatórias.

Vamos definir um conceito bastante usado em teoria da informação: a entropia de uma variável aleatória, que é um valor diretamente relacionado à quantidade de informação contida na variável aleatória em questão.

Seja p uma distribuição de probabilidade sobre um conjunto V . A *entropia de p* é definida como

$$H(p) := \sum_{v \in V} p_v \lg \frac{1}{p_v}. \quad (2.2.1)$$

Consideramos $0 \lg \frac{1}{0} = 0 \lg 0 = 0$ e $x \lg \frac{1}{0} = \infty$ para todo $x > 0$.

Definimos a *entropia de uma variável aleatória X* como $H(X) := H(\text{dist}(X))$. Podemos dizer que a entropia de X é uma medida da incerteza de X . Em outras palavras, a entropia de X pode ser interpretada como a quantidade de informação contida em X .

Sejam X e Y variáveis aleatórias que tomam seus valores em conjuntos V e U , respectivamente. A *entropia conjunta entre X e Y* é definida como

$$H(X, Y) := \sum_{x \in V} \sum_{y \in U} p_{xy} \lg \frac{1}{p_{xy}},$$

onde $p_{xy} := \mathbb{P}[X = x, Y = y]$.

A *entropia condicional de X dado Y* é definida como

$$H(X | Y) := \sum_{x \in V} \sum_{y \in U} \mathbb{P}[Y = y] H(X_y),$$

onde $X_y := (X | Y = y)$. A entropia condicional de X dado Y pode ser interpretada como a quantidade de informação contida em X mas não em Y . A seguir provamos uma relação natural entre a entropia conjunta e a entropia condicional.

Lema 2.2.1 *Sejam X e Y variáveis aleatórias. Então*

$$H(X, Y) = H(X) + H(Y | X).$$

Prova: Suponha que X e Y tomam seus valores nos conjuntos V e U , respectivamente. Abrevie $p(x) := \mathbb{P}[X = x]$ para cada $x \in V$, e $p(x, y) := \mathbb{P}[X = x, Y = y]$ e $p(y | x) := \mathbb{P}[Y = y | X = x]$ para cada $(x, y) \in V \times U$. Temos que

$$\begin{aligned}
 H(X, Y) &= - \sum_{x \in V} \sum_{y \in U} p(x, y) \lg p(x, y) \\
 &= - \sum_{x \in V} \sum_{y \in U} p(x, y) \lg (p(x)p(y | x)) \\
 &= - \sum_{x \in V} \sum_{y \in U} p(x, y) \lg p(x) - \sum_{x \in V} \sum_{y \in U} p(x, y) \lg p(y | x) \\
 &= - \sum_{x \in V} p(x) \lg p(x) - \sum_{x \in V} p(x) \sum_{y \in U} p(y | x) \lg p(y | x) \\
 &= H(X) + H(Y | X).
 \end{aligned}$$

□

Sejam p e q distribuições de probabilidade sobre um conjunto V . A *entropia de p relativa a q* é definida como

$$D(p, q) := \sum_{v \in V} p_v \lg \frac{p_v}{q_v}.$$

A entropia relativa é uma medida da distância entre duas distribuições de probabilidade. Pode-se provar que a entropia relativa entre duas distribuições de probabilidade nunca é negativa.

Lema 2.2.2 *Sejam p e q distribuições de probabilidade sobre um conjunto V . Então*

$$D(p, q) \geq 0,$$

com igualdade se e somente se $p = q$.

Prova: Tome $A := \{v \in V : p_v > 0\}$. Então

$$\begin{aligned}
 -D(p, q) &= - \sum_{a \in A} p_a \lg \frac{p_a}{q_a} = \sum_{a \in A} p_a \lg \frac{q_a}{p_a} \\
 &\leq \lg \sum_{a \in A} p_a \frac{q_a}{p_a} \leq \lg 1 = 0,
 \end{aligned} \tag{2.2.2}$$

onde a primeira desigualdade segue da desigualdade (1.1.2) de Jensen. Como $\lg x$ é uma função estritamente côncava, então (2.2.2) vale com igualdade se e somente se $p = q$. □

Sejam X e Y variáveis aleatórias que tomam seus valores em conjuntos V e U , respectivamente. A *informação mútua entre X e Y* é definida como

$$I(X \cap Y) := \sum_{x \in V} \sum_{y \in U} \mathbb{P}[X = x, Y = y] \lg \frac{\mathbb{P}[X = x, Y = y]}{\mathbb{P}[X = x]\mathbb{P}[Y = y]}.$$

A informação mútua entre X e Y pode ser interpretada como a quantidade de informação de X contida em Y . É a redução da incerteza de uma variável aleatória dado que conhecemos a outra. Essa interpretação é reforçada pelo lema a seguir.

Lema 2.2.3 *Sejam X e Y variáveis aleatórias. Então*

$$\begin{aligned} I(X \cap Y) &= H(X) - H(X | Y) \\ &= H(X) + H(Y) - H(X, Y). \end{aligned}$$

Prova: Pelo lema 2.2.1, basta provarmos a primeira igualdade. Suponha que X e Y tomam seus valores em V e U , respectivamente. Usamos as abreviações: $p(x) := \mathbb{P}[X = x]$ para cada $x \in V$ e $p(y) := \mathbb{P}[Y = y]$ para cada $y \in U$. Abreviamos também $p(x, y) := \mathbb{P}[X = x, Y = y]$ e $p(x | y) := \mathbb{P}[X = x | Y = y]$ para cada $(x, y) \in V \times U$. Vale que

$$\begin{aligned} I(X \cap Y) &= \sum_{x \in V} \sum_{y \in U} p(x, y) \lg \frac{p(x, y)}{p(x)p(y)} \\ &= \sum_{x \in V} \sum_{y \in U} p(x, y) \lg \frac{p(x | y)}{p(x)} \\ &= - \sum_{x \in V} \sum_{y \in U} p(x, y) \lg p(x) + \sum_{x \in V} \sum_{y \in U} p(x, y) \lg p(x | y) \\ &= - \sum_{x \in V} p(x) \lg p(x) + \sum_{y \in U} p(y) \sum_{x \in V} p(x | y) \lg p(x | y) \\ &= H(X) - H(X | Y). \end{aligned}$$

□

Finalmente podemos enunciar uma caracterização de entropia de grafos apresentada por Körner [15]. Omitimos a demonstração.

Teorema 2.2.4 *Seja G um grafo e p uma distribuição de probabilidade sobre $V(G)$. Seja $A(G)$ o conjunto de todos os pares ordenados de variáveis aleatórias (X, Y) que satisfazem as seguintes condições:*

- (i) X é uma variável aleatória tomando seus valores em $V(G)$ e $\text{dist}(X) = p$;
- (ii) Y é uma variável aleatória tomando seus valores em $\mathcal{S}(G)$;
- (iii) dado $X = x$, vale que Y toma seus valores em $\{S \in \mathcal{S}(G) : x \in S\}$.

Então

$$H(G, P) = \min_{(X, Y) \in A(G)} I(X \cap Y). \quad (2.2.3)$$

2.3 O politopo dos conjuntos estáveis

Nesta seção apresentamos uma caracterização de entropia de grafos provada por Csiszár, Körner, Lovász, Marton e Simonyi [2] em 1990.

O *politopo dos conjuntos estáveis de um grafo* G é definido como

$$\text{STAB}(G) := \text{conv}(\{\chi^S : S \in \mathcal{S}(G)\}).$$

Teorema 2.3.1 *Seja G um grafo e p uma distribuição de probabilidade sobre $V(G)$. Então*

$$H(G, p) = \min \left\{ -\sum_{v \in V(G)} p_v \lg a_v : a \in \text{STAB}(G) \right\}. \quad (2.3.1)$$

Prova: Tome $V := V(G)$. Primeiro vamos provar que

$$H(G, p) \geq \min \left\{ -\sum_{v \in V} p_v \lg a_v : a \in \text{STAB}(G) \right\}.$$

Sejam X e Y variáveis aleatórias tomando valores em $V(G)$ e $\mathcal{S}(G)$, respectivamente, que atingem o mínimo na caracterização (2.2.3) de $H(G, p)$. Abreviamos $r(S) := \mathbb{P}[Y = S]$ para cada $S \in \mathcal{S}(G)$ e $r(S | x) := \mathbb{P}[Y = S | X = x]$ para cada $(S, x) \in \mathcal{S}(G) \times V$. Note que

$$r(S) = \sum_{v \in V} p_v r(S | v),$$

para todo $S \in \mathcal{S}(G)$. Assim,

$$\begin{aligned} H(G, p) &= I(X \cap Y) = H(Y) - H(Y | X) \\ &= -\sum_{S \in \mathcal{S}(G)} r(S) \lg r(S) + \sum_{v \in V} p_v \sum_{S \in \mathcal{S}(G)} r(S | v) \lg r(S | v) \\ &= -\sum_{v \in V} p_v \sum_{S \in \mathcal{S}(G)} r(S | v) \lg r(S) + \sum_{v \in V} p_v \sum_{S \in \mathcal{S}(G)} r(S | v) \lg r(S | v) \\ &= -\sum_{v \in V} p_v \sum \left\{ r(S | v) \lg \frac{r(S)}{r(S | v)} : S \ni v, S \in \mathcal{S}(G) \right\} \\ &\geq -\sum_{v \in V} p_v \lg \sum \left\{ r(S) : S \ni v, S \in \mathcal{S}(G) \right\}, \end{aligned}$$

onde a última passagem segue da desigualdade (1.1.2) de Jensen. Tome $b \in \mathbb{R}_+^V$ definido como

$$b_v := \sum \left\{ r(S) : S \ni v, S \in \mathcal{S}(G) \right\},$$

para cada $v \in V$. É fácil ver que $b \in \text{STAB}(G)$. Portanto,

$$H(G, p) \geq -\sum_{v \in V} p_v \lg b_v \geq \min \left\{ -\sum_{v \in V(G)} p_v \lg a_v : a \in \text{STAB}(G) \right\}.$$

Resta provarmos que

$$H(G, p) \leq \min \left\{ - \sum_{v \in V(G)} p_v \lg a_v : a \in \text{STAB}(G) \right\}.$$

Seja d uma distribuição de probabilidade sobre $\mathcal{S}(G)$. Tome $a \in \mathbb{R}_+^V$ definido como

$$a_v := \sum \{ d_S : S \ni v, S \in \mathcal{S}(G) \}.$$

Para cada $(v, S) \in V \times \mathcal{S}(G)$, defina

$$q(S | v) := \begin{cases} d_S / a_v, & v \in S \\ 0, & \text{caso contrário.} \end{cases}$$

Para cada $S \in \mathcal{S}(G)$, tome $q(S) := \sum_{v \in V} p_v q(S | v)$. Assim,

$$H(G, p) \leq \sum_{v \in V} \sum_{S \in \mathcal{S}(G)} p_v q(S | v) \lg \frac{q(S | v)}{q(S)}.$$

Pelo lema 2.2.2,

$$\sum_{S \in \mathcal{S}(G)} q(S) \lg \frac{q(S)}{d_S} \geq 0.$$

Portanto

$$- \sum_{S \in \mathcal{S}(G)} q(S) \lg q(S) \leq - \sum_{S \in \mathcal{S}(G)} q(S) \lg d_S.$$

Concluimos que

$$H(G, p) \leq \sum_{v \in V} \sum_{S \in \mathcal{S}(G)} p_v q(S | v) \lg \frac{q(S | v)}{d_S} = - \sum_{v \in V} p_v \lg a_v,$$

como queríamos. □

2.4 Propriedades básicas

Nesta seção apresentamos algumas propriedades básicas de entropia de grafos.

Uma propriedade simples e pouco surpreendente é a monotonicidade

Lema 2.4.1 *Sejam G e F grafos tais que $V = V(G) = V(F)$ e $E(F) \subseteq E(G)$. Para qualquer distribuição de probabilidade p sobre V , vale que*

$$H(F, p) \leq H(G, p). \tag{2.4.1}$$

Prova: Segue imediatamente do seguinte fato óbvio: $\text{STAB}(G) \subseteq \text{STAB}(F)$. □

Levando em consideração a definição de entropia de grafos, a equação (2.4.1) da monotonicidade faz perfeito sentido. Basta lembrar que arestas no grafo das palavras ligam palavras distinguíveis, e portanto grafos com menos arestas têm menos palavras distinguíveis. Assim, são necessários menos bits na codificação.

A propriedade seguinte também é fácil de ser provada: vértices com probabilidade nula não influenciam na entropia do grafo.

Denotamos por $p|_U$ a restrição de p a U para qualquer distribuição de probabilidade p sobre um conjunto V e qualquer $U \subseteq V$.

Lema 2.4.2 *Seja G um grafo e p uma distribuição de probabilidade sobre $V(G)$. Seja U um subconjunto de $V(G)$ tal que $p(U) = 1$. Então*

$$H(G, p) = H(G[U], p|_U).$$

Prova: É óbvio que $H(G, p) \leq H(G[U], p|_U)$, pois todo conjunto estável de $G[U]$ é um conjunto estável de G . Para provarmos o outro lado, basta mostrarmos que, se $p_u = 0$ para algum $u \in V(G)$, então $H(G, p) = H(G - u, p')$, onde p' é a restrição de p a $V(G) \setminus \{u\}$. Seja $a \in \text{STAB}(G)$ um vetor que atinge o mínimo na caracterização (2.3.1) de $H(G, p)$. Então $a = \sum_{S \in \mathcal{S}(G)} \lambda_S \chi^S$, onde $\sum_{S \in \mathcal{S}(G)} \lambda_S = 1$. Para cada $S' \in \mathcal{S}(G - u)$, defina

$$\lambda'_{S'} := \sum \{ \lambda_S : S \in \mathcal{S}(G), S' = S \setminus \{u\} \}.$$

Tome $a' := \sum_{S' \in \mathcal{S}(G-u)} \lambda'_{S'} \chi^{S'} \in \mathcal{S}(G - u)$. Note que $a_v = a'_v$ para todo $v \neq u$. Logo,

$$H(G, p) = \sum_{v \in V(G)} p(v) \lg \frac{1}{a_v} = \sum_{v \in V(G) \setminus \{u\}} p'(v) \lg \frac{1}{a'_v} \geq H(G - u, p').$$

□

2.4.1 Subaditividade

Sejam $a, b \in \mathbb{R}_+^V$. Definimos o vetor $a \circ b$ como

$$(a \circ b)_v := a_v b_v,$$

para cada $v \in V$.

O seguinte lema segue facilmente de propriedades básicas da função $\lg x$ e de conjuntos estáveis:

Lema 2.4.3 *Sejam G e F grafos sobre um mesmo conjunto de vértices V e seja p uma distribuição de probabilidade sobre V . Então*

$$H(G \cup F, p) \leq H(G, p) + H(F, p). \tag{2.4.2}$$

Prova: Sejam $a \in \text{STAB}(G)$ e $b \in \text{STAB}(F)$ vetores que atingem o mínimo na caracterização (2.3.1) para $H(G, p)$ e $H(F, p)$, respectivamente.

O vetor a é combinação convexa de elementos de $\{\chi^S : S \in \mathcal{S}(G)\}$. Seja $a = \sum_{i \in I} \lambda_i \chi^{A_i}$ uma tal combinação. Da mesma forma, o vetor b é combinação convexa de elementos de $\{\chi^S : S \in \mathcal{S}(F)\}$. Seja $b = \sum_{j \in J} \gamma_j \chi^{B_j}$ uma tal combinação.

Note que

$$a \circ b = \sum_{i \in I} \sum_{j \in J} \lambda_i \gamma_j \cdot (\chi^{A_i} \circ \chi^{B_j})$$

e que $\chi^{A_i} \circ \chi^{B_j} = \chi^{A_i \cap B_j}$. Além disso, vale que $\sum_{i \in I} \sum_{j \in J} \lambda_i \gamma_j = 1$. Isto é, podemos escrever $a \circ b$ como combinação convexa de intersecções de conjuntos estáveis de G e F . Como a intersecção de um conjunto estável de G com um conjunto estável de F é um conjunto estável em $G \cup F$, então $a \circ b \in \text{STAB}(G \cup F)$.

Assim,

$$\begin{aligned} H(G \cup F, p) &\leq \sum_{v \in V(G)} p_v \lg \frac{1}{a_v b_v} = \sum_{v \in V(G)} p_v \lg \frac{1}{a_v} + \sum_{v \in V(G)} p_v \lg \frac{1}{b_v} \\ &= H(G, p) + H(F, p) \end{aligned}$$

□

Corolário 2.4.3.1 *Sejam G_1, G_2, \dots, G_m grafos sobre um mesmo conjunto de vértices V e seja p uma distribuição de probabilidade sobre V . Então*

$$H\left(\bigcup_{i=1}^m G_i, p\right) \leq \sum_{i=1}^m H(G_i, p)$$

Prova: Segue imediatamente do lema 2.4.3. □

Uma consequência imediata do lema 2.4.3 da subaditividade é que

$$H(G, p) + H(\overline{G}, p) \geq H(K_n, p). \quad (2.4.3)$$

No capítulo 5, vamos mostrar quais grafos satisfazem (2.4.3) com igualdade.

2.5 O lema da contração

Nesta seção mostramos como a entropia de grafos é afetada pela contração de certos subconjuntos de vértices.

Seja G um grafo e $U \subseteq V(G)$. Dizemos que U é *conjunto autônomo de G* se, para cada vértice $v \in V(G) \setminus U$, vale que v é adjacente a todos os vértices de U ou v não é adjacente a nenhum vértice de U .

Seja U um conjunto autônomo de G . Dizemos que G' é o grafo obtido a partir de G através da *contração de U* se substituímos U por um novo vértice v_U , que é ligado aos vértices de $V(G) \setminus U$ que são adjacentes aos vértices de U . Dizemos ainda que G' é obtido a partir de G através da *contração de U em v_U* .

Lema 2.5.1 *Seja G um grafo e U um conjunto autônomo de G . Seja G' o grafo obtido a partir de G através da contração de U em v_U . Seja p uma distribuição de probabilidade sobre $V(G)$. Tome a distribuição de probabilidade p' sobre $V(G')$ definida da seguinte forma:*

$$p'_v := \begin{cases} p(U), & \text{se } v = v_U; \\ p_v, & \text{caso contrário.} \end{cases}$$

Então

$$H(G, p) \geq H(G', p').$$

Prova: Seja $a \in \text{STAB}(G)$ um vetor que atinge o mínimo na caracterização (2.3.1) de $H(G, p)$. O vetor a é combinação convexa de elementos de $\{\chi^S : S \in \mathcal{S}(G)\}$. Seja $a = \sum_{i \in I} \lambda_i x^i$ uma tal combinação. Para cada i , definimos um vetor $y^i \in \mathbb{R}_+^{V(G')}$ como

$$y_v^i := x_v^i, \text{ se } v \neq v_U$$

e

$$y_{v_U}^i := \begin{cases} 1, & \text{se existe } v \in U \text{ tal que } x_v^i = 1; \\ 0, & \text{caso contrário.} \end{cases}$$

É fácil ver que cada y^i é vetor característico de algum conjunto estável de G' . Ponha

$$b := \sum_{i \in I} \lambda_i y^i.$$

É óbvio que $b \in \text{STAB}(G')$. Assim,

$$H(G, p) = \sum_{v \in V(G)} p_v \lg \frac{1}{a_v} = \sum_{v \in V(G) \setminus U} p'_v \lg \frac{1}{b_v} + \sum_{v \in U} p_v \lg \frac{1}{a_v}.$$

Se $v \in U$, então

$$a_v = \sum_{i \in I} \lambda_i x_v^i \leq \sum_{i \in I} \lambda_i y_{v_U}^i = b_{v_U}.$$

Logo,

$$\begin{aligned} H(G, p) &\geq \sum_{v \in V(G) \setminus U} p'_v \lg \frac{1}{b_v} + \sum_{v \in U} p_v \lg \frac{1}{b_{v_U}} \\ &= \sum_{v \in V(G) \setminus U} p'_v \lg \frac{1}{b_v} + p(v_U) \lg \frac{1}{b_{v_U}} \geq H(G', p'). \end{aligned}$$

□

Se acrescentarmos a hipótese de que U é um conjunto estável, é fácil provarmos que $H(G, p) = H(G', p')$.

Lema 2.5.2 (Lema da contração) *Seja G um grafo e U um conjunto autônomo estável em G . Seja G' o grafo obtido a partir de G através da contração de U em v_U . Seja p uma distribuição de probabilidade sobre $V(G)$. Tome a distribuição de probabilidade p' sobre $V(G')$ definida da seguinte forma:*

$$p'_v := \begin{cases} p(U), & \text{se } v = v_U; \\ p_v, & \text{caso contrário.} \end{cases}$$

Então,

$$H(G, p) = H(G', p').$$

Na prova do lema 2.5.1, construímos explicitamente um vetor $b \in \text{STAB}(G')$ a partir de $a \in \text{STAB}(G)$ que atinge o mínimo na caracterização (2.3.1) de $H(G, p)$. No entanto, o lema 2.5.1 segue facilmente do lema 2.5.2, pois se U é um conjunto autônomo em G , não necessariamente estável, então $H(G, p) \geq H(F, p)$, onde $F = G - E(G[U])$.

2.6 O lema da substituição

Nesta seção apresentamos o lema da substituição, que é usado em muitas provas de resultados sobre entropia de grafos.

Denotamos por p_U a distribuição de probabilidade p normalizada em U para qualquer distribuição de probabilidade p sobre um conjunto V e qualquer $U \subseteq V$.

Lema 2.6.1 *Seja T um grafo com conjunto de vértices V . Seja $U \subseteq V$ um conjunto autônomo em T . Tome $F := T[U]$ e $G := T - E(F)$. Para toda distribuição de probabilidade p sobre V , vale que*

$$H(T, p) = H(G, p) + p(U)H(F, p_U).$$

Prova: Primeiro vamos provar que

$$H(T, p) \leq H(G, p) + p(U)H(F, p_U).$$

Essa é a parte fácil da demonstração, pois é uma consequência simples do lema 2.4.3 da subaditividade.

Seja $F' := (V, E(F))$. Seja $c \in \text{STAB}(F)$ um vetor que atinge o mínimo na caracterização (2.3.1) de $H(F, p_U)$. Vamos construir um vetor $c' \in \text{STAB}(F')$. Para cada $v \in V$, defina

$$c'_v := \begin{cases} c_v, & \text{se } v \in U; \\ 1, & \text{caso contrário.} \end{cases}$$

É trivial ver que $c' \in \text{STAB}(F')$. Assim,

$$p(U)H(F, p_U) = p(U) \sum_{u \in U} \frac{p_u}{p(U)} \lg \frac{1}{c_u} = \sum_{v \in V} p_v \lg \frac{1}{c'_v} \geq H(F', p).$$

Pelo lema 2.4.3 da subaditividade, temos que

$$H(T, p) \leq H(G, p) + H(F', p) \leq H(G, p) + p(U)H(F, p_U).$$

Agora resta provarmos que

$$H(G, p) + p(U)H(F, p_U) \leq H(T, p).$$

Tome

$$\begin{aligned} \mathcal{S}_{\max}^{\subseteq \bar{U}}(G) &:= \{S \in \mathcal{S}_{\max}(G) : S \cap U = \emptyset\} \quad \text{e} \\ \mathcal{S}_{\max}^{\supseteq U}(G) &:= \{S \in \mathcal{S}_{\max}(G) : U \subseteq S\}. \end{aligned}$$

Como U é um conjunto estável de G , é fácil ver que $\mathcal{S}_{\max}(G)$ é a união disjunta de $\mathcal{S}_{\max}^{\subseteq \bar{U}}(G)$ e $\mathcal{S}_{\max}^{\supseteq U}(G)$. Além disso,

$$\mathcal{S}_{\max}(T) = \mathcal{S}_{\max}^{\subseteq \bar{U}}(G) \cup \{(S \setminus U) \cup Y : S \in \mathcal{S}_{\max}^{\supseteq U}(G) \text{ e } Y \in \mathcal{S}_{\max}(F)\}, \quad (2.6.1)$$

que também é uma união disjunta.

Seja $a \in \text{STAB}(T)$ um vetor que atinge o mínimo na caracterização (2.3.1) de $H(T, p)$. É fácil ver que podemos escrever

$$a = \sum_{A \in \mathcal{S}_{\max}(T)} \alpha_A \chi^A,$$

onde α_A é o coeficiente de $A \in \mathcal{S}_{\max}(T)$.

Agora vamos construir vetores $b \in \text{STAB}(G)$ e $c \in \text{STAB}(F)$.

Para cada $B \in \mathcal{S}_{\max}(G)$, defina

$$\beta_B := \sum_{A \subseteq B} \alpha_A.$$

Tome

$$b := \sum_{B \in \mathcal{S}_{\max}(G)} \beta_B \chi^B.$$

É claro que $\beta_B \geq 0$. É fácil ver, por (2.6.1), que todo $A \subseteq B$ para algum $B \in \mathcal{S}_{\max}(G)$ e que cada A está contido em apenas um $B \in \mathcal{S}_{\max}(G)$. Portanto,

$$\sum_{B \in \mathcal{S}_{\max}(G)} \beta_B = \sum_{B \in \mathcal{S}_{\max}(G)} \sum_{A \subseteq B} \alpha_A = \sum_{A \in \mathcal{S}_{\max}(T)} \alpha_A = 1.$$

Assim, vale que $b \in \text{STAB}(G)$.

Seja $B \in \mathcal{S}_{\max}(G)$ e $A \in \mathcal{S}_{\max}(T)$. É fácil ver que

$$v \in B \text{ e } B \supseteq A \quad \text{sse} \quad v \in A$$

para qualquer $v \in V \setminus U$. Além disso,

$$u \in B \text{ e } B \supseteq A \quad \text{sse} \quad A \cap U \neq \emptyset$$

para qualquer $u \in U$. Portanto,

$$\sum_{B \ni v} \sum_{B \supseteq A} \alpha_A = \sum_{A \ni v} \alpha_A \quad (2.6.2)$$

para todo $v \in V \setminus U$ e

$$\sum_{B \ni u} \sum_{B \supseteq A} \alpha_A = \sum_{A \cap U \neq \emptyset} \alpha_A \quad (2.6.3)$$

para todo $u \in U$.

Tome $m := \sum \{\alpha_A : A \cap U \neq \emptyset\}$. Para cada $C \in \mathcal{S}_{\max}(F)$, defina

$$\gamma_C := \frac{\sum \{\alpha_A : A \supseteq C\}}{m}.$$

Tome

$$c := \sum_{C \in \mathcal{S}_{\max}(F)} \gamma_C \chi^C.$$

Pela equação (2.6.1), é fácil ver que $A \cap U \neq \emptyset$ equivale a $A \notin \mathcal{S}_{\max}^{\subseteq \bar{U}}(G)$. Logo,

$$\sum_{C \in \mathcal{S}_{\max}(F)} \gamma_C = 1.$$

Portanto, $c \in \text{STAB}(F)$. Como cada $C \in \mathcal{S}_{\max}(F)$ está contido em exatamente um conjunto $A \in \mathcal{S}_{\max}(T)$, então

$$m \sum_{C \ni u} \gamma_C = \sum_{C \ni u} \sum_{A \supseteq C} \alpha_A = \sum_{A \ni u} \alpha_A \quad (2.6.4)$$

para todo $u \in U$.

Usando contagem dupla e as equações (2.6.2)–(2.6.4), temos

$$\begin{aligned}
 & H(G, P) + p(U)H(F, p_U) \\
 & \leq - \sum_{v \in V} p_v \lg b_v - p(U) \sum_{u \in U} \frac{p_u}{p(U)} \lg c_u \\
 & = - \sum_{v \in V} p_v \lg \left(\sum_{B \ni v} \beta_B \right) - \sum_{u \in U} \frac{p_u}{p(U)} \lg \left(\sum_{C \ni u} \gamma_C \right) \\
 & = - \sum_{v \in V} p_v \lg \left(\sum_{B \ni v} \sum_{B \supseteq A} \alpha_A \right) \\
 & \quad - \sum_{u \in U} p_u \lg \frac{\left(\sum_{C \ni u} \sum_{A \supseteq C} \alpha_A \right)}{m} \\
 & = \sum_{v \in V \setminus U} p_v \lg \left(\sum_{A \ni v} \alpha_A \right) - \sum_{u \in U} p_u \lg \left(\sum_{A \cap U \neq \emptyset} \alpha_A \right) \\
 & \quad - \sum_{u \in U} p_u \lg \frac{\left(\sum_{A \ni u} \alpha_A \right)}{m} \\
 & = \sum_{v \in V \setminus U} p_v \lg \left(\sum_{A \ni v} \alpha_A \right) - \sum_{u \in U} p_u \lg m \\
 & \quad - \sum_{u \in U} p_u \lg \frac{\left(\sum_{A \ni u} \alpha_A \right)}{m} \\
 & = - \sum_{v \in V \setminus U} p_v \lg \left(\sum_{A \ni v} \alpha_A \right) - \sum_{u \in U} p_u \lg \left(\sum_{A \ni u} \alpha_A \right) \\
 & = - \sum_{v \in V \setminus U} p_v \lg a_v - \sum_{u \in U} p_u \lg a_u \\
 & = H(T, P)
 \end{aligned}$$

e estamos feitos. \square

Sejam G e F dois grafos disjuntos nos vértices e $u \in V(G)$. Denotamos por $G_{u \leftarrow F}$ o grafo obtido pela *substituição de u por F em G* , isto é, o vértice u é removido de G e substituído por F . Cada um dos vizinhos de u deve ser ligado a cada vértice de F .

Estendemos essa operação para distribuições de probabilidade. Se p é uma distribuição de probabilidade sobre $V(G)$ e q é uma distribuição de probabilidade sobre $V(F)$, denotamos por $p_{u \leftarrow q}$ à distribuição de probabilidade sobre $V(G_{u \leftarrow F})$ definida como

$$p_{u \leftarrow q}(v) = \begin{cases} p_v, & \text{se } v \notin V(F); \\ p_u q_v, & \text{caso contrário.} \end{cases}$$

Corolário 2.6.1.1 (Lema da substituição) *Sejam G e F dois grafos disjuntos nos vértices e u um vértice de G . Sejam p e q distribuições de probabilidade sobre $V(G)$ e $V(F)$, respectivamente. Então*

$$H(G_{u \leftarrow F}, p_{u \leftarrow q}) = H(G, p) + p_u H(F, q).$$

Prova: Segue diretamente do lema 2.5.2 da contração e do lema 2.6.1. □

Corolário 2.6.1.2 *Seja G um grafo e sejam G_1, \dots, G_k seus componentes. Seja p uma distribuição de probabilidade sobre $V(G)$. Para cada G_i , definimos distribuições de probabilidade p^i sobre $V(G_i)$ como*

$$p_v^i := \frac{p_v}{p(V(G_i))}, \quad \text{para todo } v \in V(G_i).$$

Então

$$H(G, p) = \sum_i p(V(G_i)) H(G_i, p^i).$$

Prova: Segue diretamente do lema 2.6.1. □

2.6.1 A entropia de alguns grafos especiais

Lema 2.6.2 *Para todo inteiro positivo n ,*

$$H(K_n, p) = H(p),$$

onde p é uma distribuição de probabilidade sobre os vértices de K_n .

Prova: Como toda distribuição de probabilidade sobre $V(K_n)$ está em $\text{STAB}(K_n)$, então $p \in \text{STAB}(K_n)$. Seja $q \in \text{STAB}(K_n)$. Usando a desigualdade (1.1.2) de Jensen, temos que

$$\sum_{v \in V} p_v \lg \frac{1}{p_v} - \sum_{v \in V} p_v \lg \frac{1}{q_v} = \sum_{v \in V} p_v \lg \frac{q_v}{p_v} \leq \lg \sum_{v \in V} p_v \frac{q_v}{p_v} = \lg \sum_{v \in V} q_v \leq 0,$$

ou seja, p atinge o mínimo na caracterização (2.3.1) de entropia de grafos. □

Calcular a entropia do grafo vazio também é muito fácil:

Lema 2.6.3 *Para todo inteiro positivo n ,*

$$H(\overline{K_n}, p) = 0,$$

onde p é uma distribuição de probabilidade sobre os vértices de $\overline{K_n}$.

Prova: É óbvio que

$$\text{STAB}(\overline{K_n}) = \{x \in \mathbb{R}_+^{V(\overline{K_n})} : \mathbf{0} \leq x \leq \mathbf{1}\}.$$

É evidente que

$$\sum_{v \in V(\overline{K_n})} p_v \lg \frac{1}{1} = 0,$$

ou seja, $\mathbf{1}$ atinge o mínimo na caracterização (2.3.1) de entropia de grafos. □

Lema 2.6.4 Para quaisquer inteiros positivos n_1, \dots, n_m . Seja $G := K_{n_1, \dots, n_m}$. Vale que

$$H(G, p) = H(p')$$

onde p é uma distribuição de probabilidade sobre $V(G)$ e p' é a distribuição de probabilidade sobre $\mathcal{S}_{\max}(G)$ definida por $p'_S = \sum_{v \in S} p_v$, para todo $S \in \mathcal{S}_{\max}(G)$.

Prova: Segue diretamente do lema 2.5.2 da contração e do lema 2.6.2. □

Capítulo 3

Entropia de hipergrafos

Neste capítulo definimos entropia para hipergrafos. A entropia de hipergrafos é uma generalização muito natural da entropia de grafos. Enunciamos algumas propriedades básicas. Omitimos a maioria das provas, pois elas são muito parecidas com as provas das propriedades de entropia de grafos apresentadas no capítulo 2.

3.1 Definições

Um par $\mathcal{G} = (V, \mathcal{E})$ é um *hipergrafo* se V é um conjunto qualquer e \mathcal{E} é uma família de subconjuntos de V de tamanho pelo menos 2. Os elementos de V são chamados de *vértices* e os de \mathcal{E} , de *hiperarestas*. Dizemos que \mathcal{G} é um hipergrafo sobre V , e que V é o *conjunto de vértices* de \mathcal{G} e \mathcal{E} é o *conjunto de hiperarestas* de \mathcal{G} .

Dado um hipergrafo \mathcal{G} , denotamos por $V(\mathcal{G})$ o conjunto de vértices de \mathcal{G} e por $\mathcal{E}(\mathcal{G})$ o conjunto de hiperarestas de \mathcal{G} .

Seja \mathcal{G} um hipergrafo e $U \subseteq V(\mathcal{G})$. Dizemos que U é um *conjunto estável* de \mathcal{G} se não contém nenhuma hiperaresta de \mathcal{G} . Denotamos por $\mathcal{S}(\mathcal{G})$ a família de conjuntos estáveis de \mathcal{G} . O *politopo dos conjuntos estáveis do hipergrafo \mathcal{G}* é definido como

$$\text{STAB}(\mathcal{G}) := \text{conv}(\{\chi^S : S \in \mathcal{S}(\mathcal{G})\}).$$

3.2 Propriedades

Lema 3.2.1 *Sejam \mathcal{G} e \mathcal{F} hipergrafos tais que $V = V(\mathcal{G}) = V(\mathcal{F})$ e $\mathcal{E}(\mathcal{F}) \subseteq \mathcal{E}(\mathcal{G})$. Para qualquer distribuição de probabilidade p sobre V , vale que*

$$H(\mathcal{F}, p) \leq H(\mathcal{G}, p).$$

Denotamos por $p|_U$ a restrição de p a U para qualquer distribuição de probabilidade p sobre um conjunto V e qualquer $U \subseteq V$.

Seja \mathcal{G} um hipergrafo e $U \subseteq V(\mathcal{G})$. O *sub-hipergrafo induzido por U em \mathcal{G}* é o hipergrafo $\mathcal{G}[U]$ definido por

$$V(\mathcal{G}[U]) := U \quad \text{e} \quad \mathcal{E}(\mathcal{G}[U]) := \{E \in \mathcal{E}(\mathcal{G}) : E \cap (V \setminus U) = \emptyset\}.$$

Lema 3.2.2 *Seja \mathcal{G} um hipergrafo e p uma distribuição de probabilidade sobre $V(\mathcal{G})$. Seja U um subconjunto de $V(\mathcal{G})$ tal que $p(U) = 1$. Então*

$$H(\mathcal{G}, p) = H(\mathcal{G}[U], p|_U).$$

Sejam \mathcal{G} e \mathcal{F} hipergrafos sobre um conjunto V . Definimos $\mathcal{G} \cup \mathcal{F}$ como $V(\mathcal{G} \cup \mathcal{F}) := V$ e $\mathcal{E}(\mathcal{G} \cup \mathcal{F}) := \mathcal{E}(\mathcal{G}) \cup \mathcal{E}(\mathcal{F})$.

Lema 3.2.3 *Sejam \mathcal{G} e \mathcal{F} hipergrafos sobre um conjunto V e seja p uma distribuição de probabilidade sobre V . Então*

$$H(\mathcal{G} \cup \mathcal{F}, p) \leq H(\mathcal{G}, p) + H(\mathcal{F}, p).$$

Corolário 3.2.3.1 *Sejam $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_m$ grafos sobre um mesmo conjunto de vértices V e seja p uma distribuição de probabilidade sobre V . Então*

$$H\left(\bigcup_{i=1}^m \mathcal{G}_i, p\right) \leq \sum_{i=1}^m H(\mathcal{G}_i, p)$$

Agora vamos apresentar as versões para hipergrafos dos lemas da contração e da substituição.

Seja \mathcal{G} um hipergrafo. Dizemos que um conjunto $U \subseteq V(\mathcal{G})$ é um *conjunto autônomo de \mathcal{G}* , se toda hiperaresta $E \in \mathcal{E}(\mathcal{G})$ tal que $E \cap U \neq \emptyset$ e $E \cap (V \setminus U) \neq \emptyset$ satisfaz que

$$|E \cap U| = 1 \quad \text{e} \quad \{(E \setminus U) \cup \{u\} : u \in U\} \subseteq \mathcal{E}(\mathcal{G}).$$

Seja U um conjunto autônomo de \mathcal{G} . O grafo obtido a partir de \mathcal{G} através da *contração de U* é o grafo \mathcal{G}' definido por

$$\begin{aligned} V(\mathcal{G}') &:= (V(\mathcal{G}) \setminus U) \cup \{v_U\} \quad \text{e} \\ \mathcal{E}(\mathcal{G}') &:= \{E : E \cap U = \emptyset, E \in \mathcal{E}(\mathcal{G})\} \\ &\quad \cup \{(E \setminus U) \cup \{v_U\} : E \cap U \neq \emptyset, E \cap (V \setminus U) \neq \emptyset, E \in \mathcal{E}(\mathcal{G})\}, \end{aligned}$$

onde $v_U \notin V(\mathcal{G})$. Dizemos que \mathcal{G}' é obtido a partir de \mathcal{G} através da *contração de U em v_U* .

Lema 3.2.4 (Lema da contração para hipergrafos) *Seja \mathcal{G} um grafo e U um conjunto autônomo estável em \mathcal{G} . Seja \mathcal{G}' o grafo obtido a partir de \mathcal{G} através da contração de U em v_U . Seja p uma distribuição de probabilidade sobre $V(\mathcal{G})$. Tome a distribuição de probabilidade p' sobre $V(\mathcal{G}')$ definida da seguinte forma:*

$$p'_v := \begin{cases} p(U), & \text{se } v = v_U; \\ p_v, & \text{caso contrário.} \end{cases}$$

Então,

$$H(\mathcal{G}, p) = H(\mathcal{G}', p').$$

Denotamos por p_U a distribuição de probabilidade p normalizada em U para qualquer distribuição de probabilidade p sobre um conjunto V e qualquer $U \subseteq V$.

Seja \mathcal{G} um hipergrafo e $\mathcal{E} \subseteq \mathcal{E}(\mathcal{G})$. Definimos $\mathcal{G} - \mathcal{E}$ como o hipergrafo sobre $V(\mathcal{G})$ com conjunto de hiperarestas $\mathcal{E}(\mathcal{G}) \setminus \mathcal{E}$.

Lema 3.2.5 *Seja \mathcal{T} um hipergrafo com conjunto de vértices V . Seja $U \subseteq V$ um conjunto autônomo em \mathcal{T} . Tome $\mathcal{F} := \mathcal{T}[U]$ e $\mathcal{G} := \mathcal{T} - \mathcal{E}(\mathcal{F})$. Para toda distribuição de probabilidade p sobre V , vale que*

$$H(\mathcal{T}, p) = H(\mathcal{G}, p) + p(U)H(\mathcal{F}, p_U).$$

Sejam \mathcal{G} e \mathcal{F} dois hipergrafos disjuntos nos vértices e $u \in V(\mathcal{G})$. O hipergrafo obtido pela substituição de u por \mathcal{F} em \mathcal{G} é o hipergrafo $\mathcal{G}_{u \leftarrow \mathcal{F}}$ definido por

$$\begin{aligned} V(\mathcal{G}_{u \leftarrow \mathcal{F}}) &:= (V(\mathcal{G}) \setminus \{u\}) \cup V(\mathcal{F}) \\ \mathcal{E}(\mathcal{G}_{u \leftarrow \mathcal{F}}) &:= \mathcal{E}(\mathcal{F}) \cup \{E: E \not\ni u, E \in \mathcal{E}(\mathcal{G})\} \\ &\quad \cup \{(E \setminus \{u\}) \cup \{v\}: E \ni u, E \in \mathcal{E}(\mathcal{G}), v \in V(\mathcal{F})\}. \end{aligned}$$

Estendemos essa operação para distribuições de probabilidade. Se p é uma distribuição de probabilidade sobre $V(\mathcal{G})$ e q é uma distribuição de probabilidade sobre $V(\mathcal{F})$, denotamos por $p_{u \leftarrow q}$ a distribuição de probabilidade sobre $V(\mathcal{G}_{u \leftarrow \mathcal{F}})$ definida como

$$p_{u \leftarrow q}(v) = \begin{cases} p_v, & \text{se } v \notin V(\mathcal{F}); \\ p_u q_v, & \text{caso contrário.} \end{cases}$$

Corolário 3.2.5.1 (Lema da substituição para hipergrafos) *Sejam \mathcal{G} e \mathcal{F} dois grafos disjuntos nos vértices e u um vértice de \mathcal{G} . Sejam p e q distribuições de probabilidade sobre $V(\mathcal{G})$ e $V(\mathcal{F})$, respectivamente. Então*

$$H(\mathcal{G}_{u \leftarrow \mathcal{F}}, p_{u \leftarrow q}) = H(\mathcal{G}, p) + p_u H(\mathcal{F}, q).$$

Os *componentes* de um hipergrafo \mathcal{G} são os subgrafos induzidos pelas classes de equivalência de $V(\mathcal{G})$ da relação de equivalência \sim dada por: para cada $u, v \in V(\mathcal{G})$, temos $u \sim v$ se e somente se $\{u, v\} \in E$ para algum $E \in \mathcal{E}(\mathcal{G})$.

Corolário 3.2.5.2 *Seja \mathcal{G} um hipergrafo e sejam $\mathcal{G}_1, \dots, \mathcal{G}_k$ seus componentes. Seja p uma distribuição de probabilidade sobre $V(\mathcal{G})$. Para cada \mathcal{G}_i , definimos distribuições de probabilidade p^i sobre $V(\mathcal{G}_i)$ como*

$$p_v^i := \frac{p_v}{p(V(\mathcal{G}_i))}, \quad \text{para todo } v \in V(\mathcal{G}_i).$$

Então

$$H(\mathcal{G}, p) = \sum_i p(V(\mathcal{G}_i)) H(\mathcal{G}_i, p^i).$$

3.3 A entropia de alguns hipergrafos especiais

Seja $n \geq 1$ e $0 \leq k \leq n$. Dizemos que um hipergrafo é k -uniforme se todas as suas hiperarestas têm tamanho k . Dizemos que um hipergrafo \mathcal{G} é k -uniforme completo se é k -uniforme e o conjunto de hiperarestas é o conjunto $\binom{V(\mathcal{G})}{k}$. Denotamos por $\mathcal{K}_n^{(k)}$ qualquer hipergrafo k -uniforme completo com n vértices.

Seja \mathcal{G} um hipergrafo com n vértices e p uma distribuição de probabilidade sobre $V(\mathcal{G})$. Podemos supor, sem perda de generalidade, que $V(\mathcal{G}) = [n]$ e que $p_1 \geq \dots \geq p_n$. Defina os seguintes conjuntos de distribuições de probabilidade

$$R_i = \left\{ q \in \mathbb{R}_+^n : \sum_{i=1}^n q_i = 1 \text{ e } q_i \geq \frac{\sum_{j=i}^n q_j}{k-i} \right\}$$

para $1 \leq i \leq k-1$. Tome R_0 como o conjunto de todas as distribuições de probabilidade sobre $[n]$ com valores não-decrescentes.

Para toda distribuição de probabilidade p sobre $[n]$ e todo $k \leq n$, defina, para todo $0 \leq i \leq k-1$, a distribuição de probabilidade p^i sobre $[k-1]$ dada por

$$p_j^i := \begin{cases} p_j, & \text{se } j \leq i; \\ (\sum_{k=i+1}^n p_k) / (k-i-1), & \text{caso contrário,} \end{cases}$$

para todo $1 \leq j \leq k-1$.

Delmestri, Fioretto e Sgarro [3] provaram que:

Teorema 3.3.1 *Seja \mathcal{G} um hipergrafo k -uniforme completo. Suponha, sem perda de generalidade que $V(\mathcal{G}) = [n]$. Se $p \in R_i \setminus R_{i+1}$ com $0 \leq i < k-2$, então*

$$H(\mathcal{G}, p) = H(p) - H(p^i).$$

Corolário 3.3.1.1 *Vale que*

$$H(\mathcal{K}_n^{(k)}, p) = \lg \frac{n}{k-1},$$

onde p é a distribuição uniforme sobre os vértices de $\mathcal{K}_n^{(k)}$.

Prova: Seja \mathcal{G} um hipergrafo k -uniforme completo sobre $[n]$ e seja p a distribuição de probabilidade uniforme sobre $[n]$. É claro que $p \in R_0$. Por outro lado, como $k \leq n$, então $p \notin R_1$. Além disso, p^0 é a distribuição de probabilidade uniforme sobre $[k-1]$. Pelo teorema 3.3.1 e a definição (2.2.1) de entropia de variável aleatória, temos que

$$H(\mathcal{G}, p) = H(p) - H(p^0) = \lg n - \lg(k-2) = \lg \frac{n}{k-2}.$$

□

Dizemos que um hipergrafo \mathcal{G} é m -partido se existe uma partição $\{U_1, U_2, \dots, U_m\}$ de $V(\mathcal{G})$ tal que U_i é um conjunto não-vazio estável de \mathcal{G} para todo i . Denotamos por $\mathcal{K}_{n_1, \dots, n_m}^{(k)}$ a qualquer hipergrafo \mathcal{G} k -regular e m -partido tal que a partição $\{U_1, U_2, \dots, U_m\}$ em conjuntos estáveis satisfaz que $|U_i| = n_i$ para todo i e

$$\mathcal{E}(\mathcal{G}) = \binom{V(\mathcal{G})}{k} \setminus \left\{ E \in \binom{V(\mathcal{G})}{k} : |E \cap U_i| \geq 1, 1 \leq i \leq m \right\}.$$

Lema 3.3.2 *Para quaisquer inteiros positivos n_1, \dots, n_m . Vale que*

$$H(\mathcal{K}_{n_1, \dots, n_m}^{(k)}, p) \leq \lg \frac{m}{k-1},$$

onde p é a distribuição uniforme sobre os vértices de $\mathcal{K}_{n_1, \dots, n_m}^{(k)}$.

Prova: Segue diretamente do lema 3.2.4 da contração e do corolário 3.3.1.1. □

Capítulo 4

Cantos convexos

Neste capítulo definimos entropia para cantos convexos e mostramos algumas propriedades interessantes. Em seguida, relacionamos entropia de cantos convexos com conjuntos antibloqueadores e distribuições de probabilidade. Por fim, definimos o politopo fracionário dos conjuntos estáveis e mostramos algumas de suas relações com o politopo dos conjuntos estáveis.

4.1 Entropia de cantos convexos

Um conjunto $A \subseteq \mathbb{R}_+^V$ é um *canto convexo* se é fechado, limitado, convexo, tem interior não-vazio e satisfaz a propriedade de que $a' \in A$ para todo $a' \in \mathbb{R}_+^V$ tal que $0 \leq a' \leq a$ para algum $a \in A$.

Seja $A \subseteq \mathbb{R}_+^V$ um canto convexo e p uma distribuição de probabilidade sobre V . A *entropia de A com relação a p* é definida como

$$H_A(p) := \min_{a \in A} \sum_{v \in V} p_v \lg \frac{1}{a_v}. \quad (4.1.1)$$

É evidente que $\text{STAB}(G)$ é um canto convexo para todo grafo G . Além disso, é óbvio que $H(G, p) = H_{\text{STAB}(G)}(p)$.

Defina $\Lambda(A) := \{-\lg a : a \in A\}$. Note que

$$H_A(p) = \min_{x \in \Lambda(A)} \sum_{v \in V} p_v x_v = \min_{x \in \Lambda(A)} px. \quad (4.1.2)$$

Lema 4.1.1 *Seja $A \subseteq \mathbb{R}_+^V$ um canto convexo. Então $\Lambda(A)$ é convexo e $x' \in \Lambda(A)$ para todo $x' \in \mathbb{R}^V$ tal que $x' \geq x$ para algum $x \in \Lambda(A)$.*

Prova: A convexidade de $\Lambda(A)$ segue diretamente da convexidade da função $-\lg y$.

Seja $x \in \Lambda(A)$ e seja $x' \in \mathbb{R}_+^V$ tal que $x' \geq x$. Como $x \in \Lambda(A)$, então $x = -\lg a$ para algum $a \in A$. Seja $a' \in \mathbb{R}_+^V$ tal que $-\lg a' = x'$. Como $x' \geq x$, então $a' \leq a$. Logo, $a' \in A$. \square

A seguir, provamos um lema simples, mas muito poderoso, sobre entropia de cantos convexos.

Lema 4.1.2 *Seja V um conjunto finito e sejam $A, B \subseteq \mathbb{R}_+^V$ cantos convexos. Então*

$$H_A(p) \geq H_B(p)$$

para toda distribuição de probabilidade p sobre V se e somente se $A \subseteq B$.

Prova: É óbvio que $H_A(p) \geq H_B(p)$ sempre que $A \subseteq B$.

Suponha que $H_A(p) \geq H_B(p)$. Seja $b \in \Lambda(B)$ um vetor que atinge o mínimo na equação (4.1.2) de $H_B(p)$. Seja $a \in \Lambda(A)$. Então, $pb \leq pa$.

Sejam p^u , $u \in V$, distribuições de probabilidade sobre V definidas como

$$p_v^u = \begin{cases} 1, & \text{se } u = v; \\ 0, & \text{caso contrário.} \end{cases}$$

Aplicando a desigualdade $pb \leq pa$ para cada $p = p^u$, segue que $a_v \geq b_v$ para todo v , isto é, $a \geq b$. Portanto, $a \in \Lambda(B)$ pelo lema 4.1.1. Concluimos assim que $\Lambda(A) \subseteq \Lambda(B)$, de onde segue que $A \subseteq B$. \square

Corolário 4.1.2.1 *Seja $A \subseteq \mathbb{R}_+^n$ um canto convexo. Então $0 \leq H_A(p) \leq H(p)$ para toda distribuição de probabilidade $p \in \mathbb{R}_+^n$ se e somente se A está contido no n -cubo e contém o n -simplex.*

Prova: Segue imediatamente do lema 4.1.2 e dos lemas 2.6.2 e 2.6.3. \square

4.2 Pares geradores e antibloqueadores

Sejam $A, B \subseteq \mathbb{R}_+^V$ cantos convexos. Estamos interessados em saber quando podemos escrever qualquer distribuição de probabilidade p sobre V como $p = a \circ b$, onde $a \in A$ e $b \in B$.

Dizemos que um par de conjuntos $A, B \subseteq \mathbb{R}_+^n$ é um *par gerador* se toda distribuição de probabilidade p pode ser escrita como

$$p = a \circ b, \text{ para algum } a \in A \text{ e algum } b \in B.$$

Queremos saber quando dois cantos convexos A e B formam um par gerador. Para isso vamos precisar dos lemas a seguir.

Lema 4.2.1 *Sejam $A, B \subseteq \mathbb{R}_+^V$ cantos convexos e $p \in \mathbb{R}_+^V$ uma distribuição de probabilidade. Se $p = a \circ b$ para algum $a \in A$ e algum $b \in B$, então*

$$H(p) \geq H_A(p) + H_B(p),$$

com igualdade se e somente se a atinge o mínimo na definição (4.1.1) de $H_A(p)$ e b atinge o mínimo na definição (4.1.1) de $H_B(p)$.

Prova: Como $p = a \circ b$, então

$$H(p) = - \sum_{v \in V} p_v \lg a_v b_v = - \sum_{v \in V} p_v \lg a_v - \sum_{v \in V} p_v \lg b_v \geq H_A(p) + H_B(p). \quad (4.2.1)$$

É óbvio que (4.2.1) vale com igualdade se e somente se a atinge o mínimo na definição (4.1.1) de $H_A(p)$ e b atinge o mínimo na definição (4.1.1) de $H_B(p)$. \square

Seja $A \subseteq \mathbb{R}_+^V$. Definimos o *antibloqueador* de A como o conjunto

$$\text{ab}(A) := \{x \in \mathbb{R}_+^V : xa \leq 1 \text{ para todo } a \in A\}.$$

Lema 4.2.2 *Seja V um conjunto finito. Sejam $A, B \subseteq \mathbb{R}_+^V$ cantos convexos e $p \in \mathbb{R}_+^V$ uma distribuição de probabilidade. Se $\text{ab}(A) \subseteq B$, então*

$$H(p) \leq H_A(p) + H_B(p),$$

com igualdade se e somente se $p = a \circ b$ para algum $a \in A$ e algum $b \in B$.

Prova: Sejam $a \in A$ e $b \in B$ vetores que atingem o mínimo na definição (4.1.1) de $H_A(p)$ e de $H_B(p)$, respectivamente. Usando a desigualdade 1.1.2 de Jensen e o fato de que $ba \leq 1$, temos

$$H_A(p) + H_B(p) - H(p) = - \sum_{v \in V} p_v \lg \frac{a_v b_v}{p_v} \geq - \lg \left(\sum_{v \in V} a_v b_v \right) \geq 0. \quad (4.2.2)$$

Usando o lema 4.2.1, é fácil ver que (4.2.2) vale com igualdade se e somente se $p = a \circ b$ para algum $a \in A$ e algum $b \in B$. \square

O teorema que provamos a seguir é um dos principais resultados sobre pares geradores do artigo de Csiszár, Körner, Lovász, Marton e Simonyi [2].

Teorema 4.2.3 *Sejam $A, B \subseteq \mathbb{R}_+^V$ cantos convexos. As três condições a seguir são equivalentes:*

- (i) $\text{ab}(A) \subseteq B$;
- (ii) (A, B) é um par gerador;
- (iii) $H(p) \geq H_A(p) + H_B(p)$ para toda distribuição de probabilidade p sobre V .

Prova: Primeiro vamos mostrar que (i) \Rightarrow (ii). Seja p uma distribuição de probabilidade sobre V e $a \in A$ um vetor que atinge o mínimo na definição (4.1.1) de $H_A(p)$. Se $p_v > 0$, então é claro que $a_v > 0$. Então podemos definir um vetor $b \in \mathbb{R}_+^V$ como

$$b_v = \begin{cases} p_v/a_v, & \text{se } p_v > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Basta mostrarmos agora que $b \in B$. Tome

$$f(x) := - \sum_{v \in V} p_v \lg x_v \quad \text{e} \quad I := \{x \in \mathbb{R}_+^V : f(x) < f(a)\}.$$

Note que A e I são convexos e disjuntos. Portanto, existe um hiperplano que os separa. Como A e I se tocam em a e I é suave nesse ponto, então o hiperplano que os separa deve ser tangente a I e passa por a . O gradiente de $-f$ em a é $(1/\ln 2)(p/a) = (1/\ln 2)b$. Assim, o hiperplano separador é $(b/\ln 2)x = 1/\ln 2$, isto é, $bx = 1$. Logo, $bx \leq 1$ para todo $x \in A$, ou seja, $b \in \text{ab}(A) \subseteq B$. Provamos assim que (i) \Rightarrow (ii).

Segue diretamente do lema 4.2.1 que (ii) \Rightarrow (iii).

Agora vamos provar que (iii) \Rightarrow (i). Usando o fato já provado de que (i) \Rightarrow (iii) em conjunto com o lema 4.2.2, sabemos que

$$H(p) = H_A(p) + H_{\text{ab}(A)}(p),$$

para toda distribuição $p \in \mathbb{R}_+^V$. Assim, supondo que vale (iii), então $H_{\text{ab}(A)}(p) \geq H_B(p)$. Pelo lema 4.1.2, temos que $\text{ab}(A) \subseteq B$. \square

Sejam $A, B \subseteq \mathbb{R}_+^V$. Dizemos que o par (A, B) é um *par antibloqueador* se $B = \text{ab}(A)$.

É fácil provar que, se A é um canto convexo, então $\text{ab}(\text{ab}(A)) = A$. Portanto, se (A, B) é um par antibloqueador, então (B, A) também o é.

O teorema 4.2.3 e os lemas 4.2.1 e 4.2.2 implicam na seguinte caracterização de pares antibloqueadores:

Corolário 4.2.3.1 *Sejam $A, B \subseteq \mathbb{R}_+^V$ cantos convexos. Então (A, B) é um par antibloqueador se e somente se*

$$H(p) = H_A(p) + H_B(p),$$

para toda distribuição de probabilidade $p \in \mathbb{R}_+^V$.

A prova do seguinte corolário é imediata das demonstrações anteriores:

Corolário 4.2.3.2 *Seja $A \subseteq \mathbb{R}_+^V$ um canto convexo e $p \in \mathbb{R}_+^V$ uma distribuição de probabilidade. Então, vale que*

$$H(p) = H_A(p) + H_{\text{ab}(A)}(p).$$

4.3 O politopo fracionário dos conjuntos estáveis

Seja G um grafo sobre V . Definimos o *politopo fracionário dos conjuntos estáveis* de G como

$$\text{QSTAB}(G) := \left\{ b \in \mathbb{R}_+^V : \sum_{v \in K} b_v \leq 1 \text{ para toda clique } K \text{ de } G \right\}. \quad (4.3.1)$$

É óbvio que $\text{QSTAB}(G)$ é um canto convexo.

Note que todo vetor inteiro de $\text{QSTAB}(G)$ é vetor característico de um conjunto estável, e portanto está em $\text{STAB}(G)$. O lema a seguir relaciona de um modo interessante $\text{STAB}(G)$ com $\text{QSTAB}(G)$.

Teorema 4.3.1 *Seja G um grafo. Então*

$$\begin{aligned} \text{STAB}(\overline{G}) &= \text{ab}(\text{QSTAB}(G)) \quad e \\ \text{QSTAB}(\overline{G}) &= \text{ab}(\text{STAB}(G)). \end{aligned}$$

Prova: Primeiro vamos mostrar que

$$\text{ab}(X) = \text{ab}(\text{conv}(X)) \quad (4.3.2)$$

para todo $X \subseteq \mathbb{R}_+^V$. É óbvio que $\text{ab}(\text{conv}(X)) \subseteq \text{ab}(X)$. Vamos mostrar que $\text{ab}(X) \subseteq \text{ab}(\text{conv}(X))$.

Seja $y \in \text{ab}(X)$ e seja $x \in \text{conv}(X)$. O vetor x é combinação convexa de elementos de X . Seja $x = \sum_{i \in I} \lambda_i x^i$ uma tal combinação. É claro que $\lambda_i x^i y \leq \lambda_i$ para todo i . Portanto, $xy = \sum_{i \in I} \lambda_i x^i y \leq \sum_{i \in I} \lambda_i = 1$. Isso implica que $y \in \text{ab}(\text{conv}(X))$. Assim, temos que $\text{ab}(X) \subseteq \text{ab}(\text{conv}(X))$.

Agora usamos (4.3.2) para concluir que

$$\begin{aligned} \text{QSTAB}(G) &= \left\{ b \in \mathbb{R}_+^{V(G)} : \sum_{v \in K} b_v \leq 1 \text{ para toda clique } K \text{ de } G \right\} \\ &= \text{ab}(\{\chi^K : K \text{ é uma clique de } G\}) \\ &= \text{ab}(\{\chi^S : S \text{ é um conjunto estável de } \overline{G}\}) \\ &= \text{ab}(\text{STAB}(\overline{G})). \end{aligned}$$

Como $\text{STAB}(\overline{G})$ é um canto convexo, então $\text{ab}(\text{ab}(\text{STAB}(\overline{G}))) = \text{STAB}(\overline{G})$. Logo,

$$\text{STAB}(\overline{G}) = \text{ab}(\text{QSTAB}(G)).$$

□

Corolário 4.3.1.1 *Seja G um grafo. Vale que*

$$\text{STAB}(G) = \text{QSTAB}(G) \quad \text{sse} \quad \text{STAB}(\overline{G}) = \text{QSTAB}(\overline{G}).$$

Prova: Segue diretamente do lema 4.3.1.

□

Capítulo 5

Grafos perfeitos

Neste capítulo apresentamos duas caracterizações para grafos perfeitos. A primeira usa cantos convexos; a segunda, entropia de grafos.

5.1 Grafos perfeitos e cantos convexos

Nesta seção, apresentamos uma caracterização de grafos perfeitos usando cantos convexos, ou, mais especificamente, usando o politopo dos conjuntos estáveis e o politopo fracionário dos conjuntos estáveis de um grafo.

Nosso objetivo nessa seção é mostrar que um grafo G é perfeito precisamente quando $\text{QSTAB}(G) = \text{STAB}(G)$.

Seja G um grafo. Dizemos que G é um *perfeito* se, para todo subgrafo induzido G' de G , vale que

$$\omega(G') = \chi(G').$$

Existem várias definições equivalentes para grafos perfeitos. A definição que apresentamos acima foi introduzida por Claude Berge em 1961.

Primeiro vamos provar que, se G é um grafo perfeito, então $\text{QSTAB}(G) = \text{STAB}(G)$. Mas antes precisamos do seguinte lema.

Lema 5.1.1 (Lema da replicação) *Seja G um grafo perfeito e $v \in V(G)$. Seja G^+ o grafo obtido a partir de G através da replicação de v , isto é, adicionamos um novo vértice v^+ ligado a v e a todos os vizinhos de v . Então G^+ é perfeito.*

Prova: A prova é por indução em $|V(G)|$. Se $G = K_1$, então $G^+ = K_2$ é perfeito. Suponha que G é um grafo perfeito com mais de um vértice. Basta provar que $\chi(G^+) \leq \omega(G^+)$, já que todo subgrafo induzido próprio G' de G^+ ou é isomorfo a algum subgrafo induzido de G ou é obtido pela replicação de um vértice de algum subgrafo induzido próprio de G . Por hipótese de indução, G' é perfeito.

Abrevie $\omega := \omega(G)$. É claro que $\omega(G^+) \in \{\omega, \omega + 1\}$. Se $\omega(G^+) = \omega + 1$, então

$$\chi(G^+) \leq \omega + 1 = \omega(G^+).$$

Então podemos supor que $\omega(G^+) = \omega$. Neste caso, v não pertence a nenhuma clique máxima de G , pois caso contrário, sua replicação criaria uma clique de tamanho maior que ω . Considere uma coloração de G com ω cores. Seja C o conjunto de vértices que recebeu a mesma cor que v . Tome $G' := G \setminus (C \setminus \{v\})$. Como $\omega = \chi(G)$, então toda clique máxima de G tem um vértice em C , de modo que $\omega(G') < \omega$. Podemos colorir G' com $\omega - 1$ cores, já que G é perfeito. É fácil ver que $C - v + v'$ é um conjunto estável em G^+ . Assim, podemos estender a $(\omega - 1)$ -coloração de G' para uma ω -coloração de G^+ : basta atribuir a v' a mesma cor atribuída aos vértices de C e atribuir uma nova cor a v . \square

Teorema 5.1.2 *Seja G um grafo perfeito. Então*

$$STAB(G) = QSTAB(G).$$

Prova: É fácil ver que $STAB(F) \subseteq QSTAB(F)$ para todo grafo F . Então basta provarmos que $STAB(G) \supseteq QSTAB(G)$. Como $QSTAB(G)$ é um poliedro racional, então seus vértices têm coordenadas racionais. Assim, é suficiente provar que todo $x \in QSTAB(G)$ com coordenadas racionais está em $STAB(G)$.

Seja $x \in QSTAB(G)$ e suponha que αx tem coordenadas inteiras para algum $\alpha \geq 0$ inteiro. Seja G^+ o grafo obtido a partir de G da seguinte forma. Para cada $v \in V(G)$ com $\alpha x_v = 0$, remova v ; para cada $v \in V(G)$ com $\alpha x_v > 0$, replique $\alpha x_v - 1$ vezes o vértice v . Os vértices criados na replicação de v formam, junto com v , uma clique de tamanho αx_v . Chamaremos os vértices dessa clique de *clones* de v . Note que, pelo lema 5.1.1 da replicação, o grafo G^+ é perfeito.

Pela definição de $QSTAB(G)$, se K é uma clique de G então $\sum_{v \in K} x_v \leq 1$. Cada clique K^+ de G^+ está contida em uma clique de G^+ de tamanho $\sum_{v \in K} \alpha x_v$ para alguma clique K de G . Assim, vale que $\omega(G^+) \leq \alpha$. Por ser perfeito, G^+ pode ser colorido com α cores.

Seja $c: V(G) \rightarrow [\alpha]$ uma coloração dos vértices de G^+ que utiliza α cores. Para cada cor $k \in [\alpha]$ e cada vértice v de G , defina

$$y_v^k = \begin{cases} 1, & \text{se existe um clone } v' \text{ de } v \text{ tal que } c(v') = k; \\ 0, & \text{caso contrário.} \end{cases}$$

Note que cada $y^k = \chi^{S_k}$ para algum $S_k \in \mathcal{S}(G)$. Assim,

$$\frac{1}{\alpha} \sum_{k=1}^{\alpha} y^k \in STAB(G).$$

Além disso, como cada vértice de G^+ foi colorido, então

$$\sum_{k=1}^{\alpha} y_v^k = \alpha x_v$$

para todo v . Logo,

$$\frac{1}{\alpha} \sum_{k=1}^{\alpha} y^k = x$$

e estamos feitos. □

Para provar a conversa, precisamos de um resultado poliédrico.

Lema 5.1.3 *Seja $P := ab(Z)$ para algum conjunto finito $Z \subseteq \mathbb{R}_+^n$. Se $ab(P) = \emptyset$, tome $Q := \emptyset$. Caso contrário, defina*

$$Q := \{x \in P : xy = 1\}$$

para algum $y \in ab(P)$. Então ou

$$Q \subseteq \{x : xz = 1\}$$

para algum $z \in Z$, ou os conjuntos Q e Z são ambos vazios.

Prova: A prova é por indução em $|Z|$. Para a base, tome $|Z| = 0$. Neste caso, $Z = \emptyset$. Portanto $P = ab(Z) = \{x : x \geq 0\}$ e $Q = \emptyset$.

Para o passo, tome $|Z| > 0$. Suponha que z é um elemento de Z tal que, para algum $x \in P$, temos $xz \neq 1$ e $xy = 1$. É claro que $xz < 1$.

Tome $Z' := Z \setminus \{z\}$ e $P' = ab(Z')$. É fácil ver que $P \subseteq P'$. Seja $x' \in P'$. Suponha que $x'y > 1$. Tomando $x'' := (1 - \varepsilon)x + \varepsilon x'$ para $\varepsilon > 0$ suficientemente pequeno, vale que

$$x''z = (1 - \varepsilon)(xz) + \varepsilon(x'z) \leq 1,$$

isto é, $x'' \in P$. Mas

$$x''y = (1 - \varepsilon)(xy) + \varepsilon(x'y) > 1 - \varepsilon + \varepsilon = 1,$$

o que é um absurdo, já que $x'' \in P$ e $y \in ab(P)$. Portanto, temos que $x'y \leq 1$. Assim, pela hipótese de indução, vale que $Q' := \{x' \in P' : x'y = 1\} \subseteq \{x : xz' = 1\}$ para algum $z' \in Z'$. Note que $z' \in Z$. Como $P \subseteq P'$, então $Q \subseteq Q'$. Assim, $Q \subseteq Q' \subseteq \{x : xz' = 1\}$, como queríamos. □

Finalmente podemos provar a conversa do teorema 5.1.2.

Teorema 5.1.4 *Seja G um grafo. Se $STAB(G) = QSTAB(G)$, então G é perfeito.*

Prova: Abrevie $V := V(G)$. Seja $X \subseteq \mathbb{R}_+^V$. Denotaremos por $X[U]$ o conjunto de vetores indexados por U obtidos de X pela supressão dos componentes relativos a vértices de $V \setminus U$. É fácil ver que

$$QSTAB(G[U]) = QSTAB(G)[U]$$

e que

$$STAB(G[U]) = STAB(G)[U].$$

Assim, $STAB(G) = QSTAB(G)$ se e somente se $STAB(G') = QSTAB(G')$ para todo subgrafo induzido G' de G . A prova é por indução em $|V(G)|$. A base é trivial. Então, pela hipótese de indução, basta mostrar que, se $STAB(G) = QSTAB(G)$, então G pode ser colorido com $\omega := \omega(G)$ cores.

Suponha que $STAB(G) = QSTAB(G)$. Pelo corolário 4.3.1.1, vale que

$$STAB(\overline{G}) = QSTAB(\overline{G}).$$

Tome $P := \text{QSTAB}(\overline{G})$ e $y := \mathbf{1}/\omega$. Se x é vetor característico de uma clique de G , então é claro que $xy \leq 1$ e, portanto, temos que $y \in \text{ab}(P)$.

Tome $Z := \{\chi^S : S \in \mathcal{S}(G)\}$, ou seja, temos que $Z = \{\chi^K : K \text{ é uma clique de } \overline{G}\}$. Então $P = \text{QSTAB}(\overline{G}) = \text{ab}(Z)$ e $Z \neq \emptyset$. Assim, pelo lema 5.1.3,

$$Q := \{x \in P : xy = 1\} \subseteq \{x \in P : xz = 1\}$$

para algum $z \in Z$. Note que $x \in Q$ se e somente se x é vetor característico de alguma clique máxima de G . Logo, cada clique máxima intersecta o conjunto estável S tal que $z = \chi^S$. Portanto, vale que $\omega(G') = \omega(G) - 1$, onde $G' := G[V \setminus S]$. Pela hipótese de indução, podemos colorir G' com $\omega(G')$ cores. Usando uma nova cor para colorir os vértices de S , obtemos uma coloração dos vértices de G com $\omega(G)$ cores. \square

Podemos agora enunciar uma caracterização poliédrica para grafos perfeitos:

Teorema 5.1.5 *Seja G um grafo. Então*

$$G \text{ é perfeito} \quad \text{sse} \quad \text{STAB}(G) = \text{QSTAB}(G).$$

Prova: Imediato dos teoremas 5.1.2 e 5.1.4. \square

5.2 Grafos perfeitos e entropia de grafos

Nesta seção apresentaremos uma caracterização de perfeição usando entropia de grafos. Dizemos que um grafo G é *fortemente separador* se

$$H(p) = H(G, p) + H(\overline{G}, p),$$

para toda distribuição de probabilidade p sobre $V(G)$.

Teorema 5.2.1 *Seja G um grafo. Então*

$$H(p) = H(G, p) + H(\overline{G}, p)$$

para toda distribuição de probabilidade p sobre $V(G)$ se e somente se

$$\text{STAB}(G) = \text{QSTAB}(G).$$

Prova: Pelo lema 4.3.1 e pelo corolário 4.2.3.1, temos que

$$\begin{aligned} H(G, p) + H(\overline{G}, p) - H(p) &= H_{\text{STAB}(G)}(p) + H_{\text{STAB}(\overline{G})}(p) - H(p) \\ &= H_{\text{STAB}(G)}(p) - H_{\text{ab}(\text{STAB}(\overline{G}))}(p) \\ &= H_{\text{STAB}(G)}(p) - H_{\text{QSTAB}(G)}(p). \end{aligned}$$

\square

Mostramos a seguir uma caracterização de grafos perfeitos usando entropia de grafos.

Teorema 5.2.2 *Um grafo G é perfeito se e somente se é fortemente separador.*

Prova: Segue diretamente do teorema 5.2.1, do lema 4.1.2 e do teorema 5.1.5. □

Lovász [23] provou a conjectura fraca dos grafos perfeitos, que diz que um grafo é perfeito se e somente se seu complemento também o é.

Corolário 5.2.2.1 (Teorema fraco dos grafos perfeitos) *Um grafo G é perfeito se e somente se \overline{G} é perfeito.*

Prova: Segue diretamente do teorema 5.1.5 e do corolário 4.3.1.1. □

É fácil provar o seguinte corolário.

Corolário 5.2.2.2 *Um grafo G é perfeito se e somente se $\alpha(G')\omega(G') \geq |V(G')|$ para todo subgrafo induzido G' de G .*

Assim todo grafo imperfeito minimal G satisfaz $\alpha(G)\omega(G) < |V(G)|$.

Mostramos que um grafo é perfeito se e somente se é fortemente separador. Então se G é um grafo imperfeito, existe uma distribuição de probabilidade p tal que

$$H(G, p) + H(\overline{G}, p) > H(p). \quad (5.2.1)$$

A proposição a seguir mostra que se G é um grafo imperfeito minimal, então a distribuição de probabilidade uniforme satisfaz (5.2.1).

Proposição 5.2.3 *Seja G um grafo imperfeito minimal e p a distribuição de probabilidade uniforme sobre os vértices de G . Então*

$$H(G, p) + H(\overline{G}, p) > H(p).$$

Prova: Sejam a e b vetores de $\text{STAB}(G)$ e $\text{STAB}(\overline{G})$ que atingem $H(G, p)$ e $H(\overline{G}, p)$ na caracterização (2.3.1), respectivamente. Tome $V := V(G)$. Então,

$$\begin{aligned} H(G, p) + H(\overline{G}, p) &= \sum_{v \in V} \frac{1}{n} \lg \frac{1}{a_v} + \sum_{v \in V} \frac{1}{n} \lg \frac{1}{b_v} \\ &= \lg \left(1 / \left(\left(\prod_{v \in V} a_v \right)^{1/n} \left(\prod_{v \in V} b_v \right)^{1/n} \right) \right) \\ &\geq \lg \left(1 / (\alpha(G)\omega(G)/n^2) \right) \\ &> \lg n = H(p). \end{aligned}$$

A primeira desigualdade segue do lema 1.1.2; a segunda, do corolário 5.2.2.2. □

A proposição 5.2.3 implica que grafos imperfeitos não são fortemente separadores, já que podemos concentrar a distribuição de probabilidade nos vértices de um subgrafo induzido imperfeito minimal.

De acordo com a recente prova da conjectura forte dos grafos perfeitos obtida por Chudnovsky, Robertson, Seymour e Thomas [1], os grafos imperfeitos minimais são os circuitos ímpares de comprimento maior ou igual a 5 e os complementos de tais circuitos.

5.2.1 Grafos fracamente separadores

Um grafo G é dito *normal* se existe uma família \mathcal{S} de conjuntos estáveis de G e uma família \mathcal{K} de cliques de G que satisfazem as seguintes condições:

- (i) \mathcal{S} e \mathcal{K} são coberturas do conjunto de vértices de G ;
- (ii) se $S \in \mathcal{S}$ e $K \in \mathcal{K}$, então $S \cap K \neq \emptyset$.

O seguinte lema é um fato bem conhecido.

Lema 5.2.4 *Todo grafo perfeito é normal.*

Dizemos que um grafo G é *fracamente separador* se existe uma distribuição de probabilidade p sobre $V(G)$ tal que

$$H(G, p) + H(\overline{G}, p) = H(p).$$

A demonstração do teorema a seguir pode ser encontrada na resenha de Simonyi [26] sobre entropia de grafos.

Teorema 5.2.5 *Um grafo é fracamente separador se e somente se é normal.*

Capítulo 6

Casais perfeitos e normais

Neste capítulo apresentamos uma caracterização dos pares de grafos para os quais vale a aditividade da entropia da união, isto é, a soma de suas entropias é igual à entropia de sua união para qualquer distribuição de probabilidade. Tais pares de grafos são chamados de casais perfeitos. Estudamos também os casais normais, que são os pares de grafos para os quais a aditividade vale para alguma distribuição de probabilidade.

6.1 Casais perfeitos

Sejam G e F grafos sobre um conjunto V . Dizemos que G e F formam um *casal perfeito* se

$$H(G \cup F, p) = H(G, p) + H(F, p) \quad (6.1.1)$$

para toda distribuição de probabilidade p sobre V .

Apresentaremos a caracterização dos casais perfeitos fornecida por Körner, Simonyi e Tuza [21]. Para a demonstração, precisamos de um resultado sobre partições de Gallai.

Uma *partição de Gallai* é uma partição das arestas de um grafo completo tal que todo triângulo tem pelo menos 2 de suas arestas em uma mesma classe da partição.

O seguinte lema é consequência do teorema da decomposição de Gallai [9].

Lema 6.1.1 *Sejam G_1, \dots, G_k grafos sobre um conjunto V tais que $\{E(G_1), \dots, E(G_k)\}$ é uma partição de Gallai. Então no máximo 2 grafos dentre os grafos G_1, \dots, G_k são conexos.*

Denotamos por P_2 qualquer grafo sobre 3 vértices e 2 arestas. Mostramos a seguir a caracterização de casais perfeitos fornecida por Körner, Simonyi e Tuza [21].

Teorema 6.1.2 *Sejam G e F grafos sobre um conjunto V e $T := G \cup F$. Então, G e F formam um casal perfeito se e somente se as seguintes condições são satisfeitas:*

- (i) G e F são disjuntos nas arestas;
- (ii) para todo $U \subseteq V$ tal que $T[U]$ é um grafo completo, vale que $G[U]$ e $F[U]$ são perfeitos;
- (iii) para todo $U \subseteq V$ tal que $T[U] = P_2$, então $F[U] = P_2$ ou $G[U] = P_2$.

Prova: Primeiro vamos mostrar que as condições (i)–(iii) são necessárias. A necessidade de (i) é óbvia. Basta considerar o que acontece se concentramos uma distribuição de probabilidade nas extremidades de uma aresta em comum de G e F . A necessidade de (ii) segue diretamente do teorema 5.2.2 se considerarmos uma distribuição de probabilidade que se concentra nos vértices de U . Para provar a necessidade de (iii), concentre uma distribuição de probabilidade uniforme p nos vértices de $T[U] = P_2$. De fato, através de cálculos simples, obtemos que $H(P_2, p) = \lg 3 - 2/3$ e que $H(G[U], p) + H(F[U], p) = 4/3$.

Agora vamos mostrar suficiência. Esta é a parte difícil da demonstração. A prova é por indução em $|V|$. Se $|V| \leq 3$, então a suficiência é trivial. Além disso, se G ou F não têm arestas, não há nada a provar.

Tome $T := G \cup F$. Note que, por (iii), $\{E(G), E(F), E(\bar{T})\}$ é uma partição de Gallai. Portanto, pelo lema 6.1.1, pelo menos um dentre eles não é conexo. Assim, temos dois casos:

- G ou F não são conexos;
- \bar{T} não é conexo.

Suponha que vale o primeiro caso. Podemos supor que F não é conexo. Sejam F_1, \dots, F_k os componentes de F . Tome $T_i := T[V(F_i)]$ e $G_i := G[V(F_i)]$.

Seja p uma distribuição de probabilidade sobre V . Uma conseqüência fácil da condição (iii) é que $V(F_1)$ deve ser um conjunto autônomo em T . Pelo lema 2.6.1,

$$p(V(F_1))H(T_1, p_{V(F_1)}) + H(T - E(T_1), p) = H(T, p). \quad (6.1.2)$$

Além disso, como $V(F_1)$ é autônomo em $T - E(T_1)$, então, usando o lema 2.6.1 novamente, temos

$$p(V(F_1))H(G_1, p_{V(F_1)}) + H(T - E(T_1), p) = H(T - E(F_1), p). \quad (6.1.3)$$

Subtraindo a equação (6.1.3) da equação (6.1.2), obtemos

$$p(V(F_1))(H(T_1, p_{V(F_1)}) - H(G_1, p_{V(F_1)})) + H(T - E(F_1), p) = H(T, p). \quad (6.1.4)$$

Como $|V(F_1)| < |V|$ e as condições do teorema são válidas para o par G_1 e F_1 , então, pela hipótese de indução,

$$H(F_1, p_{V(F_1)}) + H(G_1, p_{V(F_1)}) = H(T_1, p_{V(F_1)}). \quad (6.1.5)$$

Usando esta última equação com a equação (6.1.4), obtemos

$$p(V(F_1))H(F_1, p_{V(F_1)}) + H(T - E(F_1), p) = H(T, p). \quad (6.1.6)$$

Fazendo isso para cada F_i para o grafo $T - \bigcup_{j < i} E(F_j)$, obtemos que

$$\sum_{i=1}^k p(V(F_i))H(F_i, p_{V(F_i)}) + H\left(T - \bigcup_{i=1}^k E(F_i), p\right) = H(T, p),$$

o que, pelo corolário 2.6.1.2 implica que

$$H(F, p) + H(G, p) = H(T, p).$$

Assim, concluímos o primeiro caso. Suponha agora que vale o segundo caso, isto é, que \bar{T} não é conexo. Se \bar{T} não tem arestas, então a suficiência segue do teorema 5.2.2. Podemos supor então que \bar{T} não é um grafo vazio. Seja \bar{T}_1 um componente de \bar{T} com pelo menos 2 vértices. Denote por G_1, F_1, T_1 os subgrafos induzidos pelos vértices de \bar{T}_1 em G, F e T , respectivamente. Note que $|V| > |V(\bar{T}_1)| \geq 2$.

Seque da condição (iii) que $V(\bar{T}_1)$ é autônomo em F e em G . Portanto, também é autônomo em T . Pelo lema 2.6.1,

$$p(V(\bar{T}_1))H(F_1, p_{V(\bar{T}_1)}) + H(F - E(F_1), p) = H(F, p), \quad (6.1.7)$$

$$p(V(\bar{T}_1))H(G_1, p_{V(\bar{T}_1)}) + H(G - E(G_1), p) = H(G, p) \quad e \quad (6.1.8)$$

$$p(V(\bar{T}_1))H(T_1, p_{V(\bar{T}_1)}) + H(T - E(T_1), p) = H(T, p). \quad (6.1.9)$$

Note que $V(\bar{T}_1)$ é um conjunto autônomo estável em $T - E(T_1), G - E(G_1)$ e $F - E(F_1)$. Portanto, se contrairmos $V(\bar{T}_1)$, usando o lema 2.5.2 da contração, vale que as entropias dos grafos resultantes da contração são iguais às dos grafos $T - E(T_1), G - E(G_1)$ e $F - E(F_1)$. Como os grafos contraídos têm menos vértices (pois $|V(\bar{T}_1)| \geq 2$) e as condições (i)–(iii) continuam sendo satisfeitas, então podemos utilizar a hipótese de indução para obtermos que

$$H(G - E(G_1), p) + H(F - E(F_1), p) = H(T - E(T_1), p). \quad (6.1.10)$$

Além disso, como $|V(\bar{T}_1)| < |V|$, usando a hipótese de indução,

$$H(G_1, p_{V(\bar{T}_1)}) + H(F_1, p_{V(\bar{T}_1)}) = H(T_1, p_{V(\bar{T}_1)}). \quad (6.1.11)$$

Somando as equações (6.1.7) e (6.1.8) e subtraindo a equação (6.1.9), pelas equações (6.1.10) e (6.1.11), temos que

$$H(G, p) + H(F, p) = H(T, p),$$

como queríamos. □

6.2 Casais normais

Já caracterizamos os pares de grafos que formam casais perfeitos, isto é, para os quais a soma de suas entropias é igual à entropia de sua união para toda distribuição de probabilidade. Enfraquecendo este conceito, obtemos a seguinte definição.

Sejam G e F grafos sobre um conjunto V . Dizemos que G e F formam um *casal normal* se

$$H(G \cup F, p) = H(G, p) + H(F, p)$$

para alguma distribuição de probabilidade positiva p sobre V .

Lema 6.2.1 *Sejam G e F grafos sobre um conjunto V . Suponha que G e F formam um casal perfeito. Então, existem coberturas $\mathcal{L}(G)$ e $\mathcal{L}(F)$ de V por conjuntos estáveis maximais de G e F , respectivamente, tais que, para todo $Y \in \mathcal{L}(G)$ e todo $Z \in \mathcal{L}(F)$, vale que $Y \cap Z$ é um conjunto estável maximal de $G \cup F$.*

Prova: Seja p uma distribuição de probabilidade positiva sobre V tal que

$$H(G \cup F, p) = H(G, p) + H(F, p).$$

Sejam $b \in \text{STAB}(G)$ e $c \in \text{STAB}(F)$ vetores que atingem o mínimo na definição (2.3.1) de $H(G, p)$ e $H(F, p)$, respectivamente.

O vetor b pode ser escrito como combinação convexa de conjuntos estáveis maximais de G . Seja $b = \sum_{Y \in \mathcal{S}_{\max}(G)} \beta_Y \chi^Y$ uma tal combinação. Tome $\mathcal{L}(G) := \{Y \in \mathcal{S}_{\max}(G) : \beta_Y > 0\}$. Similarmente, $c = \sum_{Z \in \mathcal{S}_{\max}(F)} \gamma_Z \chi^Z$ e $\mathcal{L}(F) := \{Z \in \mathcal{S}_{\max}(F) : \gamma_Z > 0\}$. Podemos supor que ambos $\mathcal{L}(G)$ e $\mathcal{L}(F)$ cobrem V , pois p é uma distribuição de probabilidade positiva.

Tome

$$\begin{aligned} \mathcal{S} &:= \{Y \cap Z : Y \in \mathcal{L}(G), Z \in \mathcal{L}(F)\} \quad \text{e} \\ \alpha_S &:= \sum \{\beta_Y \gamma_Z : Y \cap Z = S \text{ e } Y \in \mathcal{L}(G), Z \in \mathcal{L}(F)\} \end{aligned}$$

para cada S em \mathcal{S} .

Defina o vetor $a \in \mathbb{R}_+^V$ como

$$a_v := \sum \{\alpha_S : S \ni v, S \in \mathcal{S}\}$$

para cada $v \in V$. Note que a é combinação convexa de conjuntos que são estáveis tanto em G como em F . Portanto, é óbvio que $a \in \text{STAB}(G \cup F)$. Assim,

$$\begin{aligned} H(G \cup F, p) &= H(G, p) + H(F, p) \\ &= - \sum_{v \in V} p_v \lg b_v - \sum_{v \in V} p_v \lg c_v = - \sum_{v \in V} p_v \lg (b_v c_v) \\ &= - \sum_{v \in V} p_v \lg \sum \{\beta_Y \gamma_Z : v \in Y \cap Z, Y \in \mathcal{L}(G), Z \in \mathcal{L}(F)\} \\ &= - \sum_{v \in V} p_v \lg \sum \{\alpha_S : S \ni v, S \in \mathcal{S}\} = - \sum_{v \in V} p_v \lg a_v. \end{aligned}$$

É fácil ver que, se algum $S \in \mathcal{S}$ não fosse a um conjunto estável maximal de $G \cup F$, então a não atingiria o mínimo na definição (2.3.1) de $H(G \cup F, p)$. Assim, o lema está provado. \square

O corolário a seguir nos diz que, se dois grafos satisfazem as condições do teorema 6.1.2, então existem coberturas de acordo com a descrição no lema 6.2.1.

Corolário 6.2.1.1 *Sejam G e F grafos disjuntos nas arestas e sobre um mesmo conjunto de vértices. Suponha que G e F satisfazem as seguintes condições:*

- (i) para todo $U \subseteq V$ tal que $T[U]$ é um grafo completo, vale que $F[U]$ e $G[U]$ são perfeitos;
- (ii) para todo $U \subseteq V$ tal que $T[U] = P_2$, então $F[U] = P_2$ ou $G[U] = P_2$.

Então existem coberturas $\mathcal{L}(G)$ e $\mathcal{L}(F)$ de V por conjuntos estáveis maximais de G e F , respectivamente, tais que, para todo $Y \in \mathcal{L}(G)$ e todo $Z \in \mathcal{L}(F)$, vale que $Y \cap Z$ é um conjunto estável maximal de $G \cup F$.

Prova: Segue diretamente do 6.1.2, do lema 6.2.1 e do lema 5.2.4. □

Proposição 6.2.2 *Sejam G_1, \dots, G_k grafos sobre um conjunto V . Suponha que a união desses grafos é o grafo completo sobre V . Então, existe uma distribuição de probabilidade positiva p sobre V tal que*

$$H(p) = \sum_{i=1}^k H(G_i, p), \quad (6.2.1)$$

se e somente se existem coberturas $\mathcal{L}_1, \dots, \mathcal{L}_k$ de V tais que cada \mathcal{L}_i é formada por conjuntos estáveis de G_i e, para todo $Y_i \in \mathcal{L}_i$, vale que $Y_1 \cap \dots \cap Y_k \neq \emptyset$.

Prova: A prova da necessidade é semelhante à do lema 6.2.1. Vamos mostrar a suficiência. Sejam $\mathcal{L}_1, \dots, \mathcal{L}_k$ coberturas como no enunciado. Para cada i , seja $b^i = \sum_{Y \in \mathcal{L}_i} \beta_i(Y) \chi^Y$ uma combinação convexa dos conjuntos de \mathcal{L}_i com coeficientes positivos. Para cada vértice v , tome

$$p_v := \prod_{i=1}^k b^i(v) := \prod_{i=1}^k \sum_{Y \in \mathcal{L}_i} \beta_i(Y) \chi^Y(v). \quad (6.2.2)$$

Temos que mostrar que p é uma distribuição de probabilidade positiva sobre V . Como cada \mathcal{L}_i é uma cobertura de V , então, para todo $v \in V$, existe $Y \in \mathcal{L}_i$ tal que $v \in Y$. Assim, como $\beta_i(Y) > 0$ para todo $Y \in \mathcal{L}_i$, é claro que p será positiva.

Vamos provar que $\sum_{v \in V} p_v = 1$. Defina $\mathcal{L} := \mathcal{L}_1 \times \dots \times \mathcal{L}_k$ como o conjunto de todas as k -uplas (Y_1, \dots, Y_k) tais que $Y_i \in \mathcal{L}_i$. Tome

$$\beta(Y_1, \dots, Y_k) := \prod_{i=1}^k \beta_i(Y_i).$$

Para cada $v \in V$, seja $\mathcal{L}(v)$ o conjunto de k -uplas $(Y_1, \dots, Y_k) \in \mathcal{L}$ tais que $v \in \bigcap_{i=1}^k Y_i$. Assim,

$$p_v = \sum \left\{ \beta(Y_1, \dots, Y_k) : (Y_1, \dots, Y_k) \in \mathcal{L}(v) \right\}.$$

Então

$$\sum_{v \in V} p_v = \sum_{v \in V} \sum \left\{ \beta(Y_1, \dots, Y_k) : (Y_1, \dots, Y_k) \in \mathcal{L}(v) \right\}. \quad (6.2.3)$$

Temos que mostrar que toda k -upla de \mathcal{L} aparece exatamente uma vez na equação (6.2.3). Seja $(Y_1, \dots, Y_k) \in \mathcal{L}$. Como $Y_1 \cap \dots \cap Y_k \neq \emptyset$ por hipótese, então é claro que essa k -upla

aparece na equação (6.2.3). Como $G_1 \cup \dots \cup G_k = K_V$ e Y_1, \dots, Y_k são conjuntos estáveis, então para quaisquer vértices distintos u e v vale que $\{u, v\} \not\subseteq Y_1 \cap \dots \cap Y_k$. Ou seja, para cada k -upla $(Y_1, \dots, Y_k) \in \mathcal{L}$, a intersecção $\bigcap_{i=1}^k Y_i$ é unitária. Assim, provamos que p é uma distribuição de probabilidade positiva sobre V .

Como cada $b^i \in \text{STAB}(G_i)$, então

$$\begin{aligned} \sum_{i=1}^k H(G_i, p) &\leq - \sum_{i=1}^k \sum_{v \in V} p_v \lg b^i(v) = - \sum_{v \in V} \sum_{i=1}^k p_v \lg b^i(v) \\ &= - \sum_{v \in V} p_v \lg p_v = H(p). \end{aligned}$$

Por outro lado, usando o lema 2.4.3 da subaditividade,

$$\sum_{i=1}^k H(G_i, p) \geq H(p).$$

Estamos feitos. □

Proposição 6.2.3 *Se $T = G \cup F$ é um grafo perfeito, então G e F formam um casal normal se e somente se a equação (6.2.1) da proposição 6.2.2 é satisfeita com $k = 3$ e $G_1 = G$, $G_2 = F$ e $G_3 = \bar{T}$.*

Prova: Seja T um grafo perfeito sobre V . Pelo teorema 5.2.2,

$$H(T, p) + H(\bar{T}, p) = H(p), \tag{6.2.4}$$

para toda distribuição de probabilidade p sobre V . Suponha que G e F formam um casal normal. Então existe uma distribuição de probabilidade positiva p sobre V tal que

$$H(G, p) + H(F, p) = H(G \cup F, p). \tag{6.2.5}$$

Então

$$H(G, p) + H(F, p) + H(\bar{T}, p) = H(p). \tag{6.2.6}$$

Portanto, a equação (6.2.2) é satisfeita como descrito no enunciado.

Agora suponha que a equação (6.2.1) é satisfeita com $k = 3$ e $G_1 = G$, $G_2 = F$ e $G_3 = \bar{T}$, então é claro que a equação (6.2.6) vale. Pelas equações (6.2.4) e (6.2.6) e usando o lema 2.4.3 da subaditividade, a equação (6.2.5) vale. □

Capítulo 7

Modularidade

Neste capítulo apresentamos relações entre a soma da entropia de dois grafos e a soma das entropias de sua união e sua intersecção. Acrescentando a restrição de que a união dos grafos é o grafo completo, caracterizamos pares submodulares, isto é, para os quais a entropia é uma função submodular com relação a qualquer distribuição de probabilidade. Similarmemente, caracterizamos também os pares supermodulares e os modulares.

Para mais detalhes, veja o artigo de Körner e Simonyi sobre modularidade [20].

7.1 Pares submodulares

Um par de grafos G e F sobre um conjunto V é dito *submodular* se

$$H(G \cup F, p) + H(G \cap F, p) \leq H(G, p) + H(F, p)$$

para toda distribuição de probabilidade p sobre V .

Teorema 7.1.1 *Sejam G e F grafos sobre um conjunto V tais que $T := G \cup F = K_{|V|}$. Então, G e F formam um par submodular se e somente se todo triângulo em T tem pelo menos duas de suas arestas em $G - E(F)$, $F - E(G)$ ou $G \cap F$.*

Prova: Chamemos de *triângulo multicolorido* a todo triângulo em T tal que cada um dentre $G - E(F)$, $F - E(G)$ e $G \cap F$ possui exatamente uma das arestas do triângulo.

Suponha que T possui um triângulo multicolorido. Concentrando uma distribuição de probabilidade p nos 3 vértices desse triângulo, é fácil mostrar que

$$H(G \cup F, p) + H(G \cap F, p) > H(G, p) + H(F, p).$$

Provaremos a outra direção por indução no número de vértices $|V|$. Se $|V| \leq 2$, a prova é trivial. Suponha que $G \cap F$ é vazio. Então, pela propriedade 2.4.3 da subaditividade, os grafos G e F formam um par submodular. Se $G - E(F)$ ou $F - E(G)$ é um grafo vazio, então G é subgrafo de F ou vice-versa. Assim, é claro que G e F formam um par submodular. Portanto, podemos supor que $G - E(F)$, $F - E(G)$ e $G \cap F$ não são vazios.

Note que os conjuntos de arestas de $T_1 := G - E(F)$, $T_2 := F - E(G)$ e $T_3 := G \cap F$ formam uma partição de Gallai, e portanto pelo menos um dentre eles não é conexo, digamos T_i . Seja $U \subset V$ o conjunto de vértices de um componente não-vazio de T_i . Note que $1 < |U| < |V|$.

Ademais, se $v \in U$, então todas as arestas que incidem em v e têm a outra ponta em $V \setminus U$ pertencem a um mesmo T_j : caso contrário, existiria um triângulo multicolorido em T . É fácil ver que isso implica que U é um conjunto autônomo em G , F , $G \cap F$ e $F \cup G$. Denote por X qualquer um dos grafos G , F , $G \cap F$ e $G \cup F$. Denote por X_U o grafo com conjunto de vértices V e conjunto de arestas $E(X[U])$. Denote por $X_{\bar{U}}$ o grafo obtido pela contração de U em um novo vértice u . Note que o grafo obtido não contém triângulos multicoloridos. Defina uma distribuição de probabilidade p' sobre $V(X_{\bar{U}})$ como

$$p'_v := \begin{cases} p(U), & \text{se } v = u; \\ p_v, & \text{caso contrário.} \end{cases}$$

Assim, pelo corolário 2.6.1.1,

$$H(X, p) = H(X_{\bar{U}}, p') + H(X_U, p). \quad (7.1.1)$$

Assim, usando a hipótese de indução e a equação (7.1.1), obtemos que

$$\begin{aligned} H(G, p) + H(F, p) &= \\ &= H(G_{\bar{U}}, p') + H(G_U, p) + H(F_{\bar{U}}, p') + H(F_U, p) \\ &= H(G_{\bar{U}}, p') + H(F_{\bar{U}}, p') + H(G_U, p) + H(F_U, p) \\ &\leq H((G \cup F)_{\bar{U}}, p') + H((G \cap F)_{\bar{U}}, p') \\ &\quad + H((G \cup F)_U, p) + H((G \cap F)_U, p) \\ &= H(G \cup F, p) + H(G \cap F, p) \end{aligned}$$

e estamos feitos. □

7.2 Pares supermodulares

Um par de grafos G e F é dito *supermodular* se

$$H(G \cup F, p) + H(G \cap F, p) \geq H(G, p) + H(F, p)$$

para toda distribuição de probabilidade p sobre V .

Teorema 7.2.1 *Sejam G e F grafos sobre um conjunto V tais que $T := G \cup F = K_V$. Então, G e F formam um par supermodular se e somente se não existe subconjunto $U \subseteq V$ tal que $F[U]$ e $G[U]$ são disjuntos nas arestas e são imperfeitos.*

Prova: A necessidade é uma consequência imediata da proposição 5.2.3. Suponha que existe um subconjunto $U \subseteq V$ tal que $F[U]$ e $G[U]$ são disjuntos nas arestas e imperfeitos. Podemos

supor que $F[U]$ ou $G[U]$ é um grafo imperfeito minimal. Concentrando uma distribuição de probabilidade p sobre V nos vértices de U , obtemos que

$$H(G, p) + H(F, p) > H(G \cup F, p).$$

Assim, como $G \cap F$ não tem arestas em U , então $H(G \cap F, p) = 0$ e, portanto, F e G não formam um par supermodular.

Agora vamos provar a suficiência. Sejam F e G como no enunciado e suponha que não existe subconjunto $U \in V$ tal que $F[U]$ e $G[U]$ são disjuntos nas arestas e são imperfeitos. Seja p uma distribuição de probabilidade sobre V .

Seja X uma variável aleatória que toma seus valores em V com distribuição igual a p e seja Z uma variável aleatória que toma seus valores em $\mathcal{S}(G \cap F)$ tais que (X, Z) atinge o mínimo na caracterização (2.2.3) de $H(G \cap F, p)$.

Seja U um conjunto estável em $G \cap F$. Denote por p_U a probabilidade de X dado que Z vale U . Como $G[U]$ e $F[U]$ são disjuntos nas arestas, então ambos são perfeitos. Assim,

$$H(X | Z = U) = H(p_U) = H(G[U], p_U) + H(F[U], p_U).$$

Portanto,

$$H(X | Z) = \sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(G[U], p_U) + \sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(F[U], p_U).$$

Adicionando $2I(X \cap Z)$ em cada lado, obtemos

$$\begin{aligned} & \sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(G[U], p_U) + I(X \cap Z) \\ & \quad + \sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(F[U], p_U) + I(X \cap Z) \\ & = H(X | Z) + I(X \cap Z) \\ & = H(X | Z) + H(X) - H(X | Z) + I(X \cap Z) \\ & = H(p) + I(X \cap Z) \\ & = H(G \cup F, p) + H(G \cap F, p). \end{aligned}$$

Assim, basta provarmos que

$$\sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(G[U], p_U) + I(X \cap Z) \geq H(G, p) \quad \text{e} \quad (7.2.1)$$

$$\sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(F[U], p_U) + I(X \cap Z) \geq H(F, p). \quad (7.2.2)$$

Provaremos a relação (7.2.1), a prova da relação (7.2.2) é similar.

Seja Y uma variável aleatória tal que, dado $X = x$ e $Z = U$, então Y toma seus valores em $\{S \in \mathcal{S}(G[U]): x \in S \subseteq U\}$ e, para cada $U \in \mathcal{S}(G \cap F)$, vale que

$$H(G[U], p_U) = I(X \cap Y \mid Z = U).$$

Assim, temos que

$$\begin{aligned} \sum_{U \in \mathcal{S}(G \cap F)} \mathbb{P}[Z = U] H(G[U], p_U) + I(X \cap Z) \\ &= I(X \cap Y) + I(X \cap Z) \\ &= H(X) - H(X \mid Y, Z) \\ &= I(X \cap YZ) \geq I(X \cap Y) \\ &\geq H(G, p) \end{aligned}$$

e estamos feitos. □

7.3 Pares modulares

Um par de grafos G e F é dito *modular* se é ao mesmo tempo submodular e supermodular.

Teorema 7.3.1 *Sejam G e F grafos sobre um conjunto V tais que $T := G \cup F = K_V$. Então, G e F formam um par modular se e somente se as seguintes condições são satisfeitas:*

- *todo triângulo em T tem pelo menos duas de suas arestas em $G - E(F)$, $F - E(G)$ ou $G \cap F$;*
- *não existe subconjunto $U \subseteq V$ tal que $F[U]$ e $G[U]$ são disjuntos nas arestas e são imperfeitos.*

Prova: Segue diretamente dos teoremas [7.1.1](#) e [7.2.1](#). □

Parte II

Aplicações

Capítulo 8

Ordenação

Neste capítulo apresentamos uma aplicação de entropia de grafos ao problema de ordenação parcial. A aplicação foi desenvolvida por Kahn e Kim [11].

8.1 Preliminares e notação

Seja V um conjunto finito. Uma *ordem parcial* sobre V é uma relação \leq_P sobre V que é reflexiva, anti-simétrica e transitiva. Abusando da notação, chamamos $P = (V, \leq_P)$ de ordem parcial. Dizemos que $u, v \in V$ são *comparáveis em P* se $u \leq_P v$ ou $v \leq_P u$. Se $u, v \in V$ não são comparáveis, eles são *incomparáveis em P* .

Uma *ordem total* sobre V é uma ordem parcial \leq_Q tal que, para quaisquer $u, v \in V$, vale que $u \leq_Q v$ ou $v \leq_Q u$. Abusando da notação, chamamos $Q = (V, \leq_Q)$ de ordem total. Uma ordem total $Q = (V, \leq_Q)$ é uma *extensão linear* de uma ordem parcial $P = (V, \leq_P)$ se, para quaisquer $u, v \in V$, temos que $u \leq_P v$ implica $u \leq_Q v$. Denote por $e(P)$ o número de extensões lineares de P .

Seja $P = (V, \leq_P)$ uma ordem parcial. Uma *cadeia de P* é um subconjunto de V cujos elementos são dois-a-dois comparáveis. Uma *anticadeia de P* é um subconjunto de V cujos elementos são dois-a-dois incomparáveis. Para anticadeias X, Y de P , dizemos que $X \prec_P Y$ se, para todo $x \in X$, existe $y \in Y$ tal que $x \leq_P y$. Quando não houver dúvidas quanto a ordem parcial em questão usaremos apenas $X \prec Y$.

O *grafo de comparabilidade de P* é definido como o grafo sobre V no qual dois vértices são adjacentes se são comparáveis em P . Denotamos o grafo de comparabilidade de uma ordem parcial P por G_P .

Seja $U \subseteq V$. Definimos o *conjunto minimal de U com relação a P* como

$$\min_P(U) := \{u \in U : u \leq_P v \text{ ou } u \text{ é incomparável com } v, \text{ para todo } v \in U\}.$$

Definimos o *conjunto maximal de U com relação a P* como

$$\max_P(U) := \{u \in U : v \leq_P u \text{ ou } u \text{ é incomparável com } v, \text{ para todo } v \in U\}.$$

Seja $\{v_1, \dots, v_m\} \subseteq V$ tal que $v_1 < \dots < v_m$ são relações compatíveis com P , isto é, $v_1 < \dots < v_m$ vale em alguma extensão linear de P . Denotamos por $P(v_1 < \dots < v_m)$ a menor ordem parcial compatível com P que contém as relações $v_1 < \dots < v_m$. Mais formalmente, $P(v_1 < \dots < v_m)$ é a ordem parcial $P' = (V, \leq_{P'})$, onde $u \leq_{P'} w$ se e somente se $u \leq_P w$ ou, se existem $1 \leq i \leq j \leq m$, tais que $u \leq_P v_i$ e $v_j \leq_P w$.

No restante do texto, $P = (V, \leq_P)$ sempre denotará uma ordem parcial, e $n := |V|$. Algumas vezes será conveniente confundirmos o conjunto V com o par ordenado P ; por exemplo, podemos dizer que x está em P quando, na verdade, x é um elemento de V . Além disso, abreviamos $H(P) := H(G_P, p)$ e $H(\overline{P}) := H(\overline{G_P}, p)$, onde p é a distribuição de probabilidade uniforme sobre V . Denotamos por $a_{\min}(P)$ o vetor $a \in \text{STAB}(G_P)$ que atinge o mínimo na caracterização (2.3.1) de $H(P)$. Denotamos por $b_{\min}(P)$ o vetor $b \in \text{STAB}(\overline{G_P})$ que atinge o mínimo na caracterização (2.3.1) de $H(\overline{P})$.

8.2 Ordenação a partir de informação parcial

Seja $Q = (V, \leq_Q)$ uma ordem total. Um *oráculo para Q* é um oráculo capaz de responder a perguntas do tipo “ $u <_Q v$?”, para quaisquer $u, v \in V$.

O problema de *ordenação a partir de informação parcial* consiste em:

dados um conjunto V , uma ordem parcial $P = (V, \leq_P)$ e um oráculo para uma extensão linear Q de P , encontrar Q .

Chamamos esse problema de *ordenar P* .

Uma possível dificuldade para esse problema é que o oráculo pode ser considerado um adversário que tenta, a todo custo, forçar um algoritmo candidato para o problema a fazer um grande número de consultas. Por exemplo, o oráculo não precisa ter uma extensão linear pré-fixada: ele pode construir a extensão linear de acordo com as consultas feitas pelo algoritmo.

É claro que todo algoritmo que resolve o problema acima fará pelo menos $\lg e(P)$ comparações no pior caso. Esse fato é conhecido como *limite inferior da teoria da informação*. Fredman [7] mostrou que o problema pode ser resolvido com $\lg e(P) + 2n$ comparações. No entanto, a dificuldade encontra-se em como descobrir quais comparações devem ser feitas.

Uma conjectura famosa de Fredman é que, se P não é uma ordem total, então existem x e y elementos incomparáveis em P tais que

$$\frac{1}{3} \leq \frac{e(P(x < y))}{e(P)} \leq \frac{2}{3}.$$

Essa conjectura continua em aberto. No entanto, usando o teorema de Brunn-Minkowski ou as desigualdades de Aleksandrov-Fenchel, já se provou que, se P não é uma ordem total, então existem x e y elementos incomparáveis de P tais que

$$\delta \leq \frac{e(P(x < y))}{e(P)} \leq 1 - \delta$$

para valores de δ menores do que $1/3$, como por exemplo $3/11$ (vide [12, 13]). Isso já é o suficiente para mostrar que, se um algoritmo encontra x e y adequadamente, então podemos

ordenar P com $O(\lg e(P))$ comparações. Novamente, a dificuldade se encontra em descobrir tais comparações. Vamos mostrar uma aplicação de entropia de grafos para esse problema, proposta por Kahn e Kim [11].

8.3 Uma visão geral

Os principais resultados de Kahn e Kim [11] são os seguintes:

- existe um algoritmo que resolve o problema de ordenar a partir de uma ordem parcial P com $O(\lg e(P))$ comparações e que encontra as comparações em tempo polinomial no tamanho de P ;
- existe um algoritmo que computa respostas para consultas ao oráculo e roda em tempo polinomial no tamanho de P para cada consulta, que força todo algoritmo que ordena P (determinístico ou não) a usar $\Omega(\lg e(P))$ comparações.

Para prová-los, Kahn e Kim usaram uma abordagem não-convencional. Eles primeiro relacionaram o número de extensões lineares de P com a entropia de G_P de acordo com a distribuição de probabilidade uniforme. Para mostrar o primeiro resultado, eles mostraram que, se P não é uma ordem total, então existem x e y tais que, incorporando em P a resposta do oráculo relativa à consulta “ $x < y?$ ”, a entropia de G_P aumenta em pelo menos c/n , onde $c \approx 0,2$. Para o segundo resultado, eles mostraram que, para quaisquer x e y incomparáveis em P , pode-se responder a pergunta “ $x < y?$ ” de forma que a entropia de G_P não aumenta em mais que $2/n$.

Na seção 8.4, mostramos que os grafos de comparabilidade são perfeitos e apresentamos algumas conseqüências desse fato. Na seção 8.6, relacionamos $e(P)$ e $H(P)$. Nas duas outras seções, mostramos a existência dos algoritmos citados acima.

8.4 Grafos de comparabilidade

Nesta seção, mostramos que os grafos de comparabilidade são perfeitos e apresentamos algumas conseqüências importantes desse fato.

Lema 8.4.1 *Grafos de comparabilidade são perfeitos.*

Prova: Seja G o grafo de comparabilidade de uma ordem parcial (V, \leq) qualquer. Evidentemente todo subgrafo induzido de um grafo de comparabilidade também é um grafo de comparabilidade. Logo, basta mostrarmos que $\chi(G) \leq \omega(G)$. Para cada vértice v construa uma cadeia de tamanho máximo $C_v := \{u_1, \dots, u_k\}$ com $u_1 = v$ e $u_1 < \dots < u_k$. Seja ℓ o tamanho da maior cadeia assim construída. Para cada $1 \leq i \leq \ell$, tome $A_i := \{v \in V : |C_v| = i\}$. Note que dois vértices distintos pertencentes a um mesmo conjunto A_i não podem ser comparáveis. Portanto, cada A_i é um conjunto estável. Note também que $\bigcup_{i=1}^{\ell} A_i = V$. Assim, $\chi(G) \leq \ell = \omega(G)$, já que cada cadeia é uma clique. \square

Lema 8.4.2 Para toda ordem parcial P sobre V ,

$$H(P) + H(\overline{P}) = \lg |V|,$$

onde p é a distribuição de probabilidade uniforme sobre V .

Prova: Segue imediatamente do lema 8.4.1 e do teorema 5.2.2. \square

Usaremos também o seguinte resultado:

Lema 8.4.3 Existe um algoritmo polinomial para calcular $H(P)$.

Omitimos a demonstração. A idéia principal é a seguinte: como os grafos de comparabilidade são perfeitos, então o politopo dos conjuntos estáveis de um grafo de comparabilidade é separável. Isso permite que apliquemos o método dos elipsóides para calcular a entropia de grafos de comparabilidade com relação a qualquer distribuição de probabilidade. Recomendamos o artigo de Knuth [14] sobre a função ϑ de Lovász e o livro sobre o método dos elipsóides de Grötschel, Lovász e Schrijver [10].

8.5 Decomposição laminar

Nesta seção, apresentamos alguns lemas que serão muito úteis. Em particular, mostramos que podemos decompor $a_{\min}(P)$ de uma maneira especial e única, chamada de decomposição laminar de $a_{\min}(P)$.

Lema 8.5.1 Seja $a \in \text{STAB}(G_P)$ e seja $b \in \text{STAB}(\overline{G_P})$. Então $ab \leq 1$.

Prova: Pela demonstração do lema 2.4.3 da subaditividade, podemos ver que o vetor $a \circ b$, definido como

$$(a \circ b)_v := a_v b_v,$$

para todo $v \in V$, pertence a $\text{STAB}(G_P \cup \overline{G_P}) = \text{STAB}(K_V)$. Como grafos completos são perfeitos, então pelo teorema 5.1.2, $\text{STAB}(K_V) = \text{QSTAB}(K_V)$. Assim, como V é uma clique em K_V , então, pela definição 4.3.1 de $\text{QSTAB}(K_V)$, temos que $ab = \sum_{v \in V} a_v b_v \leq 1$. \square

Lema 8.5.2 Para todo $v \in P$,

$$(a_{\min}(P))_v (b_{\min}(P))_v = \frac{1}{n}. \tag{8.5.1}$$

Prova: Tome $a := a_{\min}(P)$ e $b := b_{\min}(P)$. Pelo lema 8.4.2, temos que $H(P) + H(\overline{P}) = \lg n$. Portanto, $-\sum_{v \in P} (\lg(a_v b_v)) / n = \lg n$. Isto é, o vetor $a \circ b$ (cuja definição pode ser vista no lema anterior) atinge o mínimo na caracterização (2.3.1) de $H(K_V, p)$, onde p é a distribuição de probabilidade uniforme sobre V . Ademais, pela demonstração do lema 2.4.3, podemos ver que $a \circ b \in \text{STAB}(K_V)$. Assim, pelo lema 2.2.2 e pela demonstração do lema 2.6.2, é fácil ver que $a \circ b = p$. \square

Lema 8.5.3 *Seja $a \in \mathbb{R}_+^V$. Suponha que a pode ser escrito como*

$$a = \sum_{i=1}^r \lambda_i \chi^{A_i}, \quad (8.5.2)$$

onde λ_i é um real positivo para todo i e $A_1 \prec \cdots \prec A_r$ são anticadeias maximais distintas. Então, a representação (8.5.2) é única.

Prova: Seja $P^+ := \{x \in P: a_x > 0\}$. Seja $A := \min_P(P^+)$ e $\alpha := \min\{a_x: x \in A\}$. Vamos provar que $A = A_1$ e $\alpha = \lambda_1$. Note que isso prova o lema.

É óbvio que $A \supseteq A_1$. Suponha que $A \not\subseteq A_1$. Então existe x em $A \setminus A_1$. Portanto, x está em algum A_i com $i > 1$. Se x é comparável com algum elemento de A_{i-1} , isso contradiz a hipótese de que $A_{i-1} \prec A_i$. Se x é incomparável com todo elemento de A_{i-1} , isso contradiz a maximalidade de A_{i-1} . Portanto, $A = A_1$.

Agora vamos provar que $\alpha = \lambda_1$. É óbvio que $\alpha \geq \lambda_1$. Suponha que $\alpha > \lambda_1$. Se $r < 2$, então isso é um absurdo. Se $r \geq 2$, isso implica que todo $x \in A_1$ está em mais algum A_i com $i \geq 2$. Como $A_1 \prec \cdots \prec A_r$, então $A_1 \subseteq A_2$. Isso contradiz a hipótese de que A_1 e A_2 são anticadeias maximais distintas. \square

Chamamos a representação de a na equação (8.5.2) de *decomposição laminar de a* .

A demonstração do lema a seguir utiliza uma técnica muito conhecida e poderosa: a técnica do descruzamento. Ela tem sido utilizada para a demonstração de muitos resultados célebres, como o modelo de fluxos submodulares de Edmonds e Giles [5] e um resultado de cobertura bi-supermodular de Frank e Jordan [6], usado para aumento de conectividade.

Lema 8.5.4 *Existe uma única decomposição laminar de $a_{\min}(P)$.*

Prova: Pelo lema 8.5.3, basta mostrar que existe uma decomposição laminar de $a_{\min}(P)$.

Fixe uma extensão linear α da relação \prec . Abrevie $\mathcal{S}_{\max} := \mathcal{S}_{\max}(G_P)$. Dados vetores $\lambda, \lambda' \in \mathbb{R}_+^{\mathcal{S}_{\max}}$, dizemos que λ é *lexicograficamente maior* que λ' se $\lambda_S > \lambda'_S$ para o menor $S \in \mathcal{S}_{\max}$ (sob a ordem total α) tal que $\lambda_S \neq \lambda'_S$.

Podemos escrever $a_{\min}(P)$ como combinação convexa de todos os elementos do conjunto $\{\chi^S: S \in \mathcal{S}_{\max}\}$. Seja $a_{\min}(P) = \sum \{\lambda_S \chi^S: S \in \mathcal{S}_{\max}\}$ uma tal combinação com λ lexicograficamente maximal. É fácil provar que tal combinação existe através de técnicas padrões de compacidade.

Se $\{A \in \mathcal{S}_{\max}: \lambda_A > 0\}$ é uma cadeia sob \prec , nada temos a demonstrar. Suponha então que existem $A, A' \in \mathcal{S}_{\max}$, incomparáveis sob \prec e tais que $0 < \lambda_A \leq \lambda_{A'}$. Tome

$$B := \min_P(A \cup A') \quad \text{e} \quad B' := \max_P(A \cup A')$$

e defina $\lambda' \in \mathbb{R}_+^{\mathcal{S}_{\max}}$ como

$$\lambda'_S := \begin{cases} \lambda_S - \lambda_A, & \text{se } S = A \text{ ou } S = A'; \\ \lambda_S + \lambda_A, & \text{se } S = B \text{ ou } S = B'; \\ \lambda_S, & \text{caso contrário.} \end{cases}$$

É fácil ver que B e B' são anticadeias maximais e que

$$\chi_x^B + \chi_x^{B'} = \chi_x^A + \chi_x^{A'}$$

para todo $x \in A \cup A'$. Logo,

$$a = \sum_{S \in \mathcal{S}_{\max}} \lambda'_S \chi^S.$$

No entanto, é fácil ver que λ' é lexicograficamente maior do λ , pois $B \prec A'$ e $B \prec A$, o que é um absurdo. \square

8.6 Limitantes

Nesta seção relacionamos $e(P)$ com $H(P)$. Queremos provar que

$$n(\lg n - H(P)) \geq \lg e(P) \geq \max\{\lg(n!) - nH(P), Cn(\lg n - H(P))\},$$

onde $C := (1 + 7 \lg e)^{-1}$. Primeiro, usando volumes de poliedros, provamos que

$$2^{-nH(P)} \leq \frac{e(P)}{n!} \leq \frac{n^n}{n!} 2^{-nH(P)}.$$

Essa é uma demonstração bem simples. Já a prova de que

$$\lg e(P) \geq Cn(\lg n - H(P))$$

é um pouco mais trabalhosa e ocupa a maior parte desta seção.

Definimos o *politopo da ordem P* como

$$\mathcal{O}(P) := \{y \in [0, 1]^P : y_u \leq y_v \ \forall u, v \in P \text{ com } u <_P v\}.$$

O *volume* de um poliedro $A \in \mathbb{R}_+^V$ é

$$\text{vol}(A) := \int_{x \in A} dx.$$

Linial [22] observou que $\text{vol}(\mathcal{O}(P)) = e(P)/(n!)$. Stanley [28] provou que $\text{STAB}(G_P)$ e $\mathcal{O}(P)$ têm o mesmo volume. Portanto,

$$\text{vol}(\text{STAB}(G_P)) = \frac{e(P)}{n!}. \tag{8.6.1}$$

Lema 8.6.1 *Vale que*

$$2^{-nH(P)} \leq \text{vol}(\text{STAB}(G_P)) \leq \frac{n^n}{n!} 2^{-nH(P)}.$$

Prova: Como $\text{STAB}(G_P)$ é um canto convexo e $a_{\min}(P) \in \text{STAB}(G_P)$, então

$$\text{vol}(\text{STAB}(G_P)) \geq \prod_{v \in P} a_{\min}(P)_v = 2^{-nH(P)}.$$

Resta provarmos que $\text{vol}(\text{STAB}(G_P)) \leq (n^n/n!)2^{-nH(P)}$. Tome

$$L := \left\{ s \in \mathbb{R}_+^P : \sum_{v \in P} s_v b_{\min}(P)_v \leq 1 \right\}.$$

Pelo lema 8.5.1, vale que $\text{STAB}(G_P) \subseteq L$. Portanto, pelo lema 8.5.2,

$$\text{vol}(\text{STAB}(G_P)) \leq \text{vol}(L) = \frac{1}{n!} \prod_{v \in P} \frac{1}{b_{\min}(P)_v} = \frac{n^n}{n!} \prod_{v \in P} a_{\min}(P)_v = \frac{n^n}{n!} 2^{-nH(P)}.$$

□

Corolário 8.6.1.1 *Seja c uma constante positiva. Se $e(P) \geq cn$, então*

$$nH(\bar{P}) \leq \frac{c + \lg e}{c} \lg e(P).$$

Prova: Pelo lema 8.6.1 e pela equação (8.6.1),

$$\lg e(P) - \lg(n!) \geq -nH(P).$$

Pelo lema 8.4.2,

$$\lg e(P) - \lg(n!) + n \lg n \geq nH(\bar{P}).$$

Suponha que $\lg e(P) \geq cn$. Então

$$\frac{c + \lg e}{c} \lg e(P) \geq \lg e(P) + \lg e^n \geq \lg e(P) + \lg \frac{n^n}{n!},$$

onde a última desigualdade segue do fato que $k! \geq (k/e)^k$ para todo $k \geq 1$. □

Seja $\{x_1, \dots, x_\ell\}$ uma cadeia de comprimento máximo em P , com $x_1 <_P \dots <_P x_\ell$. Seja $C := \{x_1, \dots, x_\ell\}$ e $T := \{y_1, \dots, y_t\} := P \setminus C$. Escrevemos $x \sim y$ para dizer que x e y são comparáveis em P , e $x \approx y$ caso contrário. Para cada $j \in [t]$, defina

$$\begin{aligned} K(j) &:= \{i \in [\ell] : x_i \approx y_j\}, & k_j &:= |K_j|; \\ f(j) &:= \min\{i \in [\ell] : y_j <_P x_i\}, & \text{considerando } \min \emptyset &:= \ell + 1; \\ g(j) &:= \max\{i \in [\ell] : x_i <_P y_j\}, & \text{considerando } \max \emptyset &:= 0. \end{aligned}$$

Para cada $i \in [\ell]$, defina

$$\begin{aligned} U(i) &:= \{j \in [t] : f(j) = i\}, & u_i &:= |U_i|; \\ Z(i) &:= \{j \in [t] : g(j) = i\}, & z_i &:= |Z_i|. \end{aligned}$$

É fácil provar que

$$e(P) \geq 2^t. \quad (8.6.2)$$

Dizemos que $x \in P$ é um *ponto de corte de P* se x é comparável a todos os elementos de P .

Lema 8.6.2 *Se $t < n/7$ e P não tem um ponto de corte, então existe $j \in [t]$ tal que*

$$\sum_{i \in K(j)} (u_i + z_i) \leq k_j \quad \text{e} \quad k_j \geq 3.$$

Prova: Suponha que não existe tal j . Seja $T' \subseteq T$ minimal tal que

$$\bigcup \{K(j) : j \in [t], y_j \in T'\} = [\ell]. \quad (8.6.3)$$

Note que tal T' existe, pois $\bigcup \{K_j : j \in T\} = [\ell]$. Podemos supor sem perda de generalidade que $T' = \{y_1, \dots, y_r\}$. Portanto,

$$\sum_{i \in K(j)} (u_i + z_i) \geq k_j - 2$$

para $1 \leq j \leq r$. Logo,

$$\sum_{j=1}^r \sum_{i \in K(j)} (u_i + z_i) \geq \sum_{j=1}^r k_j - 2r. \quad (8.6.4)$$

Pela equação (8.6.3) e usando o fato de que $r \leq t$, temos que

$$\sum_{j=1}^r k_j - 2r \geq \ell - 2t. \quad (8.6.5)$$

Por outro lado, como a minimalidade de T' implica que todo $i \in [\ell]$ pode estar em, no máximo, dois $K(j)$ distintos, então

$$\sum_{j=1}^r \sum_{i \in K(j)} (u_i + z_i) \leq \sum_{i=1}^{\ell} 2(u_i + z_i) \leq 2t + 2t = 4t. \quad (8.6.6)$$

Assim, usando as equações (8.6.4)–(8.6.6), temos que $4t \geq \ell - 2t$. Logo, $6t \geq \ell = n - t$, o que contradiz a hipótese de que $t < n/7$. \square

Dizemos que P é *maximal com relação à entropia* se o incorporação de qualquer relação a P aumenta a entropia, isto se, se $H(P(x < y)) > H(P)$ para quaisquer x e y incomparáveis em P .

Lema 8.6.3 *Suponha que P é maximal com relação à entropia e não tem ponto de corte. Se $t < n/7$, então existem $j \in [t]$ e $i \in [\ell]$ tais que $P' := P(x_i < y_j < x_{i+1})$ satisfaz*

$$e(P') \leq \frac{e(P)}{k_j - 1} \quad \text{e} \quad nH(\overline{P}) \leq nH(\overline{P'}) + 2\lg(2k_j + 1).$$

Prova: Seja j como no lema 8.6.2 e $K(j) = \{x_h, \dots, x_m\}$ com $x_h <_P \dots <_P x_m$. Escolha i em $\{h, \dots, m\}$ que minimize

$$\frac{e(P(x_i < y_j < x_{i+1}))}{e(P)}. \quad (8.6.7)$$

Tome $P' := P(x_i < y_j < x_{i+1})$. Note que as extensões lineares de P em que $y_j < x_i$ ou $x_{i+1} < y_j$ não são extensões lineares de P' . Portanto, como escolhemos i que minimiza (8.6.7),

$$e(P') \leq \frac{e(P)}{k_j - 1}.$$

Agora vamos provar que $nH(\overline{P}) \leq nH(\overline{P}') + 2\lg(2k_j + 1)$. Para isso, vamos provar que

$$v <_P y_j \Rightarrow v <_P x_{i+1}. \quad (8.6.8)$$

Seja $v \in P$. Suponha que $v <_P y_j$. Seja $\sum_{A \in \mathcal{A}} \lambda_A \chi^A$ uma decomposição laminar de $a_{\min}(P)$. Pela maximalidade de P com relação à entropia, existe $A \in \mathcal{A}$ tal que $x_i, y_j \in A$. Seja $A', A'' \in \mathcal{A}$ tais que $v \in A'$ e $x_{i+1} \in A''$. Como $v <_P y_j$, então $A' \prec A$. Como $x_i <_P x_{i+1}$, então $A \prec A''$. Portanto, $A' \prec A''$ e são anticadeias distintas. Novamente pela maximalidade de P com relação à entropia, vale que $v <_P x_{i+1}$, completando a prova da implicação (8.6.8). Similarmente, pode-se provar que

$$y_j <_P v \Rightarrow x_i <_P v. \quad (8.6.9)$$

Note que decorre das implicações (8.6.8) e (8.6.9) que, se $a \approx b$ em P , então $a \sim b$ em P' somente se $a = y_j$ ou $b = y_j$. Por outro lado, y_j só se tornará comparável a elementos de

$$Y := K_j \cup \left(\bigcup_{s \in K_j} U(s) \cup Z(s) \right)$$

Pelo lema 8.6.2, é fácil ver que

$$q := |Y| \leq k_j + k_j = 2k_j.$$

Seja G' o grafo sobre V com $E(G') := E(\overline{GP}) \setminus E(\overline{GP'})$. Seja p a distribuição de probabilidade uniforme sobre v . Temos que

$$\begin{aligned} nH(G', p) &\leq \lg(q + 1) + \sum_{y \in Y} \lg \frac{q + 1}{q} \\ &= \lg(q + 1) + q \lg(q + 1) - q \lg(q) \\ &\leq 2\lg(q + 1) \leq 2\lg(2k_j + 1). \end{aligned} \quad (8.6.10)$$

Pelo lema 2.4.3 da subaditividade e pela desigualdade (8.6.10),

$$nH(\overline{P}) \leq nH(\overline{P}') + nH(G', p) \leq nH(\overline{P}') + 2\lg(2k_j + 1)$$

e estamos feitos. \square

Lema 8.6.4 *Vale que*

$$nH(\overline{P}) \leq (1 + 7 \lg e) \lg(e(P)). \quad (8.6.11)$$

Prova: A prova é por indução em $n + t$. Se $n = 1$ ou $t = 0$ é claro que a inequação (8.6.11) é verdadeira.

Se P tem um ponto de corte, digamos x , é fácil ver que

$$nH(\overline{P}) = (n - 1)H(\overline{P \setminus \{x\}}) \quad \text{e} \quad e(P) = e(P \setminus \{x\}).$$

Logo, por hipótese de indução

$$nH(\overline{P}) = (n - 1)H(\overline{P \setminus \{x\}}) \leq (1 + 7 \lg e) \lg e(P \setminus \{x\}) = (1 + 7 \lg e) \lg e(P).$$

Suponha então que P não tem ponto de corte. Se $t \geq n/7$, pela inequação (8.6.2) e pelo corolário 8.6.1.1, a inequação (8.6.11) é válida. Portanto, podemos supor que $t < n/7$. Ademais, podemos supor que P é maximal com relação à entropia. Sejam i e j e P' como no lema 8.6.3. Temos que

$$\begin{aligned} nH(\overline{P}) &\leq nH(\overline{P'}) + 2 \lg(2k_j + 1) \\ &\leq (1 + 7 \lg e) \lg e(P') + 4 \lg(k_j + 1) \\ &\leq (1 + 7 \lg e) \lg e(P) + (8 - (1 + 7 \lg e)) \lg(k_j - 1) \\ &\leq (1 + 7 \lg e) \lg e(P). \end{aligned}$$

□

Teorema 8.6.5 *Vale que*

$$\begin{aligned} n(\lg n - H(P)) &\geq \lg e(P) \\ &\geq \max\{\lg(n!) - nH(P), Cn(\lg n - H(P))\}, \end{aligned}$$

onde $C := (1 + 7 \lg e)^{-1}$.

Prova: Segue diretamente do lema 8.4.2, do lema 8.6.1 e da equação (8.6.1), e do lema 8.6.4.

□

8.7 Encontrando uma boa comparação

Nesta seção mostramos um algoritmo que ordena uma ordem parcial P com $O(\lg e(P))$ comparações e encontra as comparações em tempo polinomial no tamanho de P .

Basicamente, mostramos que se, P não é uma ordem total, então existem x e y em P tais que

$$\min\{H(P(x < y)), H(P(x > y))\} \geq H(P) + \frac{c}{n}, \quad (8.7.1)$$

onde $c := 1 + 17/112$. Isso significa que ao descobirmos a relação entre x e y através de uma consulta ao oráculo, a entropia do grafo de comparabilidade da nova ordem parcial será pelo

menos a soma entre a entropia do grafo de comparabilidade da ordem parcial anterior e c/n . Assim, com, no máximo, $(n/c)(\lg n - H(P))$ comparações atingiremos a entropia do grafo completo, isto é, encontraremos a ordem total do oráculo.

Seja $a := \sum_{i=1}^r \lambda_i \chi^{A_i}$ uma decomposição laminar de $a_{\min}(P)$ com $A_1 \prec \cdots \prec A_r$. Defina

$$\alpha(x) := \min\{i \in [r] : x \in A_i\} \quad \text{e} \quad \beta(x) := \max\{i \in [r] : x \in A_i\}.$$

Lema 8.7.1 *Suponha que P não é uma cadeia. Sejam x, y incomparáveis em P e seja $\mu \in [0, 1]$. Seja $P' := P(x < y)$ e suponha que $a_y > 0$. Então*

$$nH(P') \geq nH(P) + \lg \left(1 + \mu \sum_{i=1}^{\beta(x)} \frac{\lambda_i}{a_y} \right) + \lg \left(1 + \mu \sum_{i=1}^{\alpha(y)-1} \frac{\lambda_i}{a_y} \right).$$

Prova: Seja $b := b_{\min}(P)$. O vetor b pode ser escrito como combinação convexa de elementos de $\{\chi^B : B \text{ é uma cadeia de } P\}$. Seja $\sum_{i=1}^s \xi_i \chi^{B_i}$ uma tal combinação. Podemos supor que $y \in B_i$ se e somente se $1 \leq i \leq m$, onde $m := |\{B_j : i \in B_j, 1 \leq j \leq s\}|$.

Tome

$$d(v) := \sum \{\xi_i : v \in B_i \text{ e } 1 \leq i \leq m\}.$$

Para cada $1 \leq i \leq m$, defina $C_i := B_i \setminus \{v \in P : v <_P y\}$.

Fixe $C = \{v_1, \dots, v_t\}$ com $v_i <_P \cdots <_P v_t$ uma cadeia maximal tal que $v_t = x$. Note que

$$\sum_{i=1}^t a_{v_i} = \sum_{i=1}^{\beta(x)} \lambda_i. \tag{8.7.2}$$

Defina as seguintes cadeias de P'

$$\begin{aligned} B'_i &:= B_i, & \text{se } 1 \leq i \leq s \\ B'_{i+s} &:= C \cup C_i, & \text{se } 1 \leq i \leq m. \end{aligned}$$

Defina também

$$\begin{aligned} \xi'_i &:= \xi_i, & \text{se } m+1 \leq i \leq s \\ \xi_{i+s} &:= \mu \xi_i, & \xi_i := (1 - \mu) \xi_i, & \text{se } 1 \leq i \leq m. \end{aligned}$$

Tome

$$b' := \sum_{i=1}^{s+m} \xi'_i \chi^{B'_i}.$$

É fácil ver que $b' \in \text{STAB}(\overline{G_{P'}})$. Seja $z \in P$. Se $z \in C$, então

$$b'_z = b_z - d(z) + (1 - \mu)d(z) + \mu b_y = b_z + \mu(b_y - d(z)).$$

Se $z \notin C$ e $z <_P y$, então $b'_z = b_z - \mu d(z)$. Finalmente, se $z \notin C$, e z é incomparável com y ou $y <_P z$, então $b'_z = b_z$.

Usaremos as seguintes desigualdades,

$$\lg(1 + u - v) \geq \lg(1 + u) + \lg(1 - v) \quad (8.7.3)$$

para quaisquer $u, v \in \mathbb{R}_+$ e

$$\lg(1 + u) + \lg(1 + v) \geq \lg(1 + u + v) \quad (8.7.4)$$

para quaisquer $u, v \in \mathbb{R}$ com $uv \geq 0$.

Pelo lema 8.4.2 e pelas desigualdades (8.7.3) e (8.7.4), temos que

$$\begin{aligned} nH(P') - nH(P) &= nH(\overline{P}) - nH(\overline{P}') \\ &\geq \sum_{v \in P} \lg \frac{b'_v}{b_v} = \sum \left\{ \lg \frac{b'_v}{b_v} : v \in C \right\} + \sum \left\{ \lg \frac{b'_v}{b_v} : v \in P \setminus C, v <_P y \right\} \\ &= \sum \left\{ \lg \left(1 + \mu \frac{b_y}{b_v} - \mu \frac{d(v)}{b_v} \right) : v \in C \right\} + \sum \left\{ \lg \left(1 - \mu \frac{d(v)}{b_v} \right) : v \in P \setminus C, v <_P y \right\} \\ &\geq \sum \left\{ \lg \left(1 + \mu \frac{b_y}{b_v} \right) : v \in C \right\} + \sum \left\{ \lg \left(1 - \mu \frac{d(v)}{b_v} \right) : v \in P, v <_P y \right\} \\ &\geq \lg \left(1 + \mu b_y \sum \left\{ \frac{1}{b_v} : v \in C \right\} \right) + \lg \left(1 - \mu \sum \left\{ \frac{d(v)}{b_v} : v \in P, v <_P y \right\} \right) \end{aligned}$$

Pelo lema 8.5.2 e pela equação (8.7.2),

$$\sum \left\{ \frac{1}{b_v} : v \in C \right\} = \sum \{ n a_v : v \in C \} = n \sum_{i=1}^{\beta(x)} \lambda_i.$$

Além disso,

$$\begin{aligned} \sum_{v <_P y} \frac{d(v)}{b_v} &= n \sum_{v <_P y} a_v d(v) = n \sum_{v <_P y} \sum \{ \lambda_i d(v) : v \in A_i \} \\ &= n \sum_{i=1}^{\alpha(y)-1} \lambda_i \sum \{ d(v) : v \in A_i \} \leq n \sum_{i=1}^{\alpha(y)-1} \lambda_i b_y, \end{aligned}$$

onde a última desigualdade segue do fato de que A_i é uma anticadeia para todo i . Assim,

$$\begin{aligned} nH(P') - nH(P) &\geq \lg \left(1 + \mu b_y \sum \left\{ \frac{1}{b_v} : v \in C \right\} \right) + \lg \left(1 - \mu \sum \left\{ \frac{d(v)}{b_v} : v \in P, v <_P y \right\} \right) \\ &\geq \lg \left(1 + \mu n \sum_{i=1}^{\beta(x)} \lambda_i b_y \right) + \lg \left(1 - \mu n \sum_{i=1}^{\alpha(y)-1} \lambda_i b_y \right) \\ &= \lg \left(1 + \mu \sum_{i=1}^{\beta(x)} \frac{\lambda_i}{a_y} \right) + \lg \left(1 - \mu \sum_{i=1}^{\alpha(y)-1} \frac{\lambda_i}{a_y} \right). \end{aligned}$$

□

Antes de provar a desigualdade (8.7.1), precisamos de um lema fácil.

Lema 8.7.2 *Dados $0 < \varepsilon_1 < 1$ e $0 < \varepsilon_2 < 1$, escolha x com a_x tão grande quanto possível de forma que*

$$\sum_{i=1}^{\alpha(x)-1} \lambda_i \leq \varepsilon_1 a_x.$$

Seja s o menor inteiro para o qual

$$\sum_{i=\alpha(x)}^s \lambda_i \geq \varepsilon_2 a_x.$$

Então, para todo $y \in A_s \setminus \{x\}$,

$$a_y < \frac{\varepsilon_1 + \varepsilon_2}{\varepsilon_1} a_x.$$

Prova: Se $a_y \leq a_x$, não há nada a provar. Suponha que $a_y > a_x$. Então, pela escolha de x e pelo fato de que $s \geq \alpha(y)$, temos que

$$\varepsilon_1 a_y \leq \sum_{i=1}^{\alpha(y)-1} \lambda_i = \sum_{i=1}^{\alpha(x)-1} \lambda_i + \sum_{i=\alpha(x)}^{\alpha(y)-1} \lambda_i < \varepsilon_1 a_x + \varepsilon_2 a_x.$$

□

Finalmente provamos a desigualdade (8.7.1).

Teorema 8.7.3 *Se P não é uma cadeia, então existem x, y incomparáveis em P tais que*

$$\min\{H(P(x < y)), H(P(y < x))\} \geq H(P) + \frac{c}{n}, \quad (8.7.5)$$

onde $c := (1 + 17/112)$.

Prova: Suponha P possui um ponto de corte z . Então, a prova segue por indução em n . Para $n \leq 3$, é fácil ver que a desigualdade (8.7.5) é válida. Suponha que $n > 3$. Seja p a distribuição de probabilidade uniforme sobre os elementos de P e seja p' a distribuição de probabilidade uniforme sobre os elementos de $P' := P \setminus \{z\}$. Por hipótese de indução, existem $x, y \in P'$ tais que $\min(H(P'(x < y)), H(P'(y < x))) \geq H(P') + c/(n-1)$. Usando o fato de que $nH(\overline{P}) = (n-1)H(P')$, temos que

$$\begin{aligned} nH(P) - nH(p) &= (n-1)H(P') - (n-1)H(p') \\ &\leq (n-1) \min(H(P'(x < y)), H(P'(y < x))) - (n-1)H(p') + c \\ &= -(n-1) \min(H(\overline{P'(x < y)}), H(\overline{P'(y < x)})) + c \\ &= -n \min(H(\overline{P(x < y)}), H(\overline{P(y < x)})) + c \\ &= n \min(H(P(x < y)), H(P(y < x))) - nH(p) + c. \end{aligned}$$

Suponha que P não tem um ponto de corte. Tome $\varepsilon_1 := 1/4$ e $\varepsilon_2 := 1/3$. Sejam x e y de acordo com o lema 8.7.2. Tome $\delta := (1/a_x) \sum \{\lambda_i : 1 \leq i \leq \alpha(x) - 1\}$. Note que $\delta \leq \varepsilon_1$. Pelo lema 8.7.2,

$$\mu := \frac{\varepsilon_1 a_y}{(\varepsilon_1 + \varepsilon_2) a_x} \geq 1.$$

Tome $P' := P(x < y)$. Pelo lema 8.7.1 e pelas escolhas de x e y ,

$$\begin{aligned} nH(P') - nH(P) &\geq \lg \left(1 + \mu \sum_{i=1}^{\beta(x)} \frac{\lambda_i}{a_y} \right) + \lg \left(1 + \mu \sum_{i=1}^{\alpha(y)-1} \frac{\lambda_i}{a_y} \right) \\ &= \lg \left(1 + \mu \sum_{i=1}^{\alpha(x)-1} \frac{\lambda_i}{a_y} + \mu \sum_{i=\alpha(x)}^{\beta(x)} \frac{\lambda_i}{a_y} \right) + \lg \left(1 + \mu \sum_{i=1}^{\alpha(y)-1} \frac{\lambda_i}{a_y} \right) \\ &= \lg \left(1 + \mu \frac{\delta a_x}{a_y} + \mu \frac{a_x}{a_y} \right) + \lg \left(1 + \mu \sum_{i=1}^{\alpha(y)-1} \frac{\lambda_i}{a_y} \right) \\ &= \lg \left(1 + \mu \frac{(\delta + 1) a_x}{a_y} \right) + \lg \left(1 + \mu \sum_{i=1}^{\alpha(y)-1} \frac{\lambda_i}{a_y} \right) \\ &= \lg \left(1 + \mu \frac{(\delta + 1) a_x}{a_y} \right) + \lg \left(1 + \mu \sum_{i=1}^{\alpha(x)-1} \frac{\lambda_i}{a_y} + \mu \sum_{j=\alpha(x)}^{\alpha(y)-1} \frac{\lambda_j}{a_y} \right) \\ &\geq \lg \left(1 + \mu \frac{(\delta + 1) a_x}{a_y} \right) + \lg \left(1 + \mu \frac{\delta a_x}{a_y} + \mu \frac{\varepsilon_2 a_x}{a_y} \right) \\ &= \lg \left(1 + \mu \frac{(\delta + 1) a_x}{a_y} \right) + \lg \left(1 + \mu \frac{(\delta + \varepsilon_2) a_x}{a_y} \right) \\ &\geq \lg \left(1 + \frac{\varepsilon_1 - \varepsilon_1 \varepsilon_2 - \varepsilon_2^2 - \varepsilon_1^2}{\varepsilon_1 + \varepsilon_2} \right) = \lg \left(1 + \frac{17}{112} \right). \end{aligned}$$

Por outro lado, tome $P'' := P(y < x)$. Tome $\eta := 1$. Pelo lema 8.7.1,

$$\begin{aligned} nH(P'') - nH(P) &\geq \lg \left(1 + \eta \sum_{i=1}^{\beta(y)} \frac{\lambda_i}{a_x} \right) + \lg \left(1 + \eta \sum_{i=1}^{\alpha(x)-1} \frac{\lambda_i}{a_x} \right) \\ &\geq \lg(1 + \delta + \varepsilon_2) + \lg(1 - \delta) = \lg(1 + \varepsilon_2 - \varepsilon_2 \delta - \delta^2) \\ &\geq \lg(1 + \varepsilon_2 - \varepsilon_2 \varepsilon_1 - \varepsilon_1^2) = \lg \left(1 + \frac{3}{16} \right). \end{aligned}$$

□

Vamos mostrar agora que, do teorema 8.7.3, segue facilmente a existência do algoritmo desejado.

Corolário 8.7.3.1 *Existe um algoritmo que resolve o problema de ordenar a partir de uma ordem parcial com $O(\lg e(P))$ comparações e encontra as comparações em tempo polinomial no tamanho de P .*

Prova: Considere o seguinte algoritmo.

Algoritmo

- 1 $P' \leftarrow P$
- 2 enquanto $H(P') < \lg n$ faça
- 3 encontre x, y tais que
 $\min\{H(P'(x < y)), H(P'(y < x))\} \geq H(P') + c/n,$
 onde $c = 1 + 17/112$
- 4 pergunte ao oráculo: “ $x < y?$ ”
- 5 se o oráculo responder “SIM”
- 6 então $P' \leftarrow P'(x < y)$
- 7 senão $P' \leftarrow P'(y < x)$
- 8 devolva P'

Pelo teorema 8.7.3, se P' não é uma cadeia, tais x e y existem. Além disso, pelo lema 8.4.3 podemos calcular $H(P')$, $H(P'(x < y))$ e $H(P'(y < x))$ em tempo polinomial. Note que o algoritmo só termina quando encontra uma ordem total, pois pelo lema 2.6.2, a entropia de um grafo completo com n vértice com relação à distribuição uniforme é $\lg n$.

Como em cada iteração a entropia cresce pelo menos c/n , temos que o algoritmo fará no máximo $(n/c)(\lg n - H(P))$ comparações. Pelo teorema 8.6.5, vale que $\lg(e(P)) \geq Cn(\lg n - H(P))$, onde $C := (1 + 7 \lg e)^{-1}$. Assim, o algoritmo faz $O(\lg e(P))$ comparações. \square

8.8 Computando respostas

Nesta seção mostramos um algoritmo que computa respostas a consultas a um oráculo que obriga todo algoritmo que ordena uma ordem parcial P a fazer $\Omega(e(P))$ comparações.

Basicamente, mostramos que, se P , não é uma ordem total, para quaisquer x, y incomparáveis em P ,

$$\min\{H(P(x < y)), H(P(y < x))\} \leq H(P) + \frac{2}{n}.$$

A pergunta “ $x < y?$ ” será respondida de modo a minimizar a entropia da nova ordem parcial. Isso, significa que a cada comparação, a entropia da nova ordem parcial será, no máximo, a soma entre entropia da ordem parcial anterior e $2/n$. Assim, precisaremos de pelo menos $(n/2)(\lg n - H(P))$ comparações para atingir a entropia do grafo completo, isto é, encontrar a ordem total do oráculo.

Teorema 8.8.1 *Se P não é uma cadeia e x, y são incomparáveis em P , então*

$$\min\{H(P(x < y)), H(P(y < x))\} \leq H(P) + \frac{2}{n},$$

Prova: Tome $a := a_{\min}(P)$. Defina

$$\begin{aligned} U &:= \{v \in P : v <_P x\} & \text{e} & \quad R := \{v \in P : x <_P v\}; \\ W &:= \{v \in P : v <_P y\} & \text{e} & \quad Z := \{v \in P : y <_P v\}. \end{aligned}$$

Para toda cadeia C em P , defina $w(C) := \sum_{x \in C} a_x$. Seja uma cadeia $K \subseteq U$ que maximiza $w(K)$. Escolha $L \subseteq R$, $M \subseteq W$ e $N \subseteq Z$ similarmente. Pelo lema 8.4.1 e pelo teorema 5.1.5, vale que $\text{QSTAB}(G_P) = \text{STAB}(G_P)$. Logo, pela definição (4.3.1) de $\text{QSTAB}(G_P)$,

$$w(K) + w(L) + a_x \leq 1,$$

$$w(M) + w(N) + a_y \leq 1.$$

Portanto,

$$w(K) + w(N) + \frac{a_x + a_y}{2} \leq 1 \quad \text{ou} \quad (8.8.1)$$

$$w(M) + w(L) + \frac{a_x + a_y}{2} \leq 1. \quad (8.8.2)$$

Suponha sem perda de generalidade que a inequação (8.8.1) é verdadeira. Defina $a' \in \mathbb{R}_+^P$ como

$$a'_v := \begin{cases} a_v/2, & \text{se } v = x \text{ ou } v = y; \\ a_v, & \text{caso contrário.} \end{cases}$$

Tome $P' := P(x < y)$. Vamos mostrar que $a' \in \text{QSTAB}(G_{P'})$, pelo teorema 5.1.5, isso implica que $a' \in \text{STAB}(G_{P'})$. Para toda cadeia C de P' , defina $w'(C) := \sum_{x \in C} a'_x$. Seja Q uma cadeia maximal de P' . Se $\{x, y\} \not\subseteq Q$, então é fácil ver que Q é uma cadeia em P . Portanto, como $a' \leq a$,

$$w'(Q) = \sum_{v \in Q} a'_v \leq \sum_{v \in Q} a_v \leq 1.$$

Logo, $a' \in \text{QSTAB}(G_{P'})$. Se $\{x, y\} \subseteq Q$, então tome

$$K' := \{v \in Q : v <_{P'} x\} \quad \text{e} \quad N' := \{v \in Q : y <_{P'} v\}.$$

Note que $K' \subseteq U$ e $N' \subseteq Z$. Note também que K' e N' são cadeias de P . Ademais, $Q = K' \cup N' \cup \{x, y\}$. Assim,

$$\begin{aligned} w'(Q) &= w'(K') + w'(N') + \frac{a_x + a_y}{2} = w(K') + w(N') + \frac{a_x + a_y}{2} \\ &\leq w(K) + w(N) + \frac{a_x + a_y}{2} \leq 1. \end{aligned}$$

Portanto, $a' \in \text{QSTAB}(G_{P'}) = \text{STAB}(G_{P'})$. Assim, como $a'_x = a_x/2$ e $a'_y = a_y/2$,

$$\begin{aligned} H(P') &\leq -\frac{1}{n} \sum_{v \in P'} \lg a'_v \\ &= -\frac{1}{n} \sum_{v \in P' \setminus \{x, y\}} \lg a_v - \frac{1}{n} \lg \frac{a_x}{2} - \frac{1}{n} \lg \frac{a_y}{2} \\ &= -\frac{1}{n} \sum_{v \in P} \lg a_v + \frac{1}{n} \lg 2 + \frac{1}{n} \lg 2 \\ &= -\frac{1}{n} \sum_{v \in P} \lg a_v + \frac{2}{n} = H(P) + \frac{2}{n} \end{aligned}$$

□

Corolário 8.8.1.1 *Existe um algoritmo que computa respostas para perguntas ao oráculo e roda em tempo polinomial no tamanho de P , que força todo algoritmo que ordena P a usar $\Omega(\lg e(P))$ comparações.*

Prova: O algoritmo que computa as respostas do oráculo deve conhecer a ordem parcial P . O oráculo deverá consultar esse algoritmo para responder as consultas de um algoritmo candidato a ordenar P .

Considere o seguinte algoritmo.

Algoritmo

- 1 $P' \leftarrow P$
- 2 enquanto o oráculo faz uma pergunta “ $x < y$?” faça
- 3 se x, y são comparáveis em P'
- 4 então se $x <_{P'} y$
- 5 então devolva “SIM”
- 6 senão devolva “NÃO”
- 7 senão se $H(P'(x < y)) \leq H(P'(y < x))$
- 8 então $P' \leftarrow P'(x < y)$ e devolva “SIM”
- 9 senão $P' \leftarrow P'(y < x)$ e devolva “NÃO”

Pelo teorema 8.8.1, se x e y são incomparáveis em P' , então $H(P'(x < y)) \leq H(P') + 2/n$ ou $H(P'(y < x)) \leq H(P') + 2/n$. Assim, a cada comparação a entropia de $G_{P'}$ aumentará no máximo $2/n$. Pelo teorema 8.6.5, $\lg e(P') \leq n(\lg n - H(P'))$. Isso significa, que o algoritmo que ordena P fará $\Omega(e(P))$ comparações. □

Capítulo 9

Funções de hashing

Neste capítulo apresentamos uma aplicação de entropia de grafos para encontrar delimitações para o tamanho mínimo famílias perfeitas de funções de hashing. Apresentamos uma delimitação superior e duas delimitações inferiores. A entropia de grafos é usada na prova das delimitações inferiores.

9.1 O problema

Seja V um conjunto com n elementos e seja $W = \{W_1, \dots, W_b\}$ uma partição de V . Dizemos que um subconjunto $S \subseteq V$ é *separado por* W se cada elemento de W contém no máximo um elemento de S , isto é, se $|W_i \cap S| \leq 1$ para todo i . Uma família \mathcal{W} de partições de V em b conjuntos é chamada de (b, k) -sistema se todo subconjunto de V com tamanho k é separado por algum elemento de \mathcal{W} . Definimos $Y(b, k, n)$ como o tamanho mínimo de um (b, k) -sistema de um conjunto com n elementos, isto é,

$$Y(b, k, n) = \min\{|\mathcal{W}| : \mathcal{W} \text{ é um } (b, k)\text{-sistema de um conjunto de tamanho } n\}.$$

Uma função h de V em $[b]$ é uma *função de hashing perfeita sobre* $S \subseteq V$ se h é injetora sobre S . Uma família C de funções de V em $[b]$ é uma (b, k) -família perfeita de funções de hashing se, para todo subconjunto $S \subseteq V$ com tamanho k , existe uma função em C que é injetora sobre S . Existe uma correspondência óbvia entre partições de V em b partes e funções de V em $[b]$. Por isso, é claro que $Y(b, k, n)$ é também o tamanho mínimo de uma (b, k) -família perfeita de funções de hashing.

Para todo inteiro não-negativo x , definimos $x^{\underline{k}} := x!/k!$.

9.2 Limitando superiormente

Nesta seção mostramos um limite superior para o tamanho mínimo de uma (b, k) -família perfeita de funções de hashing. Mostramos uma cota provada por Fredman e Komlós [8]. A demonstração não usa entropia de grafos.

Teorema 9.2.1 *Vale que*

$$Y(b, k, n) = O\left(\binom{-k \lg n}{\lg\left(1 - \frac{b^k}{n}\right)}\right).$$

Prova: Seja m o maior número possível de conjuntos com k elementos que podem ser separados por uma partição fixada. Diremos que uma partição é *maximal* se separa m conjuntos com k elementos. É fácil ver que $Y(b, k, n) \geq N/m$, onde $N := \binom{n}{k}$. Seja $t > 0$ um inteiro. Note que o conjunto de todas as partições maximais é um (b, k) -sistema. Considere uma família composta por t partições maximais escolhidas aleatoriamente. A probabilidade dessa família não ser um (b, k) -sistema é limitada por $p := N(1 - m/N)^t$. Logo, se escolhermos t de forma que $p < 1$, então $Y(b, k, n) \leq t$. É fácil ver que $p < 1$, se

$$t > -\frac{\lg N}{\lg(1 - m/N)}.$$

Portanto,

$$Y(b, k, n) = O\left(-\frac{\lg N}{\lg(1 - m/N)}\right).$$

Por outro lado, toda partição maximal é composta por conjuntos de tamanho o mais uniforme possível. Assim, para n grande o suficiente,

$$m \sim \binom{b}{k} \left(\frac{n}{b}\right)^k.$$

Ademais,

$$\lim_{n \rightarrow \infty} \frac{n^k (n - k)!}{n!} = 1.$$

Logo,

$$Y(b, k, n) = O\left(\binom{-k \lg n}{\lg\left(1 - \frac{b^k}{n}\right)}\right),$$

como queríamos. □

9.3 Limitando inferiormente

Nesta seção, mostramos limites inferiores para o tamanho mínimo de uma (b, k) -família perfeita de funções de hashing.

Primeiro, mostramos uma cota provada por Fredman e Komlós [8]. Körner [16] provou essa mesma cota usando entropia de grafos. Apresentamos a demonstração de Körner. Depois, mostramos uma cota melhor provada por Körner e Marton [19]. A demonstração usa entropia de hipergrafos.

Seja t um inteiro positivo. Definimos $N(b, k, t)$ como o tamanho do maior conjunto para o qual t funções com contra-domínio $[b]$ podem formar uma (b, k) -família perfeita de funções de

hashing. Seja B um conjunto com b elementos. Seja $A \subseteq B^t$ e seja A' um conjunto qualquer de mesmo tamanho que A . Como A e A' têm mesmo tamanho, então existe uma bijeção de A' em A . Assim, podemos considerar cada elemento de A' como o valor de t funções de um elemento de A . Logo, $N(b, k, t)$ é o tamanho do maior $A \subseteq B^t$ tal que, para todo $S \subset A$ com $|S| = k$, vale que, se $x, y \in S$, então existe uma coordenada i tal que $x_i \neq y_i$. Dizemos que tal A é k -separado.

Teorema 9.3.1 *Vale que*

$$\frac{1}{t} \lg N(b, k, t) = O\left(\frac{b^{k-1}}{b^{k-1}} \lg(b - k + 2)\right).$$

Prova: Seja $A \subseteq B^t$ k -separado de tamanho $N := N(b, k, t)$, onde $|B| = b$. Vamos construir um grafo em que todo vértice corresponde a um subconjunto de A de tamanho $k - 1$ e cada aresta corresponde a um subconjunto de tamanho k . Seja G o grafo definido como

$$\begin{aligned} V(G) &:= \{(D, x) : D \subseteq A, |D| = k - 2, x \in A \setminus D\} \\ E(G) &:= \left\{ \{(D, x), (D, x')\} \in \binom{V(G)}{2} : D = D', x \neq x' \right\}. \end{aligned}$$

É fácil ver que G tem $\binom{N}{k-2}$ componentes e que cada componente é um grafo completo com $N - k + 2$ vértices. Para cada $1 \leq i \leq t$, defina o grafo G_i sobre $V(G)$ onde dois vértices (D, x) e (D', x') são adjacentes em G e $D \cup \{x, x'\}$ é separado na i -ésima coordenada, isto é, todos os elementos de $D \cup \{x, x'\}$ diferem na i -ésima coordenada. Note que $G = \bigcup_{i=1}^t G_i$. Seja p a distribuição de probabilidade uniforme sobre $V(G)$. Pelo corolário 2.4.3.1,

$$H(G, p) \leq \sum_{i=1}^t H(G_i, p),$$

portanto,

$$\frac{H(G, p)}{t} \leq \max_{i=1}^t H(G_i, p). \quad (9.3.1)$$

É fácil calcular a entropia de G . Usando o corolário 2.6.1.2 e o lema 2.6.2, temos que

$$H(G, p) = \lg(N - k + 2). \quad (9.3.2)$$

Agora vamos procurar um limitante superior para $\max_{i=1}^t H(G_i, p)$. Para cada $D \subseteq A$ com tamanho $k - 2$, defina $V_D := \{(D', x') \in V(G) : D' = D\}$. Note que, para toda coordenada i , o subgrafo induzido $G_i[V_D]$ ou é vazio ou é q -partido, com $q \leq b - k + 2$. Seja $1 \leq \ell \leq t$. Pode-se provar que o número de vértices isolados em G_ℓ é mínimo quando todos os elementos de B aparecem da forma mais balanceada possível na ℓ -ésima coordenada dos elementos de A . Note que, nesse caso, o subgrafo induzido $G_\ell[D]$ é $b - k + 2$ -partido para todo $D \subseteq A$ com tamanho $k - 2$. Seja m o número de vértices não-isolados. Temos que

$$m \sim \binom{b}{k-1} \left(\frac{N}{b}\right)^{k-1} (k-1).$$

Como temos $\binom{N}{k-2}(N-k+2)$ vértices, então a soma das probabilidades dos vértices não-isolados é

$$\frac{b^{k-1}N^{k-1}}{b^{k-1}N^{k-1}}.$$

Com t grande o suficiente,

$$\frac{b^{k-1}N^{k-1}}{b^{k-1}N^{k-1}} \sim \frac{b^{k-1}}{b^{k-1}}.$$

Assim, usando o corolário 2.6.1.2 e o lema 2.6.4, temos que

$$H(G_i, p) \leq \frac{b^{k-1}}{b^{k-1}} \lg(b-k+2). \quad (9.3.3)$$

Usando a desigualdade (9.3.1) e a equação (9.3.2) e a desigualdade (9.3.3), temos que

$$\frac{\lg N(b, k, t)}{t} \leq O\left(\frac{b^{k-1}}{b^{k-1}} \lg(b-k+2)\right).$$

e estamos feitos. \square

Corolário 9.3.1.1 *Vale que*

$$Y(b, k, n) = \Omega\left(\frac{b^{k-1} \lg n}{b^{k-1} \lg(b-k-2)}\right). \quad (9.3.4)$$

Prova: Segue diretamente do teorema 9.3.1. \square

Körner e Marton [19] conseguiram melhorar a cota (9.3.4) usando entropia de hipergrafos. A demonstração segue um esquema similar ao da demonstração do teorema 9.3.1.

Teorema 9.3.2 *Vale que*

$$\frac{1}{t} \lg N(b, k, t) = O\left(\min_{0 \leq j \leq k-2} \frac{b^{j+1}}{b^{j+1}} \lg \frac{b-j}{k-j-1}\right).$$

Prova: Seja $A \subseteq B^t$ k -separado de tamanho $N := N(b, k, t)$, onde $|B| = b$. Seja $0 \leq j \leq k-2$. Vamos construir um hipergrafo $(k-j)$ -uniforme em que todo vértice corresponde a um subconjunto de A de tamanho $j+1$ e cada hiperaresta corresponde a um subconjunto de tamanho k . Seja \mathcal{G} o hipergrafo definido como

$$\begin{aligned} V(\mathcal{G}) &:= \left\{ (D, x) : D \subseteq A, |D| = j, x \in A \setminus D \right\} \\ E(\mathcal{G}) &:= \left\{ \{(D_1, x^1), \dots, (D_{k-j}, x^{k-j})\} \in \binom{V(\mathcal{G})}{k} : D_1 = \dots = D_{k-j}, \right. \\ &\quad \left. x^i \neq x^j \text{ para quaisquer } 1 \leq i < j \leq k-j \right\}. \end{aligned}$$

É fácil ver que \mathcal{G} tem $\binom{N}{j}$ componentes e que cada componente é um grafo completo com $N-j$ vértices. Para cada $1 \leq i \leq t$, defina o hipergrafo \mathcal{G}_i sobre $V(\mathcal{G})$ onde vértices

$(D_1, x^1), \dots, (D_{k-j}, x^{k-j})$ formam uma hiperaresta se $\{(D_1, x^1), \dots, (D_{k-j}, x^{k-j})\} \in \mathcal{E}(\mathcal{G})$ e $S := D_1 \cup \{x^1, \dots, x^{k-j}\}$ é separado na i -ésima coordenada, isto é, todos os elementos de S diferem na i -ésima coordenada. Note que $\mathcal{G} = \bigcup_{i=1}^t \mathcal{G}_i$. Seja p a distribuição de probabilidade uniforme sobre $V(\mathcal{G})$. Pelo corolário 3.2.3.1,

$$H(\mathcal{G}, p) \leq \sum_{i=1}^t H(\mathcal{G}_i, p),$$

portanto,

$$\frac{H(\mathcal{G}, p)}{t} \leq \max_{i=1}^t H(\mathcal{G}_i, p). \quad (9.3.5)$$

É fácil calcular a entropia de \mathcal{G} . Usando o corolário 3.2.5.2 e o corolário 3.3.1.1, temos que

$$H(\mathcal{G}, p) = \lg \frac{N-j}{k-j+1}. \quad (9.3.6)$$

Agora vamos procurar um limitante superior para $\max_{i=1}^t H(\mathcal{G}_i, p)$. Para cada $D \subseteq A$ com tamanho j , defina $V_D := \{(D', x') \in V(\mathcal{G}) : D' = D\}$. Note que, para toda coordenada i , o sub-hipergrafo induzido $\mathcal{G}_i[V_D]$ ou é vazio ou é q -partido, com $q \leq b-j$. Seja $1 \leq \ell \leq t$. Como na prova anterior, pode-se provar que o número de vértices isolados em \mathcal{G}_ℓ é mínimo quando todos os elementos de B aparecem da forma mais balanceada possível na ℓ -ésima coordenada dos elementos de A . Note que, nesse caso, o sub-hipergrafo induzido $\mathcal{G}_\ell[D]$ é $(b-j)$ -partido para todo $D \subseteq A$ com tamanho j . Seja m o número de vértices não-isolados. Temos que

$$m \sim \binom{b}{j+1} \left(\frac{N}{b}\right)^{j+1} (j+1).$$

Como temos $\binom{N}{j}(N-j)$ vértices, então a soma das probabilidades dos vértices não-isolados é

$$\frac{b^{j+1}N^{j+1}}{b^{j+1}N^{j+1}}.$$

Com t grande o suficiente,

$$\frac{b^{j+1}N^{j+1}}{b^{j+1}N^{j+1}} \sim \frac{b^{j+1}}{b^{j+1}}.$$

Assim, usando o corolário 3.2.5.2 e o lema 3.3.2, temos que

$$H(\mathcal{G}_i, p) \leq \frac{b^{j+1}}{b^{j+1}} \lg \frac{b-j}{k-j-1}. \quad (9.3.7)$$

Usando a desigualdade (9.3.5), a equação (9.3.6) e a desigualdade (9.3.7), temos que

$$\frac{\lg N(b, k, t)}{t} \leq O\left(\frac{b^{j+1}}{b^{j+1}} \lg \frac{b-j}{k-j-1}\right).$$

e estamos feitos. □

Corolário 9.3.2.1 *Vale que*

$$Y(b, k, n) = \Omega\left(\lg n \cdot \left(\min_{0 \leq j \leq k-2} \frac{b^{j+1}}{b^{j+1}} \lg \frac{b-j}{k-j-1}\right)^{-1}\right).$$

Prova: Segue diretamente do teorema 9.3.2.

□

Referências Bibliográficas

- [1] M. Chudnovsky, N. Robertson, P. D. Seymour, and R. Thomas. The strong perfect graph theorem. *Ann. Math.*, 164:51–229, 2006.
- [2] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi. Entropy splitting for antiblocking corners and perfect graphs. *Combinatorica*, 10(1):27–40, 1990.
- [3] A. Delmestri, A. Fioretto, and A. Sgarro. An explicit formula for fractional entropy. In B. Bouchon-Meunier, L. Valverde, and R.R. Yager, editors, *Uncertainty in intelligent systems*, pages 413–420. Elsevier, 1993.
- [4] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2000.
- [5] J. Edmonds and R. Giles. A min-max relation for submodular functions on graphs. In *Studies in integer programming (Proceedings Workshop on Integer Programming, Bonn, 1975)*, volume 1 of *Annals of Discrete Mathematics*, pages 185–204. North-Holland, Amsterdam, 1977.
- [6] A. Frank and T. Jordán. Minimal edge-coverings of pairs of sets. *J. Combin. Theory Ser. B*, 65(1):73–110, 1995.
- [7] M. L. Fredman. How good is the information theory bound in sorting? *Theoret. Comput. Sci.*, 1(4):355–361, 1975/76.
- [8] M. L. Fredman and J. Komlós. On the size of separating systems and families of perfect hash functions. *SIAM J. Algebraic Discrete Methods*, 5(1):61–68, 1984.
- [9] T. Gallai. Transitiv orientierbare Graphen. *Acta Math. Acad. Sci. Hungar.*, 18:25–66, 1967.
- [10] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1988.
- [11] J. Kahn and J. H. Kim. Entropy and sorting. *J. Comput. System Sci.*, 51(3):390–399, 1995. 24th Annual ACM Symposium on the Theory of Computing (Victoria, BC, 1992).

- [12] J. Kahn and N. Linial. Balancing extensions via Brunn-Minkowski. *Combinatorica*, 11(4):363–368, 1991.
- [13] J. Kahn and M. Saks. Balancing poset extensions. *Order*, 1(2):113–126, 1984.
- [14] D. E. Knuth. The sandwich theorem. *Electron. J. Combin.*, 1:Article 1, approx. 48 pp. (electronic), 1994.
- [15] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech Univ., Prague, 1971; dedicated to the memory of Antonín Špaček)*, pages 411–425. Academia, Prague, 1973.
- [16] J. Körner. Fredman-Komlós bounds and information theory. *SIAM J. Algebraic Discrete Methods*, 7(4):560–570, 1986.
- [17] J. Körner and G. Longo. Two-step encoding for finite sources. *IEEE Trans. Information Theory*, IT-19:778–782, 1973.
- [18] J. Körner and K. Marton. Graphs that split entropies. *SIAM J. Discrete Math.*, 1(1):71–79, 1988.
- [19] J. Körner and K. Marton. New bounds for perfect hashing via information theory. *European J. Combin.*, 9(6):523–530, 1988.
- [20] J. Körner and G. Simonyi. Graph pairs and their entropies: modularity problems. *Combinatorica*, 20(2):227–240, 2000.
- [21] J. Körner, G. Simonyi, and Z. Tuza. Perfect couples of graphs. *Combinatorica*, 12(2):179–192, 1992.
- [22] N. Linial. The information-theoretic bound is good for merging. *SIAM J. Comput.*, 13(4):795–801, 1984.
- [23] L. Lovász. A characterization of perfect graphs. *J. Combin. Theory Ser. B*, 13:95–98, 1972.
- [24] J. Radhakrishnan. $\Sigma\Pi\Sigma$ threshold formulas. *Combinatorica*, 14(3):345–374, 1994.
- [25] J. Radhakrishnan. Better lower bounds for monotone threshold formulas. *J. Comput. System Sci.*, 54(2, part 1):221–226, 1997. 32nd Annual Symposium on Foundations of Computer Science (San Juan, PR, 1991).
- [26] G. Simonyi. Graph entropy: a survey. In *Combinatorial optimization (New Brunswick, NJ, 1992–1993)*, volume 20 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 399–441. Amer. Math. Soc., Providence, RI, 1995.
- [27] G. Simonyi. Perfect graphs and graph entropy. An updated survey. In *Perfect graphs*, Wiley-Intersci. Ser. Discrete Math. Optim., pages 293–328. Wiley, Chichester, 2001.

- [28] R. P. Stanley. Two poset polytopes. *Discrete Comput. Geom.*, 1(1):9–23, 1986.