Tre definizioni equivalenti di entropia di grafo

Jacopo Notarstefano

11 aprile 2012

1 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 $\{v_1, \dots, v_n\}$, 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 certe 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 Kolmos 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 un grafo comporta alcune difficoltà tecniche, la cui risoluzione costituisce l'obiettivo principale di questo lavoro. La prima sezione sarà dedicata a proporre tre possibili definizioni di entropia, la cui equivalenza sarà dimostrata nella sezione successiva.

2 Tre definizioni di entropia di grafo

2.1 Definizione in termini di Numero Cromatico

Definizione. Chiameremo *grafo probabilistico* una coppia (G, P) dove G è un grafo e P una distribuzione di probabilità discreta sui vertici, cioè 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. Supporremo inoltre che tali simboli non siano a due a due distinguibili, ma anzi che $(v_i, v_j) \in E$ se e solo se v_i e v_j sono distinguibili.

Vogliamo valutare la bontà della migliore codifica possibile dell'informazione emessa dalla sorgente. Per fare questo sia 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

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

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 ε.

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

$$\liminf_{\varepsilon \to 0} \liminf_{t \to \infty} R(G, P, t, \varepsilon) \tag{2}$$

misurerà quindi la complessità del grafo.

La definizione (2) non consente però il calcolo dell'entropia di 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. Abbiamo bisogno di una delle possibili definizioni di potenza di grafo.

Definizione. Sia t un intero positivo e G un grafo. Chiameremo *$t-$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. TODO

Osserviamo che la coppia $(G^t, \mathbf{P}^t)$, dove $\mathbf{P}^t$ è definita da (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.

Dovendo codificare propriamente un sottoinsieme $U \subset V^t$ avremo bisogno di almeno tante parole quanti insiemi indipendenti al minimo partizionano U. Ma questo è proprio il numero cromatico del sottografo indotto da G^t su U, che denoteremo $\chi(G^t(U))$. Siamo ora pronti per enunciare la prima definizione di entropia di grafo.

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

$$H(G, P) = \lim_{t \to \infty} \min_{\substack{U \subset V^t \\ \mathbf{P}^t(U) > 1-\varepsilon}} \frac{1}{t} \log \chi(G^t(U)) \tag{3}$$

Tale limite non dipende da $0 < \varepsilon < 1$.

Perché la precedente risulti una buona definizione dovremmo dimostrarne l'indipendenza da ε. Salteremo questa verifica e dimostreremo direttamente nella prossima sezione l'equivalenza con la seconda definizione, la quale non dipende da ε. Ne seguirà la buona definizione.

2.2 Definizione in termini di Mutua Informazione

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

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

Definizione. Siano X ed Y due variabili aleatorie discrete. Chiameremo *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. Chiameremo *entropia di grafo* il seguente minimo:

$$H(G, P) = \min I(X; Y) \tag{4}$$

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

Perché la precedente risulti una buona definizione dovremmo dimostrare che tale minimo esiste. Rinviamo alla prossima sezione per una dimostrazione di questo fatto.

2.3 Definizione in termini di Politopo dei Vertici

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

Esempio. Sia G il grafo semplice non diretto su 3 vertici con 2 archi. A meno di isomorfismo siano $V = \{1, 2, 3\}$ ed $E = \{(1, 2), (1, 3)\}$. Gli insiemi indipendenti di G sono allora $\{\varnothing, \{1\}, \{2\}, \{3\}, \{2, 3\}\}$. 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 $\mathrm{VP}(G)$.

TODO: Disegno

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

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

$$H(G, P) = \min_{\substack{\mathbf{a} \in \mathrm{VP}(G) \\ \mathbf{a} > 0}} \sum_{i=1}^{n} p_i \log \frac{1}{a_i} \tag{5}$$

Osservazione. La funzione obiettivo da minimizzare in (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 segue che tale minimo esiste finito sul bordo di $\mathrm{VP}(G)$ e che $\mathbf{a}$ che realizza il minimo ha tutte le coordinate maggiori di 0.

3 Equivalenza delle tre definizioni

3.1 Numero Cromatico e Mutua Informazione

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

Lemma 3.1 *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 possa 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 3.2 *Sia G un grafo e siano v, w suoi vertici. Diciamo che*

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

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 e 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. Chiameremo *$t-$esima potenza normale* di G il grafo $G_t = (V^t, E_t)$, dove V^t è il prodotto cartesiano di t copie di V, mentre

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

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

Esempio. TODO

Osserviamo che la coppia $(G_t, \mathbf{P}^t)$, dove $\mathbf{P}^t$ è definita da (1), è un grafo probabilistico. Sia $0 < \varepsilon < 1$, 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 necessarie in una codifica propria di quasi tutte le sequenze di lunghezza t, in analogia con quanto discusso nel paragrafo 2.1.

Enunciamo due lemmi omettendone la dimostrazione. Diamo inoltre alcune necessarie definizioni.

Lemma 3.3 *Sia G unione di grafi completi a due a due disgiunti e $\sim$ relazione d'equivalenza come nel precedente lemma. Denotiamo con $[v]$ la classe d'equivalenza di v e definiamo*

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

Allora avremo

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

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

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

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

per una qualche costante K.

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

Lemma 3.4 *Sia $T^t(P)$ l'insieme delle sequenze $P-$tipiche. Allora*

1. *Per ogni λ 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 $2^{-tH(P) - C\sqrt{n}} \leq \mathbf{P}^t(\mathbf{v}) \leq 2^{-tH(P) + C\sqrt{n}}$*
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. Chiameremo *grafo dei nuclei* il grafo Γ d'insieme dei vertici le coppie del tipo (v, A) con A nucleo di G contenente v. 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. Chiameremo *distribuzione ausiliaria ammissibile* una distribuzione di probabilità Q su $V(\Gamma)$ che soddisfi

$$\sum_{A:\ v\in A} Q(v, A) = P(v). \tag{6}$$

Converremo di chiamare $\mathcal{N}$ l'insieme dei nuclei di G. Una distribuzione ausiliaria ammissibile definisce su $\mathcal{N}$ una distribuzione marginale R tramite

$$R(A) = \sum_{v:\ v\in A} Q(v, A). \tag{7}$$

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

$$I(Q) = I(X;Y) = \sum_{v\in V}\sum_{A\in\mathcal{N}} Q(v, A) \log \frac{Q(v, A)}{P(v)R(A)}. \tag{8}$$

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 (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 (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.

Teorema 3.5 *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 in una codifica propria rispetto al prodotto conormale, tranne per un insieme di probabilità totale ε.

Dimostrazione. Per il punto 1 del lemma 3.4, 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 $\mathbf{P}^t(U_t) \geq 1 - \varepsilon$. Semplici considerazioni insiemistiche danno allora

$$1 - \varepsilon - \lambda \leq \mathbf{P}^t(U_t \cap T^t(P)). \tag{9}$$

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

$$\mathbf{P}^t(U_t \cap T^t(P)) \leq \chi(U_t \cap T^t(P)) \cdot \max_{\mathbf{N}\in\mathcal{N}^t} \mathbf{P}^t(\mathbf{N} \cap T^t(P)) \tag{10}$$

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

$$\max_{\mathbf{N} \in \mathcal{N}^t} \mathbf{P}^t(\mathbf{N} \cap T^t(P)) \leq \max_{\mathbf{v} \in T^t(P)} \mathbf{P}^t(\mathbf{v}) \cdot \max_{\mathbf{N} \in \mathcal{N}^t} \left|\mathbf{N} \cap T^t(P)\right|. \tag{11}$$

Sia ora $\mathbf{N}$ che renda massima $\left|\mathbf{N} \cap T^t(P)\right|$ e sia $\mathbf{v} \in T^t(P) \cap \mathcal{N}$. Siano $w \in V$ e M un nucleo di G contenente w, denotiamo allora con $N(w, M \mid \mathbf{v}, \mathbf{N})$ 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 $V(\Gamma)$

$$Q(w, M) = \frac{N(w, M \mid \mathbf{v}, \mathbf{N})}{N(w \mid \mathbf{v})} \cdot P(w). \tag{12}$$

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

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

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

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

dove la diseguaglianza centrale segue dalla concavità della radice.

Nel corso della dimostrazione del lemma 3.3 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 3.2 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 (12), è un grafo probabilistico. Perciò siamo nelle condizioni di applicare il lemma 3.3 e ottenere che il numero di sequenze $Q-$tipiche in ogni nucleo è compreso nell'intervallo

$$\left[2^{tH(Q|\sim)-K_2\sqrt{t}},\ 2^{tH(Q|\sim)+K_2\sqrt{t}}\right]. \tag{13}$$

Esistono al più $(t+1)^{|V(\Gamma)|}$ distribuzioni ausiliarie di probabilità del tipo (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 (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

$$\max_{\mathbf{N}\in\mathcal{N}^t} \left|\mathbf{N} \cap T^t(P)\right| \leq (t+1)^{|V(\Gamma)|} \cdot 2^{t \max_{Q\in\mathcal{A}} H(Q|\sim)+K_1\sqrt{t}}, \tag{14}$$

perché le Q sono distribuzioni ammissibili nel senso della definizione (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 la parte 2 del lemma 3.4 abbiamo

$$\max_{\mathbf{v}\in T^t(P)} \mathbf{P}^t(\mathbf{v}) \leq 2^{-tH(P)-C\sqrt{t}}. \tag{15}$$

Possiamo ora concludere. Dalle disuguaglianze (9) $-$ (11), (14) e (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 3.6 *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 $\mathbf{P}^t(U_t) \geq 1-\varepsilon$ ed inoltre*

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

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

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

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

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

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

Osserviamo che vale

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

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

$$\sum_{\mathbf{v}\in V^t} R^*_M(C_{\mathbf{v}})\cdot\mathbf{P}^t(\mathbf{v}) = \sum_{\mathbf{v}\in T^t(P)} R^*_M(C_{\mathbf{v}})\cdot\mathbf{P}^t(\mathbf{v}) \;+\; \sum_{\mathbf{v}\in\overline{T^t(P)}} R^*_M(C_{\mathbf{v}})\cdot\mathbf{P}^t(\mathbf{v}) \qquad (16)$$

ed osservare che il secondo termine è maggiorato da $\mathbf{P}^t\left(\overline{T^t(P)}\right)$, che per il punto 1 del lemma 3.4 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 abbiano massa totale 1. Maggioriamo ancora questo ultimo 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 3.3 il numero di sequenze $Q-$tipiche soddisfa

$$\big|T^t(Q)\big| \geq 2^{tH(Q|\sim)-K_3\sqrt{t}}. \tag{17}$$

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 (6) la probabilità totale di una classe d'equivalenza è $P(v)$, quindi vale

$$H(Q\,|\sim) = \sum_{v\in V}\sum_{A\in\mathcal{N}} Q(v,A)\log\frac{P(v)}{Q(v,A)}. \tag{18}$$

Inoltre il punto 2 lemma 3.4 implica che se $\mathbf{A}$ è una sequenza di nuclei $R-$tipica allora

$$R^*(\mathbf{A}) \geq 2^{-tH(R)-K_4\sqrt{t}}. \tag{19}$$

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 (17). Abbiamo stimato la probabilità di ciascuna in (19), per cui possiamo concludere che $R^*(\mathcal{N}_\mathbf{v})$ sia maggiore del loro prodotto.

Per quanto appena visto otteniamo

$$\max_{\mathbf{v}\in T^t(P)} R^*_M(C_\mathbf{v}) \leq \left(1-\exp_2\left[-tH(R)-K_4\sqrt{t}+tH(Q\,|\sim)-K_3\sqrt{t}\right]\right)^M \tag{20}$$

Mettendo insieme le equazioni (7) e (18) osserviamo che $H(Q\,|\sim)-H(R)$ si semplifica in $H(G,P)$. Possiamo quindi riscrivere

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

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 (16) sia infinitesimo per $t\to\infty$. Quindi per ogni ε e δ esistono M nuclei che ricoprono un sottografo di probabilità almeno $1-\varepsilon$. Ma per il lemma 3.1 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

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

□

3.2 Mutua Informazione e Politopo dei Vertici

Teorema 3.7 *Sia G un grafo e siano $S(G)$ i suoi insiemi indipendenti. Sia P una densità discreta sui vertici di G. Avremo allora:*

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

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 $\mathrm{VP}(G)$ perché combinazione convessa di vettori delle caratteristiche degli insiemi indipendenti. Ne segue

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

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 $\mathbf{a} \in \mathrm{VP}(G)$. 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:

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

□