

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

Jacopo Notarstefano

5 maggio 2012

Capitolo 0

Introduzione e definizioni preliminari

Il concetto di entropia di grafo fu introdotto da Janos Körner 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 $\{v_1, \dots, v_n\}$, con probabilità p_1, \dots, 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, \dots, 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 subadditività 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 adattivamente 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 [?]. L'algoritmo di Kahn e Kim è polinomiale nel numero di operazioni elementari, ma richiede di eseguire ad ogni passo il metodo dell'ellissoide. Invece, in un recente articolo, Cardinal et al. hanno esibito tre nuovi algoritmi, comunque polinomiali, che non ne fanno uso [?].

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. Infine nel terzo capitolo verranno presentati tre algoritmi basati sul calcolo 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 $p_1 + \dots + p_n = 1$.

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 .

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 $\mathbf{v} = v_{i_1} \dots v_{i_t}$ la probabilità

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

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 vengono mandate in simboli distinti. È intuitivamente preferibile una codifica che faccia uso del minor numero possibile di simboli. Cerchiamo dunque di minimizzare la quantità

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

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

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

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 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 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:

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

Affinchè la precedente 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 scriviamo $H(P)$ 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:

$$H(G, P) = \min I(X; Y) \quad (1.4)$$

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 $\text{STAB}(G)$.

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 $\text{STAB}(G)$.

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:

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

Osservazione. La funzione obiettivo di cui cerchiamo il minimo in (1.5) è convessa, tende ad ∞ 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 $\text{STAB}(G)$ 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 $\chi(G) \leq \xi(G)$. \square

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 $\text{diam}(G) = 1$. 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 $(v, w) \in E$. \square

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

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

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

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

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

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

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

1. Per ogni λ esiste K tale che $\mathbf{P}^t(\overline{T^t(P)}) < \lambda$ per questo K .
2. Se \mathbf{v} è P -tipica allora $2^{-tH(P)-C\sqrt{n}} \leq \mathbf{P}^t(\mathbf{v}) \leq 2^{-tH(P)+C\sqrt{n}}$.
3. 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

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

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

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

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

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

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 ε .

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

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

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

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

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 disuguaglianza

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

Sia ora \mathbf{N} tale che $|\mathbf{N} \cap T^t(P)|$ 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 $V(\Gamma)$

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

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

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

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

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

dove la disuguaglianza 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 $\bar{\Gamma}$ 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 $(\bar{\Gamma}, Q)$, 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

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

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

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

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

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

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

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

□

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 $(\mathbf{A}_1, \dots, \mathbf{A}_M)^c$ 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 $\mathbf{P}^t((\mathbf{A}_1, \dots, \mathbf{A}_M)^c)$ 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

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

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

$$\sum_{\mathbf{v} \in V^t} R_M^*(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) = \sum_{\mathbf{v} \in T^t(P)} R_M^*(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) + \sum_{\mathbf{v} \in \overline{T^t(P)}} R_M^*(C_{\mathbf{v}}) \cdot \mathbf{P}^t(\mathbf{v}) \quad (1.16)$$

ed osservare che il secondo termine è maggiorato da $\mathbf{P}^t(\overline{T^t(P)})$, che per il lemma 1.2.4 (1) possiamo supporre essere più piccolo di $\varepsilon/2$. Per stimare il primo termine osserviamo innanzitutto che

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

maggiorando con il massimo e sfruttando il fatto che hanno massa totale 1. Possiamo migliorare 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

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

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

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

Inoltre il lemma 1.2.4 (2) implica che se \mathbf{A} è una sequenza di nuclei R -tipica allora

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

Osserviamo quindi che possiamo stimare $R^*(\mathcal{N}_v)$ 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 $R^*(\mathcal{N}_v)$ sia maggiore del loro prodotto.

Per quanto appena visto otteniamo

$$\max_{v \in T^t(P)} R_M^*(C_v) \leq \left(1 - \exp_2 \left[-tH(R) - K_4\sqrt{t} + tH(Q | \sim) - K_3\sqrt{t} \right] \right)^M \quad (1.20)$$

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

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

e inoltre applicare la diseguaglianza di Bernoulli per ottenere

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

Poniamo $M = \lfloor 2^{tH(G, P) + \delta} \rfloor$. È immediato allora che il primo termine del membro di destra di (1.16) sia infinitesimo per $t \rightarrow \infty$. 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 $\delta > 0$

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

o equivalentemente

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

□

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

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

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

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

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

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

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

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

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

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

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

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

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 disuguaglianza

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

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 $Q^*(J)$,

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

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

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



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 $VP(G) \subset VP(F)$. 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. \square

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 $\mathbf{a} \in VP(F)$ e $\mathbf{b} \in VP(G)$ 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 $VP(F \cup G)$. Ma allora possiamo scrivere

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

\square

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 -simpleso, ma poiché sappiamo che la funzione obiettivo è minima sul bordo deduciamo che

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

Applichiamo ora la disuguaglianza 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*

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

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 $\{S_1, \dots, S_k\}$ 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 $\text{STAB}(G)$. 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, \dots, 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 $\{1, \dots, k\}$ 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 $z \in \text{STAB}(G)$. 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 $z \in \text{STAB}(G)$, possiamo scrivere la disuguaglianza

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

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. Sia $P = (V, \leq_P)$ un insieme parzialmente ordinato di cardinalità n . 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 di cardinalità n . 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*.

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

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.

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 $e(P)$ [?]. Il seguente teorema afferma ad esempio che $nH(\overline{P}) = \Theta(\log e(P))$.

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

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

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

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

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 2.3.2 [?]. 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*

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

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

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

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

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

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

$$\{(y_{v-}^1, y_{v+}^1)\}_{v \in V} \quad \text{e} \quad \{(y_{v-}^2, y_{v+}^2)\}_{v \in V}$$

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

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

e infine

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

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

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

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

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

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

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

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. □

Algoritmo 1 “Insertion sort” con informazione parziale

```
1: // Preparazione
2: trova una catena  $C$  di lunghezza massima in  $P$ 
3: // Ordinamento
4: while  $P - C \neq \emptyset$  do
5:   toglì un elemento da  $P - C$  e inseriscilo in  $C$  con una ricerca binaria
6: end while
7: return  $C$ 
```

3.2 Insertion sort

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 $-\log \frac{\alpha}{n}$ [?]. La tesi segue applicando questo fatto a $G = \overline{G}(P)$. \square

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. \square

3.3 Merge sort naive

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

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.*

Algoritmo 2 “Merge sort naive” con informazione parziale

```
1: // Preparazione
2: trova una decomposizione golosa di  $P$  in catene  $C_1, \dots, C_k$ 
3:  $\mathcal{C} \leftarrow \{C_1, \dots, C_k\}$ 
4: // Ordinamento
5: while  $|\mathcal{C}| > 1$  do
6:   seleziona da  $\mathcal{C}$  due catene di lunghezza minima  $C$  e  $C'$ 
7:   fondi  $C$  e  $C'$  in tempo lineare, ottenendo  $C''$ 
8:   cancella  $C$  e  $C'$  da  $\mathcal{C}$ , aggiungi  $C''$ 
9: end while
10: return l'unica catena di  $\mathcal{C}$ 
```

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 $\tilde{h} + 1$ [?]. 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 precedente lemma otteniamo

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

inoltre abbiamo

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

dove abbiamo applicato i teoremi 2.3.2 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 [?]. 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.

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*

1. G è bipartito.
2. G è biconvesso, cioè i vertici adiacenti ad un vertice di una catena formano un intervallo nella catena opposta.
3. 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 □

Sia $h(x) = -x \log x - (1-x) \log (1-x)$ l'entropia binaria, in cui per convenzione poniamo $h(0) = h(1) = 0$.

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

$$A = A_1 \cup \dots \cup A_k \quad e \quad B = B_1 \cup \dots \cup B_k$$

tali che

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

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 $A' = A - (A_1 \cup \dots \cup A_{i-1})$ 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 $B' = B - (B_1 \cup \dots \cup B_{i-1})$ non lo è allora scegliamo $v \in B'$ e poniamo $A_i = \emptyset$ e $B_i = \{v\}$.

Per evitare casi degeneri denoteremo nel seguito con $\text{STAB}^*(G)$ i punti di $\text{STAB}(G)$ di coordinate tutte positive. Infatti se così non fosse la funzione obiettivo dell'equazione 1.5 non sarebbe definita.

Nel caso di grafi bipartiti possiamo semplificare il lemma 2.3.1. Abbiamo allora che

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

Definizione. Diciamo che un arco uv è *stretto* rispetto ad $x \in \text{STAB}(G)$ 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 3.4.3 *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 $\text{STAB}(G)$ 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. \square

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

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

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

Definizione. Sia $x \in \text{STAB}^*(G)$. 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 $x \in \text{STAB}^*(G)$. 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 $u \leq_P w \leq_P u''$.

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 $x \in \text{STAB}^*(G)$ 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 $w, u', u'' \in A$ 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 (3) 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 2.3.1.

2. 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

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

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. \square

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

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

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

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**
 - 5: **return** x'
-

Quando una componente connessa K viene ribilanciata può accadere che un arco non stretto di $G(x)$ diventi stretto in $G(x')$. In questo 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 $x \in \text{STAB}(G)$. Diciamo che due componenti connesse K ed L di $G(x)$ si *toccano* 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 $L \cap B \leq_P \{v\} \leq_P K \cap B$.

Lemma 3.4.5 *Sia $x \in \text{STAB}^*(G)$, 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 si toccano.*

Dimostrazione. Supponiamo per assurdo che K ed L non si toccano. 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 sono contenuti in una stessa componente connessa di $G(x)$, una contraddizione. \square

Definizione. Sia $x \in \text{STAB}^*(G)$. 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 $x \in \text{STAB}^*(G)$. 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

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

Lemma 3.4.6 *Sia $x \in \text{STAB}^*(G)$ 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 . \square

Osservazione. TODO

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 [?]. 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. \square

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.

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. \square

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.

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  $X = A \cap K$  e  $Y = B \cap K$  con l'algoritmo di Hwang-Lin
4:   for  $v \in K$  do
5:      $x_v = \max \{x_v, \frac{1}{2}\}$  se  $v \in A$ ,  $x_v = \max \{x_v, \frac{1}{2} + \frac{1}{2n}\}$  se  $v \in B$ 
6:   end for
7:   ribilancia  $x$  con l'algoritmo 3
8:   for  $v \in K$  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  $C$ 

```

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:

- $v \in A$ 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

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

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 $G'(x')$.

TODO

□

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 $x \in \text{STAB}(G)$ 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:

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

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

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

Osserviamo inoltre che vale

$$r_j - r_{j+1} \geq s_j, \quad (3.3)$$

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

Supponiamo invece che K sia blu in $G_j(x')$. TODO

TODO

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 $x \in \text{STAB}(G)$ 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 “naive” 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 $|A| \geq n \cdot 2^{-H(\bar{P})}$, dunque

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

in cui abbiamo usato l’ovvia disuguaglianza $1 - 2^x \leq \ln 2 \cdot x$. Grazie ai teoremi 3.3.2 e 3.4.11 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,

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

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

Fine della dimostrazione del Teorema 3.1.2

Dimostrazione. Resta da dimostrare 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

$$\frac{e(P_1)}{e(P)} + \frac{e(P_2)}{e(P)} \leq 1 \quad (\text{A.1})$$

$$\frac{1}{\sqrt{\alpha_1 \beta_1}} + \frac{1}{\sqrt{\alpha_2 \beta_2}} \geq 1. \quad (\text{A.2})$$

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

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

e

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

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

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

e

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

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

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

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

la cui derivata è

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

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. \square

