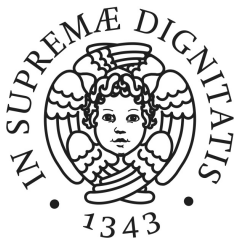

UNIVERSITÀ DEGLI STUDI DI PISA

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea in Matematica

Tesi di Laurea

Entropia di grafo e il problema dell'ordinamento con informazione parziale

Relatore:
Prof. Roberto Grossi

Candidato:
Jacopo Notarstefano

ANNO ACCADEMICO 2011/2012

Indice

Introduzione

Il concetto di entropia di una sorgente fu introdotto da Claude E. Shannon nel fondamentale articolo "A Mathematical Theory of Communication" [23].

Definizione. Sia X una sorgente che a ogni istante discreto emetta un simbolo v_i nell'alfabeto finito $\{v_1, \ldots, v_n\}$ con probabilità p_i per $1 \leq i \leq n$. Chiamiamo *entropia di* X la quantità

$$H(X) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

Dalla sua introduzione sono state proposte più generalizzazioni, fra cui ricordiamo in particolare l'entropia di Rényi [22]. In questa tesi andremo a esporre un'ulteriore generalizzazione dovuta a Janos Körner, nota come "entropia di grafo" [16]. Infatti, oltre a generalizzare la nozione di entropia di una sorgente, questa entropia consente di assegnare a un grafo un numero che ne rappresenti la complessità. Ciò si ottiene interpretando il grafo come rappresentante la relazione di distinguibilità dei simboli emessi da una sorgente discreta. Più precisamente, detta X una sorgente come nella precedente definizione, associamo a essa un grafo di insieme dei vertici $V = \{v_1, \ldots, v_n\}$ e un arco (v_i, v_j) ogni volta che i simboli v_i e v_j sono distinguibili. Vorremmo quindi riottenere la classica nozione di entropia di una sorgente nel caso di totale distinguibilità, ovvero di grafo completo.

La definizione rigorosa di entropia di grafo comporta alcune difficoltà tecniche, la cui soluzione costituisce l'obiettivo del primo capitolo. Nel secondo capitolo passeremo a esporre le principali proprietà dell'entropia di grafo, fra cui la monotonia e una forma di subadditività. Enunceremo inoltre l'interessante relazione con i grafi perfetti, cioè quei grafi per cui numero cromatico e numero di cricca coincidono. Ne dedurremo alcuni lemmi sui grafi di confrontabilità associati agli insiemi parzialmente ordinati. Questi risultati verranno sfruttati nel terzo capitolo per descrivere algoritmi che risolvano il problema dell'ordinamento con informazione parziale. Tale problema consiste nel determinare adattivamente un ordine totale fissato ma ignoto a partire da un insieme parzialmente ordinato. Jeff Kahn e Jeong H. Kim nell'articolo "Entropy and Sorting" evidenziarono per primi il collegamento esistente fra il problema dell'ordinamento con informazione parziale e l'entropia del grafo a esso associato [12]. Essi esibirono inoltre un algoritmo per la soluzione di tale problema in un numero asintoticamente ottimo di confronti e per giunta polinomiale nelle operazioni elementari, ma che a ogni passo fa uso del metodo dell'ellissoide. In un recente articolo Jean Cardinal et al. hanno invece esibito tre algoritmi per la soluzione dello stesso problema in un numero asintoticamente ottimo di confronti e comunque polinomiali nelle operazioni elementari, senza però far uso del metodo dell'ellissoide [4]. Nel terzo capitolo verranno illustrati in dettaglio questi ultimi tre algoritmi.

Sull'entropia di grafo è stata scritta un'esauriente rassegna da Gábor Simonyi nel 1995 e una versione più aggiornata nel 2001 [24], [25].

Capitolo 1

Tre definizioni equivalenti di entropia di grafo

1.1 Tre definizioni di entropia di grafo

1.1.1 Definizione in termini di Numero Cromatico

Definizione. Un *grafo probabilistico* consiste in una coppia (G, P), dove $G = (V, E)$ è un grafo e P una distribuzione di probabilità discreta sui vertici, ovvero se $|V| = n$ allora P è una n-upla $(p_1, \ldots, p_n)$ tale che $p_1 + \cdots + p_n = 1$.

Come accennato nell'introduzione, questo oggetto modella una sorgente priva di memoria e stazionaria che emette ad ogni istante discreto un simbolo v_i in un insieme finito $V = \{v_1, \ldots, v_n\}$ con la corrispondente probabilità p_i. Nel nostro caso supponiamo che due simboli v_i e v_j siano distinguibili se e solo se $(v_i, v_j) \in E$.

Sia t un intero positivo fissato, e sia V^t l'insieme delle stringhe lunghe t di simboli in V. Poiché la sorgente è priva di memoria, è naturale assegnare alla stringa $\mathbf{v} = v_{i_1} \ldots v_{i_t}$ la probabilità

$$\mathbf{P}^t(\mathbf{v}) = \prod_{j=1}^{t} p_{i_j}, \tag{1.1}$$

e al sottoinsieme $U \subset V^t$ la probabilità

$$\mathbf{P}^t(U) = \sum_{\mathbf{v} \in U} \mathbf{P}^t(\mathbf{v}).$$

Sia $0 < \varepsilon < 1$ fissato e sia $U \subset V^t$ tale che $\mathbf{P}^t(U) > 1 - \varepsilon$. Una *codifica propria* è una mappa dalle stringhe di V^t a un insieme di simboli, tale che stringhe distinguibili di U, cioè distinguibili in almeno un elemento, vengono mandate in simboli distinti.[1] È intuitivamente preferibile una codifica che faccia uso del minor numero possibile di simboli. Cerchiamo dunque di minimizzare la quantità

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

[1] Trascuriamo parte delle stringhe, come tipico in teoria dei codici, perché ad esse non ci interessa assegnare un significato e dunque un simbolo.

Figura 1.1: Esempio di potenza conormale per $t = 2$.

dove M è la cardinalità dell'immagine secondo la codifica. Sia quindi $R(G, P, t, \varepsilon)$ il minimo di tale frazione al variare delle codifiche. Il limite

$$\liminf_{\varepsilon \to 0} \liminf_{t \to \infty} R(G, P, t, \varepsilon) \tag{1.2}$$

misurerà quindi la complessità del grafo. La definizione (1.2) non consente però il calcolo esplicito dell'entropia del grafo, perché la quantità $R(G, P, t, \varepsilon)$ è un minimo al variare in un insieme che non sappiamo descrivere esplicitamente. Dobbiamo quindi semplificarla ulteriormente, e per fare questo sfrutteremo la struttura di grafo data.

Definizione. Sia t un intero positivo e G un grafo. Chiamiamo *t-esima potenza conormale* di G il grafo $G^t = (V^t, E^t)$, dove V^t è il prodotto cartesiano di t copie di V, mentre

$$E^t = \{(\mathbf{v}, \mathbf{w}) \mid \exists i \ (v_i, w_i) \in E\}.$$

In altri termini fra due vertici $\mathbf{v}$ e $\mathbf{w}$ di G^t esiste un arco se e soltanto se le corrispondenti sequenze di lunghezza t sono distinguibili in almeno un elemento.

Esempio. Sia L il grafo su tre vertici con due archi, rappresentato a sinistra nella Figura 1.1. Allora L^2 è rappresentato a destra.

Osserviamo che la coppia $(G^t, \mathbf{P}^t)$, dove $\mathbf{P}^t$ è definita da (1.1), è ancora un grafo probabilistico. Volendo codificare propriamente un sottoinsieme $U \subset V^t$ abbiamo bisogno di almeno tante parole quanti insiemi indipendenti servono per partizionare U. Ma questo è proprio il numero cromatico del sottografo indotto da G^t su U, che denotiamo $\chi(G^t(U))$. Siamo ora pronti per enunciare la prima definizione di entropia di grafo.

Definizione. Sia (G, P) un grafo probabilistico. Chiamiamo *entropia di grafo* il seguente limite:

$$H(G, P) = \lim_{t \to \infty} \min_{\substack{U \subset V^t \\ \mathbf{P}^t(U) > 1 - \varepsilon}} \frac{1}{t} \log \chi(G^t(U)). \tag{1.3}$$

Affinché la (1.3) risulti una buona definizione dovremmo dimostrarne l'indipendenza da ε. Salteremo questa verifica e dimostreremo direttamente nella prossima sezione l'equivalenza con una seconda definizione, la quale non dipende da ε.

1.1.2 Definizione in termini di Mutua Informazione

Definizione. Sia X una variabile aleatoria discreta di densità $P = (p_1 \dots p_n)$. Chiamiamo *entropia di Shannon di* X la somma

$$H(X) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

Con abuso di notazione scriveremo talvolta $H(P)$ per indicare l'entropia di Shannon di una variabile aleatoria discreta di densità P.

Definizione. Siano X ed Y due variabili aleatorie discrete. Chiamiamo *mutua informazione di* X *ed* Y la quantità

$$I(X;Y) = H(X) + H(Y) - H((X,Y))$$

dove (X,Y) è la variabile aleatoria congiunta.

Possiamo già enunciare la seconda definizione di entropia di grafo.

Definizione. Sia (G,P) un grafo probabilistico. Chiamiamo *entropia di grafo* il seguente minimo:

$$H(G,P) = \min_{\substack{X \sim P \\ X \in Y \in S(G)}} I(X;Y) \tag{1.4}$$

dove $S(G)$ è l'insieme degli insiemi indipendenti di G.

Affinché la (1.4) risulti una buona definizione occorre dimostrare che tale minimo esiste. Rinviamo alla prossima sezione per una dimostrazione di questo fatto.

1.1.3 Definizione in termini di Politopo dei Vertici

Definizione. Sia G un grafo. Chiamiamo *politopo dei vertici di* G l'involucro convesso dei vettori caratteristici degli insiemi indipendenti di G e lo denotiamo $\mathrm{STAB}(G)$.

Esempio. Sia G il grafo semplice non diretto su 3 vertici con 2 archi. A meno di isomorfismo possiamo supporre $V = \{1,2,3\}$ ed $E = \{(1,2),(1,3)\}$. Gli insiemi indipendenti di G sono allora $\{\emptyset, \{1\}, \{2\}, \{3\}, \{2,3\}\}$. I loro vettori caratteristici sono quindi

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

e il loro involucro convesso è rappresentato nella figura 1.2.

Non abbiamo bisogno di altro per enunciare la terza definizione di entropia di grafo.

Definizione. Sia (G,P) un grafo probabilistico. Chiamiamo *entropia di grafo* il seguente minimo:

$$H(G,P) = \min_{\substack{\mathbf{a} \in \mathrm{STAB}(G) \\ \mathbf{a} > 0}} \sum_{i=1}^{n} p_i \log \frac{1}{a_i}. \tag{1.5}$$

Osservazione. La funzione obiettivo di cui cerchiamo il minimo in (1.5) è convessa, tende a ∞ quando una delle coordinate di $\mathbf{a}$ tende a 0, tende monotonamente a $-\infty$ lungo le rette per l'origine. Ne deriva che tale minimo esiste finito sul bordo di $\mathrm{STAB}(G)$ e che il vettore che lo realizza ha tutte le coordinate maggiori di 0.

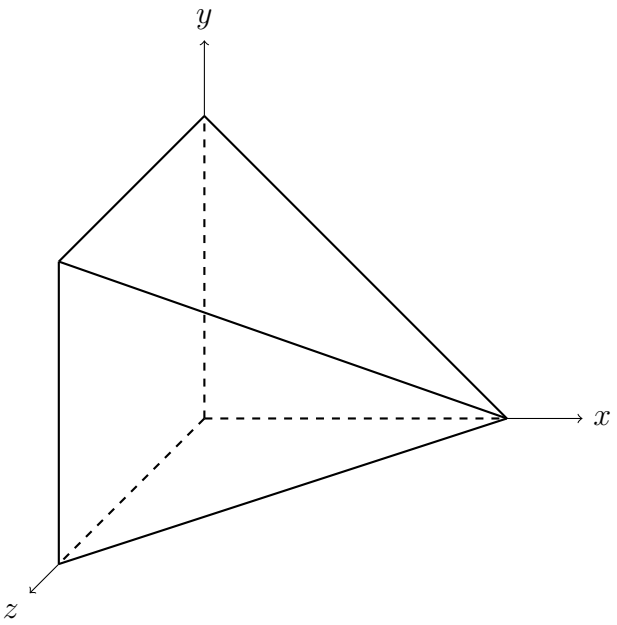

Figura 1.2: Esempio di STAB(G).

1.2 Equivalenza delle tre definizioni

1.2.1 Numero Cromatico e Mutua Informazione

Sia G un grafo. Chiamiamo *nucleo* un sottoinsieme dei vertici indipendente e massimale per l'inclusione. Avremo bisogno di due lemmi di teoria dei grafi e di una seconda nozione di potenza di grafo.

Lemma 1.2.1 *Il minimo numero di nuclei necessari a ricoprire V coincide con il numero cromatico $\chi(G)$.*

Dimostrazione. Sia $\xi(G)$ il numero minimo di nuclei che ricoprono V. Vogliamo mostrare che $\xi(G) = \chi(G)$. È evidente che una colorazione può essere estesa ad un ricoprimento di nuclei: basta prendere per ogni colore un nucleo che lo contenga, dunque $\xi(G) \leq \chi(G)$. Dato invece un ricoprimento minimo osserviamo che ogni nucleo possiede almeno un vertice non contenuto in alcun altro nucleo. Se per assurdo così non fosse potremmo rimuovere dal ricoprimento un nucleo i cui elementi sono contenuti in un qualche altro nucleo, contraddicendo la minimalità. Assegnando gli elementi in comune fra più nuclei ad uno qualunque di essi estraiamo una colorazione dei vertici in $\xi(G)$ colori, e quindi $\chi(G) \leq \xi(G)$. □

Lemma 1.2.2 *Sia G un grafo e siano v, w suoi vertici. Diciamo che*

$$v \sim w \iff (v, w) \in E \quad \vee \quad v = w.$$

Allora $\sim$ è una relazione d'equivalenza su V se e soltanto se G è unione di grafi completi a due a due disgiunti.

Figura 1.3: Esempio di potenza normale per $t = 2$.

Dimostrazione. Senza perdita di generalità possiamo supporre che G sia connesso: basta infatti dimostrare il risultato per ogni componente connessa. Supponiamo che G sia completo. È allora ovvio che $\sim$ sia d'equivalenza. Supponiamo invece che $\sim$ sia d'equivalenza e siano v, w vertici di G. Poiché G è connesso esiste un cammino da v a w, quindi applicando induttivamente la transitività otteniamo che $(v, w) \in E$. □

Definizione. Sia t un intero positivo e G un grafo. Chiamiamo *t-esima potenza normale* di G il grafo $G_t = (V^t, E_t)$, dove V^t è il prodotto cartesiano di t copie di V, mentre

$$E_t = \{(\mathbf{v}, \mathbf{w}) \mid \forall i \ v_i = w_i \vee (v_i, w_i) \in E, \ \exists i \ (v_i, w_i) \in E\}.$$

In altri termini fra due vertici $\mathbf{v}$ e $\mathbf{w}$ di G_t esiste un arco se le corrispondenti sequenze di lunghezza t sono distinguibili in almeno un elemento, nei restanti sono distinguibili oppure coincidono.

Esempio. Sia L il grafo su tre vertici con due archi, rappresentato a sinistra nella Figura 1.3. Allora L_2 è rappresentato a destra.

Osserviamo che la coppia $(G_t, \mathbf{P}^t)$, dove $\mathbf{P}^t$ è definita da (1.1), è un grafo probabilistico. Sia $0 < \varepsilon < 1$, allora la quantità

$$\chi_n(t, \varepsilon) = \min_{\substack{U \subset V^t \\ \mathbf{P}^t(U) \geq 1-\varepsilon}} \chi(G_t(U))$$

conta il numero minimo di parole che sono necessarie per dare una codifica propria di quasi tutte le sequenze di lunghezza t, in analogia con quanto discusso nel paragrafo 1.1.1.

Enunciamo i seguenti due lemmi omettendone la dimostrazione [15].

Lemma 1.2.3 *Sia G unione di grafi completi a due a due disgiunti e $\sim$ la relazione d'equivalenza definita nel Lemma 1.2.2. Denotiamo con $[v]$ la classe d'equivalenza di v per tale relazione e definiamo*

$$H(P \mid \sim) = \sum_{v \in V} P(v) \log \frac{P([v])}{P(v)}.$$

Allora

$$2^{tH(P|\sim) - K\sqrt{N}} \leq \chi_n(t, \varepsilon) \leq 2^{tH(P|\sim) + K\sqrt{N}}$$

per una qualche costante K non dipendente da t o da P, ma solo da $|V|$ ed ε.

Definizione. Siano V un insieme finito e P una densità discreta su di esso. Chiamiamo *P-tipica* una sequenza $\mathbf{v}$ di lunghezza t se $\forall w \in V$ il numero $N(w|\mathbf{v})$ di occorrenze di w nella sequenza soddisfa

$$\Big|N(w|\mathbf{v}) - tP(w)\Big| \leq K\sqrt{tP(w)}$$

per una qualche costante K.

In altre parole una sequenza è P-tipica se il numero di occorrenze di ogni simbolo w non differisce significativamente dal valore atteso $tP(w)$.

Lemma 1.2.4 *Sia $T^t(P)$ l'insieme delle sequenze P-tipiche. Allora*

(i) Per ogni λ esiste K tale che $\mathbf{P}^t\big(\overline{T^t(P)}\big) < \lambda$ per questo K.

(ii) Se $\mathbf{v}$ è P-tipica allora $2^{-tH(P)-C\sqrt{n}} \leq \mathbf{P}^t(\mathbf{v}) \leq 2^{-tH(P)+C\sqrt{n}}$.

(iii) La cardinalità di $T^t(P)$ è limitata da $2^{tH(P)-C\sqrt{n}} \leq |T^t(P)| \leq 2^{tH(P)+C\sqrt{n}}$.

Inoltre la costante C non dipende da t o da P, ma soltanto da $|V|$ ed λ.

Definizione. Sia (G, P) un grafo probabilistico. Chiamiamo *grafo dei nuclei* il grafo Γ, in cui l'insieme dei vertici sono le coppie del tipo (v, A) con A nucleo di G contenente v e due vertici (v, A) e (w, B) sono adiacenti se e soltanto se $A \neq B$.

Definizione. Sia (G, P) un grafo probabilistico e sia Γ il suo grafo dei nuclei. Una distribuzione di probabilità Q su $V(\Gamma)$ è detta *distribuzione ausiliaria ammissibile* se soddisfa

$$\sum_{A:\ v\in A} Q(v, A) = P(v). \tag{1.6}$$

Se $\mathcal{N}$ denota l'insieme dei nuclei di G, una distribuzione ausiliaria ammissibile definisce su $\mathcal{N}$ una distribuzione marginale R tramite

$$R(A) = \sum_{v:\ v\in A} Q(v, A). \tag{1.7}$$

Definendo le variabili aleatorie $X \sim P$ e $Y \sim R$ rispettivamente a valori in V e in $\mathcal{N}$ abbiamo che

$$I(Q) = I(X;Y) = \sum_{v\in V}\sum_{A\in\mathcal{N}} Q(v, A)\log\frac{Q(v, A)}{P(v)R(A)}. \tag{1.8}$$

Sia $\mathcal{A}$ l'insieme di tutte le distribuzioni ammissibili. Osserviamo che si tratta di un insieme chiuso e limitato nello spazio vettoriale delle funzioni di dominio $V \times \mathcal{N}$. La limitatezza segue dall'essere un insieme di distribuzioni di probabilità, quindi a somma 1. La chiusura invece è conseguenza dell'essere preimmagine di un insieme finito di punti tramite le equazioni (1.6) al variare di $v \in V$. Inoltre tale spazio vettoriale è di dimensione finita essendo sia V sia $\mathcal{N}$ finiti, per cui possiamo applicare il Teorema di Heine-Borel e concludere la compattezza di $\mathcal{A}$. Ma allora la funzione continua (1.8) possiede minimo per il Teorema di Weierstrass, da cui segue la buona definizione della seconda definizione di entropia di grafo. Nel seguito denoteremo con $H(G, P)$ tale minimo.

Abbiamo ora tutti gli strumenti necessari per dimostrare l'equivalenza delle prime due definizioni di entropia di grafo. Ne divideremo la dimostrazione nei successivi due teoremi, come in [16].

Teorema 1.2.5 (Körner) *Sia (G, P) un grafo probabilistico e sia $\delta > 0$. Per ogni $0 < \varepsilon < 1$ e per t abbastanza grande abbiamo*

$$\chi(t, \varepsilon) \geq 2^{t(H(G,P)-\delta)},$$

dove $\chi(t, \varepsilon)$ è il minimo numero di parole necessarie ad una codifica che sia propria rispetto alla potenza conormale, tranne per un insieme di probabilità totale ε.

Dimostrazione. Per il Lemma 1.2.4 (i), per ogni $\lambda > 0$ esiste K tale che, per t sufficientemente grande, $\mathbf{P}^t(T^t(P)) \geq 1 - \lambda$. Per un tale t sia U_t un sottoinsieme di V^t tale che $\mathbf{P}^t(U_t) \geq 1 - \varepsilon$. Ne deriva allora che

$$1 - \varepsilon - \lambda \leq \mathbf{P}^t(U_t \cap T^t(P)). \tag{1.9}$$

Ogni insieme indipendente di una colorazione nel minimo numero di colori può essere espanso ad un nucleo, quindi possiamo scrivere

$$\mathbf{P}^t(U_t \cap T^t(P)) \leq \chi(U_t \cap T^t(P)) \cdot \max_{\mathbf{N} \in \mathcal{N}^t} \mathbf{P}^t(\mathbf{N} \cap T^t(P)) \tag{1.10}$$

dove abbiamo indicato con $\mathcal{N}^t$ l'insieme dei nuclei di G^t.

Si tratta in effetti del prodotto cartesiano di t copie di $\mathcal{N}$. Supponiamo infatti che $\mathbf{N} \in \mathcal{N}^t$, e che p_i per $i \in \{1 \dots t\}$ siano le proiezioni di una sequenza di lunghezza t sull'i-esimo elemento. Supponiamo per assurdo che $p_i(\mathcal{N}^t)$ non sia un nucleo. Allora deve esistere v non connesso ad alcun vertice di $p_i(\mathcal{N}^t)$. Definiamo $\mathbf{w}$ come un qualunque elemento di $\mathbf{N}$ in cui abbiamo rimpiazzato l'i-esimo elemento con v. Questo elemento non è adiacente ad alcun elemento del nucleo, contro la massimalità di $\mathbf{N}$.

Stimando con il massimo la probabilità di ogni $\mathbf{v} \in \mathbf{N} \cap T^t(P)$ abbiamo la diseguaglianza

$$\max_{\mathbf{N} \in \mathcal{N}^t} \mathbf{P}^t(\mathbf{N} \cap T^t(P)) \leq \max_{\mathbf{v} \in T^t(P)} \mathbf{P}^t(\mathbf{v}) \cdot \max_{\mathbf{N} \in \mathcal{N}^t} \left|\mathbf{N} \cap T^t(P)\right|. \tag{1.11}$$

Sia ora $\mathbf{N}$ tale che $\left|\mathbf{N} \cap T^t(P)\right|$ sia massimo e sia $\mathbf{v} \in T^t(P) \cap \mathcal{N}$. Siano $w \in V$ e M un nucleo di G contenente w, e sia $N(w, M \mid \mathbf{v}, \mathbf{N})$ il numero di occorrenze di (w, M) nella sequenza

$$(v_1, N_1) \ \dots \ (v_t, N_t),$$

inoltre denotiamo con $N(w \mid \mathbf{v})$ il numero di occorrenze di w in $\mathbf{v}$. Allora la sequenza $(\mathbf{v}, \mathbf{N})$ è Q-tipica per la distribuzione di probabilità su $V(\Gamma)$

$$Q(w, M) = \frac{N(w, M \mid \mathbf{v}, \mathbf{N})}{N(w \mid \mathbf{v})} \cdot P(w). \tag{1.12}$$

Infatti, sfruttando la P-tipicità di $\mathbf{v}$, abbiamo

$$\begin{aligned}
\left|N(w, M \mid \mathbf{v}, \mathbf{N}) - tQ(w, M)\right| &= \left|\frac{tQ(w, M)}{tP(w)}\right| \cdot \left|N(w \mid \mathbf{v}) - tP(w)\right| \\
&\leq \left|\frac{Q(w, M)}{P(w)}\right| \cdot K\sqrt{tP(w)} = K\sqrt{t \cdot \frac{Q^2(w, M)}{P(w)}} \\
&\leq K\sqrt{tQ(w, M)}.
\end{aligned}$$

Sia poi R la distribuzione marginale su $\mathcal{N}$ definita da

$$R(M) = \sum_{w:\ w\in M} Q(w, M).$$

Allora $\mathbf{N}$ è P-tipica rispetto ad essa. Infatti abbiamo

$$N(M \mid \mathbf{N}) - tR(M) = \sum_{w\in M} N(w, M \mid \mathbf{v}, \mathbf{N}) - tQ(w, M)$$

e, sfruttando la Q-tipicità di $(\mathbf{v}, \mathbf{N})$, vale

$$\big|N(M \mid \mathbf{N}) - tR(M)\big| \le \sum_{w\in V} K\sqrt{tQ(w, M)} \le K_1\sqrt{t\sum_{w\in V} Q(w, M)} = K_1\sqrt{tR(M)},$$

dove la diseguaglianza centrale segue dalla concavità della radice.

Nel corso della dimostrazione del Lemma 1.2.3 viene dimostrato che una sequenza tipica di classi d'equivalenza per $\sim$ contiene T sequenze $P-$tipiche di G^t, dove T soddisfa

$$2^{tH(P|\sim)-C\sqrt{t}} \le T \le 2^{tH(P|\sim)+C\sqrt{t}}.$$

Osserviamo che $(v, A) \sim (w, B) \iff ((v, A), (w, B)) \notin E(\Gamma)$ è una relazione d'equivalenza, infatti è l'eguaglianza della seconda coordinata. Quindi per il Lemma 1.2.2 il grafo complementare $\overline{\Gamma}$ è unione di grafi completi a due a due disgiunti, e le sue componenti connesse sono i nuclei di G. Osserviamo inoltre che la coppia $(\overline{\Gamma}, Q)$, dove Q è definita dall'equazione (1.12), è un grafo probabilistico. Perciò siamo nelle condizioni di applicare il Lemma 1.2.3 e ottenere che il numero di sequenze $Q-$tipiche in ogni nucleo è compreso nell'intervallo

$$\Big[2^{tH(Q|\sim)-K_2\sqrt{t}},\ 2^{tH(Q|\sim)+K_2\sqrt{t}}\Big]. \tag{1.13}$$

Esistono al più $(t+1)^{|V(\Gamma)|}$ distribuzioni ausiliarie di probabilità del tipo (1.12), infatti ogni coppia (w, M) comparirà da 0 a t volte in $(\mathbf{v}, \mathbf{N})$. Ogni sequenza $\mathbf{v}$ in $\mathbf{N} \cap T^t(P)$ sarà Q-tipica per una probabilità del tipo (1.12); infatti dalla dimostrazione è evidente che basta che $\mathbf{v}$ sia P-tipica e poi porre $\mathbf{A} = \mathbf{N}$. Allora abbiamo la stima

$$\max_{\mathbf{N}\in\mathcal{N}^t} \big|\mathbf{N} \cap T^t(P)\big| \le (t+1)^{|V(\Gamma)|} \cdot 2^{t\max_{Q\in\mathcal{A}} H(Q|\sim)+K_1\sqrt{t}}, \tag{1.14}$$

perché le Q sono distribuzioni ammissibili nel senso della definizione (1.6), infatti

$$\sum_{M:\ w\in M} Q(w, M) = \frac{P(w)}{N(w \mid \mathbf{v})} \cdot \sum_{M:\ w\in M} N(w, M \mid \mathbf{v}, \mathbf{A}) = P(w).$$

Ricordando inoltre il Lemma 1.2.4 (ii) abbiamo

$$\max_{\mathbf{v}\in T^t(P)} \mathbf{P}^t(\mathbf{v}) \le 2^{-tH(P)-C\sqrt{t}}. \tag{1.15}$$

Possiamo ora concludere. Dalle disuguaglianze (1.9) − (1.11), (1.14) e (1.15) otteniamo

$$1 - \varepsilon - \lambda \le \chi(U_t \cap T^t(P)) \cdot \exp_2\left[t(\max_{Q\in\mathcal{A}} H(Q\ |\sim) - H(P)) + K\sqrt{t} + |V(\Gamma)| \cdot \log(t+1)\right]$$

ed equivalentemente

$$\chi(U_t \cap T^t(P)) \geq (1-\varepsilon-\lambda)\cdot \exp_2\left[t(H(P) - \max_{Q\in\mathcal{A}} H(Q\,|\sim)) - K\sqrt{t} - |V(\Gamma)|\cdot\log(t+1)\right].$$

D'altra parte vale

$$H(P) - \max_{Q\in\mathcal{A}} H(Q\,|\sim) = \min_{Q\in\mathcal{A}} \sum_{v\in V}\sum_{A\in\mathcal{N}} Q(v,A)\log\frac{Q(v,A)}{P(v)R(A)}$$

e quindi riconosciamo in questo addendo $H(G,P)$. Inoltre chiaramente $\chi(U_t) \geq \chi(U_t \cap T^t(P))$ intersecando gli insiemi indipendenti di una colorazione. Perciò possiamo scrivere

$$\chi(U_t) \geq (1-\varepsilon-\lambda)\cdot \exp_2\left[tH(G,P) - K\sqrt{t} - |V(\Gamma)|\cdot\log(t+1)\right]$$

per ogni U_t che soddisfi $\mathbf{P}^t(U_t) \geq 1-\varepsilon$. Prendendo i logaritmi e dividendo per t abbiamo

$$\frac{1}{t}\log \min_{\mathbf{P}^t(U_t)\geq 1-\varepsilon} \chi(U_t) \geq \frac{1}{t}\log(1-\varepsilon-\lambda) + H(G,P) - \frac{K}{\sqrt{t}} - \frac{|V(\Gamma)|}{t}\cdot\log(t+1)$$

e quindi otteniamo la tesi

$$\liminf_{t\to\infty} \frac{1}{t}\log\chi(t,\varepsilon) \geq H(G,P).$$

□

Teorema 1.2.6 (Körner) *Sia (G,P) un grafo probabilistico e sia $\delta > 0$. Per ogni $0 < \varepsilon < 1$ e per t abbastanza grande esiste U_t, sottografo di G^t tale che $\mathbf{P}^t(U_t) \geq 1-\varepsilon$ e inoltre*

$$\chi(U_t) \leq 2^{t(H(G,P)+\delta)}.$$

Dimostrazione. Sia R la distribuzione marginale sull'insieme dei nuclei della Q che realizza il minimo $H(G,P)$. Poiché i nuclei di G^t sono prodotti cartesiani di t nuclei di G possiamo definire una distribuzione di probabilità su $\mathcal{N}^t$ tramite

$$R^*(\mathbf{A}) = \prod_{i=1}^{t} R(A_i).$$

Similmente sull'insieme delle successioni di M nuclei di G^t la formula

$$R^*_M(\mathbf{A}_1,\dots\mathbf{A}_M) = \prod_{j=1}^{M} R^*(\mathbf{A}_j)$$

definisce una misura di probabilità sugli insiemi di M nuclei di G^t. Denoteremo con $(\mathbf{A}_1,\dots\mathbf{A}_M)^{\mathsf{c}}$ l'insieme dei $\mathbf{v}\in V^t$ non contenuti in alcun nucleo $\mathbf{A}_1,\dots\mathbf{A}_M$. Il nostro obiettivo è trovare un M per cui il valore atteso di $\mathbf{P}^t\big((\mathbf{A}_1,\dots\mathbf{A}_M)^{\mathsf{c}}\big)$ sia minore di ε, in modo che esista un sottografo U_t ricoperto dai nuclei $\mathbf{A}_1,\dots\mathbf{A}_M$ e tale che $\mathbf{P}^t(U_t) \geq 1-\varepsilon$.

Osserviamo che vale

$$\sum_{\mathbf{A}_1,\dots\mathbf{A}_M} R^*_M(\mathbf{A}_1,\dots\mathbf{A}_M)\cdot\mathbf{P}^t\big((\mathbf{A}_1,\dots\mathbf{A}_M)^{\mathsf{c}}\big) = \sum_{\mathbf{v}\in V^t} R^*_M(C_{\mathbf{v}})\cdot\mathbf{P}^t(\mathbf{v})$$

dove $C_{\mathbf{v}}$ è l'evento $\{\mathbf{v} \notin \mathbf{A}_1, \dots \mathbf{A}_M\}$. Possiamo spezzare il membro di destra in

$$\sum_{\mathbf{v}\in V^t} R^*_M(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) = \sum_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) \; + \sum_{\mathbf{v}\in \overline{T^t(P)}} R^*_M(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) \tag{1.16}$$

ed osservare che il secondo termine è maggiorato da $\mathbf{P}^t\Big(\overline{T^t(P)}\Big)$, che per il Lemma 1.2.4 (i) possiamo supporre essere più piccolo di $\varepsilon/2$. Per stimare il primo termine osserviamo innanzitutto che

$$\sum_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) \leq \mathbf{P}^t\big(T^t(P)\big) \cdot \max_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}}) \leq \max_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}})$$

maggiorando con il massimo e sfruttando il fatto che hanno massa totale 1. Possiamo maggiorare ulteriormente questo termine con

$$\max_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}}) \leq \max_{\mathbf{v}\in T^t(P)} (1 - R^*(\mathcal{N}_{\mathbf{v}}))^M,$$

dove $\mathcal{N}_{\mathbf{v}}$ indica l'insieme dei nuclei che contengono $\mathbf{v}$, infatti ognuno degli M termini della produttoria $R^*_M(C_{\mathbf{v}})$ è maggiorato da $1 - R^*(\mathcal{N}_{\mathbf{v}})$. Vogliamo ora stimare $R^*(\mathcal{N}_{\mathbf{v}})$.

Ricordiamo che Q è la distribuzione su $V(\Gamma)$ che realizza il minimo $H(G, P)$, e che se $\mathbf{v}$ è $P-$tipica allora $(\mathbf{v}, \mathbf{N})$ è $Q-$tipica. Osserviamo che l'uguaglianza della prima coordinata è una relazione d'equivalenza su $V(\Gamma)$, quindi applicando il Lemma 1.2.3 il numero di sequenze $Q-$tipiche soddisfa

$$\big|T^t(Q)\big| \geq 2^{tH(Q|\sim) - K_3\sqrt{t}}. \tag{1.17}$$

In questo caso la classe d'equivalenza di un vertice (v, A) è l'insieme delle coppie (v, B) con B nucleo contenente v. Per l'identità (1.6) la probabilità totale di una classe d'equivalenza è $P(v)$, quindi vale

$$H(Q\,|\sim) = \sum_{v\in V} \sum_{A\in\mathcal{N}} Q(v, A) \log \frac{P(v)}{Q(v, A)}. \tag{1.18}$$

Inoltre il Lemma 1.2.4 (ii) implica che se $\mathbf{A}$ è una sequenza di nuclei $R-$tipica allora

$$R^*(\mathbf{A}) \geq 2^{-tH(R) - K_4\sqrt{t}}. \tag{1.19}$$

Possiamo allora stimare $R^*(\mathcal{N}_{\mathbf{v}})$ contando il numero di sequenze $R-$tipiche in esso. Queste saranno almeno tante quante le sequenze $Q-$tipiche di prima componente v, che abbiamo stimato con (1.17). Abbiamo stimato la probabilità di ciascuna in (1.19), per cui possiamo concludere che $R^*(\mathcal{N}_{\mathbf{v}})$ sia maggiore del loro prodotto.

Per quanto appena visto otteniamo

$$\max_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}}) \leq \Big(1 - \exp_2\Big[-tH(R) - K_4\sqrt{t} + tH(Q\,|\sim) - K_3\sqrt{t}\Big]\Big)^M \tag{1.20}$$

Mettendo insieme le equazioni (1.7) e (1.18) osserviamo che $H(Q\,|\sim) - H(R)$ si semplifica in $H(G, P)$. Possiamo quindi riscrivere

$$\max_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}}) \leq \Big(1 - \exp_2\Big[-tH(G, P) - K_5\sqrt{t}\Big]\Big)^M$$

e inoltre, con facili calcoli, ottenere

$$\max_{\mathbf{v}\in T^t(P)} R_M^*(C_{\mathbf{v}}) \leq \exp_2\left(-M\cdot 2^{-tH(G,P)-K_5\sqrt{t}}\right).$$

Poniamo $M = \left\lfloor 2^{tH(G,P)+\delta}\right\rfloor$. È immediato allora che il primo termine del membro di destra di (1.16) sia infinitesimo per $t\to\infty$. Quindi per ogni ε e δ esistono M nuclei che ricoprono un sottografo di probabilità almeno $1-\varepsilon$. Ma per il Lemma 1.2.1 il numero cromatico di un grafo è uguale al minimo numero di nuclei che lo ricoprono, quindi per ogni $\delta > 0$

$$\min_{\substack{U_t\subset V^t\\ \mathbf{P}^t(U_t)\geq 1-\varepsilon}} \chi(U_t) \leq 2^{tH(G,P)+\delta}$$

o equivalentemente

$$\limsup_{t\to\infty}\frac{1}{t}\log\chi(t,\varepsilon) \leq H(G,P).$$

□

1.2.2 Mutua Informazione e Politopo dei Vertici

Teorema 1.2.7 (Csiszár, Körner, Lovász, Marton, Simonyi) *Sia G un grafo e siano $S(G)$ i suoi insiemi indipendenti. Sia P una densità discreta sui vertici di G. Allora abbiamo:*

$$\min_{\substack{X\in Y\in S(G)\\ X\sim P}} I(X;Y) \;=\; \min_{\substack{\mathbf{a}\in STAB(G)\\ \mathbf{a}>0}} \sum_{i=1}^{n} p_i\log\frac{1}{a_i}.$$

Dimostrazione. Siano X, Y variabili aleatorie che realizzino il minimo del membro di sinistra, e sia Q la distribuzione marginale di Y. Denotiamo con R la distribuzione condizionale di Y nota X. Abbiamo

$$I(X;Y) = -\sum_{i=1}^{n} p_i \sum_{i\in J\in S(G)} R(J\mid i)\log\frac{Q(J)}{R(J\mid i)} \geq -\sum_{i=1}^{n} p_i\log\sum_{i\in J\in S(G)} Q(J)$$

utilizzando nel primo passaggio la definizione di mutua informazione, nel secondo la concavità del logaritmo. Poniamo

$$a_i = \sum_{i\in J\in S(G)} Q(J)$$

ed osserviamo che $\mathbf{a}$ è contenuto in STAB(G) perché combinazione convessa di vettori delle caratteristiche degli insiemi indipendenti. Ne segue

$$\min_{\substack{X\in Y\in S(G)\\ X\sim P}} I(X;Y) \;\geq\; \min_{\substack{\mathbf{a}\in \mathrm{STAB}(G)\\ \mathbf{a}>0}} \sum_{i=1}^{n} p_i\log\frac{1}{a_i}.$$

Sia allora $\mathbf{a}$ che realizzi il minimo nel membro di destra. Similmente a prima possiamo porre

$$a_i = \sum_{i\in J\in S(G)} Q'(J)$$

poiché $\mathbf{a} \in \text{STAB}(G)$. Possiamo pensare i $Q'(J)$ sia come pesi di una combinazione convessa sia come una distribuzione di probabilità su $S(G)$. Definiamo

$$R'(J \mid i) = \begin{cases} \frac{Q'(J)}{a_i} & \text{se } i \in J \\ 0 & \text{altrimenti} \end{cases}$$

e, grazie ad essa, una nuova distribuzione su $S(G)$ tramite la formula

$$Q^*(J) = \sum_{i=1}^{n} p_i \; R'(J \mid i).$$

Siano allora X di legge P ed Y di legge Q^*. Per come le abbiamo definite esse soddisfano $X \in Y \in S(G)$, quindi vale la diseguaglianza

$$\min_{\substack{X \in Y \in S(G) \\ X \sim P}} I(X;Y) \leq -\sum_{i=1}^{n} p_i \sum_{i \in J \in S(G)} R'(J \mid i) \log \frac{Q^*(J)}{R'(J \mid i)}.$$

Scrivendo la disuguaglianza di concavità del logaritmo con pesi i $Q^*(J)$ abbiamo

$$\sum_{J \in S(G)} Q^*(J) \log \frac{(Q'(J))}{Q^*(J)} \leq 0,$$

da cui, sostituendo la definizione di $Q^*(J)$,

$$-\sum_{i,J} p_i \; R'(J \mid i) \log(Q^*(J)) \leq -\sum_{i,J} p_i \; R'(J \mid i) \log(Q'(J)).$$

Sostituendo e ricordando le definizioni di $Q'(J)$ e $R'(J \mid i)$ abbiamo la tesi:

$$\begin{aligned} \min_{\substack{X \in Y \in S(G) \\ X \sim P}} I(X;Y) &\leq -\sum_{i=1}^{n} p_i \sum_{i \in J \in S(G)} R'(J \mid i) \log \frac{Q^*(J)}{R'(J \mid i)} \\ &\leq -\sum_{i=1}^{n} p_i \sum_{i \in J \in S(G)} R'(J \mid i) \log \frac{Q'(J)}{R'(J \mid i)} \\ &= -\sum_{i=1}^{n} p_i \log a_i. \end{aligned}$$

□

Capitolo 2

Proprietà e teoremi principali

2.1 Proprietà dell'entropia di grafo

2.1.1 Monotonia

Lemma 2.1.1 *Siano F e G grafi tali che $V(F) = V(G)$ e $E(F) \subset E(G)$. Allora per ogni scelta di P densità discreta sui vertici si ha $H(F, P) \leq H(G, P)$.*

Dimostrazione. Osserviamo che se $E(F) \subset E(G)$ allora $\mathrm{STAB}(G) \subset \mathrm{STAB}(F)$. Sfruttando la terza definizione di entropia di grafo abbiamo immediatamente la tesi, infatti stiamo prendendo il minimo della stessa funzione obiettivo su un insieme più grande. □

2.1.2 Subadditività

Lemma 2.1.2 *Siano F e G grafi aventi $V(F) = V(G)$. Sia $F \cup G$ il grafo di vertici V ed insieme degli archi $E(F) \cup E(G)$. Per ogni scelta di P densità discreta sui vertici si ha*

$$H(F \cup G, P) \leq H(F, P) + H(G, P).$$

Dimostrazione. Siano $\mathbf{a} \in \mathrm{STAB}(F)$ e $\mathbf{b} \in \mathrm{STAB}(G)$ i vettori che realizzino il minimo delle rispettive entropie. Osserviamo che l'intersezione di un insieme indipendente di F e di un insieme indipendente di G è un insieme indipendente in $F \cup G$. In altri termini il prodotto componente a componente $\mathbf{a} \cdot \mathbf{b}$ dei loro vettori caratteristici è il vettore caratteristico di un insieme indipendente di $F \cup G$. Pertanto, sfruttando la convessità del politopo dei vertici, $\mathbf{a} \cdot \mathbf{b}$ appartiene a $\mathrm{STAB}(F \cup G)$. Ma allora possiamo scrivere

$$H(F, P) + H(G, P) = \sum_{i=1}^{n} p_i \log \frac{1}{a_i} + \sum_{i=1}^{n} p_i \log \frac{1}{b_i} = \sum_{i=1}^{n} p_i \log \frac{1}{a_i b_i} \geq H(F \cup G, P).$$

□

2.1.3 Additività per sostituzioni

Siano F e G grafi tali che $V(F) \cap V(G) = \emptyset$ e sia v un vertice di G. Chiamiamo grafo ottenuto sostituendo F a v, e scriviamo $G_{v \leftarrow F}$, il grafo ottenuto da G cancellando v e connettendo ogni vertice adiacente a v con ciascun vertice di una copia isomorfa di F. Supponiamo inoltre che P sia una densità sui vertici di G e che Q sia una densità sui vertici di F. Allora possiamo definire una densità $P_{v \leftarrow Q}$ in modo che la coppia $(G_{v \leftarrow F}, P_{v \leftarrow Q})$ sia un grafo probabilistico. Per fare questo poniamo

$$P_{v \leftarrow Q}(x) = \begin{cases} P(x) & \text{se } x \in V(G) - \{v\} \\ P(v)Q(x) & \text{se } x \in V(F). \end{cases}$$

Vale allora il seguente lemma sull'entropia di grafi della forma $G_{v \leftarrow F}$ per certi grafi F e G, di cui omettiamo la dimostrazione [18].

Lemma 2.1.3 *Siano F e G grafi su insiemi di vertici disgiunti, e sia v un vertice di G. Siano inoltre P una densità sui vertici di G e Q una densità sui vertici di F. Allora abbiamo*

$$H(G_{v \leftarrow F}, P_{v \leftarrow Q}) = H(G, P) + P(v)H(F, Q).$$

Il precedente lemma consente inoltre di ricondurre il calcolo dell'entropia di un grafo al calcolo delle entropie delle componenti connesse.

Corollario 2.1.4 *Sia (G, P) un grafo probabilistico e siano $G_1 \dots G_k$ le sue componenti connesse. Poniamo $P_i(x) = \frac{P(x)}{P(V(G_i))}$ per $x \in V(G_i)$. Allora abbiamo*

$$H(G, P) = \sum_{i=1}^{k} P(V(G_i))H(G_i, P_i).$$

Dimostrazione. Consideriamo il grafo su k vertici $\{v_1 \dots v_k\}$ privo di archi, e sia Q la densità discreta definita da $Q(v_i) = P(V(G_i))$. Allora otteniamo la tesi applicando k volte il lemma precedente, sostituendo ad ogni passo il vertice v_i con la componente connessa G_i. □

2.1.4 Entropia di grafo completo

Proposizione 2.1.5 *Sia K_n il grafo completo su n vertici. Per ogni P densità discreta sui vertici abbiamo*

$$H(K_n, P) = H(P).$$

Dimostrazione. Sfruttando la terza definizione di entropia di grafo sappiamo che

$$H(K_n, P) = \sum_{i=1}^{n} p_i \log \frac{1}{q_i}$$

per certi $q_1 \dots q_n$ positivi. Osserviamo inoltre che nel grafo completo gli insiemi indipendenti sono soltanto $\emptyset$ e i singoletti dei vertici. Pertanto il politopo dei vertici è l'n-simplesso, ma poiché sappiamo che la funzione obiettivo è minima sul bordo deduciamo che

$$\sum_{i=1}^{n} q_i = 1. \tag{2.1}$$

Dalla disuguaglianza di Jensen segue la ben nota "Log sum inequality", cioè la disuguaglianza

$$\sum_{i=1}^{n}\left(a_i \log \frac{a_i}{b_i}\right) \geq \sum_{i=1}^{n} a_i \cdot \log \frac{\sum_{j=1}^{n} a_j}{\sum_{j=1}^{n} b_j}$$

per due n-uple di numeri non negativi $a_1, \ldots, a_n$ e $b_1, \ldots, b_n$ [7]. Ponendo allora $p_i = a_i$ e $b_i = q_i$ per $1 \leq i \leq n$ otteniamo

$$\sum_{i=1}^{n}\left(p_i \log \frac{p_i}{q_i}\right) \leq \log \sum_{i=1}^{n} q_i = 0,$$

dove nell'ultimo passaggio abbiamo sfruttato l'equazione (2.1). Ma allora deduciamo

$$\sum_{i=1}^{n} p_i \log q_i - \sum_{i=1} p_i \log p_i \leq 0,$$

cioè

$$\sum_{i=1}^{n}\left(p_i \log \frac{1}{q_i}\right) \geq \sum_{i=1}^{n}\left(p_i \log \frac{1}{p_i}\right).$$

Osserviamo infine che tale minimo viene realizzato ponendo $p_i = q_i$ per $1 \leq i \leq n$, da cui la tesi. □

Osservazione. Come anticipato nell'introduzione abbiamo riottenuto l'entropia di Shannon come caso particolare dell'entropia di grafo.

2.2 Entropia e grafi perfetti

Definizione. Sia G un grafo. Chiamiamo *cricca* un sottografo completo di G, e chiamiamo *numero di cricca* il massimo numero di vertici $\omega(G)$ in una cricca di G.

Il numero di cricca di un grafo G fornisce una stima dal basso del numero cromatico. È infatti evidente che $\omega(G) \leq \chi(G)$, poiché sono necessari almeno tanti colori quanti sono i vertici della massima cricca. Possiamo dunque porci il problema di caratterizzare quei grafi per cui tale disuguaglianza sia in realtà una uguaglianza.

Definizione. Sia G un grafo. Diciamo che G è *perfetto* se, per ogni sottografo H, vale

$$\omega(H) = \chi(H).$$

Nella precedente definizione, dovuta a Berge [1], abbiamo richiesto che l'uguaglianza valga per ogni sottografo al fine di non considerare perfette le unioni disgiunte di componenti per cui valga l'uguaglianza e componenti per cui non valga. I grafi perfetti hanno interesse combinatoriale e algoritmico poiché per essi è possibile esibire algoritmi polinomiali per problemi NP-completi nel caso generale [10]. La loro caratterizzazione è stata oggetto di più congetture, la più importante delle quali dimostrata nel 2002 [5]. Esiste un sorprendente collegamento fra l'entropia di grafo e i grafi perfetti [8].

Teorema 2.2.1 (Csiszár, Körner, Lovász, Marton, Simonyi) *Sia G un grafo. G è perfetto se e soltanto se, per ogni distribuzione di probabilità P sui vertici, vale*

$$H(G,P) + H\left(\overline{G},P\right) = H(P).$$

Come corollario del precedente teorema otteniamo che G è perfetto se e soltanto se $\overline{G}$ è perfetto. Questo enunciato, un tempo noto come congettura debole dei grafi perfetti, è noto anche come Teorema di Lovász [20].

È infine possibile dare una caratterizzazione più precisa di STAB(G) quando G è un grafo perfetto. Vale infatti il seguente lemma, di cui omettiamo la dimostrazione [6].

Lemma 2.2.2 (Chvatál) *Sia G un grafo perfetto di insieme di vertici V. Allora*

$$STAB(G) = \left\{ x \in \mathbb{R}_+^V \ : \ \sum_{v \in K} x_v \leq 1 \quad \forall K \text{ cricca di } G \right\}.$$

2.3 Grafi associati a ordini parziali

Definizione. Chiamiamo *ordine parziale* una relazione binaria $\leq$ che sia riflessiva, antisimmetrica e transitiva. Inoltre chiamiamo *insieme parzialmente ordinato* la coppia di un insieme e un ordine parziale su di esso.

Siano P insieme parzialmente ordinato e a, b due suoi elementi. Diciamo che a e b sono *confrontabili* se $a \leq b$ oppure $b \leq a$, *inconfrontabili* altrimenti. Chiamiamo *catena* un sottoinsieme di elementi a due a due confrontabili, e *anticatena* un sottoinsieme di elementi a due a due inconfrontabili. Infine diciamo che b è *massimale* se vale $a \leq b$ per ogni a confrontabile con b, e definiamo in modo analogo gli elementi *minimali*. È immediato osservare che gli elementi massimali e gli elementi minimali formano anticatene.

Associamo a P un grafo di vertici gli elementi di P e un arco fra $a, b \in P$ se e soltanto se a e b sono confrontabili, che denotiamo $G(P)$. Nel seguito faremo uso anche del grafo complementare, che denoteremo invece $\overline{G}(P)$.

I grafi associati agli ordini parziali godono di interessanti proprietà. Ad esempio il seguente teorema fornisce condizioni sufficienti perché un tale grafo sia partizionabile in insiemi indipendenti [21].

Teorema 2.3.1 (Mirsky) *Siano P un insieme parzialmente ordinato ed m un intero positivo. Se P non possiede alcuna catena di lunghezza $m+1$ allora può essere scritto come unione disgiunta di m anticatene.*

Dimostrazione. Dimostriamo l'enunciato per induzione su m. Se $m = 1$ non c'è niente da dimostrare: in tal caso infatti non esistono catene di lunghezza 2, quindi tutti gli elementi sono fra loro inconfrontabili, perciò P è una anticatena.

Sia allora $m \geq 2$ e supponiamo la tesi vera per ogni intero positivo minore di m. Sia P un insieme parzialmente ordinato privo di una catena di lunghezza $m+1$, sia M l'anticatena di P degli elementi massimali. M è non vuota perché contiene almeno i

massimi delle catene di lunghezza massima di P. Inoltre $P - M$ non contiene alcuna catena di cardinalità m. Se infatti così non fosse allora esisterebbe una catena

$$x_1 < x_2 < \cdots < x_m, \qquad \text{con} \quad x_k \in P - M \quad \text{per} \quad 1 \leq k \leq m,$$

di cardinalità m e quindi massimale, perciò deve essere $x_k \in M$, contro l'ipotesi. Possiamo dunque applicare l'ipotesi induttiva a $P - M$, che può quindi essere scritto come unione disgiunta di $m-1$ anticatene. Ma allora abbiamo scritto P come unione di m anticatene: le precedenti $m - 1$ ed M. □

Corollario 2.3.2 *Sia P un insieme parzialmente ordinato. Allora il grafo associato $G(P)$ è perfetto.*

Dimostrazione. Gli elementi di una catena di P sono fra loro confrontabili, dunque inducono in $G(P)$ un sottografo completo. Gli elementi di una anticatena, invece, inducono un insieme indipendente. Il Teorema 2.3.1 afferma dunque che, se $G(P)$ possiede una cricca massimale di m vertici, allora si può scrivere come unione di m insiemi indipendenti, cioè ha numero cromatico al più m. Quindi $\chi(G(P)) \leq \omega(G(P))$, e perciò $\chi(G(P)) = \omega(G(P))$. Osserviamo ora che le restrizioni di un ordine parziale ai sottoinsiemi dell'insieme di sostegno sono ancora ordini parziali. Ai grafi ad essi associati si applica il precedente argomento, cioè numero cromatico e numero di cricca coincidono. Abbiamo quindi dimostrato che, per ogni sottografo H di $G(P)$, si ha $\chi(H) = \omega(H)$, perciò $G(P)$ è perfetto. □

2.4 Entropia approssimata

Abbiamo finora sviluppato la teoria dell'entropia di grafo per una distribuzione di probabilità qualunque sui vertici. Nel seguito sarà sufficiente considerare il caso in cui tale distribuzione sia uniforme. Sia dunque G un grafo di insieme dei vertici V. Allora definiamo

$$H(x) = -\frac{1}{n} \sum_{v \in V} \log x_v$$

per $x \in \mathbb{R}^V$, e denotiamo con $H(G)$ l'entropia rispetto alla distribuzione uniforme, cioè

$$H(G) = \min_{x \in \mathrm{STAB}(G)} -\frac{1}{n} \sum_{v \in V} \log x_v = \min_{x \in \mathrm{STAB}(G)} H(x).$$

Questa forma semplificata permette di darne più facilmente stime. Consideriamo infatti l'algoritmo goloso che ad ogni passo sceglie un insieme indipendente e massimale e procede ricorsivamente sul complementare, e sia $\{S_1, \ldots, S_k\}$ una partizione dei vertici in insiemi indipendenti ottenuta con tale algoritmo. Allora chiamiamo *punto goloso* il punto $\tilde{x}$ definito da

$$\tilde{x} = \sum_{i=1}^{k} \frac{|S_i|}{n} \chi^{S_i}.$$

Dalla costruzione è evidente che tale punto appartenga a $\mathrm{STAB}(G)$, e dunque valga $H(G) \leq H(\tilde{x})$. Nel prossimo teorema viene invece dimostrata una stima superiore [3].

Teorema 2.4.1 (Cardinal, Fiorini, Joret, Jungers, Munro) *Sia G un grafo perfetto su n vertici e sia $\tilde{x}$ un suo punto goloso. Allora, comunque fissato $\varepsilon > 0$, vale*

$$H(\tilde{x}) \leq (1+\varepsilon)H(G) + (1+\varepsilon)\log\left(1+\frac{1}{\varepsilon}\right).$$

Dimostrazione. Sia $S_1, \ldots, S_k$ la sequenza di insiemi indipendenti prodotta dall'algoritmo goloso. In altri termini S_1 è un insieme indipendente e massimale in G, mentre S_2 è indipendente e massimale in $G - S_1$ e così via. Sia $\delta > 0$ fissato. Per ogni vertice $v \in V$ denotiamo con $m(v)$ l'unico indice in $\{1, \ldots, k\}$ tale che $v \in S_{m(v)}$. Definiamo allora un punto z di componenti date da

$$z_v = \frac{\delta}{n}\left(\frac{1}{\tilde{x}_v}\right)^{1-\delta} = \frac{\delta}{n}\left(\frac{n}{|S_{m(v)}|}\right)^{1-\delta} = \frac{\delta}{n^\delta}\left(\frac{1}{|S_{m(v)}|}\right)^{1-\delta}$$

e dimostriamo che $z \in \mathrm{STAB}(\overline{G})$. A tale scopo mostreremo che, per ogni insieme indipendente S, vale

$$\sum_{v\in S} z_v \leq 1;$$

infatti segue immediatamente dal Lemma 2.2.2 che

$$\mathrm{STAB}(\overline{G}) = \left\{ x \in \mathbb{R}_+^V \ : \ \sum_{v\in K} x_v \leq 1 \quad \forall K \text{ insieme indipendente e massimale di } G \right\}.$$

L'insieme indipendente S è ricoperto da l insiemi fra gli $S_1, \ldots, S_k$. Senza perdita di generalità possiamo assumere che questi siano i primi l. Scriviamo allora $S = T_1 \cup T_2 \cup \cdots \cup T_l$, dove T_i è l'intersezione fra S_i ed S. Per ogni $v \in T_1$ abbiamo $S_{m(v)} = S_1$, dunque $|S_{m(v)}| > |S|$, altrimenti l'algoritmo goloso avrebbe selezionato S al posto di $S_{m(v)}$. Analogamente otteniamo che per $1 \leq i \leq l$ e $v \in T_i$ abbiamo $|S_{m(v)}| \geq |S| - \sum_{j=1}^{i-1} |T_j|$. In particolare, poiché i T_i sono non vuoti, possiamo numerare i punti di S in modo che

$$|S_{m(v_i)}| \geq |S| - i + 1 \qquad \forall i \in \{1, 2, \ldots, s\}.$$

Ma allora abbiamo

$$\begin{aligned}
\sum_{v\in S} z_v &\leq \frac{\delta}{n^\delta}\left(\left(\frac{1}{|S|}\right)^{1-\delta} + \left(\frac{1}{|S|-1}\right)^{1-\delta} + \cdots + 1\right) \\
&\leq \frac{\delta}{n^\delta}\left(\int_0^{|S|} \frac{1}{x^{1-\delta}}\,\mathrm{d}x\right) \\
&\leq 1.
\end{aligned}$$

Possiamo ora concludere. Poiché G è perfetto possiamo applicare il Teorema 2.2.1; inoltre, essendo $z \in \mathrm{STAB}(G)$, possiamo scrivere la disuguaglianza

$$\begin{aligned}
H(G) &= \log(n) - H(\overline{G}) \\
&\geq \log(n) + \frac{1}{n}\sum_{v\in V} \log z_v.
\end{aligned}$$

Con semplici passaggi algebrici otteniamo

$$\begin{aligned}\log(n) + \frac{1}{n}\sum_{v\in V}\log z_v &= \log(n) + \frac{1}{n}\sum_{v\in V}\log\left(\frac{\delta}{n}\left(\frac{1}{\tilde{x}_v}\right)^{1-\delta}\right)\\ &= -\frac{1-\delta}{n}\sum_{v\in V}\log(\tilde{x}_v) - \log\frac{1}{\delta}\\ &= (1-\delta)H(\tilde{x}) - \log\frac{1}{\delta},\end{aligned}$$

da cui deduciamo

$$H(\tilde{x}) \leq \frac{1}{1-\delta}H(G) + \frac{1}{1-\delta}\log\frac{1}{\delta}.$$

Basta ora porre $\delta = \frac{\varepsilon}{\varepsilon+1}$ e ricaviamo la tesi. □

2.5 Formulazione per ottimizzazione convessa

Sia $P = (V, \leq_P)$ un insieme parzialmente ordinato e sia $G = G(P)$ il grafo ad esso associato. Nel seguito scriveremo $H(P)$ e $H(\overline{P})$ invece di $H(G(P))$ e $H(\overline{G}(P))$.

Siano $v, w \in V$. Diciamo che v *è ricoperto da* w se $v \leq_P w$, $v \neq w$ e $v \leq_P z \leq_P w$ implica $z = v$ oppure $z = w$. Chiamiamo *diagramma di Hasse* il grafo diretto di insieme dei nodi V e insieme delle frecce $\{(v, w) : v \text{ è ricoperto da } w \text{ in } P\}$.

Costruiamo ora un grafo diretto $D = D(P)$ di insieme dei nodi

$$N(D) \;=\; \{s, t\} \cup \{v^- : v \in V\} \cup \{v^+ : v \in V\}$$

e insieme delle frecce

$$\begin{aligned}A(D) \;=& \{(s, v^-) : v \in V,\, v \text{ è minimale in } P\} \cup \{(v^-, v^+) : v \in V\} \cup\\ & \{(v^+, w^-) : v \text{ è ricoperto da } w \text{ in } P\} \cup \{(v^+, t) : v \in V,\, v \text{ è massimale in } P\}.\end{aligned}$$

In altri termini abbiamo espanso ogni nodo v in un arco diretto da v^- a v^+. Abbiamo poi definito un vertice s da cui parta un arco verso v^- per ogni v minimale, e un arco t in cui arrivi un arco da v^+ per ogni v massimale.

Lemma 2.5.1 *Sia $P = (V, \leq_P)$ un insieme parzialmente ordinato e siano $G = G(P)$ e $D = D(P)$ rispettivamente il grafo e il grafo diretto ad esso associato. Allora $x \in \mathbb{R}^V$ appartiene a STAB(G) se e solo se esiste un vettore $y \in \mathbb{R}^{N(D)}$ tale che $y_s = 0$, $y_t = 1$, y non diminuisce lungo gli archi di D e $y_{v^+} - y_{v^-} = x_v$ per ogni $v \in V$.*

Dimostrazione. Supponiamo che esista un tale $y \in \mathbb{R}^{N(D)}$. Sia $v_1 \leq_P v_2 \leq_P \cdots \leq_P v_c$ una catena di P. Abbiamo allora

$$\begin{aligned}\sum_{i=1}^{c} x_v &= (y_{v_1^+} - y_{v_1^-}) + \cdots + (y_{v_c^+} - y_{v_c^-})\\ &\leq (y_{v_1^-} - y_s) + (y_{v_1^+} - y_{v_1^-}) + \cdots + (y_{v_c^+} - y_{v_c^-}) + (y_t - y_{v_c^+})\\ &= y_t - y_s = 1,\end{aligned}$$

dunque dal Lemma 2.2.2 segue che $x \in \mathrm{STAB}(G)$.

Sia invece $x \in \mathrm{STAB}(G)$. Per ogni vertice $v \in V$ poniamo y_{v^+} il massimo peso totale di una catena di P il cui massimo sia v, dove abbiamo assegnato al vertice w il peso x_w. Poniamo inoltre $y_{v^-} = y_{v^+} - x_v$, e infine $y_s = 0$ e $y_t = 1$. È immediato verificare che un tale y soddisfa la tesi. □

In conclusione abbiamo ottenuto che $H(P)$ è la soluzione del seguente problema di ottimizzazione convessa:

$$\begin{array}{rlllll} \min & -\dfrac{1}{n}\displaystyle\sum_{v\in V} \log x_v & & & & \\ \text{tale che} & x_v & = & y_{v^+} - y_{v^-} & \forall v \in V \\ & y_p & \leq & y_q & \forall (p,q) \in A(D) \\ & y_s & = & 0 & \\ & y_t & = & 1. & \end{array}$$

Capitolo 3

Tre algoritmi per ordinare con informazione parziale

3.1 Ordinamento con informazione parziale

Definizione. Sia $P = (V, \leq_P)$ un insieme parzialmente ordinato. Diciamo che un ordine totale $\leq$ è una *estensione lineare* di $\leq_P$ se $(\forall v_i, v_j \in V)\ v_i \leq_P v_j \Rightarrow v_i \leq v_j$. Denotiamo inoltre con $e(P)$ il numero di estensioni lineari di P.

Definizione. Sia $P = (V, \leq_P)$ un insieme parzialmente ordinato. Il *problema dell'ordinamento con informazione parziale* consiste nel determinare una estensione lineare $\leq$ fissata ma ignota per mezzo di domande del tipo "è vero che $v_i \leq v_j$?", detti *confronti.*

È evidente che siano necessari $\Omega(\log e(P))$ confronti: dobbiamo infatti discriminare fra $e(P)$ possibili risultati, e ogni confronto ci fornisce esattamente un bit di informazione. Servono quindi $\log e(P)$ bit, da cui il precedente limite inferiore.

Il problema dell'ordinamento con informazione parziale fu originariamente posto da Fredman nel 1976 [9]. Sempre Fredman dimostrò l'esistenza di un algoritmo che lo risolva compiendo $\log e(P) + 2n$ confronti, dove n è la cardinalità dell'insieme parzialmente ordinato. Osserviamo che, quando $e(P)$ cresce subesponenzialmente, tale quantità è dominata dall'addendo $2n$, dunque l'algoritmo non è asintoticamente ottimo nel numero di confronti. L'algoritmo proposto da Fredman è inoltre superpolinomiale nel numero di operazioni elementari. Nel 1984 Kahn e Saks dimostrarono l'esistenza di un algoritmo che richiede $O(\log e(P))$ confronti [13]. Essi mostrarono infatti che esiste sempre un confronto tale che le estensioni lineari per cui la risposta sia affermativa siano una parte compresa fra 3/11 e 8/11 del totale.[1]

La questione dell'esistenza di un algoritmo che compia $O(\log e(P))$ confronti e che richieda tempo polinomiale rimase aperta fino al 1995, quando un articolo di Kahn e Kim evidenziò il collegamento esistente fra l'entropia del grafo associato a P ed $e(P)$ [12]. Il seguente teorema afferma ad esempio che $nH(\overline{P}) = \Theta(\log e(P))$.

[1] Questo enunciato è un rilassamento della congettura $1/3 - 2/3$, indipendentemente posta da Kislitsyn nel 1968 [14], da Fredman nel 1975 e da Linial nel 1984 [19].

Teorema 3.1.1 (Kahn, Kim) *Sia P un insieme parzialmente ordinato di cardinalità n. Allora vale*

$$\log e(P) \leq nH(\overline{P}) \leq \min\left\{\log e(P) + \log e \cdot n,\ c \log e(P)\right\},$$

dove $c = 1 + 7 \log e \approx 11.1$.

L'Algoritmo di Kahn e Kim calcola inizialmente l'entropia del grafo associato a P. Successivamente stima la variazione dell'entropia dei grafi associati agli ordini parziali P', ottenuti da P aggiungendo il risultato di un confronto. Viene quindi selezionato quel confronto che avvicini maggiormente l'entropia a $\log n$, entropia del grafo completo, associato all'insieme totalmente ordinato. Ad ogni passo è dunque necessario il calcolo dell'entropia di un grafo, un problema di minimizzazione su un insieme convesso per l'equazione (1.5). Possiamo quindi applicare il metodo dell'ellissoide, ottenendo un algoritmo polinomiale ma non utile nella pratica.

In un recente articolo Cardinal et al. hanno proposto tre algoritmi che non richiedono il calcolo esatto dell'entropia, ma sfruttano la versione approssimata presentata nel Teorema 2.4.1 [4]. Questo consente di ottenere algoritmi che richiedono $O(\log e(P))$ confronti e che sono contemporaneamente polinomiali e pratici. In questo capitolo andremo ad esporre tali algoritmi.

Concludiamo questa sezione con una versione più precisa della stima superiore del Teorema 3.1.1, un risultato che sarà utile nel seguito.

Teorema 3.1.2 (Cardinal, Fiorini, Joret, Jungers, Munro) *Sia P un insieme parzialmente ordinato di cardinalità n. Allora vale*

$$nH\left(\overline{P}\right) \leq 2 \log e(P).$$

Dimostrazione. La dimostrazione procede per induzione su n, e, per n fissato, sul numero di elementi inconfrontabili di P. Essendo la tesi banalmente vera per $n = 1$ supponiamo $n \geq 2$. Sia $x \in \mathbb{R}_+^V$ un vettore che realizzi il minimo dell'entropia. Sia inoltre $\{(y_{v^-}, y_{v^+})\}_{v \in V}$ la corrispondente collezione di intervalli. Sia infine $a \in V$ tale che y_{a^+} sia massimo. Se a fosse confrontabile con tutti gli elementi di V avremmo per ipotesi induttiva che

$$nH\left(\overline{P}\right) = (n-1)H\left(\overline{P - a}\right) \leq 2 \log e(P - a) = 2 \log e(P).$$

Sia allora b non confrontabile con a e tale inoltre che y_{b^+} sia massimo. Per come abbiamo scelto a deve per forza valere $y_{b^+} \leq y_{a^+}$. In realtà vale l'uguaglianza: supponiamo infatti per assurdo che $y_{b^+} < y_{a^+}$, ed estendiamo a destra l'intervallo corrispondente a b di $y_{a^+} - y_{b^+}$. Questa nuova collezione di intervalli è ancora consistente con P, ma il punto $x' \in \mathbb{R}_+^V$ da essa definito realizzerebbe un valore dell'entropia più piccolo del minimo. Abbiamo infatti

$$-\frac{1}{n}\sum_{v \in V} \log x'_v = -\frac{1}{n}\sum_{v \in V} \log x_v + \frac{1}{n}\left(\log x_b - \log x'_b\right) < -\frac{1}{n}\sum_{v \in V} \log x_v,$$

contro l'ipotesi che x realizzi il minimo dell'entropia. A meno di scambiare a e b possiamo ora supporre che $x_a \geq x_b$. Il nostro obiettivo ora è definire due nuove famiglie di intervalli

$$\left\{\left(y^1_{v^-}, y^1_{v^+}\right)\right\}_{v \in V} \qquad \text{e} \qquad \left\{\left(y^2_{v^-}, y^2_{v^+}\right)\right\}_{v \in V}$$

tali che gli insiemi parzialmente ordinati P_1 e P_2 ad esse associati estendano P, e tali inoltre che le quantità $e(P_1)$ ed $e(P_2)$ varino in modo controllato. Per fare questo poniamo

$$\lambda = \frac{x_b}{x_a},$$

compreso fra 0 e 1 per come abbiamo scelto a e b. Poniamo inoltre

$$\alpha_1 = \begin{cases} \frac{1}{1-\lambda} & \text{se } \lambda \leq \frac{1}{2} \\ 2 & \text{altrimenti} \end{cases} \qquad \text{e} \qquad \beta_1 = \begin{cases} 1 & \text{se } \lambda \leq \frac{1}{2} \\ 2\lambda & \text{altrimenti} \end{cases}$$

e infine

$$\alpha_2 = \frac{2}{\lambda} \qquad \text{e} \qquad \beta_2 = 2.$$

Allora la famiglia di intervalli $\left\{\left(y^1_{v^-}, y^1_{v^+}\right)\right\}_{v\in V}$ coincide con $\{(y_{v^-}, y_{v^+})\}_{v\in V}$ tranne per

$$\begin{aligned} y^1_{a^+} &= y_{a^-} + \frac{x_a}{\alpha_1} \\ y^1_{b^-} &= y_{b^+} - \frac{x_b}{\beta_1}, \end{aligned}$$

e analogamente $\left\{\left(y^2_{v^-}, y^2_{v^+}\right)\right\}_{v\in V}$ coincide con $\{(y_{v^-}, y_{v^+})\}_{v\in V}$ eccetto per

$$\begin{aligned} y^2_{a^-} &= y_{a^+} - \frac{x_a}{\alpha_2} \\ y^2_{b^+} &= y_{b^-} + \frac{x_b}{\beta_2}. \end{aligned}$$

Siano rispettivamente P_1 e P_2 gli insiemi parzialmente ordinati definiti dalla prima e dalla seconda famiglia di intervalli. Allora esiste un indice $i \in \{1, 2\}$ tale che

$$\frac{e(P_i)}{e(P)} \leq \frac{1}{\sqrt{\alpha_i \beta_i}}. \tag{3.1}$$

Questo fatto verrà dimostrato in appendice. Assumendo che esista un tale i sia $x' \in \mathbb{R}^V_+$ il vettore definito dalla corrispondente famiglia di intervalli. Abbiamo allora che

$$H(P_i) \leq -\frac{1}{n}\sum_{v\in V} \log x'_v = -\frac{1}{n}\sum_{v\in V} \log x_v + \frac{1}{n}\log \alpha_i + \frac{1}{n}\log \beta_i,$$

dunque

$$nH(P_i) \leq nH(P) + \log \alpha_i \beta_i.$$

Possiamo ora concludere. Per il Teorema 2.2.1 e per la disuguaglianza appena dimostrata possiamo scrivere

$$\begin{aligned} nH(\overline{P}) &= n\log n - nH(P) \\ &\leq n\log n - nH(P_i) + \log \alpha_i \beta_i \\ &= nH(\overline{P_i}) + \log \alpha_i \beta_i, \end{aligned}$$

mentre per ipotesi induttiva e per la disuguaglianza (3.1) abbiamo

$$\begin{aligned} nH(\overline{P_i}) + \log \alpha_i\beta_i &\leq 2\log e(P_i) + \log \alpha_i\beta_i \\ &\leq 2\log \frac{e(P)}{\sqrt{\alpha_i\beta_i}} + \log \alpha_i\beta_i \\ &\leq 2\log e(P), \end{aligned}$$

cioè la tesi. □

3.2 Insertion sort

Algoritmo 1 "Insertion sort" con informazione parziale

```
1: // Preparazione
2: trova una catena C di lunghezza massima in P
3: // Ordinamento
4: while P − C ≠ ∅ do
5:     togli un elemento da P − C e inseriscilo in C con una ricerca binaria
6: end while
7: return C
```

Lemma 3.2.1 *Sia P un insieme parzialmente ordinato di cardinalità n e sia C una catena di lunghezza massima in P. Vale allora $|C| \geq n \cdot 2^{-H(\overline{P})}$.*

Dimostrazione. È noto che l'entropia di un grafo su n vertici e dimensione massima di un insieme indipendente α è maggiore o uguale a $-\log \frac{\alpha}{n}$ [2]. La tesi segue applicando questo fatto a $G = \overline{G}(P)$. □

Teorema 3.2.2 *Sia P un insieme parzialmente ordinato di cardinalità n. Allora l'Algoritmo 1 risolve il problema dell'ordinamento con informazione parziale in $O(\log n \cdot \log e(P))$ confronti.*

Dimostrazione. Sia $g(P)$ il numero di confronti necessario per ordinare P. È chiaro che

$$g(P) \leq \log n \cdot (n - |C|),$$

inoltre per il Lemma 3.2.1

$$g(P) \leq \log n \cdot (n - 2^{-H(\overline{P})}n).$$

Usando l'ovvia disuguaglianza $1 - 2^{-x} \leq \ln 2 \cdot x$ deduciamo

$$g(P) \leq \log n \cdot \ln 2 \cdot nH(\overline{P}),$$

e applicando il Teorema 3.1.2 abbiamo

$$g(P) = O(\log n \cdot \log e(P)),$$

cioè la tesi. □

3.3 Merge sort naïve

Algoritmo 2 "Merge sort naïve" con informazione parziale

```
// Preparazione
trova una decomposizione golosa di P in catene C_1, ..., C_k
C ← {C_1, ..., C_k}
// Ordinamento
while |C| > 1 do
    seleziona da C due catene di lunghezza minima C e C'
    fondi C e C' in tempo lineare, ottenendo C''
    cancella C e C' da C, aggiungi C''
end while
return l'unica catena di C
```

Lemma 3.3.1 *Sia P un insieme parzialmente ordinato di cardinalità n. Allora l'Algoritmo 2 risolve il problema dell'ordinamento parziale compiendo al più $(\tilde{h}+1)n$ confronti, dove $\tilde{h}$ è l'entropia di Shannon della probabilità discreta $\left\{\frac{|C_1|}{n}, \ldots, \frac{|C_k|}{n}\right\}$.*

Dimostrazione. Per fondere due catene useremo l'ovvio algoritmo lineare che a ogni passo rimuove e copia nell'output l'elemento minore fra i minimi delle catene. Nel caso peggiore tale algoritmo richiede tanti confronti quanti sono gli elementi della catena di lunghezza maggiore. La sequenza di fusioni delle catene forma un albero, detto di Huffman. È noto che l'altezza media di tale albero è maggiorata da $\tilde{h}+1$ [7]. Denotata con t_i l'altezza della catena C_i, le precedenti osservazioni permettono di stimare il numero di confronti con

$$\sum_{i=1}^{k} t_i|C_i| = n\sum_{i=1}^{k} t_i\frac{|C_i|}{n} \leq n(\tilde{h}+1),$$

cioè la tesi. □

Teorema 3.3.2 *Sia P un insieme parzialmente ordinato di cardinalità n. Allora, per ogni $\varepsilon > 0$, l'Algoritmo 2 risolve il problema dell'ordinamento parziale impiegando al più $(1+\varepsilon)\log e(P) + (1+\varepsilon)\left(\log e + \log\left(1+\frac{1}{\varepsilon}\right)+1\right)\cdot n$ confronti.*

Dimostrazione. Sia $g(P)$ il numero di confronti richiesto per ordinare P. Grazie al Lemma 3.3.1 otteniamo

$$g(P) \leq n(\tilde{h}+1),$$

inoltre abbiamo

$$\begin{aligned} g(P) &\leq (1+\varepsilon)nH(\overline{P}) + (1+\varepsilon)n\log\left(1+\frac{1}{\varepsilon}\right) + n \\ &\leq (1+\varepsilon)(\log e(P) + \log e\cdot n) + (1+\varepsilon)n\log\left(1+\frac{1}{\varepsilon}\right) + n \\ &= (1+\varepsilon)\log e(P) + (1+\varepsilon)\left(\log e + \log\left(1+\frac{1}{\varepsilon}\right)+1\right)\cdot n, \end{aligned}$$

dove abbiamo applicato i Teoremi 2.4.1 e 3.1.1. □

3.4 Merge con informazione parziale

Scopo di questa sezione è risolvere il problema dell'ordinamento con informazione parziale per un insieme partizionabile in due catene disgiunte. Tale caso particolare è noto come *problema della fusione con informazione parziale* e fu studiato da Linial nel 1984 [19]. Egli propose un algoritmo che richiede al più $C \log e(P)$ confronti con $C = \left(\log \frac{1+\sqrt{5}}{2}\right)^{-1}$ e dimostrò l'ottimalità di tale costante. Il suo algoritmo necessita però del calcolo di n determinanti di matrici $n \times n$. In questa sezione proporremo un algoritmo basato sull'entropia di grafo che risolva tale problema, e dimostreremo che richiede al più $6 \log e(P)$ confronti. È possibile inoltre dimostrare che questo algoritmo necessita di $O(n^2 \log n)$ operazioni elementari.

Il seguente risultato descrive la struttura dei grafi associati agli insiemi parzialmente ordinati partizionabili in due catene disgiunte. La dimostrazione è immediata e quindi omessa.

Lemma 3.4.1 *Sia P un insieme parzialmente ordinato ricoperto da due catene disgiunte A e B e sia $G = \overline{G}(P)$ il grafo ad esso associato. Allora*

(i) *G è bipartito.*

(ii) *G è biconvesso, cioè i vertici adiacenti ad un vertice di una catena formano un intervallo nella catena opposta.*

(iii) *Siano u e v vertici appartenenti alla stessa catena, che supporremo senza perdita di generalità essere A, tali che $u \leq_P v$. Siano $[c_u, d_u]$ e $[c_v, d_v]$ gli intervalli dei vertici B adiacenti rispettivamente a u e v. Allora $c_u \leq_P c_v$ e $d_u \leq_P d_v$, e in particolare se $u \leq_P w \leq_P v$ i vertici adiacenti a w sono un intervallo contenuto in $[c_u, d_v]$.*

Nel caso di grafi bipartiti è possibile calcolare esattamente l'entropia. Vale infatti il seguente teorema dovuto a Körner e Marton, di cui omettiamo la dimostrazione [17].

Teorema 3.4.2 (Körner, Marton) *Sia G un grafo bipartito di ordine n, di bipartizione A e B. Allora possiamo trovare partizioni*

$$A = A_1 \cup \cdots \cup A_k \qquad e \qquad B = B_1 \cup \cdots \cup B_k$$

tali che

$$H(G) = \sum_{i=1}^{k} \frac{|A_i| + |B_i|}{n} h\left(\frac{|A_i|}{|A_i| + |B_i|}\right),$$

dove $h(x) = -x \log x - (1-x) \log (1-x)$ è la funzione di entropia binaria, nella quale per convenzione poniamo $h(0) = h(1) = 0$.

Possiamo costruire tali partizioni iterativamente. Dato $S \subset V$ sottoinsieme dei vertici, denotiamo con $N_G(S)$ l'insieme dei vertici adiacenti ad almeno un vertice di S. Sia allora A_i un sottoinsieme di $A' = A - (A_1 \cup \cdots \cup A_{i-1})$ che renda massimo

$$\frac{|A_i|}{|N_{G'}(A_i)|},$$

dove G' è ottenuto da G rimuovendo tutti i vertici contenuti in un qualche A_j o B_j per $j < i$, e $B_i = N_{G'}(A_i)$. Se esiste un vertice isolato u in A' allora poniamo $A_i = \{u\}$ e $B_i = \emptyset$. Se invece A' è vuoto e $B' = B - (B_1 \cup \cdots \cup B_{i-1})$ non lo è allora scegliamo $v \in B'$ e poniamo $A_i = \emptyset$ e $B_i = \{v\}$.

Al fine di evitare casi degeneri denotiamo nel seguito con $\mathrm{STAB}^*(G)$ i punti di $\mathrm{STAB}(G)$ di coordinate tutte positive. Se infatti così non fosse la funzione obiettivo dell'equazione (1.5) non sarebbe definita.

Nel caso di grafi bipartiti possiamo semplificare il Lemma 2.2.2. Abbiamo allora che

$$\mathrm{STAB}(G) = \left\{x \in \mathbb{R}^V : \quad x_u + x_v \leq 1 \text{ per ogni } uv \in E, \quad 0 \leq x_v \leq 1 \text{ per ogni } v \in V\right\}.$$

Definizione. Diciamo che un arco uv è *stretto* rispetto ad $x \in \mathrm{STAB}(G)$ se vale $x_u + x_v = 1$. Denotiamo con $G(x)$ il grafo i cui vertici siano gli stessi di G e i cui archi siano gli archi di G stretti rispetto ad x.

Definizione. Siano uv e $u'v'$ archi di G tali che $u, u' \in A$ e $v, v' \in B$. Diciamo che *si incrociano* se $u <_P u'$ e $v' <_P v$ oppure se $u' <_P u$ e $v <_P v'$.

Lemma 3.4.3 *Sia P un insieme parzialmente ordinato ricoperto da due catene disgiunte A e B e sia $G = \overline{G}(P)$ il grafo ad esso associato. Sia x un punto di $STAB(G)$ e siano uv e $u'v'$ archi stretti rispetto ad x tali inoltre che $u, u' \in A$ e $v, v' \in B$. Se uv e $u'v'$ si incrociano allora sia $u'v$ sia uv' sono archi di G, entrambi stretti rispetto ad x.*

Dimostrazione. Dal Lemma 3.4.1 (*iii*) segue che $u'v$ e uv' sono archi di G. Supponiamo per assurdo che uv' non sia stretto. Avremmo allora:

$$\begin{aligned} x_v &= 1 - x_u && \text{(poiché } uv \text{ è stretto)} \\ &> x_{v'} && \text{(poiché } uv' \text{ non è stretto)} \\ &= 1 - x_{u'} && \text{(poiché } u'v' \text{ è stretto)} \\ &\geq x_v, && \text{(poiché } u'v \text{ è un arco e per il Lemma 2.2.2)} \end{aligned}$$

chiaramente un assurdo. Possiamo procedere analogamente per $u'v$, da cui la tesi. □

Definizione. Diciamo che $x \in \mathrm{STAB}^*(G)$ è *localmente ottimo* se per ogni componente connessa K di $G(x)$ valgono

$$x_u = \frac{|A \cap K|}{|K|} \quad \text{per ogni } u \in A \cap K \qquad \text{e} \qquad x_v = \frac{|B \cap K|}{|K|} \quad \text{per ogni } v \in B \cap K.$$

Diciamo che K è *bilanciata* se per essa valgono le precedenti condizioni di ottimalità, *sbilanciata* altrimenti.

Definizione. Sia $x \in \mathrm{STAB}^*(G)$. Una componente connessa K di $G(x)$ è detta banale se consiste di un unico vertice, *non banale* altrimenti. Inoltre chiamiamo *libera* una componente che sia banale e sbilanciata.

Definizione. Sia $x \in \mathrm{STAB}^*(G)$. Una componente connessa L di $G(x)$ è detta *incastonata* in un'altra componente connessa K se esistono un vertice $w \in L$ e due vertici $u, u'' \in K$ tutti appartenenti ad un'unica catena e tali inoltre che $u \leq_P w \leq_P u''$.

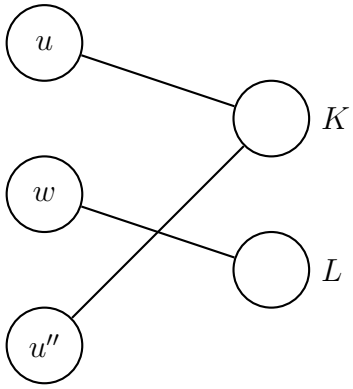

Figura 3.1: Esempio di componenti connesse incastonate.

Esempio. Nella figura 3.1 la componente connessa L è incastonata in K.

La relazione d'ordine $\leq_P$ induce una relazione d'ordine sui sottoinsiemi dell'insieme di sostegno. Diciamo che $S \leq_P T$ se S e T sono sottoinsiemi tali che $u \leq_P v$ per ogni $u \in S$ e per ogni $v \in T$.

Lemma 3.4.4 *Sia P un insieme parzialmente ordinato ricoperto da due catene disgiunte A e B e sia $G = \overline{G}(P)$ il grafo ad esso associato. Dato $x \in STAB^*(G)$ allora:*

(i) Se in $G(x)$ una componente connessa L è incastonata in K allora L è libera.

(ii) Se K ed L sono componenti connesse non banali di $G(x)$ allora vale $K \leq_P L$ oppure $L \leq_P K$.

Dimostrazione. (i) Supponiamo per assurdo che L non sia libera ma sia incastonata in K. Siano allora $w \in L$ e $u, u'' \in K$ come nella definizione. Senza perdita di generalità possiamo assumere $w, u', u'' \in A$ e che $u' \notin K$ se $u \leq_P u' \leq_P u''$. K è una componente connessa che contiene sia u sia u'', dunque esiste $v \in K \cap B$ adiacente ad entrambi. Per il Lemma 3.4.1 (*iii*) abbiamo che vw è un arco di G, ma poiché K ed L sono componenti connesse distinte vw non è un arco di $G(x)$. Se L non fosse banale allora esisterebbe un arco di L incidente in w che incrocia uv o $u''v$, dunque per il Lemma 3.4.3 avremmo che vw è stretto, assurdo. Se invece L fosse bilanciata avremmo $x_w + x_v = 1 + x_v > 1$, contro il Lemma 2.2.2.

(ii) Supponiamo per assurdo che siano false entrambe le disuguaglianze. Per il punto precedente K e L, essendo non banali, non possono essere incastonate l'una nell'altra. Senza perdita di generalità possiamo allora assumere che valgano

$$K \cap A \leq_P L \cap A \qquad \text{e} \qquad L \cap B \leq_P K \cap B.$$

Siano quindi $u, u' \in A$ e $v, v' \in B$ tali che uv e $u'v'$ siano archi di G. Dunque tali archi si incrociano, perciò in particolare uv' è un arco di G. Ma allora K ed L sarebbero la stessa componente connessa, una contraddizione. □

Definizione. Sia $x \in \mathrm{STAB}^*(G)$ e sia K una componente connessa di $G(x)$. Chiamiamo *scarto* di K il reale σ che renda minimo

$$\operatorname*{argmin}_{\sigma} \max_{v \in A \cap K} \left| x_v + \sigma - \frac{|A \cap K|}{|K|} \right|,$$

e tale inoltre che, se

$$x'_v = \begin{cases} x_v + \sigma \text{ per } v \in A \cap K \\ x_v - \sigma \text{ per } v \in B \cap K, \end{cases}$$

allora $x' \in \mathrm{STAB}^*(G)$.

Algoritmo 3 Ribilanciamento

1: **while** esiste una componente sbilanciata K di $G(x)$ **do**
2: calcola lo scarto σ di K
3: poni $x_v := x_v + \sigma$ per $v \in A \cap K$, $x_v := x_v - \sigma$ per $v \in B \cap K$
4: **end while**

Denotiamo con x' il vettore x dopo una esecuzione del corpo del while. Ad ogni esecuzione accade che o una componente risulta bilanciata o un arco di G fra componenti connesse distinte diviene stretto. In entrambi i casi il numero di componenti sbilanciate diminuisce, perciò l'algoritmo termina.

Nel secondo caso diciamo che K *si fonde* con un'altra componente connessa di $G(x')$. Possiamo dare una condizione necessaria affinché ciò accada in termini della seguente definizione.

Definizione. Sia $x \in \mathrm{STAB}(G)$. Diciamo che due componenti connesse K ed L di $G(x)$ *sono a contatto* se sono collegate da un arco di G e se non esiste una componente connessa non banale M diversa da K ed L tale che un arco uv di M sia compreso fra K ed L. In altri termini, supposto $u \in A$ e $v \in B$, non devono valere $K \cap A \leq_P \{u\} \leq_P L \cap A$ e $K \cap B \leq_P \{v\} \leq_P L \cap B$, oppure $L \cap A \leq_P \{u\} \leq_P K \cap A$ e $L \cap B \leq_P \{v\} \leq_P K \cap B$.

Esempio. Nella figura 3.2 le componenti K ed L sono a contatto. Invece le componenti K' e L' non lo sono per via della componente M e dell'arco uv.

Lemma 3.4.5 *Sia $x \in STAB^*(G)$, e sia x' il punto ottenuto ribilanciando x. Se la componente connessa K si fonde con la componente connessa L di $G(x)$ allora K ed L sono a contatto.*

Dimostrazione. Supponiamo per assurdo che K ed L non siano a contatto. Siano allora M una componente connessa e uv un suo arco come nella precedente definizione. Sia inoltre $u'v'$ l'arco di estremi rispettivamente in K ed L che sia diventato stretto in $G(x')$. Gli archi uv e $u'v'$ si incrociano: possiamo allora applicare il Lemma 3.4.3 e concludere che uv' è un arco di $G(x')$, quindi M ed L sono contenuti in una stessa componente connessa di $G(x')$. Ma l'Algoritmo 3 non altera i pesi dei vertici non appartenenti a K, pertanto M ed L erano contenuti in una stessa componente connessa di $G(x)$, una contraddizione. □

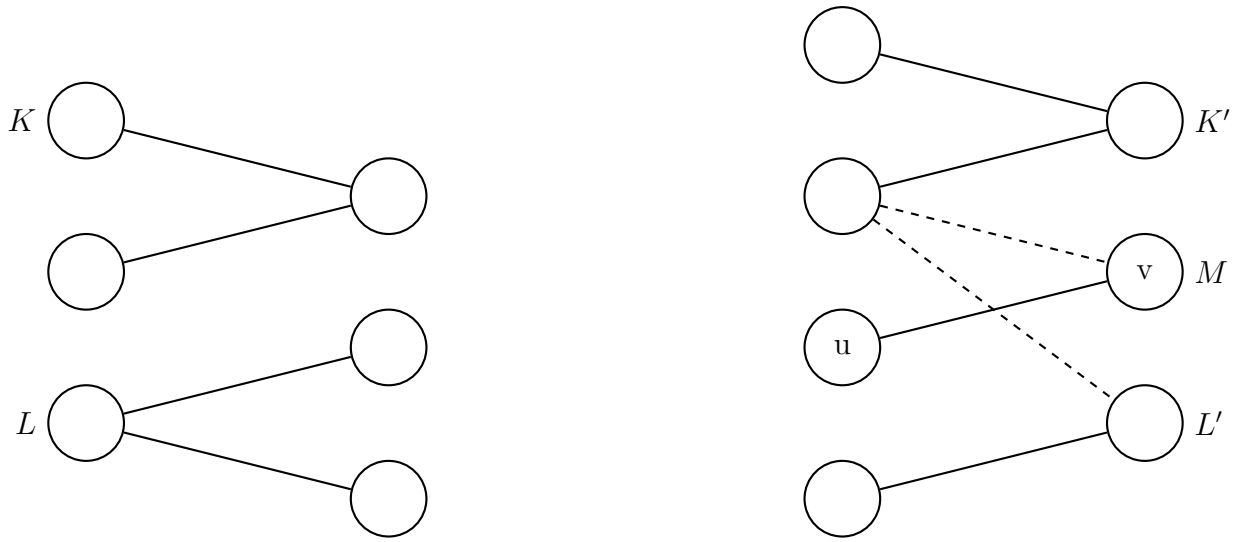

Figura 3.2: Esempio di componenti connesse a contatto e non a contatto.

Definizione. Sia $x \in \mathrm{STAB}^*(G)$. Diciamo che una componente connessa K di $G(x)$ è *rossa* se si ha $|A \cap K| \geq |B \cap K|$, altrimenti diciamo che K è *blu*.

Definizione. Sia $x \in \mathrm{STAB}^*(G)$. Diciamo che *rispetta i colori* se per ogni componente connessa K di $G(x)$ e per ogni scelta di $u \in A \cap K$ e $v \in B \cap K$ abbiamo

$$\begin{aligned} x_u \geq \frac{1}{2} \quad \text{e} \quad x_v \leq \frac{1}{2} \qquad & \text{se } K \text{ è rossa,} \\ x_u < \frac{1}{2} \quad \text{e} \quad x_v > \frac{1}{2} \qquad & \text{se } K \text{ è blu.} \end{aligned}$$

Osserviamo che la precedente definizione è un rilassamento della definizione di componente connessa bilanciata.

Lemma 3.4.6 *Sia $x \in STAB^*(G)$ e sia x' ottenuto ribilanciando x. Sia poi K una componente connessa di $G(x)$. Allora se x rispetta i colori anche x' rispetta i colori. Inoltre non è possibile che K si sia fusa con componenti connesse di colore diverso, perciò la componente di $G(x')$ contenente K è dello stesso colore.*

Dimostrazione. Poiché x rispetta i colori lo scarto σ è non negativo, dunque per $v \in A \cap K$ abbiamo $x'_v \geq \frac{1}{2}$ se e solo se $x_v \geq \frac{1}{2}$, mentre per $v \in B \cap K$ abbiamo $x'_v > \frac{1}{2}$ se e solo se $x_v > \frac{1}{2}$. Pertanto se K non si fonde con altre componenti connesse allora anche x' rispetta i colori. Supponiamo invece che K si fonda con una componente L, e sia vw un arco di $G(x)$ che sia diventato stretto rispetto a x', e tale inoltre che $v \in K$ e $w \in L$. Da $x'_w = x_w$ e $x_v + x_w < x'_v + x'_w = 1$ segue che $x'_v > x_v$. Abbiamo quattro casi:

- $v \in A$ e K è rossa in $G(x)$. Allora $x'_v > x_v \geq \frac{1}{2}$ e $x_w = x'_w < \frac{1}{2}$.
- $v \in A$ e K è blu in $G(x)$. Allora $x_v < \frac{1}{2}$, quindi $x'_v < \frac{1}{2}$ e $x_w = x'_w > \frac{1}{2}$.
- $v \in B$ e K è rossa in $G(x)$. Allora $x_v \leq \frac{1}{2}$, quindi $x'_v \leq \frac{1}{2}$ e $x_w = x'_w \geq \frac{1}{2}$.

- $v \in B$ e K è blu in $G(x)$. Allora $x'_v > x_v > \frac{1}{2}$ e $x'_w = x_w < \frac{1}{2}$.

In tutti i casi concludiamo che L abbia lo stesso colore di K, quindi a maggior ragione la componente connessa di $G(x')$ contenente K avrà lo stesso colore di K. □

Osservazione. Ad ogni passo dell'algoritmo l'entropia di x' è minore dell'entropia di x. Infatti la funzione

$$\xi \mapsto \frac{|A \cap K|}{|K|} \log \xi + \frac{|B \cap K|}{|K|} \log (1 - \xi)$$

è convessa sull'intervallo $(0, 1)$ con minimo in $\xi = \frac{|A \cap K|}{|K|}$.

La fusione di catene verrà effettuata con l'Algoritmo di Hwang-Lin. Date due catene X e Y con $|X| \geq |Y|$, tale algoritmo divide la catena maggiore in blocchi di grandezza $2^{\lfloor \log \frac{|X|}{|Y|} \rfloor}$. Ogni vertice di Y è inserito eseguendo prima una scansione lineare fra i blocchi e poi per bisezione all'interno del blocco. Osserviamo che, poiché gli elementi di Y sono ordinati, una volta scartato un blocco questo non dovrà essere considerato per i successivi elementi. Avremo bisogno della seguente stima sul numero di confronti effettuato da tale algoritmo.

Lemma 3.4.7 *Siano X e Y due catene disgiunte. Supponiamo che $|X| \geq |Y|$. Allora il numero di confronti richiesto dall'Algoritmo di Hwang-Lin è maggiorato da $|Y| \log(\frac{4|X|}{|Y|})$.*

Dimostrazione. È noto che l'Algoritmo di Hwang-Lin compie al più

$$|Y| \left(1 + \left\lfloor \log \frac{X}{Y} \right\rfloor\right) + \left\lfloor \frac{|X|}{2^{\lfloor \log \frac{|X|}{|Y|} \rfloor}} \right\rfloor - 1$$

confronti [11]. Sia allora $\xi \in [0, 1)$ tale che

$$\left\lfloor \log \frac{|X|}{|Y|} \right\rfloor = \log \frac{|X|}{|Y|} - \xi.$$

È facile verificare che per $\xi \in [0, 1)$ vale la disuguaglianza

$$1 - \xi + 2^{\xi} \leq 2.$$

Semplici passaggi algebrici danno

$$\frac{|X|}{2^{\lfloor \log \frac{|X|}{|Y|} \rfloor}} = \frac{|X|}{2^{\log \frac{|X|}{|Y|} - \xi}} = \frac{|X|}{2^{\log \frac{|X|}{|Y|}}} \cdot 2^{\xi} = |Y| \cdot 2^{\xi}.$$

Possiamo infine mettere insieme le precedenti due equazioni per ottenere

$$\begin{aligned} |Y| \left(1 + \left\lfloor \log \frac{X}{Y} \right\rfloor\right) + \left\lfloor \frac{|X|}{2^{\lfloor \log \frac{|X|}{|Y|} \rfloor}} \right\rfloor - 1 &\leq |Y| \left(1 - \xi + \log \frac{|X|}{|Y|} + 2^{\xi}\right) \\ &\leq |Y| \left(\log \frac{|X|}{|Y|} + 2\right) \\ &= |Y| \left(\log \frac{4|X|}{|Y|}\right), \end{aligned}$$

cioè la tesi. □

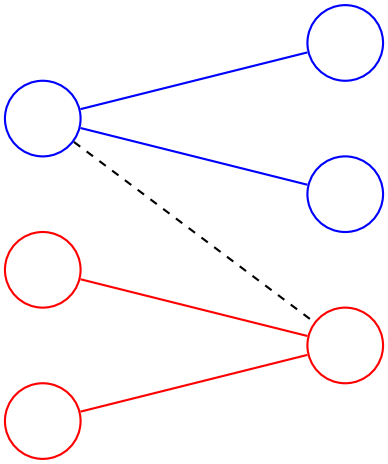

Figura 3.3: Un esempio di componente connessa buona.

Definizione. Sia K una componente connessa di $G(x)$. Se K è rossa chiamiamo $A \cap K$ *catena maggiore* e $B \cap K$ *catena minore*. Se K è blu il contrario.

Definizione. Sia K una componente connessa di $G(x)$. Diciamo che K è *buona* se ogni arco di G che possiede un estremo nella catena minore di K ha l'altro estremo nella catena maggiore oppure in una componente connessa di colore opposto.

Esempio. Nella figura 3.3 sia la componente rossa sia la componente blu sono buone.

Lemma 3.4.8 *Sia $x \in STAB(G)$ localmente ottimo. Se $G(x)$ possiede almeno una componente rossa non banale allora una di esse è buona.*

Dimostrazione. Sia K una componente connessa rossa non banale tale che $\frac{|A \cap K|}{|K|}$ sia minimo. Vogliamo dimostrare che K è buona. Sia $v \in B \cap K$ e sia w adiacente a v in G ma non in $G(x)$. Per definizione l'arco di estremi v e w non è stretto, quindi $x_v + x_w < 1$. In particolare $x_w < 1$, quindi w appartiene ad una qualche componente connessa L non banale. Se per assurdo L fosse rossa per ipotesi $\frac{|A \cap L|}{|L|} \geq \frac{|A \cap K|}{|K|}$, dunque per ottimalità di x avremmo

$$x_v + x_w = \frac{|B \cap K|}{|K|} + \frac{|A \cap L|}{|L|} \geq \frac{|B \cap K|}{|K|} + \frac{|A \cap K|}{|K|} \geq 1$$

da cui dedurremmo che l'arco di estremi v e w è stretto, una contraddizione. Segue quindi che L è blu oppure non esiste w adiacente a v in G ma non in $G(x)$, cioè la tesi. □

È immediato osservare che la precedente dimostrazione si applica, *mutatis mutandis*, all'insieme delle componenti blu non banali. Pertanto in analoghe ipotesi esiste una componente blu che sia buona.

Lemma 3.4.9 *Denotiamo con G e x rispettivamente il grafo e il punto all'inizio di un ciclo while. Denotiamo invece con G' e x' il grafo e il punto appena eseguita la riga 6 dell'Algoritmo 4. Allora x' appartiene a $STAB(G')$ e rispetta i colori.*

Dimostrazione. Osserviamo che se $v \in K$ allora $\{v\}$ è una componente libera di $G'(x)$. Il nostro obiettivo è dimostrare che questo è vero anche in $G'(x')$. Possiamo distinguere quattro casi:

Algoritmo 4 Parte essenziale del "merge" con informazione parziale

```
1: while G(x) possiede una componente connessa non banale do
2:    scegli una componente buona K, con precedenza a quelle rosse
3:    fondi le catene X = A ∩ K e Y = B ∩ K con l'Algoritmo di Hwang-Lin
4:    for v ∈ K do
5:       x_v = max{x_v, 1/2} se v ∈ A, x_v = max{x_v, 1/2 + 1/2n} se v ∈ B
6:    end for
7:    ribilancia x con l'Algoritmo 3
8:    for v ∈ K do
9:       if x_v = 1 then
10:         copia v alla sua posizione finale in C
11:      end if
12:   end for
13: end while
14: return C
```

- $v \in A$ e K è rossa. Allora $x_v \geq \frac{1}{2}$, quindi $x'_v = x_v$. Pertanto $\{v\}$ rimane libera in $G'(x')$.

- $v \in B$ e K è rossa. Allora $x_v \leq \frac{1}{2}$, quindi $x'_v = \frac{1}{2} + \frac{1}{2n}$. Poiché K è buona un vertice w che sia adiacente a v in G' apparteneva ad una componente connessa blu di $G(x)$, dunque $x_w < \frac{1}{2}$. Essendo x localmente ottimo deve essere $x_w \leq \frac{1}{2} - \frac{1}{n}$, e da questo segue

$$\begin{aligned} x'_v + x'_w &= (\frac{1}{2} + \frac{1}{2n}) + x_w \\ &\leq (\frac{1}{2} + \frac{1}{2n}) + (\frac{1}{2} - \frac{1}{n}) \\ &\leq 1, \end{aligned}$$

  cioè vw non è stretto in $G'(x')$, quindi $\{v\}$ è libera in $G'(x')$.

- $v \in B$ e K è blu. Allora $x_v > \frac{1}{2}$, quindi, poiché x è localmente ottimo, $x'_v = x_v \geq \frac{1}{2} + \frac{1}{n}$.

- $v \in A$ e K è blu. Allora $x_v < \frac{1}{2}$, quindi $x'_v = \frac{1}{2}$. Se stiamo considerando una componente blu allora tutte le componenti rosse di $G(x)$ sono banali. Inoltre, poiché x è localmente ottimo, nessuna di esse può essere libera. Essendo K buona v è adiacente in G soltanto a vertici della catena opposta, dunque a nessun vertice in G'. Di conseguenza $\{v\}$ è libera in $G'(x')$.

In tutti i casi abbiamo ottenuto che $\{v\}$ è libera in $G'(x')$, dunque la tesi. □

Lemma 3.4.10 *Sia P un insieme parzialmente ordinato ricoperto da due catene disgiunte A e B e sia $G = \overline{G}(P)$ il grafo ad esso associato. Supponiamo che $x \in STAB(G)$ sia localmente ottimo e tale che il contributo ad $H(x)$ delle componenti rosse superi quello delle componenti blu. Allora l'Algoritmo 4 fonde A e B impiegando al più $3nH(x)$ confronti.*

Dimostrazione. Sia k il numero di esecuzioni del corpo del while, e fissiamo j compreso fra 1 e k. Siano poi G_j e x' rispettivamente il grafo e il punto all'inizio della $j-$esima esecuzione, e siano x'' il punto dopo la riga 6 e x''' alla fine dell'esecuzione. Siano inoltre K la componente buona scelta a tale passo, s_j e t_j rispettivamente il numero di elementi nella catena minore e maggiore. Denotiamo infine con r_j il numero di vertici nella catena minore di qualche componente rossa in $G_j(x')$, e definiamo $\phi_j = nH(x')+r_j$, dove abbiamo posto $r_{k+1} = 0$ e $\phi_{k+1} = 0$.

Supponiamo che K sia rossa in $G_j(x')$ e sia v un vertice della catena minore di K. Il contributo a $H(x')$ corrispondente a tale vertice è $x'_v = -\frac{1}{n}\log\frac{s_j}{s_j+t_j}$. Dopo la riga 6, poiché x''_v è diventato almeno $\frac{1}{2}$, il contributo a $H(x'')$ è al più $x''_v = -\frac{1}{n}\log\frac{1}{2} = \frac{1}{n}$. Pertanto, sommando su tutti i vertici della catena minore di K, abbiamo:

$$H(x') - H(x'') \geq -\frac{s_j}{n}\log\frac{s_j}{s_j+t_j} - \frac{s_j}{n} = \frac{s_j}{n}\log\frac{s_j+t_j}{s_j} - \frac{s_j}{n}.$$

Abbiamo in precedenza osservato che il ribilanciamento diminuisce l'entropia, dunque vale la stima

$$H(x') - H(x''') \geq \frac{s_j}{n}\log\frac{s_j+t_j}{s_j} - \frac{s_j}{n}. \tag{3.2}$$

Osserviamo inoltre che vale

$$r_j - r_{j+1} \geq s_j, \tag{3.3}$$

poiché dopo il passo j non avremo più almeno i vertici rossi nella catena minore di K.

Supponiamo invece che K sia blu in $G_j(x')$. Allora non possono esistere componenti connesse rosse, altrimenti una di esse sarebbe buona per il Lemma 3.4.8 e l'algoritmo avrebbe dato precedenza a questa. Poiché K è buona non può accadere che K si fonda con un'altra componente blu, pertanto durante il ribilanciamento ogni componente x_v corrispondente a un elemento della catena minore è diventato almeno 1. Il contributo a $H(x''')$ diventa quindi nullo e, sommando su tutti i vertici della catena minore di K, abbiamo

$$H(x') - H(x''') \geq \frac{s_j}{n}\log\frac{s_j+t_j}{s_j}. \tag{3.4}$$

Infine, poiché non ci sono più componenti rosse, deve valere

$$r_j - r_{j+1} = 0. \tag{3.5}$$

Dalle equazioni (3.2), (3.3) nel caso di componente rossa e (3.4), (3.5) nel caso di componente blu otteniamo

$$\phi_j - \phi_{j+1} \geq s_j \log\frac{s_j+t_j}{s_j}.$$

Possiamo dunque scrivere la somma telescopica

$$\phi_1 = \sum_{j=i}^{k}(\phi_j - \phi_{j+1}) \geq \sum_{j=1}^{k} s_j \log\frac{s_j+t_j}{s_j} \geq \sum_{j=1}^{k}\frac{1}{2}s_j\log\frac{4t_j}{s_j},$$

dove nell'ultimo passaggio abbiamo sfruttato l'ovvia disuguaglianza

$$\left(\frac{s_j+t_j}{s_j}\right)^2 = \frac{(s_j-t_j)^2}{s_j^2} + \frac{4t_j s_j}{s_j} \geq \frac{4t_j}{s_j}.$$

Sia $g(P)$ il numero di confronti necessario per ordinare P. Dal Lemma 3.4.7 otteniamo

$$\frac{1}{2}\sum_{j=1}^{k} s_j \log \frac{4t_j}{s_j} \geq \frac{1}{2} g(P),$$

da cui segue

$$g(P) \leq 2\phi_1 = 2nH(x) + 2r_1. \tag{3.6}$$

Osserviamo che r_1 è uguale a $nH(\tilde{x})$, dove $\tilde{x}$ è definito ponendo $x_v = \frac{1}{2}$ per i v nella catena minore di qualche componente rossa di G ed 1 altrimenti. L'entropia di $H(\tilde{x})$ è maggiorata dal contributo delle componenti rosse all'entropia del grafo, dunque da $\frac{H(x)}{2}$ per ipotesi. Perciò abbiamo

$$g(P) \leq 2nH(x) + 2r_1 = 2nH(x) + 2nH(\tilde{x}) \leq 3nH(x),$$

cioè la tesi. □

Algoritmo 5 "Merge" con informazione parziale

1: calcola $x \in \mathrm{STAB}(G)$ che realizzi il minimo dell'entropia con l'Algoritmo di Körner e Marton
2: **if** il contributo delle componenti rosse ad $H(x)$ supera quelle delle componenti blu **then**
3: scambia le catene A e B
4: **end if**
5: **for** $v \in A \cup B$ **do**
6: **if** v è un taglio **then**
7: copia v alla sua posizione finale in C
8: **end if**
9: **end for**
10: invoca l'Algoritmo 4
11: **return** C

Teorema 3.4.11 *Sia P un insieme parzialmente ordinato ricoperto da due catene disgiunte A e B e sia $G = \overline{G}(P)$ il grafo ad esso associato. Allora l'Algoritmo 5 fonde A e B impiegando al più $6 \log e(P)$ confronti.*

Dimostrazione. Sia $g(P)$ il numero di confronti necessario a fondere A e B. Il calcolo dell'entropia alla riga 1 non comporta alcun confronto. Nelle righe da 2 a 9 ci assicuriamo di essere nelle condizioni di applicare il Lemma 3.4.10, e poter quindi dedurre che, eseguendo la riga 10, vengano compiuti al più $3nH(x)$ confronti. Per il Teorema 3.1.2 abbiamo allora

$$g(P) \leq 3nH(x) \leq 6 \log e(P),$$

cioè la tesi. □

3.5 Merge sort

Abbiamo definito "naïve" l'Algoritmo 2 poiché, ogni volta che effettua una fusione fra due catene, non fa uso dell'informazione contenuta nell'insieme parzialmente ordinato. In questa sezione dimostreremo che è sufficiente compiere con cautela solo l'ultima di queste fusioni per ottenere un algoritmo che ordini compiendo $O(\log e(P))$ confronti. Per fare questo sfrutteremo l'Algoritmo 5 della precedente sezione.

Algoritmo 6 "Merge sort" con informazione parziale

1: trova una catena A di lunghezza massima in P
2: applica l'Algoritmo 2 a $P - A$, ottenendo una catena B
3: applica l'Algoritmo 5 all'ordine parziale corrente P'
4: **return** la catena risultante

Teorema 3.5.1 *Sia P un insieme parzialmente ordinato di cardinalità n. Allora l'Algoritmo 6 risolve il problema dell'ordinamento con informazione parziale impiegando al più $c \log e(P)$ confronti, dove $c \approx 15.08$.*

Dimostrazione. Sia $g(P)$ il numero di confronti necessario ad ordinare P. Per il Lemma 3.2.1 abbiamo $|A| \geq n \cdot 2^{-H(\overline{P})}$, dunque

$$|B| = |P - A| \leq n \left(1 - 2^{-H(\overline{P})}\right) \leq \ln 2 \cdot nH(\overline{P}),$$

in cui abbiamo usato l'ovvia disuguaglianza $1 - 2^{-x} \leq \ln 2 \cdot x$ per $x \geq 0$. Grazie ai Teoremi 3.3.2 e 3.4.11 possiamo maggiorare il numero di confronti compiuti con

$$g(P) \leq (1+\varepsilon) \log e(P - A) + \left((1+\varepsilon)\left(\log e + \log\left(1 + \frac{1}{\varepsilon}\right)\right) + 1\right) |P - A| + 6 \log e(P')$$

e, per la disuguaglianza appena dimostrata,

$$g(P) \leq (1+\varepsilon) \log e(P) + \left((1+\varepsilon)\left(1 + \ln\left(1 + \frac{1}{\varepsilon}\right)\right) + \ln 2\right) nH(\overline{P}) + 6 \log e(P').$$

Possiamo quindi applicare il Teorema 3.1.2 ed ottenere la stima

$$g(P) \leq \left(1 + \varepsilon + 2\left((1+\varepsilon)\left(1 + \ln\left(1 + \frac{1}{\varepsilon}\right)\right) + \ln 2\right) + 6\right) \log e(P).$$

Infine, ponendo $\varepsilon \approx 0.351198$, abbiamo

$$g(P) \leq c \log e(P)$$

dove $c \approx 15.08$, cioè la tesi. □

Capitolo 4

Conclusioni

Nel primo capitolo di questa tesi abbiamo presentato tre definizioni di una generalizzazione dell'entropia di Shannon nota come entropia di grafo. La prima di esse è motivata da un problema di codifica, ma risulta in un'espressione difficile da calcolare e non evidentemente ben definita. Abbiamo quindi dimostrato l'equivalenza con una seconda forma, comunque espressa in termini di teoria dell'informazione, per la quale risulta facile affermare la buona definizione. È infine immediato dimostrare l'equivalenza con una terza, più utile nella pratica, che consente di vedere il calcolo dell'entropia come un problema di minimizzazione di una funzione convessa su un convesso.

Nel secondo capitolo sono state presentate le proprietà principali, fra cui una forma di subadditività per coppie di grafi aventi lo stesso insieme di vertici. Abbiamo evidenziato il sorprendente collegamento esistente fra l'entropia e i grafi perfetti, una vasta classe di grafi dalle interessanti proprietà algoritmiche e combinatoriali. Sono in particolare perfetti i grafi di confrontabilità associati a ordini parziali; questo ci ha consentito di riformulare l'entropia di grafo in termini puramente combinatoriali e di ottenere un utile risultato sulla propria approssimabilità.

Il terzo capitolo di questa tesi ha applicato i precedenti risultati alla soluzione del problema dell'ordinamento con informazione parziale. Tale problema consiste nell'estensione di un ordine parziale a un ordine totale facendo uso del minimo numero di ulteriori confronti. L'entropia di grafo, come mostrato da Kahn e Kim, consente di esibire algoritmi ottimi per questo problema. Abbiamo presentato in particolare tre algoritmi dovuti a Cardinal et al. che, sfruttando la versione approssimata dell'entropia presentata nel precedente capitolo, riescono a evitare l'uso del metodo dell'ellissoide, essenziale invece nell'articolo di Kahn e Kim.

Bibliografia

[1] C. Berge. "Les problemes de coloration en théorie des graphes". In: *Publ. Inst. Stat. Univ. Paris* 9 (1960), pp. 123–160.

[2] J. Cardinal e S. Fiorini. "Minimum entropy coloring". In: *Algorithms and Computation* X (2005), pp. 819–828.

[3] J. Cardinal, S. Fiorini, G. Joret, R.M. Jungers e J.I. Munro. "An Efficient Algorithm for Partial Order Production". In: *Proceedings of the 41st annual ACM symposium on Theory of computing*. 2009, pp. 93–100.

[4] J. Cardinal, S. Fiorini, G. Joret, R.M. Jungers e J.I. Munro. "Sorting under Partial Information (without the Ellipsoid Algorithm)". In: *Proceedings of the 42nd ACM symposium on Theory of computing*. 2010, pp. 359–368.

[5] M. Chudnovsky, N. Robertson, P. Seymour e R. Thomas. "The Strong Perfect Graph Theorem". In: *The Annals of Mathematics*. Second Series 164.1 (2006), pp. 51–229.

[6] V. Chvátal. "On certain polytopes associated with graphs". In: *Journal of Combinatorial Theory, Series B* 18.2 (1975), pp. 138–154.

[7] T.M. Cover e J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[8] I. Csiszár, J. Körner, L. Lovász, K. Marton e G. Simonyi. "Entropy splitting for antiblocking corners and perfect graphs". In: *Combinatorica* 10.1 (1990), pp. 27–40.

[9] M.L. Fredman. "How good is the information theory bound in sorting?" In: *Theoretical Computer Science* 1.4 (1976), pp. 355 –361.

[10] M.C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Annals of Discrete Mathematics. Elsevier, 2004.

[11] F.K. Hwang e S. Lin. "A Simple Algorithm for Merging Two Disjoint Linearly Ordered Sets". In: *SIAM Journal on Computing* 1.1 (1972), pp. 31–39.

[12] J. Kahn e J.H. Kim. "Entropy and Sorting". In: *Journal of Computer and System Sciences* 51.3 (1995), pp. 390–399.

[13] J. Kahn e M. Saks. "Balancing poset extensions". In: *Order* 1 (2 1984), pp. 113–126.

[14] S.S. Kislitsyn. "A finite partially ordered set and its corresponding set of permutations". In: *Mathematical Notes* 4 (5 1968), pp. 798–801.

[15] J. Körner. “A property of conditional entropy”. In: *Stud. Sci. Math. Hung.* 6 (1971), pp. 355–359.

[16] J. Körner. “Coding of an information source having ambiguous alphabet and the entropy of graphs”. In: *6th Prague Conference on Information Theory*. 1973, pp. 411–425.

[17] J. Körner e K. Marton. “Graphs that Split Entropies”. In: *SIAM J. on Discrete Mathematics* 1 (1 1988), pp. 71–79.

[18] J. Körner, G. Simonyi e Z. Tuza. “Perfect couples of graphs”. In: *Combinatorica* 12.2 (1992), pp. 179–192.

[19] N. Linial. “The information-theoretic bound is good for merging”. In: *SIAM J. Comput.* 13.4 (1984), pp. 795–801.

[20] L. Lovász. “Normal hypergraphs and the perfect graph conjecture”. In: *Discrete Mathematics* 2.3 (1972), pp. 253 –267.

[21] L. Mirsky. “A Dual of Dilworth’s Decomposition Theorem”. In: *The American Mathematical Monthly* 78.8 (1971), pp. 876–877.

[22] A. Rényi. “On measures of information and entropy”. In: *Proc. 4th Berkeley Symp. Math. Statist. and Prob.* Vol. 1. 1961, pp. 547–561.

[23] C.E. Shannon. “A Mathematical Theory of Communication”. In: *Bell Syst. Tech. J.* 27 (1948), pp. 379–423, 623–656.

[24] G. Simonyi. “Graph Entropy: A Survey”. In: *Combinatorial Optimization: Papers from the DIMACS Special Year* (1995), pp. 399–441.

[25] G. Simonyi. “Perfect graphs and graph entropy: An updated survey”. In: *Perfect Graphs*. A cura di J.L.R. Alfonsín e B.A. Reed. Wiley, 2001, pp. 293–328.

Fine della dimostrazione del Teorema 3.1.2

Dimostrazione. Resta da dimostrare la disuguaglianza (3.1), cioè l'esistenza di un indice $i \in \{1, 2\}$ tale che

$$\frac{e(P_i)}{e(P)} \leq \frac{1}{\sqrt{\alpha_i \beta_i}}.$$

Dimostreremo che

$$\frac{e(P_1)}{e(P)} + \frac{e(P_2)}{e(P)} \leq 1 \tag{1}$$

$$\frac{1}{\sqrt{\alpha_1 \beta_1}} + \frac{1}{\sqrt{\alpha_2 \beta_2}} \geq 1. \tag{2}$$

Per dimostrare (1) è sufficiente dimostrare che $a <_{P_1} b$ e $a >_{P_2} b$. Sfruttando le definizioni di α_1 e β_1 abbiamo

$$y^1_{a^+} = y_a + \frac{x_a}{\alpha_1} = \begin{cases} y_{a^-} + x_a - x_b & \text{se} \quad \lambda \leq \frac{1}{2} \\ y_{a^-} + \frac{x_a}{2} & \text{altrimenti} \end{cases}$$

e

$$y^1_{b^-} = y_{b^+} - \frac{x_b}{\beta_1} = y_a + x_a - \frac{x_b}{\beta_1} = \begin{cases} y_{a^-} + x_a - x_b & \text{se} \quad \lambda \leq \frac{1}{2} \\ y_{a^-} + \frac{x_a}{2} & \text{altrimenti,} \end{cases}$$

cioè $a <_{P_1} b$. Per le analoghe definizioni di α_2 e β_2 abbiamo

$$y^2_{a^-} = y_{a^+} - \frac{x_a}{\alpha_2} = y_{a^+} - \frac{x_b}{2}$$

e

$$y^2_{b^+} = y_{b^-} + \frac{x_b}{\beta_2} = y_{a^+} - x_b + \frac{x_b}{\beta_2} = y_{a^+} - \frac{x_b}{2},$$

cioè $a >_{P_2} b$. Poiché P_1 e P_2 sono estensioni di P otteniamo la tesi.

Per dimostrare (2) basta studiare la funzione di λ definita dal membro sinistro. Abbiamo infatti

$$f(\lambda) = \begin{cases} \sqrt{1-\lambda} + \frac{\sqrt{\lambda}}{2} & \text{se} \quad \lambda \leq \frac{1}{2} \\ \frac{1}{2\sqrt{\lambda}} + \frac{\sqrt{\lambda}}{2} & \text{altrimenti,} \end{cases}$$

la cui derivata è

$$f'(\lambda) = \begin{cases} \frac{1}{4\sqrt{\lambda}} - \frac{1}{2\sqrt{1-\lambda}} & \text{se} \quad \lambda \leq \frac{1}{2} \\ \frac{1}{4\sqrt{\lambda}} - \frac{1}{4\lambda^{\frac{3}{2}}} & \text{altrimenti.} \end{cases}$$

È facile inoltre verificare che la derivata sia positiva per $\lambda < \frac{1}{5}$ e negativa altrimenti. Poiché $f(0) = f(1) = 1$ possiamo dedurre che f è maggiore di 1 per ogni $\lambda \in [0, 1]$, cioè la tesi. □