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

Jacopo Notarstefano

18 aprile 2012

# Capitolo 0

# Introduzione e definizioni preliminari

Il concetto di entropia di grafo fu introdotto da Janos Köner come possibile risposta al problema di assegnare ad un grafo un numero che ne rappresenti la complessità, ma che allo stesso tempo ammetta un'interpretazione in termini di teoria dell'informazione. [**?**] Questo si ottiene interpretando il grafo come rappresentante la relazione di distinguibilità dei simboli emessi da una sorgente discreta. Più precisamente, consideriamo $X$ una sorgente la quale ad ogni istante discreto emetta un simbolo $v_i$ dall'alfabeto finito $\{v_1, \ldots, v_n\}$, con probabilità $p_1, \ldots, p_n$ rispettivamente, ma supponiamo di non essere in grado di distinguere $v_i$ da $v_j$ per ogni $i, j$. Rappresentiamo dunque tale relazione di distinguibilità con un grafo con vertici i simboli $v_1, \ldots, v_n$ e con un arco $(v_i, v_j)$ ogni volta che i simboli $v_i$ e $v_j$ sono distinguibili; questa relazione, come è immediato verificare, è simmetrica ma non necessariamente transitiva nè riflessiva. Vogliamo inoltre che, una volta definita, l'entropia sia, per un grafo completo (ovvero nel caso di totale distinguibilità), equivalente alla classica nozione di entropia per una sorgente; mostreremo infatti che essa generalizza l'entropia di Shannon.

Grazie ad alcune sue proprietà peculiari, ad esempio una particolare forma di subadditività, l'entropia di grafo ha trovato applicazione in campo combinatorico e algoritmico, in particolare per dimostrare alcune disuguaglianze. Lo stesso Körner l'ha usata per mostrare nuovamente una stima sul numero di funzioni di hash perfetto di un insieme dovuta a Fredman e Komlós [**?**] e successivamente, insieme a Marton, per migliorare tale stima sfruttando le disuguaglianze più precise che si ottengono estendendo il concetto di entropia agli ipergrafi [**?**]. Sempre Körner, in collaborazione con altri, ha analizzato i casi in cui la subbaditività è in realtà un'uguaglianza, deducendone una caratterizzazione dei grafi perfetti in termini di entropia e quindi una nuova dimostrazione della congettura debole che li riguarda [**?**]. Kahn e Kim hanno invece esibito un algoritmo che utilizza l'entropia per determinare additivamente un ordine totale fissato in precedenza ma ignoto, tramite il calcolo della stessa su una particolare successione di grafi associati [**?**]. Questo approccio ha anche consentito di stimare il numero di estensioni lineari di un dato ordine parziale, la cui determinazione esatta è un problema #P-completo [**?**].

Per tutto il seguito con $G = (V, E)$ indicheremo un grafo semplice non diretto di insieme di vertici $V$ ed insieme degli archi $E$. Ricordiamo inoltre che un sottoinsieme di vertici di $G$ è un *insieme indipendente* se essi sono due a due non adiacenti e che una *colorazione* di $G$ è una partizione di $V$ in insiemi indipendenti. Definiamo quindi il *numero cromatico* di $G$ il numero minimo $\chi(G)$ di insiemi indipendenti necessari a

colorare $G$. Infine, assumiamo che tutti i logaritmi siano in base 2.

La rigorosa definizione di entropia di grafo comporta alcune difficoltà tecniche, la cui risoluzione costituisce l'obiettivo principale del primo capitolo. Nel secondo capitolo verranno enunciate e dimostrate alcune proprietà di rilievo dell'entropia.

# 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 $G$ è un grafo e $P$ una distribuzione di probabilità discreta sui vertici, ovvero se $|V| = n$ allora $P$ è una $n-$upla $(p_1 \dots p_n)$ 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 $V = \{v_1 \dots v_n\}$ con la corrispondente probabilità $p_i$. Nel nostro caso supponiamo che tali simboli non siano a due a due distinguibili, in particolare lo sono se e solo se $(v_i, v_j) \in E$, insieme degli archi di $G$.

Vogliamo valutare la bontà della migliore codifica possibile dell'informazione emessa dalla sorgente. Per fare questo fissiamo $t$ un intero positivo e consideriamo tutte le stringhe di simboli in $V$ di lunghezza $t$. Poiché la sorgente è priva di memoria, è naturale assegnare alla stringa $\mathbf{v} = v_{i_1} \dots v_{i_t}$ la probabilità

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

Una *codifica propria dell'informazione* in esse contenute è una mappa dalle stringhe a un insieme finito di simboli tale che a stringhe distinguibili vengano assegnati simboli distinti.

È però tipico dei problemi di codifica trascurare parte delle stringhe possibili, a cui non ci interessa assegnare un significato e quindi un simbolo. Sia quindi $0 < \varepsilon < 1$ fissato, ed ammettiamo che la condizione di codifica propria debba essere soddisfatta all'esterno di un insieme di probabilità totale $\varepsilon$.

Una definizione ragionevole di bontà della codifica è la quantità

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

dove $M$ è la cardinalità dell'insieme immagine della 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. Allora $L^2$, prodotto normale di $L$ con se stesso, è rappresentato a destra. TODO: Disegno

Osserviamo che la coppia $(G^t, \mathbf{P}^t)$, dove $\mathbf{P}^t$ è definita da (1.1), è un grafo probabilistico. Se $U \subset V^t$ possiamo denotare con $\mathbf{P}^t(U)$ la somma delle probabilità dei vertici in $U$.

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:

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

Affinchè la precedente risulti una buona definizione dovremmo dimostrarne l'indipendenza da $\varepsilon$. Salteremo questa verifica e dimostreremo direttamente nella prossima sezione l'equivalenza con una seconda definizione, la quale non dipende da $\varepsilon$.

### 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 scriviamo $H(P)$ per indicare l'entropia di Shannon di una variabile aleatoria discreta di densitá $P$.

**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:

$$H(G,P) = \min I(X;Y) \tag{1.4}$$

dove $X$ è una variabile aleatoria a valori nei vertici di $G$ con densità $P$ e $Y$ è una variabile aleatoria a valori negli insiemi indipendenti di $G$ tale che l'evento $\{X \in Y\}$ abbia probabilità 1.

Affinchè la precedente 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 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

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

ed il loro involucro convesso è STAB($G$).

TODO: Disegno

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* 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 ad $\infty$ 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 STAB($G$) e che il vettore che lo realizza il minimo ha tutte le coordinate maggiori di 0.

## 1.2 Equivalenza delle tre definizioni

### 1.2.1 Numero Cromatico e Mutua Informazione

Sia $G$ un grafo. Chiamiamo *nucleo* un insieme 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.*

**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. Otteniamo immediatamente che la relazione $\sim$ sia d'equivalenza: simmetria e riflessività seguono dalla definizione di $\sim$, la transitività da $\mathrm{diam}(G) = 1$. Supponiamo 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.

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. TODO: citazione

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

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

*Allora*

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

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

**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

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

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*

1. *Per ogni $\lambda$ esiste $K$ tale che $\mathbf{P}^t\big(\overline{T^t(P)}\big) < \lambda$ per questo $K$.*
2. *Se $\mathbf{v}$ è $P-$tipica allora $2^{-tH(P)-C\sqrt{n}} \le \mathbf{P}^t(\mathbf{v}) \le 2^{-tH(P)+C\sqrt{n}}$.*
3. *La cardinalità di $T^t(P)$ è limitata da $2^{tH(P)-C\sqrt{n}} \le |T^t(P)| \le 2^{tH(P)+C\sqrt{n}}$.*

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

**Definizione.** Sia $(G,P)$ un grafo probabilistico. Chiamiamo *grafo dei nuclei* il grafo $\Gamma$, 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 $\Gamma$ 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 [**?**].

**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 $\varepsilon$.*

**Dimostrazione.** Per il lemma 1.2.4 (1), 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

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

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

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

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

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

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

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

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 (2)abbiamo

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

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

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

ed equivalentemente

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

D'altra parte vale

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

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 $\varepsilon$, 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 (1) 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 (2) implica che se $\mathbf{A}$ è una sequenza di nuclei $R-$tipica allora

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

Osserviamo quindi che possiamo 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

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

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

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

e inoltre applicare la diseguaglianza di Bernoulli per 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 $\varepsilon$ e $\delta$ 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 (Simonyi)** *Sia $G$ un grafo e siano $S(G)$ i suoi insiemi indipendenti. Sia $P$ una densità discreta sui vertici di $G$. Avremo allora:*

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

**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 $\mathbf{a} \in$ 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:

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

□

# Capitolo 2

# Proprietà e teoremi notevoli

## 2.1 Proprietà notevoli dell'entropia di grafo

### 2.1.1 Monotonia

**Lemma 2.1.1** *Siano $F$ e $G$ grafi tali che $V(G) = V(F)$ 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 $\mathrm{VP}(G) \subset \mathrm{VP}(F)$. 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 di comune insieme dei vertici $V$. 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 $\mathbf{a} \in \mathrm{VP}(F)$ e $\mathbf{b} \in \mathrm{VP}(G)$ 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 scalare dei loro vettori caratteristici è il vettore caratteristico di un insieme indipendente di $F \cup G$. Pertanto, sfruttando la convessità del politopo dei vertici, il prodotto scalare $\mathbf{a} \cdot \mathbf{b}$ appartiene a $\mathrm{VP}(F \cup G)$. 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 su insiemi di vertici disgiunti, 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 $P_{v\leftarrow Q}$ in modo che la coppia $(G_{v\leftarrow F}, P_{v\leftarrow Q})$ sia un grafo probabilistico. Per fare questo poniamo

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

**Lemma 2.1.3** *Siano $F$ e $G$ grafi su insiemi di vertici disgiunti, 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*

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

**Dimostrazione.** TODO □

**Corollario 2.1.4** *Sia $(G, P)$ un grafo probabilistico e siano $G_1 \dots G_n$ le sue componenti connesse. Poniamo $P_i(x) = P(x)[P(V(G_i))]^{-1}$ per $x \in V(G_i)$. Allora abbiamo*

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

**Dimostrazione.** Consideriamo il grafo su $n$ vertici $\{v_1 \dots v_n\}$ privo di archi, e sia $Q$ la densità discreta definita da $Q(v_i) = P(V(G_i))$. Otteniamo la tesi applicando $n$ 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. Comunque scelta $P$ densità discreta sui vertici avremo*

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

**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

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

Applichiamo ora la diseguaglianza sulle somme dei logaritmi (TODO: spiegare) e otteniamo che il minimo è realizzato per

$$q_i = p_i \quad \forall i$$

□

**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 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. Esiste un sorprendente collegamento fra l'entropia di grafo e i grafi perfetti [**?**].

**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.

## 2.3 Grafi associati ad ordini parziali

Abbiamo finora sviluppato la teoria dell'entropia di grafo per una distribuzione di probabilità $P$ qualunque sui vertici. Nel seguito sarà sufficiente considerare il caso in cui tale distribuzione sia uniforme.

**Lemma 2.3.1 (Chvatál)** *TODO*

**Dimostrazione.** TODO □

Sia $G$ un grafo. Possiamo trovare una partizione dei suoi vertici in insiemi indipendenti con l'algoritmo goloso che ad ogni passo trova un insieme indipendente e massimale e procede ricorsivamente sul complementare.

**Definizione.** Sia $G$ un grafo perfetto e sia $\{S_1, \ldots, S_k\}$ una partizione dei vertici ottenuta con il precedente algoritmo goloso. Chiameremo *punto goloso* il punto $x$ definito da

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

Dalla costruzione è evidente che tale punto appartiene a STAB($G$). Il seguente teorema TODO

**Teorema 2.3.2 (Cardinal, Fiorini, Joret, Jungers, Munro)** *Sia $G$ un grafo perfetto su $n$ vertici e sia $x$ un suo punto goloso. Allora, comunque fissato $\varepsilon > 0$, vale*

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

**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

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

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

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

TODO

Abbiamo ora tutti gli strumenti necessari per concludere. Poiché $G$ è perfetto possiamo applicare il teorema 2.2.1; inoltre, essendo $z \in$ 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

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

da cui deduciamo

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

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

# Capitolo 3

# Tre algoritmi per ordinare con informazione parziale

## 3.1 Ordinamento con informazione parziale

**Definizione.** TODO: scrivere meglio questa parte! Sia $P$ un insieme parzialmente ordinato di sostegno $V = \{v_1, \ldots, v_n\}$, a propria volta dotato di un ordine totale $\leq$ fissato ma ignoto. Il *problema dell'ordinamento con informazione parziale* consiste nel determinare l'ordine totale per mezzo di domande del tipo "*è vero che* $v_i \leq v_j$ *?*", detti *confronti*. Chiamiamo *estensione lineare* un qualunque ordine totale che sia compatibile con $P$ e denotiamo con $e(P)$ il numero di estensioni lineari.

Tale problema fu originariamente posto da Fredman nel 1976 [**?**]. Sempre Fredman dimostrò l'esistenza di un algoritmo che lo risolva compiendo $\log e(P) + 2n$ confronti, il quale tuttavia richiede tempo di esecuzione superpolinomiale. Nel 1984 Kahn e Saks dimostrarono l'esistenza di un algoritmo che compia $O(\log e(P))$ confronti [**?**]. 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]

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

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

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

**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 $n = 1$

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

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

```
// Preparazione
trova una catena C di lunghezza massima in P
// Ordinamento
while P − C ≠ ∅ do
    togli un elemento da P − C e inseriscilo in C con una ricerca binaria
end while
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.** (TODO: spiegare un po' meglio) È noto che l'entropia di un grafo su $n$ vertici e dimensione massima di un insieme indipendente $\alpha$ è maggiore o uguale a $-\log \frac{\alpha}{n}$ [**?**]. 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 naive

**Algoritmo 2** "Merge sort naive" 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
```

Sia $\tilde{h}$ l'entropia di Shannon della probabilità discreta $\left\{ \frac{|C_1|}{n}, \ldots, \frac{|C_k|}{n} \right\}$.

**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.*

**Dimostrazione.** TODO □

**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.** TODO □

## 3.4 Merge con informazione parziale

**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*

- *$G$ è bipartito*
- *$G$ è biconvesso (TODO: spiegare cosa significhi)*
- *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 $[c_u, d_v]$.*

**Dimostrazione.** TODO □

**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 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 3.4.2** *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 (3) segue che $u'v$ e $uv'$ sono archi di $G$. Supponiamo per assurdo che $uv'$ non sia stretto. Avremmo allora:

$$\begin{aligned} x_v &= 1 - x_u && \text{(poiché } uv \text{ è stretto)} \\ &> x_{v'} && \text{(poiché } uv' \text{ non è stretto)} \\ &= 1 - x_{u'} && \text{(poiché } u'v' \text{ è stretto)} \\ &\geq x_v && \text{(poiché } u'v \text{ è un arco e per il lemma 2.3.1),} \end{aligned}$$

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.

TODO: definizione di $\mathrm{STAB}^*(G)$

**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 $u \leq_P w \leq_P u''$.

**Lemma 3.4.3** *Sia $P$ un insieme parzialmente ordinato ricoperto da due catene disgiunte $A$ e $B$, sia inoltre $G = \overline{G}(P)$. Dato $x \in STAB^*(G)$ allora*

1. *se in $G(x)$ una componente connessa $L$ è incastonata in $K$ allora $L$ è libera.*
2. *se $K$ ed $L$ sono componenti connesse non banali di $G(x)$ allora vale $K \leq_P L$ oppure $L \leq_P K$.*

**Dimostrazione.** 1. 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 $w, u', u'' \in A$ TODO

2. TODO □

**Lemma 3.4.4** *TODO*

**Dimostrazione.** TODO □

**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*.

**Lemma 3.4.5** *TODO*

**Dimostrazione.** TODO □

**Lemma 3.4.6** *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

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

confronti [**?**]. 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

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

Possiamo infine mettere insieme le precedenti due equazioni per ottenere

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

cioé la tesi. □

**Definizione.** Sia $K$ una componente connessa di $G(x)$. Se $K$ è rossa chiamo $A \cap K$ *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.

**Lemma 3.4.7** *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 $x$ avremo

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

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 3.4.8** *TODO*

**Dimostrazione.** TODO □

**Lemma 3.4.9** *TODO*

**Dimostrazione.** TODO □

**Teorema 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. Allora l'algoritmo 3 fonde $A$ e $B$ impiegando al più $6 \log e(P)$ confronti.*

**Dimostrazione.** TODO □

**Algoritmo 3** "Merge" con informazione parziale

1: TODO

## 3.5 Merge sort

**Algoritmo 4** "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 3 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 4 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 $1 - 2^x \leq \ln 2 \cdot x$. Grazie ai teoremi 3.3.2 e 3.4.10 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. □

# Appendice A

# Dimostrazione di 3.1