ENTROPIA DE GRAFOS

CRISTIANE MARIA SATO
ORIENTADOR: YOSHIHARU KOHAYAKAWA

Sumário

1. Introdução 2

Parte I. Entropia de grafos 4
2. Preliminares e notação 5
3. Definição e caracterizações 7
4. Propriedades básicas 13
5. Cantos convexos 15
6. O politopo fracionário dos conjuntos estáveis 18
7. Grafos perfeitos 19

Parte II. Uma aplicação à ordenação 24
8. Preliminares e notação 25
9. Ordenação a partir de informação parcial 25
10. Uma visão geral 26
11. Grafos de comparabilidade 27
12. Decomposição laminar 27
13. Limitantes 29
14. Encontrando uma boa comparação 33
15. Computando respostas 37
Referências 40

Date: 1 de dezembro de 2006.
Financiado parcialmente pela FAPESP. Processo número 05/60504-6.

1. Introdução

1.1. Breve histórico.

1.1. 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 [12] 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. Esta é usualmente interpretada como uma medida da quantidade de informação contida na variável aleatória.

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 [14]. Em 1988, Körner e Marton [15] 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 [17] 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 [8]); para a determinação de cotas do tipo Fredman-Komlós para funções de espalhamento (*hashing*) perfeitas e sistemas separadores (Körner [13] e Körner e Marton [16]); e em complexidade computacional (Radhakrishnan [20, 21]).

1.2. **Organização do texto.** Este texto é dividido em duas partes.

Na primeira, apresentamos a definição de entropia de grafos, caracterizações, algumas propriedades básicas e uma caracterização de grafos perfeitos usando entropia de grafos. Definimos também entropia para cantos convexos, que são conjuntos com algumas características especiais. Apresentamos algumas relações entre o politopo dos conjuntos estáveis e o politopo fracionário dos conjuntos estáveis. Essas relações são essenciais para a caracterização de grafos perfeitos.

Na segunda parte, apresentamos uma aplicação de entropia de grafos ao problema de ordenação a partir de informação parcial.

Os principais artigos estudados para o desenvolvimento dessa monografia são:

(1) as resenhas sobre entropia de grafos de Simonyi [22, 23];
(2) o artigo de Knuth [11] sobre a função ϑ de Lovász;
(3) o artigo de Csiszár, Körner, Lóvasz, Marton e Simonyi [2] em que são apresentadas a caracterização de entropia de grafos usando o politopo fracionário dos conjuntos estáveis e a caracterização de grafos perfeitos usando entropia de grafos;
(4) o artigo de Kahn e Kim [8] sobre o problema de ordenação a partir de informação parcial.

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.

Parte I

Entropia de grafos

2. Preliminares e notação

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

2.1. **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 $[n] := \{1, \ldots, n\}$.

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

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \tag{2.1}$$

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 (2.1) 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 2.1 (Desigualdade de Jensen) *Seja $f : R \to \mathbb{R}$ uma função convexa e sejam $x_1, \ldots, x_k \in R$. Então*

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) \leq \sum_{i=1}^{k} \lambda_i f(x_i), \tag{2.2}$$

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 $x^1, \ldots, x^k \in \mathbb{R}^n$. Sejam $\lambda_1, \ldots, \lambda_k$ 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 $x^1, \ldots, x^k$.

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 $\mathrm{conv}(A)$.

O seguinte resultado é bem conhecido:

Lema 2.2 **(Média geométrica e média aritmética)** *Sejam $x_1, \ldots, x_k \in \mathbb{R}$. Então*

$$\left(\prod_{i=1}^{k} x_i\right)^{1/k} \leq \frac{1}{k}\sum_{i=1}^{k} x_i. \tag{2.3}$$

2.2. **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 $\overline{G} := \left(V, \binom{V}{2} \setminus E(G)\right)$.

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.

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 $c \colon V(G) \to C$ 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

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

2.3. **Probabilidade.** Um *espaço de probabilidade finito* consiste de um conjunto finito Ω e de uma função $\mathbb{P}\colon \Omega \to [0,1]$ 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

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

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 $v \in V$.

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

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

para todo $v \in V$.

3. Definição e caracterizações

Nesta seção apresentamos a definição de entropia de grafos dada por Körner em 1973. Mostramos também duas caracterizações com as quais é mais fácil trabalhar.

3.1. **Codificação e entropia de grafos.** A entropia de grafos surgiu naturalmente de um problema proposto por Körner [12] 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 $(u_1, \ldots, u_t)$, onde $u_i \in U$ para todo i. Uma *palavra* de comprimento t (emitida pela fonte) é uma t-upla $(v_1, \ldots, v_t) \in V^t$ de símbolos emitidos consecutivamente pela fonte. Duas palavras $x = (x_1, \ldots, x_t)$ e $y = (y_1, \ldots, y_t)$ 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 $u = (u_1, \ldots, u_t)$ 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 [12]. Seja G um grafo e p uma distribuição de probabilidade sobre $V(G)$. A *entropia de G com relação a p* é definida como

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

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 subseção 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 $tH(G,p)$.

3.2. **Uma caracterização alternativa.** Nesta subseção apresentamos uma caracterização de entropia de grafos dada por Körner [12]. 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

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

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 $H(X) := H(\mathrm{dist}(X))$. 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

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

onde $X_y := (X \mid Y = y)$. 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 3.1 *Sejam X e Y variáveis aleatórias. Então*

$$H(X,Y) = H(X) + H(Y \mid X).$$

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 $p(y \mid x) := \mathbb{P}[Y = y \mid X = x]$ para cada $(x, y) \in V \times U$. Temos que

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

□

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 3.2 *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

$$\begin{aligned}
-D(p,q) &= -\sum_{a\in A} p_a \lg \frac{p_a}{q_a} = \sum_{a\in A} p_a \lg \frac{q_a}{p_a} \\
&\leq \lg \sum_{a\in A} p_a \frac{q_a}{p_a} \leq \lg 1 = 0,
\end{aligned} \tag{3.1}$$

onde a primeira desigualdade segue da desigualdade (2.2) de Jensen. Como $\lg x$ é uma função estritamente côncava, então (3.1) 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 3.3 *Sejam X e Y variáveis aleatórias. Então*

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

Prova: Pelo lema 3.1, 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 $p(x \mid y) := \mathbb{P}[X = x \mid Y = y]$ para cada $(x,y) \in V \times U$. Vale que

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

□

Finalmente podemos enunciar uma caracterização de entropia de grafos apresentada por Körner [12]. Omitimos a demonstração.

Teorema 3.4 *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 $dist(X) = p$;*
(ii) *Y é uma variável aleatória tomando seus valores em $\mathcal{S}(G)$;*
(iii) *dado $X = x$, vale que Y toma seus valores em $\{S \in \mathcal{S}(G) \colon x \in S\}$.*

Então

$$H(G,P) = \min_{(X,Y) \in A(G)} I(X \cap Y). \tag{3.2}$$

3.3. **O politopo dos conjuntos estáveis.** Nesta subseção 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

$$\mathrm{STAB}(G) := \mathrm{conv}\left(\left\{\chi^S \colon S \in \mathcal{S}(G)\right\}\right).$$

Teorema 3.5 *Seja G um grafo e p uma distribuição de probabilidade sobre $V(G)$. Então*

$$H(G,p) = \min\Big\{-\textstyle\sum_{v \in V(G)} p_v \lg a_v \colon a \in STAB(G)\Big\}. \tag{3.3}$$

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

$$H(G,p) \geq \min\Big\{-\textstyle\sum_{v \in V} p_v \lg a_v \colon a \in \mathrm{STAB}(G)\Big\}.$$

Sejam X e Y variáveis aleatórias tomando valores em $V(G)$ e $\mathcal{S}(G)$, respectivamente, que atingem o mínimo na caracterização (3.2) de $H(G,p)$. Abreviamos $r(S) := \mathbb{P}[Y = S]$ para cada $S \in \mathcal{S}(G)$ e $r(S \mid x) := \mathbb{P}[Y = S \mid X = x]$ para cada $(S,x) \in \mathcal{S}(G) \times V$. Note que

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

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

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

onde a última passagem segue da desigualdade (2.2) de Jensen. Tome $b \in \mathbb{R}^V_+$ definido como

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

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

$$H(G,p) \geq -\sum_{v \in V} p_v \lg b_v \geq \min \Big\{ -\textstyle\sum_{v \in V(G)} p_v \lg a_v \colon a \in \mathrm{STAB}(G) \Big\}.$$

Resta provarmos que

$$H(G,p) \leq \min \Big\{ -\textstyle\sum_{v \in V(G)} p_v \lg a_v \colon a \in \mathrm{STAB}(G) \Big\}.$$

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

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

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

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

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

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

Pelo lema 3.2,

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

Portanto
$$-\sum_{S\in\mathcal{S}(G)} q(S)\lg q(S) \le -\sum_{S\in\mathcal{S}(G)} q(S)\lg d_S.$$
Concluímos que
$$H(G,p) \le \sum_{v\in V}\sum_{S\in\mathcal{S}(G)} p_v q(S\mid v)\lg\frac{q(S\mid v)}{d_S} = -\sum_{v\in V} p_v \lg a_v,$$
como queríamos. □

4. Propriedades básicas

Nesta seção, apresentamos algumas propriedades básicas de entropia de grafos.

Uma propriedade simples e pouco surpreendente é a monotonicidade:

Lema 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*
$$H(F,p) \le H(G,p). \tag{4.1}$$

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

Levando em consideração a definição de entropia de grafos a equação (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 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 $H(G,p) \le H(G[U],p|_U)$, 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 $a \in \mathrm{STAB}(G)$ um vetor que atinge o mínimo na caracterização (3.3) 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 $v \ne u$. Logo,
$$H(G,p) = \sum_{v\in V(G)} p(v)\lg\frac{1}{a_v} = \sum_{v\in V(G)\setminus\{u\}} p'(v)\lg\frac{1}{a'_v} \ge H(G-u,p').$$
□

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

$$H(G \cup F, p) \leq H(G, p) + H(F, p). \tag{4.2}$$

Prova: Sejam $a \in \mathrm{STAB}(G)$ e $b \in \mathrm{STAB}(F)$ vetores que atingem o mínimo na caracterização (3.3) para $H(G, p)$ e $H(F, p)$, respectivamente.

O vetor a é combinação convexa de elementos de $\{\chi^S \colon S \in \mathcal{S}(G)\}$. 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 $\{\chi^S \colon S \in \mathcal{S}(F)\}$. 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 $a \circ b \in \mathrm{STAB}(G \cup F)$.

Assim,

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

$\square$

Uma conseqüência imediata do lema anterior é que

$$H(G, p) + H(\overline{G}, p) \geq H(K_n, p). \tag{4.3}$$

Na seção 7, vamos mostrar quais grafos satisfazem (4.3) com igualdade.

4.2. **A entropia do grafo completo e a do grafo vazio.**

Lema 4.4 *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 $\mathrm{STAB}(K_n)$, então $p \in \mathrm{STAB}(K_n)$. Seja $q \in \mathrm{STAB}(K_n)$. Usando a desigualdade (2.2) 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 (3.3) de entropia de grafos. $\square$

Calcular a entropia do grafo vazio também é muito fácil:

Lema 4.5 *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

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

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 (3.3) de entropia de grafos. $\square$

5. Cantos convexos

5.1. **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

$$H_A(p) := \min_{a \in A} \sum_{v \in V} p_v \lg \frac{1}{a_v}. \tag{5.1}$$

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

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

$$H_A(p) = \min_{x \in \Lambda(A)} \sum_{v \in V} p_v x_v = \min_{x \in \Lambda(A)} px. \tag{5.2}$$

Lema 5.1 *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 5.2 *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 (5.2) 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 5.1. Concluímos assim que $\Lambda(A) \subseteq \Lambda(B)$, de onde segue que $A \subseteq B$. □

Corolário 5.2.1 *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 5.2 e dos lemas 4.4 e 4.5. □

5.2. **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 5.3 *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 (5.1) de $H_A(p)$ e b atinge o mínimo na definição (5.1) de $H_B(p)$.

Prova: Como $p = a \circ b$, então

$$H(p) = -\sum_{v \in V} p_v \lg a_v b_v = -\sum_{v \in V} p_v \lg a_v - \sum_{v \in V} p_v \lg b_v \geq H_A(p) + H_B(p). \tag{5.3}$$

É óbvio que (5.3) vale com igualdade se e somente se a atinge o mínimo na definição (5.1) de $H_A(p)$ e b atinge o mínimo na definição (5.1) de $H_B(p)$. □

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

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

Lema 5.4 *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 $\mathrm{ab}(A) \subseteq B$, 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 (5.1) de $H_A(p)$ e de $H_B(p)$, respectivamente. Usando a desigualdade 2.2 de Jensen e o fato de que $ba \leq 1$, temos

$$H_A(p) + H_B(p) - H(p) = -\sum_{v \in V} p_v \lg \frac{a_v b_v}{p_v} \geq -\lg\Big(\sum_{v \in V} a_v b_v\Big) \geq 0. \tag{5.4}$$

Usando o lema 5.3, é fácil ver que (5.4) vale com igualdade se e somente se $p = a \circ b$ para algum $a \in A$ e algum $b \in B$. □

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

(i) $ab(A) \subseteq B$;
(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 (5.1) 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, $b \in \mathrm{ab}(A) \subseteq B$. Provamos assim que (i) $\Rightarrow$ (ii).

Segue diretamente do lema 5.3 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 5.4, sabemos que

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

para toda distribuição $p \in \mathbb{R}^V_+$. Assim, supondo que vale (iii), então $H_{\mathrm{ab}(A)}(p) \geq H_B(p)$. Pelo lema 5.2, temos que $\mathrm{ab}(A) \subseteq B$. □

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

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

O teorema 5.5 e os lemas 5.3 e 5.4 implicam na seguinte caracterização de pares antibloqueadores:

Corolário 5.5.1 *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 5.5.2 *Seja $A \subseteq \mathbb{R}_+^V$ um canto convexo e $p \in \mathbb{R}_+^V$ uma distribuição de probabilidade. Então, vale que*
$$H(p) = H_A(p) + H_{ab(A)}(p).$$

6. O politopo fracionário dos conjuntos estáveis

Seja G um grafo sobre V. Definimos o *politopo fracionário dos conjuntos estáveis* de G como
$$\mathrm{QSTAB}(G) := \Big\{ b \in \mathbb{R}_+^V : \sum_{v \in K} b_v \leq 1 \text{ para toda clique } K \text{ de } G \Big\}. \tag{6.1}$$

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

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

Teorema 6.1 *Seja G um grafo. Então*
$$STAB(\overline{G}) = ab(QSTAB(G)) \quad e$$
$$QSTAB(\overline{G}) = ab(STAB(G)).$$

Prova: Primeiro vamos mostrar que
$$\mathrm{ab}(X) = \mathrm{ab}(\mathrm{conv}(X)) \tag{6.2}$$
para todo $X \subseteq \mathbb{R}_+^V$. É óbvio que $\mathrm{ab}(\mathrm{conv}(X)) \subseteq \mathrm{ab}(X)$. Vamos mostrar que $\mathrm{ab}(X) \subseteq \mathrm{ab}(\mathrm{conv}(X))$.

Seja $y \in \mathrm{ab}(X)$ e seja $x \in \mathrm{conv}(X)$. 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 $y \in \mathrm{ab}(\mathrm{conv}(X))$. Assim, temos que $\mathrm{ab}(X) \subseteq \mathrm{ab}(\mathrm{conv}(X))$.

Agora usamos (6.2) para concluir que
$$\begin{aligned}
\mathrm{QSTAB}(G) &= \Big\{ b \in \mathbb{R}_+^{V(G)} : \sum_{v \in K} b_v \leq 1 \text{ para toda clique } K \text{ de } G \Big\} \\
&= \mathrm{ab}\big(\{\chi^K : K \text{ é uma clique de } G\}\big) \\
&= \mathrm{ab}\big(\{\chi^S : S \text{ é um conjunto estável de } \overline{G}\}\big) \\
&= \mathrm{ab}(\mathrm{STAB}(\overline{G})).
\end{aligned}$$

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

□

Corolário 6.1.1 *Seja G um grafo. Vale que*

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

Prova: Segue diretamente do lema 6.1. □

7. Grafos perfeitos

7.1. **Grafos perfeitos e cantos convexos.** Nesta subseção 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 subseção é mostrar que um grafo G é perfeito precisamente quando $\mathrm{QSTAB}(G) = \mathrm{STAB}(G)$.

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 Berge em 1961.

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

Lema 7.1 (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. □

Teorema 7.2 *Seja G um grafo perfeito. Então*

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

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

Seja $x \in \mathrm{QSTAB}(G)$ 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 7.1 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 $c\colon V(G) \to [\alpha]$ 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,

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

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

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

para algum $y \in ab(P)$. Então ou

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

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 $P = \mathrm{ab}(Z) = \{x\colon x \geq 0\}$ 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 $P' = \mathrm{ab}(Z')$. 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 $y \in \mathrm{ab}(P)$. 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. $\square$

Finalmente podemos provar a conversa do teorema 7.2.

Teorema 7.4 *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

$$\mathrm{QSTAB}(G[U]) = \mathrm{QSTAB}(G)[U]$$

e que

$$\mathrm{STAB}(G[U]) = \mathrm{STAB}(G)[U].$$

Assim, $\mathrm{STAB}(G) = \mathrm{QSTAB}(G)$ se e somente se $\mathrm{STAB}(G') = \mathrm{QSTAB}(G')$ 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 $\mathrm{STAB}(G) = \mathrm{QSTAB}(G)$, então G pode ser colorido com $\omega := \omega(G)$ cores.

Suponha que $\mathrm{STAB}(G) = \mathrm{QSTAB}(G)$. Pelo corolário 6.1.1, vale que

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

Tome $P := \mathrm{QSTAB}(\overline{G})$ 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 $y \in \mathrm{ab}(P)$.

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 $P = \mathrm{QSTAB}(\overline{G}) = \mathrm{ab}(Z)$ e $Z \neq \emptyset$. Assim, pelo lema 7.3,

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

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 7.5 *Seja G um grafo. Então*

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

Prova: Imediato dos teoremas 7.2 e 7.4. $\square$

7.2. **Grafos perfeitos e entropia de grafos.** Nesta subseção apresentamos 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 7.6 *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

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

Prova: Pelo lema 6.1 e pelo corolário 5.5.1, temos que

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

□

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

Teorema 7.7 *Um grafo G é perfeito se e somente se é fortemente separador.*

Prova: Segue diretamente do teorema 7.6, do lema 5.2 e do teorema 7.5. □

Lovász [19] provou a conjectura fraca dos grafos perfeitos, que diz que um grafo é perfeito se e somente se seu complemento também o é.

Corolário 7.7.1 (Teorema fraco dos grafos perfeitos) *Um grafo G é perfeito se e somente se $\overline{G}$ é perfeito.*

Prova: Segue diretamente do teorema 7.5 e do corolário 6.1.1. □

É fácil provar o seguinte corolário.

Corolário 7.7.2 *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

$$H(G,p) + H(\overline{G},p) > H(p). \tag{7.1}$$

A proposição a seguir mostra que se G é um grafo imperfeito minimal, então a distribuição de probabilidade uniforme satisfaz (7.1).

Proposição 7.8 *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 $\mathrm{STAB}(G)$ e $\mathrm{STAB}(\overline{G})$ que atingem $H(G,p)$ e $H(\overline{G},p)$ na caracterização (3.3), respectivamente. Tome $V := V(G)$. Então,

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

A primeira desigualdade segue do lema 2.2; a segunda, do corolário 7.7.2. □

A proposição 7.8 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.

Parte II

Uma aplicação à ordenação

8. 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

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

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

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

Seja $\{v_1, \ldots, v_m\} \subseteq V$ tal que $v_1 < \cdots < v_m$ são relações compatíveis com P, isto é, $v_1 < \cdots < v_m$ vale em alguma extensão linear de P. Denotamos por $P(v_1 < \cdots < v_m)$ a menor ordem parcial compatível com P que contém as relações $v_1 < \cdots < v_m$. Mais formalmente, $P(v_1 < \cdots < v_m)$ 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 $a \in \mathrm{STAB}(G_P)$ que atinge o mínimo na caracterização (3.3) de $H(P)$. Denotamos por $b_{\min}(P)$ o vetor $b \in \mathrm{STAB}(\overline{G_P})$ que atinge o mínimo na caracterização (3.3) de $H(\overline{P})$.

9. 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 [6] 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 [9, 10]). 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 [8].

10. Uma visão geral

Os principais resultados de Kahn e Kim [8] 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 11, mostramos que os grafos de comparabilidade são perfeitos e apresentamos algumas conseqüências desse fato. Na seção 13, relacionamos $e(P)$ e $H(P)$. Nas duas outras seções, mostramos a existência dos algoritmos citados acima.

11. Grafos de comparabilidade

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

Lema 11.1 *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 $C_v := \{u_1, \ldots, u_k\}$ com $u_1 = v$ e $u_1 < \cdots < u_k$. 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. □

Lema 11.2 *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 11.1 e do teorema 7.7. □

Usaremos também o seguinte resultado:

Lema 11.3 *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 [11] sobre a função ϑ de Lovász e o livro sobre o método dos elipsóides de Grötschel, Lovász e Schrijver [7].

12. 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 12.1 *Seja $a \in STAB(G_P)$ e seja $b \in STAB(\overline{G_P})$. Então $ab \leq 1$.*

Prova: Pela demonstração do lema 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 $\mathrm{STAB}(G_P \cup \overline{G_P}) = \mathrm{STAB}(K_V)$. Como grafos completos são perfeitos, então pelo teorema 7.2, $\mathrm{STAB}(K_V) = \mathrm{QSTAB}(K_V)$. Assim, como V é uma clique em K_V, então, pela definição 6.1 de $\mathrm{QSTAB}(K_V)$, temos que $ab = \sum_{v \in V} a_v b_v \leq 1$. $\square$

Lema 12.2 *Para todo $v \in P$,*

$$(a_{\min}(P))_v (b_{\min}(P))_v = \frac{1}{n}. \tag{12.1}$$

Prova: Tome $a := a_{\min}(P)$ e $b := b_{\min}(P)$. Pelo lema 11.2, 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 (3.3) de $H(K_V, p)$, onde p é a distribuição de probabilidade uniforme sobre V. Ademais, pela demonstração do lema 4.3, podemos ver que $a \circ b \in \mathrm{STAB}(K_V)$. Assim, pelo lema 3.2 e pela demonstração do lema 4.4, é fácil ver que $a \circ b = p$. $\square$

Lema 12.3 *Seja $a \in \mathbb{R}^V_+$. Suponha que a pode ser escrito como*

$$a = \sum_{i=1}^{r} \lambda_i \chi^{A_i}, \tag{12.2}$$

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 (12.2) 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 (12.2) 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 muito resultados célebres, como o modelo de fluxos submodulares de Edmonds e Giles [4] e um resultado de cobertura bi-supermodular de Frank e Jordan [5], usado para aumento de conexidade.

Lema 12.4 *Existe uma única decomposição laminar de $a_{\min}(P)$.*

Prova: Pelo lema 12.3, basta mostrar que existe uma decomposição laminar de $a_{\min}(P)$.

Fixe uma extensão linear $\propto$ 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 $\propto$) 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. □

13. 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

$$\mathrm{vol}(A) := \int_{x \in A} \mathrm{d}x.$$

Linial [18] observou que $\mathrm{vol}(\mathcal{O}(P)) = e(P)/(n!)$. Stanley [24] provou que $\mathrm{STAB}(G_P)$ e $\mathcal{O}(P)$ têm o mesmo volume. Portanto,

$$\mathrm{vol}(\mathrm{STAB}(G_P)) = \frac{e(P)}{n!}. \tag{13.1}$$

Lema 13.1 *Vale que*

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

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

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

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

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

Pelo lema 12.1, vale que $\mathrm{STAB}(G_P) \subseteq L$. Portanto, pelo lema 12.2,

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

□

Corolário 13.1.1 *Seja c uma constante positiva. Se $e(P) \geq cn$, então*

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

Prova: Pelo lema 13.1 e pela equação (13.1),

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

Pelo lema 11.2,

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

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

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

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

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

É fácil provar que

$$e(P) \geq 2^t. \tag{13.2}$$

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

Lema 13.2 *Se $t < n/7$ e P não tem um ponto de corte, então existe $j \in [t]$ tal que*

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

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

$$\bigcup \{K(j) \colon j \in [t],\ y_j \in T'\} = [\ell]. \tag{13.3}$$

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

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

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

$$\sum_{j=1}^{r} \sum_{i \in K(j)} (u_i + z_i) \geq \sum_{j=1}^{r} k_j - 2r. \tag{13.4}$$

Pela equação (13.3) e usando o fato de que $r \leq t$, temos que

$$\sum_{j=1}^{r} k_j - 2r \geq \ell - 2t. \tag{13.5}$$

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

$$\sum_{j=1}^{r} \sum_{i \in K(j)} (u_i + z_i) \leq \sum_{i=1}^{\ell} 2(u_i + z_i) \leq 2t + 2t = 4t. \tag{13.6}$$

Assim, usando as equações (13.4)–(13.6), temos que $4t \geq \ell - 2t$. Logo, $6t \geq \ell = n - t$, o que contradiz a hipótese de que $t < n/7$. □

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 13.3 *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*

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

Prova: Seja j como no lema 13.2 e $K(j) = \{x_h, \ldots, x_m\}$ com $x_h <_P \cdots <_P x_m$. Escolha i em $\{h, \ldots, m\}$ que minimize

$$\frac{e\big(P(x_i < y_j < x_{i+1})\big)}{e(P)}. \tag{13.7}$$

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 (13.7),

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

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

$$v <_P y_j \Rightarrow v <_P x_{i+1}. \tag{13.8}$$

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 (13.8). Similarmente, pode-se provar que

$$y_j <_P v \Rightarrow x_i <_P v. \tag{13.9}$$

Note que decorre das implicações (13.8) e (13.9) que, se $a \nsim b$ 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 13.2, é fácil ver que

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

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

$$\begin{aligned} nH(G', p) &\leq \lg(q+1) + \sum_{y \in Y} \lg \frac{q+1}{q} \\ &= \lg(q+1) + q\lg(q+1) - q\lg(q) \\ &\leq 2\lg(q+1) \leq 2\lg(2k_j+1). \end{aligned} \tag{13.10}$$

Pelo lema 4.3 da subaditividade e pela desigualdade (13.10),

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

e estamos feitos. □

Lema 13.4 *Vale que*

$$nH(\overline{P}) \leq (1 + 7\lg e)\lg(e(P)). \tag{13.11}$$

Prova: A prova é por indução em $n + t$. Se $n = 1$ ou $t = 0$ é claro que a inequação (13.11) 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 (13.2) e pelo corolário 13.1.1, a inequação (13.11) 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 13.3. 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 13.5 *Vale que*

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

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

Prova: Segue diretamente do lema 11.2, do lema 13.1 e da equação (13.1), e do lema 13.4. □

14. 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

$$\min\{H(P(x < y)), H(P(x < y))\} \geq H(P) + \frac{c}{n}, \tag{14.1}$$

onde $c := 1 + 17/112$. Isso significa que ao descobrirmos 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 14.1 *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 $\{\chi^B : B$ 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 $1 \le i \le m$, defina $C_i := B_i \setminus \{v \in P : v <_P y\}$.

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

$$\sum_{i=1}^{t} a_{v_i} = \sum_{i=1}^{\beta(x)} \lambda_i. \tag{14.2}$$

Defina as seguintes cadeias de P'

$$B'_i := B_i, \quad \text{se } 1 \le i \le s$$
$$B'_{i+s} := C \cup C_i, \quad \text{se } 1 \le i \le m.$$

Defina também

$$\xi'_i := \xi_i, \quad \text{se } m+1 \le i \le s$$
$$\xi_{i+s} := \mu\xi_i, \quad \xi_i := (1-\mu)\xi_i, \quad \text{se } 1 \le i \le m.$$

Tome

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

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

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

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,

$$\lg(1+u-v) \ge \lg(1+u) + \lg(1-v) \tag{14.3}$$

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

$$\lg(1+u) + \lg(1+v) \ge \lg(1+u+v) \tag{14.4}$$

para quaisquer $u, v \in \mathbb{R}$ com $uw \ge 0$.

Pelo lema 11.2 e pelas desigualdades (14.3) e (14.4), temos que

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

Pelo lema 12.2 e pela equação (14.2),

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

Além disso,

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

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

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

□

Antes de provar a desigualdade (14.1), precisamos de um lema fácil.

Lema 14.2 *Dados $0<\varepsilon_1<1$ e $0<\varepsilon_2<1$, escolha x com a_x tão grande quanto possível de forma que*

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

Seja s o menor inteiro para o qual

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

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

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

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

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

□

Finalmente provamos a desigualdade (14.1).

Teorema 14.3 *Se P não é uma cadeia, então existem x, y incomparáveis em P tais que*

$$\min\{H(P(x<y)), H(P(y<x))\} \geq H(P) + \frac{c}{n}, \tag{14.5}$$

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 (14.5) 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 $P' := P \backslash \{z\}$. 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 $nH(\overline{P}) = (n-1)H(\overline{P'})$, 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 14.2. Tome $\delta := (1/a_x)\sum\{\lambda_i \colon 1 \leq i \leq \alpha(x) - 1\}$. Note que $\delta \leq \varepsilon_1$. Pelo lema 14.2,

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

Tome $P' := P(x<y)$. Pelo lema 14.1 e pelas escolhas de x e y,

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

Por outro lado, tome $P'' := P(y < x)$. Tome $\eta := 1$. Pelo lema 14.1,

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

□

Vamos mostrar agora que, do teorema 14.3, segue facilmente a existência do algoritmo desejado.

Corolário 14.3.1 *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

$P' \leftarrow P$
enquanto $H(P') < \lg n$ faça
  encontre x, y tais que
    $\min\{H(P'(x < y)), H(P'(y < x))\} \geq H(P') + c/n$,
    onde $c = 1 + 17/112$
  pergunte ao oráculo: "$x < y$?"
  se o oráculo responder "SIM"
    então $P' \leftarrow P'(x < y)$
    senão $P' \leftarrow P'(y < x)$
devolva P'

Pelo teorema 14.3, se P' não é uma cadeia, tais x e y existem. Além disso, pelo lema 11.3 podemos calcular $H(P')$, $H(P'(x < y))$ e $H(P'(x < y))$ em tempo polinomial. Note que o algoritmo só termina quando encontra uma ordem total, pois pelo lema 4.4, 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 13.5, vale que $\lg(e(P)) \geq Cn(\log n - H(P))$, onde $C := (1 + 7\lg e)^{-1}$. Assim, o algoritmo faz $O(\lg e(P))$ comparações. □

15. 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 15.1 *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

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

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 11.1 e pelo teorema 7.5, vale que $\mathrm{QSTAB}(G_P) = \mathrm{STAB}(G_P)$. Logo, pela definição (6.1) de $\mathrm{QSTAB}(G_P)$,

$$\begin{aligned} w(K) + w(L) + a_x &\leq 1, \\ w(M) + w(N) + a_y &\leq 1. \end{aligned}$$

Portanto,

$$w(K) + w(N) + \frac{a_x + a_y}{2} \leq 1 \quad \text{ou} \tag{15.1}$$

$$w(M) + w(L) + \frac{a_x + a_y}{2} \leq 1. \tag{15.2}$$

Suponha sem perda de generalidade que a inequação (15.1) 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 $a' \in \mathrm{QSTAB}(G_{P'})$, pelo teorema 7.5, isso implica que $a' \in \mathrm{STAB}(G_{P'})$. Para toda cadeia C de P', defina $w'(C) := \sum_{x \in C} a'_x$. Seja Q uma cadeia maximal de P'. Se $\{x, y\} \nsubseteq Q$, 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, $a' \in \mathrm{QSTAB}(G_{P'})$. Se $\{x, y\} \subseteq Q$, então tome

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

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, $a' \in \mathrm{QSTAB}(G_{P'}) = \mathrm{STAB}(G_{P'})$. 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 15.1.1 *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

$P' \leftarrow P$
enquanto o oráculo faz uma pergunta "$x < y$?" faça
  se x, y são comparáveis em P'
  então se $x <_{P'} y$
    então devolva "SIM"
    senão devolva "NÃO"
  senão se $H(P'(x < y)) \leq H(P'(y < x))$
    então $P' \leftarrow P'(x < y)$ e devolva "SIM"
    senão $P' \leftarrow P'(y < x)$ e devolva "NÃO"

Pelo teorema 15.1, 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 13.5, $\lg e(P') \leq n(\lg n - H(P'))$. Isso significa, que o algoritmo que ordena P fará $\Omega(e(P))$ comparações. □

Referências

[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] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2000.

[4] 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.

[5] A. Frank and T. Jordán. Minimal edge-coverings of pairs of sets. *J. Combin. Theory Ser. B*, 65(1):73–110, 1995.

[6] M. L. Fredman. How good is the information theory bound in sorting? *Theoret. Comput. Sci.*, 1(4):355–361, 1975/76.

[7] 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.

[8] 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).

[9] J. Kahn and N. Linial. Balancing extensions via Brunn-Minkowski. *Combinatorica*, 11(4):363–368, 1991.

[10] J. Kahn and M. Saks. Balancing poset extensions. *Order*, 1(2):113–126, 1984.

[11] D. E. Knuth. The sandwich theorem. *Electron. J. Combin.*, 1:Article 1, approx. 48 pp. (electronic), 1994.

[12] 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.

[13] J. Körner. Fredman-Komlós bounds and information theory. *SIAM J. Algebraic Discrete Methods*, 7(4):560–570, 1986.

[14] J. Körner and G. Longo. Two-step encoding for finite sources. *IEEE Trans. Information Theory*, IT-19:778–782, 1973.

[15] J. Körner and K. Marton. Graphs that split entropies. *SIAM J. Discrete Math.*, 1(1):71–79, 1988.

[16] J. Körner and K. Marton. New bounds for perfect hashing via information theory. *European J. Combin.*, 9(6):523–530, 1988.

[17] J. Körner, G. Simonyi, and Z. Tuza. Perfect couples of graphs. *Combinatorica*, 12(2):179–192, 1992.

[18] N. Linial. The information-theoretic bound is good for merging. *SIAM J. Comput.*, 13(4):795–801, 1984.

[19] L. Lovász. A characterization of perfect graphs. *J. Combin. Theory Ser. B*, 13:95–98, 1972.

[20] J. Radhakrishnan. $\Sigma\Pi\Sigma$ threshold formulas. *Combinatorica*, 14(3):345–374, 1994.

[21] 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).

[22] 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.

[23] 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.

[24] R. P. Stanley. Two poset polytopes. *Discrete Comput. Geom.*, 1(1):9–23, 1986.

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–900 São Paulo, SP

Endereços Eletrônicos: `cmsato@gmail.com, yoshi@ime.usp.br`

URL: `http://www.ime.usp.br/~csato`