



Teoria dei grafi

Andrea Cosentino

13 agosto 2024

Indice

1	Prima lezione	4
2	Seconda lezione	9
2.1	Connettività di un grafo	11
3	Terza lezione	13
3.1	Cammino euleriano	13
3.2	Ciclo hamiltoniano	14
3.3	Grafo bipartito	15
4	Quarta lezione	17
4.1	Parametri dei grafi	17
4.1.1	Numero di indipendenza	17
4.1.2	Numero di clique	18
4.1.3	Numero cromatico	19
5	Quinta lezione	22
5.1	Numero di dominazione	25
6	Sesta lezione	26
6.1	Proprietà numero di dominazione	26
6.2	Grafi casuali	29
7	Settima lezione	31
7.1	Strumenti di probabilità	31
7.2	Proprietà grafi	31
8	Ottava lezione	34
8.1	Proprietà asintotiche	36

9	Nona lezione	39
9.1	Introduzione	39
9.2	Ripasso di algebra lineare	40
10	Decima lezione	43
10.1	Clustering Spettrale	45
10.2	Matrice laplaciana	47
11	Undicesima lezione	49
11.1	Legame matrice laplaciana e autovalori	49
11.2	Disuguaglianza di Cheeger	51
12	Dodicesima lezione	53
13	Tredicesima lezione	57
13.1	Grafi non regolari	59
13.2	Introduzione clique piantate	61
14	Quattordicesima lezione	62
15	Quindicesima lezione	67
15.1	Caso $k = \Omega(\sqrt{n})$	67
15.2	Algoritmo Improved Clique Finder	70
16	Sedicesima lezione	72
16.1	Introduzione camminate casuali su grafi	72
16.2	Relazione tra autovalori e grado del grafo	72
16.3	Risultati utili su autovettori e autovalori	73
17	Diciassettesima lezione	77
17.1	Definizione camminata casuale	77
17.2	Distribuzione stazionaria	78
17.3	Convergenza della camminata casuale alla distribuzione stazionaria	78
18	Diciottesima lezione	81
18.1	Dimostrazione	81
18.2	Velocità di convergenza	83
18.3	Camminata pigra	84

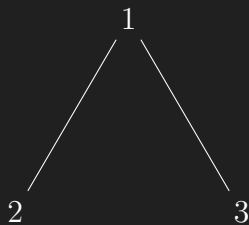
19 Diciannovesima lezione	86
19.1 Tempo di Mixing	86
19.1.1 Caratterizzazione autovalori	87
19.1.2 Tempo di Mixing su clique	88
19.1.3 Tempo di Mixing su grafo a stella	88
19.1.4 Tempo di mixing grafo a manubrio	90
20 Ventesima lezione	92
20.1 Algoritmi distribuiti	92
21 Ventunesima lezione	97
21.1 Decisioni sequenziali	97

Capitolo 1

Prima lezione

Un grafo $G = (V, E)$ è una struttura algebrica dove V è l'insieme finito di vertici e E è l'insieme finito di archi. Inoltre, vale che $E \subseteq [V]^2$. Dato un insieme S e un qualunque intero $k \in \{2, \dots, |S|\}$, diciamo che $[S]^k$ è la collezione di tutti i sottoinsiemi di S formati da k elementi. Per esempio, dato l'insieme $S = \{1, 2, 3\}$ l'insieme $[S]^2$ contiene $\{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$.

Esempio 1. Un esempio di grafo $G = (V, E)$ è $V = \{1, 2, 3\}$ e $E = \{\{1, 2\}, \{1, 3\}\}$.

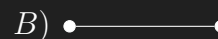


Notare che ci concentriamo su grafi con archi non orientati.

La nomenclatura che utilizzeremo per indicare dei vertici generici

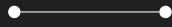
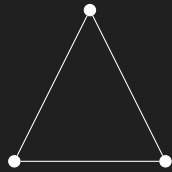
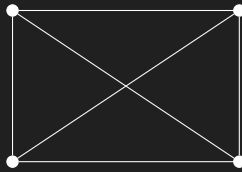
è i, j, u, v , mentre per indicare degli archi generici è (i, j) . Dire (i, j) implicherebbe un ordine, per evitare di scrivere $\{i, j\}$ useremo (i, j) senza implicare che l'arco sia orientato.

Il numero di nodi del grafo è detto **ordine**, e corrisponde a $|V|$. Un grafo di ordine 0 è detto grafo **vuoto**, mentre un grafo di ordine ≤ 1 è detto grafo **banale**. Esistono solamente due grafi di ordine 2:



Dato un arco $e = (i, j) \in E$ diciamo che i, j sono vertici incidenti all'arco e . Due vertici i, j con $i \neq j$ tali che $(i, j) \in E$ sono detti vertici **adiacenti** in $G(V, E)$. Se $E \equiv [V]^2$ diciamo che il grafo è **completo** oppure che è una **clique** (o cricca in italiano). Un grafo completo su n verti-

ci è chiamato K_n . Alcuni esempi di grafi completi sono

 K_2  K_3  K_4

Un grafo completo su n vertici ha un numero di archi pari a

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

I grafi che consideriamo sono non orientati e **semplici**. Un grafo è semplice se non ha loops (o cappi)

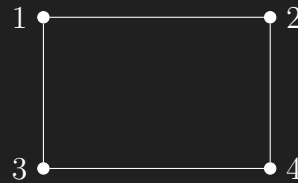


e non ha archi multipli, ovvero tra due nodi o c'è un arco non ce n'è neanche io. Quindi la situazione in figura non è ammessa.

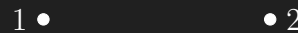


Il sotto-grafo di un grafo $G = (V, E)$ è $G' = (V', E')$ tale che $V' \subseteq V$ e $E' \subseteq E[V']^2$. Nella seconda condizione imponiamo che se vogliamo avere l'arco (i, j) nel grafo, allora $i, j \in V'$. Senza questa condizione non otterremmo un grafo.

Esempio 2. Dato il grafo



Se la seconda condizione fosse solamente $E' \subseteq E$ potremmo scegliere $V' = \{1, 2\}$ ed $E' = \{(1, 2), (2, 3)\}$, ma siccome 3 non è un nodo, il risultato non è un grafo. Un esempio di sotto-grafo è $V' = \{1, 2, 3, 4\}$, $E' = \{(3, 4)\}$



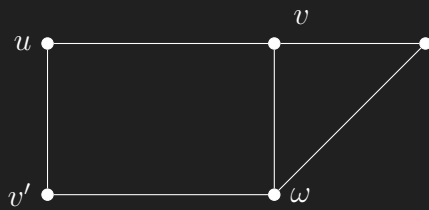
Dato $V' \subseteq V$ il sotto-grafo G' **indotto** da V' è $G'(V', E')$ con $E' = E[V']^2$. Ovvero, se seleziono i vertici seleziono anche gli archi su cui sono incidenti. Dato l'insieme di vertici V' c'è solo un sotto-grafo indotto.

Dato il grafo $G(V, E)$ il vicinato di $N(v)$ di $v \in V$ in G è

$$N(v) = \{j \in V : (v, j) \in E\}$$

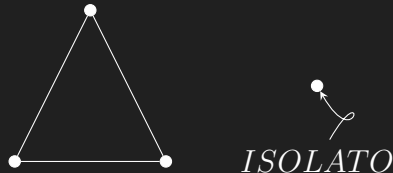
Cioè tutti i nodi connessi a v con un arco.

Esempio 3. Dato il grafo



Il vicinato di v è $N(v) = V \setminus \{v\}$ mentre il vicinato di v' è $N(v') = \{u, v, \omega\}$.

Il grado di v in G è $d(v) = |N(v)|$. Se v ha $d(v) = 0$ in G allora si dice **isolato**.



Definiamo il grado minimo come

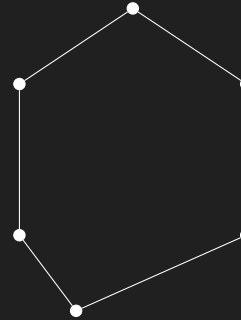
$$\delta(G) = \min d(v) : v \in V$$

e il grado massimo

$$\Delta(G) = \max\{d(v) : v \in V\}$$

Se $\Delta(G) = \delta(G) = k$ allora G è k -regolare.

Esempio 4. Il seguente grafo è 2-regolare



Il grado medio è

$$D(G) = \frac{1}{|V|} \sum_{v \in V} d(v)$$

Vale che $\delta(G) \leq D(G) \leq \Delta(G)$. La **densità** è invece definita come

$$\varepsilon(G) = \frac{|E|}{|V|}$$

La densità ci dice quanti archi ha, in media, ciascun vertice. Assomiglia al grado medio ma in quest'ultimo contiamo due volte ogni arco. Infatti vale che

$$\begin{aligned} |E| &= \frac{1}{2} \sum_{v \in V} d(v) \\ &= \frac{1}{2} D(G) |V| \end{aligned}$$

e quindi

$$\varepsilon(G) = \frac{|E|}{|V|} = \frac{1}{2} D(G)$$

Fatto 1. In ogni grafo il numero di vertici di grado dispari è pari.

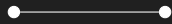
Dimostrazione 1. Cominciamo con l'osservare che $|E|$ è un numero intero, e siccome vale che $|E| = \frac{1}{2} \sum_{v \in V} d(v)$ allora anche $\frac{1}{2} \sum_{v \in V} d(v)$ è intero. Il valore $\sum_{v \in V} d(v)$ deve essere per forza pari, dato che la sua metà è intera. Dividiamo la sommatoria in due sommatorie:

$$\begin{aligned} & \sum_{v \in V: d(v) \text{ è pari}} d(v) \\ + & \sum_{v \in V: d(v) \text{ è dispari}} d(v) \end{aligned}$$

La sommatoria pari ha come risultato sicuramente un numero pari. Questo vuol dire che, se come risultato finale vogliamo un numero pari, anche la sommatoria dispari deve risultare pari. Ciò è possibile se e solo se il numero di elementi è pari. Infatti, sommando un numero pari di numero dispari otteniamo un numero pari. Quindi il numero di vertici di grado dispari è pari.

Ci poniamo adesso la domanda se la densità può scendere sotto il grado minimo. Vediamolo prima con un esempio

Esempio 5. Il seguente grafo



Ha $\delta(G) = 1$ e $\varepsilon(G) = \frac{1}{2}$, quindi $\delta(G) > \varepsilon(G)$

Fatto 2. $\forall G$ con almeno un arco, ha un sotto-grafo indotto H tale che

$$\delta(H) > \varepsilon(H) \geq \varepsilon(G)$$

Dimostrazione 2. Consideriamo una sequenza di grafi

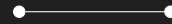
$$G = G_0, G_1, G_2, \dots$$

Dove $G_i = (V_i, E_i)$ e $V_0 \supseteq V_1 \supseteq V_2$, con G_i grafo indotto da V_i . Se $V_0 (= V)$ ha v_0 tale che $d(v_0) \leq \varepsilon(G_0)$ creiamo $V_1 = V_0 \setminus \{v_0\}$. Notiamo che se non esiste v_0 che rispetta la condizione, allora

$$\forall v \in V d(v) > \varepsilon(G_0)$$

e quindi $d(G_0) > \varepsilon(G_0)$. In questo caso avremmo già dimostrato il teorema con $H = G$.

Consideriamo adesso G_1 indotto da V_1 (ricordiamo che $V_1 = V_0 \setminus v_0$). Iteriamo svolgendo la stessa operazione di prima fino a quando V_i è tale che $\forall v \in V_i d(v) > \varepsilon(G_i)$. Notiamo che ci fermeremo prima di svuotare il grafo, infatti arriveremo al caso base



dove sappiamo che vale $\delta(G) > \varepsilon(G)$. Se G_{i+1} viene creato, allora

$$\begin{aligned} \varepsilon(G_{i+1}) &= \frac{|E_{i+1}|}{|V_{i+1}|} \\ &= \frac{|E_i - d(v_i)|}{|V_i - 1|} \geq \frac{|E_i - \varepsilon(G_i)|}{|V_i - 1|} \end{aligned}$$

Dove la disuguaglianza vale per la condizione con cui costruiamo il sotto-grafo.

$$= \frac{|E_i| - \frac{|E_i|}{|V_i|}}{|V_i| - 1} = \frac{|E_i||V_i| - |E_i|}{|V_i|(|V_i| - 1)}$$

dove abbiamo portato a fattore comune il numeratore.

$$= \frac{|E_i|(|V_i| - 1)}{|V_i|(|V_i| - 1)} = \varepsilon(G_i)$$

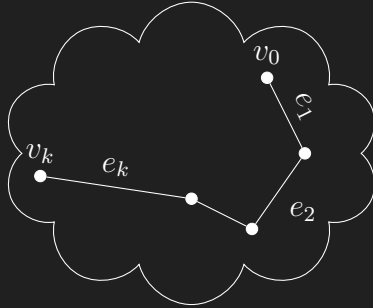
Quindi quando ci fermiamo avremo G_k tale che

$$\delta(G_k) > \varepsilon(G_k) \geq \varepsilon(G_0)$$

Capitolo 2

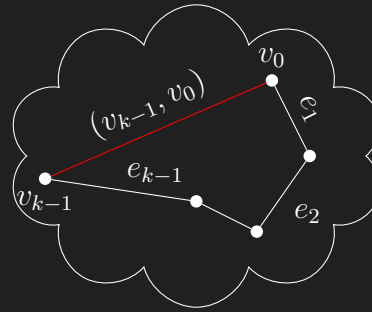
Seconda lezione

Un **cammino** di lunghezza $k \geq 0$ in $G = (V, E)$ è un sotto-grafo P_k con k archi e $k+1$ vertici distinti tale che $e_i = (v_{i-1}, v_i)$. Indichiamo gli archi con $e_1 \dots e_k$ e i nodi con v_0, \dots, v_k .



Usiamo la nuvoletta quando non ci interessa la struttura del grafo. Evidenziamo solo una certa parte. Nel caso in cui P_0 non abbiamo archi nel cammino ma un singolo vertice.

Un **ciclo** C_k di lunghezza $k \geq 3$ è formato da un cammino P_{k-1} che può essere esteso in G includendo l'arco (v_{k-1}, v_0) .

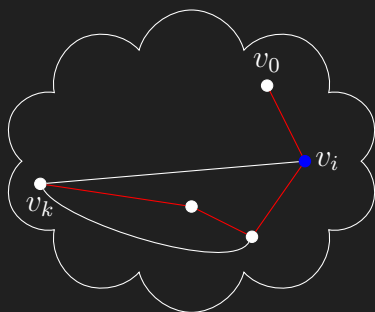


In un grafo G , il **calibro** $g(G)$ è la lunghezza del ciclo più breve. La **circonferenza** è la lunghezza del ciclo più lungo.

Fatto 3. $\forall G$ con $\delta(G) > 2$ contiene un cammino di lunghezza $\delta(G)$ e un ciclo di lunghezza almeno $\delta(G) + 1$.

Dimostrazione 3. Prendiamo il cammino più lungo del grafo, P_k . Allora tutti i vicini di P_k fanno parte del cammino, altrimenti potrei aggiungerli e allungarlo, P_k non sarebbe il più lungo. Quindi il cammino P_k è almeno lungo $|N(v_k)|$, dove v_k è l'ultimo nodo del cammino. Siccome per ipotesi $|N(v_k)| \geq \delta(G)$ allora esiste un cammino di lunghez-

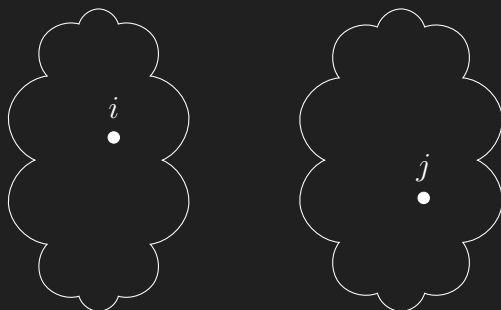
za $\delta(G)$. Consideriamo ora il primo vertice che è un vicino di v_k .



In rosso è evidenziato il cammino P_k e in blu il primo vertice che è vicino di v_k . Se consideriamo il cammino in rosso da v_i fino a v_k e aggiungiamo (v_k, v_i) troviamo un ciclo, ciò vale sempre per il fatto $\delta(G) \geq 2$. Il ciclo C è lungo almeno $N(v_k) + 1 \geq \delta(G) + 1$.

Dato $G = (V, E) \forall i, j \in V \exists d(i, j)$ se i, j sono connessi in G da almeno 1 cammino allora $d(i, j)$ è la lunghezza del cammino più breve, altrimenti è ∞ .

Esempio 6. Dato il grafo



La distanza tra i, j è $d(i, j) = \infty$.

Il **diametro** è definito come

$$diam(G) = \max_{i, j \in V} d(i, j) = \max_{i \in V} \max_{j \in V} d(i, j)$$

e il raggio

$$rad(G) = \min_{i \in V} \max_{j \in V} d(i, j)$$

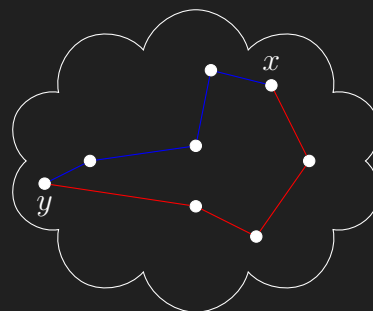
Il raggio lo possiamo vedere come il punto "più centrale". Sia x questo punto centrale, vale che $\forall v \in V d(x, v) \leq rad(G)$. Inoltre $rad(G) \leq diam(G)$ e questo è ovvio dato che il diametro è una massimizzazione del massimo, mentre il raggio è una minimizzazione del massimo. Possiamo anche dire che $diam(G) \leq 2rad(G)$, dato che

$$\begin{aligned} \forall u, v \in V d(u, v) &\leq d(u, x) + d(x, v) \\ &\leq rad(G) + rad(G) = 2rad(G) \end{aligned}$$

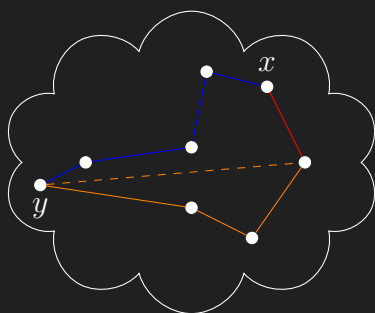
Fatto 4. $\forall G$ che ha almeno un ciclo soddisfa

$$g(G) \leq 2diam(G) + 1$$

Dimostrazione 4. Consideriamo il grafo



dove il ciclo C è il più corto, con lunghezza $g(G)$. I due vertici x, y sono vertici opposti, cioè tagliano il ciclo in due parti il più possibile uguali. Chiamiamo il percorso in rosso p_1 e il percorso in blu p_2 . Assumiamo per assurdo che $g(G) \geq 2diam(G) + 2$. Allora p_1, p_2 sono lunghi ciascuno almeno $diam(G) + 1$. Però $d(x, y) \leq diam(G)$ per la definizione stessa di diametro. Non tutti gli archi di P (cioè del percorso più breve) stanno su C , altrimenti il ciclo avrebbe lunghezza $2diam(G) + 1$. Quindi, possiamo costruire un ciclo più piccolo, prendendo gli archi che non stanno né su P_1 né su P_2 .



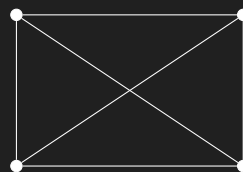
Il ciclo in arancione è più piccolo di C , quindi deve per forza valere che $g(G) \leq 2diam(G) + 1$.

2.1 Connettività di un grafo

Un grafo è **sconnesso** se $\exists i, j \in V \mid d(i, j) = \infty$. Una **componente** di un grafo è un qualunque insieme massimale di vertici connessi. Se un grafo è connesso il componente è

il grafo stesso. G è k -connesso se $|V| > k$ e $\forall X \subset V$ con $|X| < k$ il sotto-grafo indotto $V \setminus X$ è connesso. Se un grafo è k -connesso non possiamo sconnettere il grafo rimuovendo al più $k - 1$ vertici. Tutti i grafi sono 0-connessi. Se G è connesso è anche 1-connesso, tranne il caso K_1 (cricca di un elemento) perché non rispetta la condizione $|V| > 1$. Il massimo intero k tale che G è k -connesso è detta **connettività** di G , che denotiamo con $K(G)$. Vale che $K(K_n) = n - 1$.

Esempio 7. Nel caso di K_4

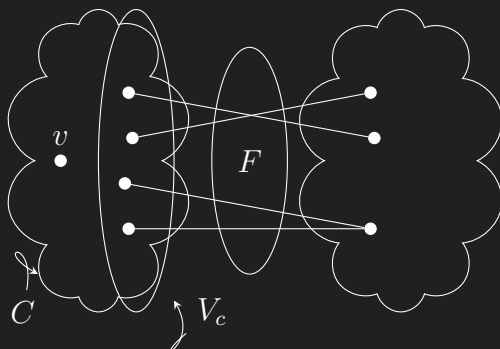


il numero di nodi che possiamo rimuovere è 3.

Teorema 1. Se $G \notin \{K_0, K_1\}$ (ovvero G non è un grafo banale), allora $K(G) \leq F \leq \delta(G)$ dove k è qualsiasi insieme minimo di archi la cui rimozione sconnette il grafo.

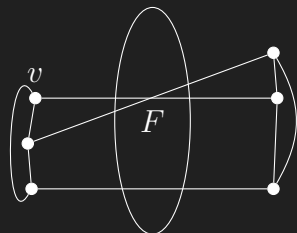
Dimostrazione 5. La disequazione $F \leq \delta(G)$ è banale. Infatti se sconnetto tutti gli archi attorno a un nodo ho sconnesso il grafo. Concentriamoci su $K(G) \leq F$ e distinguiamo due casi:

- G ha un vertice v che non è incidente a F .



dove C è la componente del grafo che ottengo quando rimuovo F e V_c è l'insieme dei nodi connessi agli archi in F . Siccome rimuovendo V_c sconnetto il grafo allora $K(G) \leq |V_c| \leq |F|$

□ G è tale che tutti i vertici sono incidenti con qualche arco in F .



Il grafo G ha connettività $K(G) \leq d(v)$. Siccome $d(v) = |F| = \delta(G)$ vale che $K(G) \leq |F|$.

Capitolo 3

Terza lezione

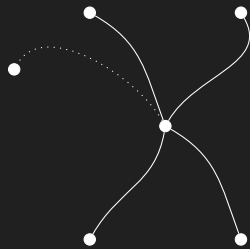
3.1 Cammino euleriano

Un cammino **chiuso** (in inglese closed walk) è un ciclo in cui i vertici non sono distinti. Un cammino chiuso si dice **euleriano** se attraversa tutti gli archi del grafo esattamente una volta. Un grafo è euleriano se ammette un cammino euleriano.

Teorema 2. Teorema di Eulero (1746)

Un grafo connesso è euleriano se e solo se ogni vertice ha grado pari.

Dimostrazione 6. Cominciamo con dimostrare il lato \Rightarrow del teorema. Quindi, dato un grafo connesso euleriano questo ogni vertice ha grado pari. Prendiamo un vertice che si trova sul cammino euleriano.



Se il cammino passa per il vertice, allora deve sia entrare che uscire. Non può esserci un arco che collega un vicino che non sia nel cammino. Quindi o un vertice è isolato oppure il cammino esce ed entra. Allora devono avere grado pari.

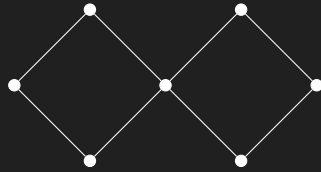
L'altro verso necessita un po' più di lavoro per essere dimostrato. Quello che vogliamo dimostrare è che se ogni vertice ha grado pari allora il grafo è euleriano. Facciamo una dimostrazione per induzione su $|E|$.

Caso base $|E| = 0$, banale. Implica che $|V| = 1$ perché parliamo di grafi connessi.

Ipotesi induttiva $|E| \geq 1$. Enunciamo un fatto utile.

Fatto 5. Se G ha tutti i vertici con grado pari con $E \leq 1$, posso trovare in G un cammino chiuso che non contiene un arco **più** di una volta.

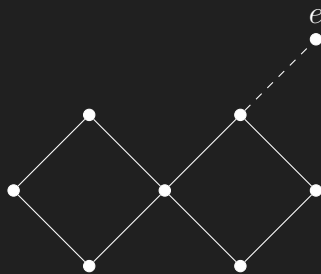
Sia ω un tale cammino di lunghezza massima. Ne rappresentiamo uno da esempio in figura.



Definiamo come F l'insieme degli archi di ω . Se $F \equiv E$ allora abbiamo finito, dato che tutti gli archi di G fanno parte del cammino ω . Assumiamo per assurdo che non sia così. Allora deve valere

$$E' \equiv E \setminus F \neq \emptyset$$

Notiamo che $\forall v \in V$ un numero pari di $u \in N(v)$ appartiene a F . Allora il sotto-grafo $G' = (V, E')$ ha tutti i vertici di grado pari (ricordiamo che 0 è pari). E' evidente che ci debba essere almeno un nodo e attaccato al cammino, altrimenti il grafo non sarebbe connesso.



Sia C la componente di G' che contiene e . C ha un numero di archi $< |E|$, dato che almeno un arco l'abbiamo rimosso. Per ipotesi induttiva C contiene un cammino euleriano. Ma allora possiamo costruire un cammino euleriano per G unendo ω

e il cammino trovato in C . Quindi abbiamo costruito un cammino più lungo di ω , contraddicendo l'ipotesi che sia massimo. Allora $F \equiv E$ e così abbiamo dimostrato il teorema.

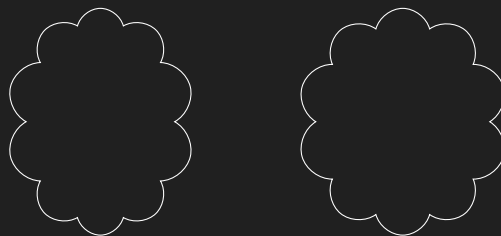
Se un grafo è euleriano possiamo trovare un cammino euleriano in tempo $O(|E|)$, i.e. in tempo lineare nella descrizione del grafo (algoritmo di Hierholzer).

3.2 Ciclo hamiltoniano

Un **ciclo hamiltoniano** è un ciclo che contiene tutti i vertici. Un grafo si dice hamiltoniano se contiene un ciclo hamiltoniano. Non è nota alcuna condizione necessarie e sufficienti affinché un grafo sia hamiltoniano. Sono note solamente condizioni sufficienti.

Teorema 3. Teorema di Dirac (1952). Un grafo $G = (V, E)$ con $|V| \geq 3$ e $\delta(G) \geq \frac{|V|}{2}$ è hamiltoniano.

Dimostrazione 7. Cominciamo con dimostrare che G deve essere connesso. Se per assurdo non lo fosse allora ha almeno due componenti.



Ogni componente è tale che $|C| \leq \frac{|V|}{2}$. Questo è ovvio, perché se una componente ne avesse più di $\frac{|V|}{2}$, un'altra dovrebbe averne di meno, e quindi non varrebbe la condizione $\delta(G) \geq \frac{|V|}{2}$. Notiamo ora che

$$\forall v \in C \quad d(v) \leq |C| - 1$$

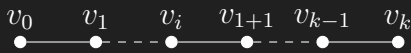
poiché al massimo un nodo può avere un arco con tutti gli altri nodi nella componente. Questa affermazione ci porta a poter dire che

$$d(v) \leq |C| - 1 < \frac{|V|}{2}$$

ovvero

$$d(v) < \frac{|V|}{2}$$

che viola le ipotesi. Sappiamo che G è connesso. Sia ora p un cammino di lunghezza massima in G con nodi v_0, v_1, \dots, v_k , con archi (v_i, v_{i+1}) , dove v_i viene detto **vertice sinistro** e v_{i+1} **vertice destro**.

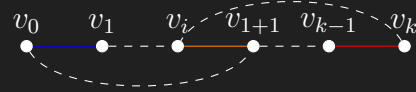


Tutti i vicini di v_0 e v_k sono nel cammino, altrimenti posso allungarlo aggiungendoli. Ricordiamo che vale anche $d(v_i) \geq \frac{n}{2}$ dove $n = |V|$. p non può avere più di n archi, quindi la sua lunghezza k è tale che

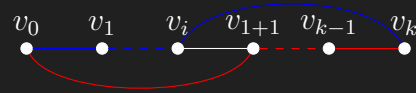
$$k \leq n - 1$$

dove non può essere $k \geq n$ sennò ripeterei dei nodi (e quindi non sarebbe un cammino). Ora associamo a

ogni vicino di v_0 l'arco a sinistra (per esempio a v_i associamo (v_{i-1}, v_i)) e a ogni vicino di v_k l'arco a destra. Per il principio della piccionaia c'è almeno un arco che è preso sia da un vicino di v_0 che da un vicino di v_k



Possiamo costruire un ciclo che va da v_0 a v_i , poi da v_i raggiunge v_k , da v_k a v_{i+1} e poi v_0 .



Se esistesse un vertice che non facesse parte di questo ciclo, sarebbe sicuramente un vicino, dato che il grafo è connesso. Ma allora potrei usarlo per allungare il percorso p violando l'ipotesi di massimalità. Quindi il ciclo passa per tutti i nodi, i.e. è hamiltoniano.

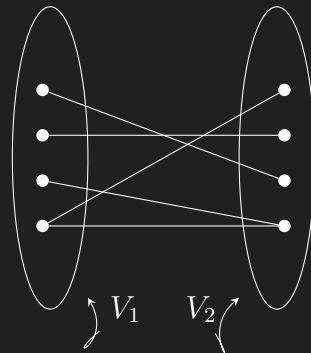
Il problema di determinare se un grafo G contenga un cammino hamiltoniano è NP-completo, questo spiega il motivo per cui non ci siano delle condizioni necessarie e sufficienti.

3.3 Grafo bipartito

Un grafo $G = (V, E)$ è detto **bipartito** se \exists una partizione V_1, V_2 di V

tali che $\forall (i, j) \in E \ i \in V_1 \wedge j \in V_2$ o viceversa. Ricordiamo che V_1 e V_2 in quanto partizione di V sono tali che $V_1 \cap V_2 \equiv V_1 \cup V_2 = \emptyset$.

Esempio 8. I grafi bipartiti sono usati per esempio su Tinder, Amazon e Netflix.



Possono contenere cicli, che sono sempre pari! Vale anche il viceversa, ovvero un grafo che contiene solo cicli di lunghezza pari è bipartito.

Capitolo 4

Quarta lezione

4.1 Parametri dei grafi

Possiamo definire informalmente un **parametro** come una proprietà. Ne abbiamo già viste alcune:

- Taglia
- Numero di lati
- Diametro
- Calibro

Formalmente un parametro è una funzione

$$\phi : \mathcal{G} \rightarrow \mathbf{R}$$

dove \mathcal{G} è la classe dei grafi non orientati e semplici. Altre possibili proprietà possono essere:

- G è euclideo?
- $K_3 \subseteq G$?

I parametri sono detti **invarianti**, ovvero mantengono lo stesso valore tra **isomorfismi** di grafi.

Definizione 1. Dati due grafi $G, H \in \mathcal{G}$ sono isomorfismi se $\exists f : V(G) \rightarrow V(H)$ con f biettiva tale che

$$\{(x, y)\} \in E(G) \iff \{f(x), f(y)\} \in E(H)$$

f è detto isomorfismo tra G e H .

Tutti i ϕ sono isomorfismi.

4.1.1 Numero di indipendenza

Definizione 2. $U \subseteq V$ è indipendente in $G = (V, E)$ se

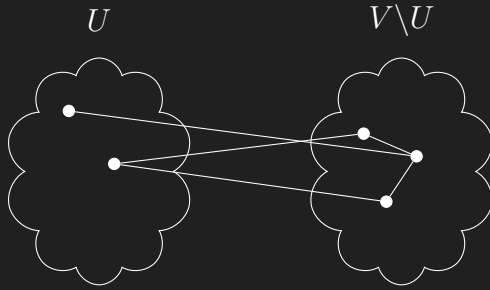
$$\forall x, y \in U \{x, y\} \notin E$$

In altre parole $G[U] = (V, \emptyset)$ è un grafo privo di lati.

Definizione 3. $\alpha(G)$ è detto **numero di indipendenza** ed è tale che

$$\alpha(G) := \max\{k \in \mathbb{N} \mid \exists U \subseteq V \text{ indipendente in } G \wedge |U| = k\}$$

Il problema di trovare il numero di indipendenza maggiore è NP-completo. Notiamo che $\alpha(G)$ sommato alla dimensione più piccola di vertex cover in G è uguale a $|V|$.



Notiamo infatti che i nodi in U hanno almeno un arco con un nodo in $V \setminus U$, altrimenti il grafo sarebbe sconnesso. Notiamo anche che, per definizione, i nodi in U non sono collegati tra loro. Inoltre, ogni nodo in $V \setminus U$ è collegato ad almeno un nodo in U , altrimenti U non sarebbe massimale. E' quindi evidente che per coprire tutti gli archi ho bisogno di tutti i nodi di $V \setminus U$. Assumiamo per assurdo che non sia così, e ci sia un nodo ω di $V \setminus U$ che non usiamo. Allora gli archi tra ω e U non sono coperti. Per metterci in un caso favorevole assumiamo anche che gli archi $(j, \omega) \in E$ tali che $j \in V \setminus U$ siano già coperti da j (se così non fosse avremmo bisogno di ω per coprirli e quindi avremmo dimostrato la sua necessità). I nodi in $i \in U$ tali che $(i, \omega) \in E$ devono essere parte della vertex cover, altrimenti non

copriamo degli archi. Tuttavia, se invece di scegliere ω scegliamo i , la cardinalità resterebbe la stessa. Ma se i coprisse due archi, (i, ω) e (i, ω_1) tale che $\omega_1 \in V \setminus U$, allora potremmo scegliere i per coprirli entrambi, ottenendo una cardinalità minore. Ma se ω, ω_1 hanno altri archi con nodi in U , allora questi sarebbero scoperti, e andrebbero coperti aggiungendo dei nodi, vanificando il vantaggio ottenuto. Se però non ne hanno altri allora ω, ω_1 formano, insieme a $U \setminus i$ un insieme indipendente più grande, e ciò non è possibile.

4.1.2 Numero di clique

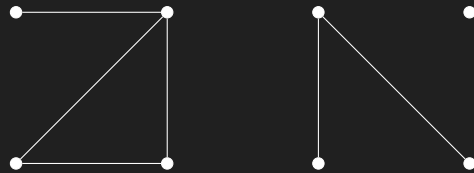
Definizione 4. Il numero di clique è definito come

$$\omega(G) := \max\{k \in \mathbb{N} \mid \exists U \subseteq V, \\ G[U] \text{ completo con } |U| = k\}$$

Il problema di trovare il numero di cricca massimo è NP-completo. Una cricca in $G = (V, E)$ è un insieme indipendente in \bar{G} (G **complemento**), dove \bar{G} è definito come

$$\bar{G} = (V, [V]^2 \setminus E)$$

Esempio 9. In figura G e il suo complemento.



4.1.3 Numero cromatico

Definizione 5. Una **coloratura** dei vertici $G = (V, E)$ è una funzione

$$c : V \rightarrow \{1, \dots, k\}$$

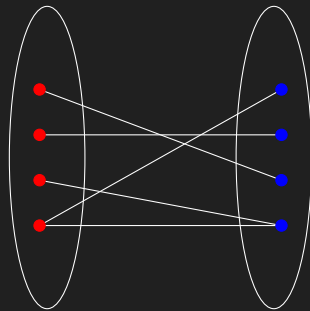
tale che $\{x, y\} \in E \Rightarrow c(x) \neq c(y)$.

La funzione c associa a ogni nodo un colore (indicato con un numero). Una coloratura con k colori è detta **k-coloratura**. Un grafo si dice **k-colorabile** se \exists k-coloratura c . Il **numero cromatico** è definito come

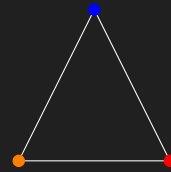
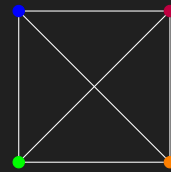
$$\chi(G) = \min \{k \in \mathbb{N} | G \text{ è k-colorabile}\}$$

Anche il problema di trovare $\chi(G)$ è NP-completo.

Esempio 10. Il caso in $k = 2$ è possibile se e solo se il grafo è bipartito.

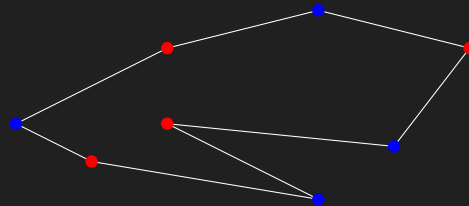


Esempio 11. Quanti colori ci servono per una cricca di n elementi?

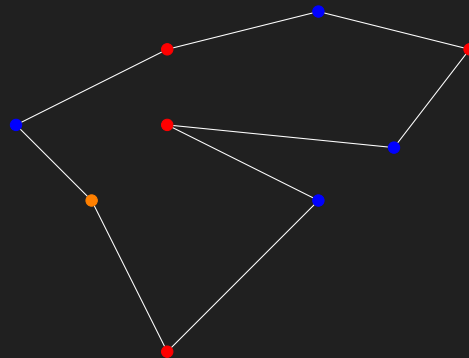

 K_3

 K_4

In generale per una cricca di n elementi ci servono n colori.

Esempio 12. Quanti colori ci servono per un ciclo? Se il ciclo è pari, C_{2n}



Ci servono 2 colori, se invece il ciclo è dispari, C_{2n+1}



Ci servono 3 colori.

Teorema 4. Teorema dei quattro colori.

$$\chi(G) \leq 4 \quad \forall G \text{ planare}$$

dove con **planare** intendiamo uno grafo tale per cui esiste una rappresentazione grafica in cui gli archi non si intersecano.

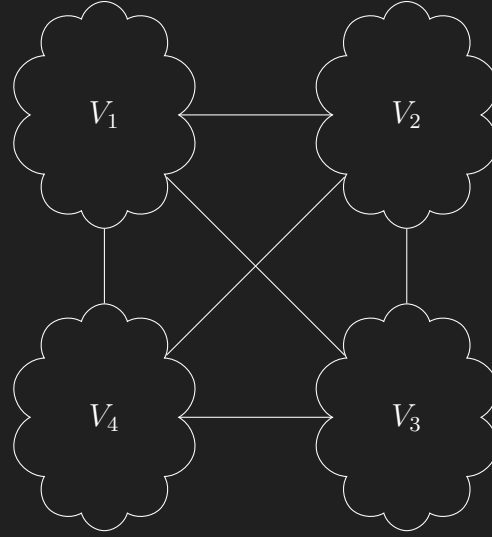
Fatto 6. Diamo un upper-bound per $\chi(G)$

$$\chi(G) \leq \frac{1}{2} + \sqrt{2|E| + \frac{1}{4}}$$

Dimostrazione 8. Poniamo $x = \chi(G)$. Se G è k -colorabile allora \exists partizione V_1, \dots, V_k tale che V_i è indipendente e $\forall i \ V_i = \{v \in V | c(v) = i\} = C^{-1}(i)$, dove $C^{-1}(i)$ è la preimmagine. c è la funzione che assegna a ogni partizione un colore,

$$c : V \rightarrow \{1, \dots, k\}$$

Non è possibile che esistano due partizioni non collegate tra loro, altrimenti avremmo un numero di insiemi indipendenti minore e k non sarebbe massimo.



Ovvero $\forall i \neq j \ \exists \geq 1$ lato tra V_i, V_j . E' ovvio che il numero di lati nel grafo è maggiore uguale del numero di coppie di partizioni presenti

$$|E| \geq \#coppie(V_i, V_j)$$

dove il numero di coppie è

$$\binom{k}{2} = \frac{k(k-1)}{2}$$

Quindi....

$$|E| \geq \frac{k(k-1)}{2}$$

$$2|E| \geq k(k-1)$$

$$k^2 - k - 2|E| \geq 0$$

prendiamo l'equazione associata $k^2 - k - 2|E| = 0$ e la risolviamo

$$k_{1,2} = \frac{1 \pm \sqrt{1 + 8|E|}}{2}$$

$$= \frac{1}{2} \pm \frac{\sqrt{1+8|E|}}{2} = \frac{1}{2} \pm \frac{\sqrt{4(\frac{1}{4} + 2|E|)}}{2}$$

$$\frac{1}{2} \pm \frac{2\sqrt{(\frac{1}{4} + 2|E|)}}{2} = \frac{1}{2} \pm \sqrt{(\frac{1}{4} + 2|E|)}$$

Quindi

$$\frac{1}{2} - \sqrt{(\frac{1}{4} + 2|E|)} \leq \chi(G)$$

$$\leq \frac{1}{2} + \sqrt{(\frac{1}{4} + 2|E|)}$$

Abbiamo dimostrato il teorema.

Notiamo grafi con gradi alti richiedono più colori rispetto a grafi con gradi più bassi.

Fatto 7. Vale quanto segue

$$\forall G \chi(G) \leq \Delta(G) + 1$$

Se G è una cricca o un ciclo di lunghezza dispari allora è un'uguaglianza.

Dimostrazione 9. Supponiamo v_1, \dots, v_n arbitrario. Allora procediamo in questo modo:

1. Assegniamo 1 a v_1
2. Se v_2 è vicino di v_1 assegniamo 2, altrimenti 1.

e così via, fino ad assegnare tutti gli n nodi. Siccome sappiamo che vale $\forall i \delta(v_i) \leq \Delta(G)$ per definizione,

allora se arriviamo all' i -esimo nodo, avendo assegnato già $\Delta(G) + 1$ colori diversi, allora non ne abbiamo bisogno di uno nuovo per v_i dato che al massimo $\Delta(G)$ vicini e ci sono $\Delta(G) + 1$ colori disponibili.

Fatto 8. Vale che

$$\forall G \chi(G) \cdot \alpha(G) \geq |V|$$

Dimostrazione 10. Poniamo

$k = \chi(G)$. Sia V_1, \dots, V_k una partizione indotta da una k -colorazione di G . Allora

$$\sum_{i=1}^k |V_i| = |V|$$

Sapendo che $|V_i| \leq \alpha(G)$ possiamo scrivere

$$|V| \leq \sum_i \alpha(G) = k\alpha(G) = \chi(G)\alpha(G)$$

Abbiamo dimostrato il fatto.

Capitolo 5

Quinta lezione

Terminiamo la lezione precedente enunciando un fatto:

Fatto 9. Vale quanto segue

$$\chi(G) \geq \omega(G)$$

Proseguiamo presentando il teorema di **Turán**.

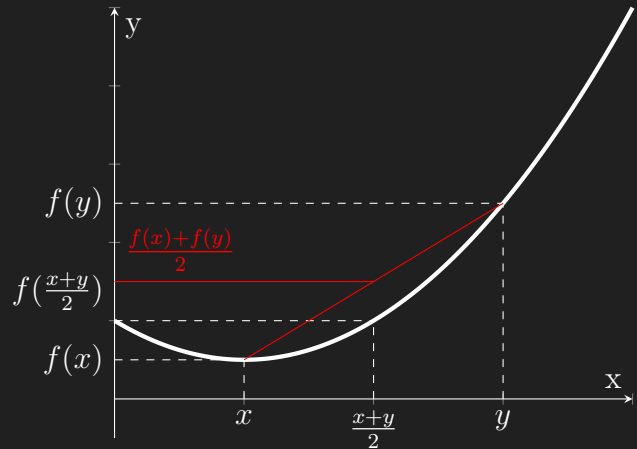
Teorema 5. Teorema di Turán. $\forall G = (V, E)$ vale che

$$\alpha(G)(d(G) + 1) \geq |V|$$

Dividendo per $d(G)+1$ ambo i membri otteniamo un minorante per $\alpha(G)$. Prima di dimostrare il teorema introduciamo la **disuguaglianza di Jensen**. Data una variabile aleatoria $X \in \mathbb{R}$ tale che $X \sim P$ (dove P possiamo pensarla come una distribuzione empirica. Diamo probabilità uniforme agli elementi. Il valore atteso diventa la media). Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione convessa. Allora vale che

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(x)]$$

Intuitivamente possiamo pensare



Notiamo che, visivamente, la disuguaglianza di Jensen ha senso, dato che

$$\frac{f(x) + f(y)}{2} \geq f\left(\frac{x+y}{2}\right)$$

Dimostrazione 11. Possiamo ora dimostrare il teorema di Turán. Facciamo una dimostrazione costruttiva, ovvero dimostriamo l'esistenza di un oggetto matematico creando un metodo per costruire tale oggetto. Nel nostro caso costruiremo una se-

rie di grafi, $G = G_1 = G_2 = \dots = G_i$ dove ogni grafo è tale che $G_i = (V_i, E_i)$.

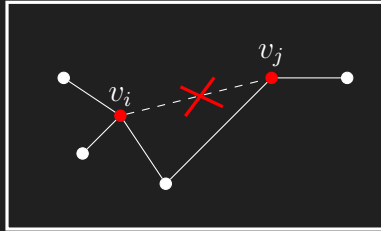
Algorithm 1

```

1:  $i \leftarrow 1$ 
2: while  $G_i \neq 0$  do
3:   Sia  $v_i \in \operatorname{argmin}_{v \in V_i} d_i(v)$ 
   ovvero un vertice di grado
   minimo
4:    $G_{i+1} \leftarrow G_i - C_i(v_i)$ 
5:    $i \leftarrow i + 1$ 
6: return  $\{v_1, \dots, v_i\}$ 

```

$C_i(v) = N_i(v) \cup \{v\}$ è definito come il **vicinato esteso** di v in G_i . Quello che fa l'algoritmo sopra descritto è prendere, a ogni iterazione, uno dei nodi con grado minimo e rimuoverlo, insieme a tutti i nodi vicini. I nodi v_i che scegliamo formano un insieme indipendente. Infatti non può esistere una situazione del tipo



Se tra v_i e v_j esistesse un arco, allora scegliendo v_i avremmo eliminato v_j (o viceversa), e quindi non avremmo potuto sceglierlo successivamente.

Sia m l'iterazione dell'algoritmo. Sappiamo che $m \leq |V|$ e che $m \leq \alpha(G)$, dato che $\{v_1, \dots, v_m\}$ forma-

no un insieme indipendente. Definiamo

$$Q(G) = \sum_{v \in V_g} \frac{1}{1 + D_G(v)}$$

A ogni passo G_i decresce, quindi

$$Q(G_1) - Q(G_2) \geq 0$$

Cerchiamo un upper-bound per questa sottrazione. Per definizione scriviamo

$$Q(G_1) - Q(G_2) = \sum_{u \in G_1(v_1)} \frac{1}{1 + d(u)}$$

Sappiamo che a ogni iterazione il numero di vicini di un grafo può solo diminuire, quindi

$$\sum_{u \in G_1(v_1)} \frac{1}{1 + d(u)} \leq \sum_{u \in G_1(v_1)} \frac{1}{1 + d_1(u)}$$

dove abbiamo utilizzato d_1 al posto di d . Sappiamo inoltre che, per ogni scelta di v_i , vale $d_i(u) \geq d_i(v_i) \forall u \in V$ per la definizione stessa di v_i .

$$\sum_{u \in G_1(v_1)} \frac{1}{1 + d_1(u)} \leq \sum_{u \in G_1(v_1)} \frac{1}{1 + d_1(v_1)}$$

$$= \frac{|C_1(v_1)|}{1 + d_1(v_1)} = \frac{1 + d_1(v_1)}{1 + d_1(v_1)} = 1$$

Riprendiamo la definizione di $Q(G)$

$$Q(G) = \sum_{v \in V_g} \frac{1}{1 + d_G(v)}$$

e riscriviamola come

$$\begin{aligned} &= \sum_{i=1}^m \sum_{u \in C(v_i)} \frac{1}{1 + d(u)} \\ &\leq \sum_{i=1}^m 1 = m \leq \alpha(G) \end{aligned}$$

Se dividiamo l'equazione originale per $|V|$ otteniamo

$$\frac{Q(G)}{|V|} = \frac{\sum_{v \in V_g} \frac{1}{1 + d(v)}}{|V|}$$

Se poniamo $f(x) = \frac{1}{1+x}$

$$\frac{Q(G)}{|V|} = \frac{\sum_{v \in V_g} f(d(v))}{|V|}$$

Siccome $\frac{1}{1+x}$ è convessa per $x > -1$ possiamo applicare la disuguaglianza di Jensen. Poniamo $X \sim \text{Unif}(V)$ e $Y = d(X)$, quindi

$$\mathbb{E}[Y] = \sum_{v \in V} d(v) \cdot \frac{1}{|V|} = \frac{1}{|V|} \sum_{v \in V} d(v)$$

e

$$\mathbb{E}[f(Y)] = \sum_{v \in V} \frac{1}{|V|} \cdot f(d(v))$$

$$\begin{aligned} &= \frac{1}{|V|} \sum_{v \in V} \frac{1}{1 + d(v)} \geq f(\mathbb{E}[Y]) \\ &= \frac{1}{1 + \frac{1}{|V|} \sum_{v \in V} d(v)} = \frac{1}{1 + d(G)} \end{aligned}$$

Mettendo tutto assieme

$$\alpha(G) \geq Q(G) \geq \frac{|V|}{1 + d(G)}$$

quindi

$$\alpha(G)(d(G) + 1) \geq |V|$$

Inoltre possiamo dire che $d(G)$ è approssimabile in tempo polinomiale con fattore $O(\log n)$

Presentiamo il seguente fatto:

Fatto 10. La seguente disuguaglianza vale $\forall G = (V, E)$

$$\omega(G)(|V| - d(G)) \geq |V|$$

Dimostrazione 12. Poniamo $n = |V|$ e $\bar{G} = (V, \bar{E})$. Sappiamo che

$$d_{\bar{G}}(v) = n - 1 - d_G(v)$$

Quindi

$$d(\bar{G}) = \frac{1}{n} \sum_{v \in V} d_{\bar{G}}(v) = \frac{1}{n} \sum_{v \in V} (n - 1 - d(G))$$

Siccome un insieme indipendente in \bar{G} corrisponde a una clique in G , possiamo applicare il teorema di Turan

$$\omega(G) = \alpha(\bar{G}) \geq \frac{n}{1 + d(\bar{G})} = \frac{n}{n - d(G)}$$

E a questo punto abbiamo concluso la dimostrazione, infatti muovendo il denominatore

$$\omega(G)(|V| - d(G)) \geq |V|$$

5.1 Numero di dominazione

Definizione 6. Il numero di dominazione è

$$\gamma(G) := \min\{k \in \mathbb{N} \mid \exists U \subseteq V \text{ dominante, } |U| = k\}$$

Dove un **insieme di dominazione** è $U \subseteq V$ tale che ogni vertice in $V \setminus U$ ha un unico vicino in U . Un vertice domina se stesso.

Il problema di trovare $\gamma(G)$ è NP-completo.

Notiamo, inoltre, che se un insieme indipendente è dominante allora non domina alcun vertice dell'insieme indipendente.

Capitolo 6

Sesta lezione

6.1 Proprietà numero di dominazione

Continuiamo la trattazione del numero di dominazione, iniziata nella scorsa lezione.

Fatto 11. $\forall G = (V, E)$ vale che

$$\gamma(G) \leq \alpha(G)$$

Dimostrazione 13. Sia $U \subseteq V$ indipendente e $|U| = \alpha(G)$. Per assurdo, supponiamo che U non sia dominante. Allora $\exists x \in V \setminus U$ tale che non ha vicini in U . Questo implica che x non è dominato da U . Allora $U \cup \{x\}$ è indipendente e ha cardinalità $\alpha(G) + 1$, ma questo è assurdo, perché U non sarebbe l'insieme indipendente massimo. Allora U deve essere dominante. Per definizione stessa di $\gamma(G)$ vale che

$$|U| = \alpha(G) \leq \gamma(G)$$

Infatti, l'insieme dominante minimo al massimo ha come cardinalità U , dato che questo è dominante, e il

minimo può essere solo uguale o più piccolo.

Il prossimo teorema mostra che se tutti i vertici di un grafo hanno un grado alto, allora il numero di dominazione deve essere piccolo.

Teorema 6. (Arnaudov, 1974; Pagan, 1975; Lovász 1966) Vale la seguente disequazione

$$\gamma(G) \frac{1 + \delta(G)}{1 + \ln(1 + \delta(G))} \leq |V|$$

Dimostrazione 14. Facciamo una dimostrazione per costruzione. Cominciamo con il porre $n = |V|$ e $\delta = \delta(G)$.

Algorithm 2

```

1:  $S \leftarrow \emptyset$ 
2:  $U \leftarrow V$ 
3: while  $U \neq \emptyset$  do
4:    $v' \in \operatorname{argmax}_{v \in V} |U \cap C(v)|$ 
5:    $S \leftarrow S \cup \{v'\}$ 
6:    $U \leftarrow U \setminus C(v')$ 
7: return  $S$ 
```

Consideriamo un'iterazione qualsiasi dell'algoritmo. Sia U l'insieme dei vertici non ancora dominati all'inizio di tale iterazione e poniamo $r = |U|$. Allora

$$|U \cap C(v')| = \sum_{u \in U} \mathbb{I}\{u \in C(v')\}$$

Esplicitiamo v' come il nodo che massimizza la sommatoria

$$\max_{v \in V} \sum_{u \in U} \mathbb{I}\{u \in C(v)\}$$

Definiamo la variabile aleatoria $X \sim \text{Unif}(V)$, dove la distribuzione è uniforme su V . Sappiamo che il massimo è sempre maggiore o uguale della media. Allora possiamo scrivere

$$\geq \mathbb{E}\left[\sum_{v \in U} \mathbb{I}\{u \in C(X)\}\right]$$

Per la linearità del valore atteso scriviamo

$$= \sum_{v \in U} \mathbb{E}[\mathbb{I}\{u \in C(X)\}]$$

Il valore atteso di una funzione indicatrice è la probabilità che avvenga l'evento.

$$= \sum_{v \in U} P(u \in C(X))$$

Notiamo ora che se u fa parte del vicinato esteso di X , allora vale anche l'opposto, ovvero X fa parte del vicinato esteso di u .

$$= \sum_{v \in U} P(X \in C(u))$$

La probabilità che X faccia parte del vicinato esteso di u è semplicemente il numero di elementi in $C(u)$ fratto il numero totale di nodi.

$$= \sum_{v \in U} \frac{|C(u)|}{n} = \sum_{v \in U} \frac{1 + d(u)}{n}$$

Siccome $d(u) \geq \delta(G)$ per definizione

$$\geq \sum_{v \in U} \frac{1 + \delta(G)}{n}$$

Ci siamo liberati di v all'interno della sommatoria, quindi

$$= r \frac{1 + \delta(G)}{n}$$

Possiamo concludere che al termine dell'iterazione rimangono al massimo

$$r - r \frac{1 + \delta(G)}{n} = r \left(1 - \frac{1 + \delta(G)}{n}\right)$$

nodi. Notiamo che, partendo dall'inizio, dopo m iterazioni rimarranno

$$n \left(1 - \frac{1 + \delta(G)}{n}\right) \dots \left(1 - \frac{1 + \delta(G)}{n}\right)$$

$$= n \left(1 - \frac{1 + \delta(G)}{n}\right)^m$$

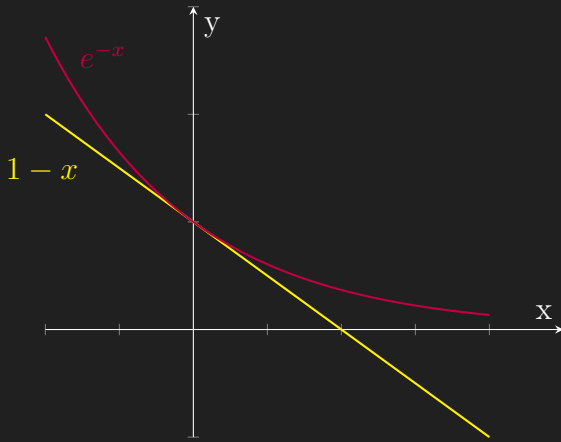
nodi. Ci chiediamo ora quando m è sufficiente affinché rimangano \leq

$\frac{n}{1+\delta(G)}$ vertici ancora da dominare.
Ovvero

$$n(1 - \frac{1+\delta(G)}{n})^m \leq \frac{n}{1+\delta(G)}$$

Applichiamo la nota disequazione

$$1 - x \leq e^{-x} \quad \forall x \in \mathbb{R}$$



$$\begin{aligned} & n(1 - \frac{1+\delta(G)}{n})^m \\ & \leq n \cdot \exp(-\frac{1+\delta(G)}{n}m) \end{aligned}$$

Usiamo \exp per evitare di avere e con esponente una frazione, ma è la stessa cosa.

$$n \cdot \exp(-\frac{1+\delta(G)}{n}m) \leq \frac{1}{1+\delta(G)}$$

Applichiamo la funzione logaritmo su ambo i lati

$$-\frac{1+\delta(G)}{n}m \leq -\ln(1+\delta(G))$$

$$m \geq n \frac{\ln(1+\delta(G))}{1+\delta(G)}$$

Assumiamo di essere arrivati all'iterazione $m = n \frac{\ln(1+\delta(G))}{1+\delta(G)}$. Sia s il numero di vertici ancora da dominare dopo m passi. Siccome m è la risposta alla domanda che ci siamo posti prima, vale che

$$s \leq \frac{n}{1+\delta(G)}$$

Nel caso peggiore l'algoritmo sceglierà altri s vertici per completare la costruzione dell'insieme dominante. Ma allora la cardinalità dell'insieme finale sarà

$$\gamma(G) \leq |S| \leq m + s$$

$$\leq n \frac{\ln(1+\delta(G))}{1+\delta(G)} + \frac{n}{1+\delta(G)}$$

Sapendo che $n = |V|$ e rigirando l'equazione

$$\gamma(G) \frac{1+\delta(G)}{1+\ln(1+\delta(G))} \leq |V|$$

6.2 Grafi casuali

Un **modello generativo** per grafi è una distribuzione di probabilità su tutti i grafi di un certo ordine. Il più famoso è il modello di **Erdős-Rényi**. Indichiamo con $\mathcal{G}(n, p)$ la distribuzione di probabilità su grafi di ordine n . Dato il grafo $G = (V, E)$, l'arco $\{i, j\} \in [V]^2$ è tale che $\{i, j\} \in E$ con probabilità p . L'estrazione è indipendente $\forall i \neq j$. Quindi

$$\mathbb{I}\{\{i, j\} \in E\} \sim \text{Bern}(p)$$

- Se $p = 0$ allora il grafo ottenuto ha sempre 0 lati.
- Se $p = 1$ allora il grafo ottenuto ha sempre tutti i lati in $[V]^2$.
- Se $0 < p < 1$ allora ogni grafo di ordine n ha probabilità > 0 di essere estratto da $\mathcal{G}(n, p)$

Notiamo che p regola la densità del grafo. La distribuzione $\mathcal{G}(n, \frac{1}{2})$ è uniforme su tutti i grafi di ordine n . Sia $H = (V, E)$ un grafo di ordine n . La probabilità di estrarre H è

$$P(H) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$$

Invece la probabilità di estrarre un grafo con k lati è

$$\begin{aligned} P(|E| = k) &= \sum_{\substack{E' \in [V]^2 \\ \wedge |E'| = k}} p^k (1-p)^{\binom{n}{2}-k} \\ &= \binom{|[V]^2|}{k} p^k (1-p)^{\binom{n}{2}-k} \\ &= \binom{\binom{n}{2}}{k} p^k (1-p)^{\binom{n}{2}-k} \end{aligned}$$

ed è una binomiale di parametri $\binom{n}{2}$ e p . Il valore atteso è

$$\mathbb{E}[|E|] = \binom{n}{2} p$$

ovvero il prodotto tra i parametri della distribuzione. Siccome a volte la dimostrazione costruttiva, usiamo il modello di Erdős-Rényi per dimostrare delle proprietà attraverso il **metodo probabilistico**. Ovvero, dato lo spazio di probabilità definito dalla tripla

$$(\Omega, F, P)$$

dove Ω è lo spazio campione (contiene gli eventi elementari), F lo spazio degli eventi ($F \subseteq \Omega$) e P la misura (distribuzione) di probabilità. Per esempio, sia Ω l'insieme dei grafi di ordine n . Ci chiediamo se ci siano dei grafi bipartiti. Sia A l'insieme dei grafi bipartiti di ordine n ($A \subseteq \Omega$) ci chiediamo

$$\exists \omega \in \Omega \text{ t.c. } \omega \in A?$$

Se $A \neq \emptyset$ allora esiste. Questo equivale a chiedersi $P(A) > 0$, che implica $\exists \omega \in \Omega$ che soddisfa la proprietà (notare che non è una co-implicazione!).

Capitolo 7

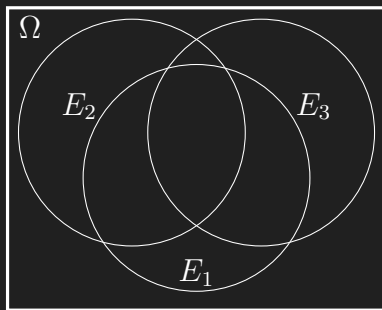
Settima lezione

7.1 Strumenti di probabilità

Introduciamo due importanti disequazioni che saranno utili per diverse dimostrazioni usando il metodo probabilistico. Sia (Ω, P) uno spazio di probabilità e siano E_1, \dots, E_n degli eventi qualsiasi. Per la **regola dell'unione** vale la disuguaglianza

$$P(E_1 \cup \dots \cup E_n) \leq \sum_{i=1}^n P(E_i)$$

Graficamente possiamo convincerci della correttezza della disequazione.



Le zone che sono in comune, sommando le probabilità dei singoli E_i ,

vengono contate più volte, quindi è ovvio che la sommatoria sia maggiore (o al massimo uguale) alla probabilità dell'unione. Data una variabile aleatoria X non negativa, $\forall a > 0$ vale

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Questa disuguaglianza è detta **disuguaglianza di Markov**. La dimostrazione è omessa.

7.2 Proprietà grafi

Fatto 12. $\forall n \geq 4$ e $\forall K \geq 2 \log n$ $\exists G$ di ordine n tale che $\alpha(G) < k$ e $\omega(G) < k$.

Notiamo che le due proprietà sembrano quasi contrastanti. Un insieme di indipendenza grande preclude una grande clique, e viceversa.

Dimostrazione 15. Consideriamo $G \sim \mathcal{G}(n, \frac{1}{2})$. Qual è la probabilità che $U \subset V$, con $|U| = k$, sia indipendente in G ? $(1 - \frac{1}{2})^{\binom{k}{2}}$. Quindi,

$$\begin{aligned}
& P(\alpha(G) > k) \\
&= P(\exists U \subseteq V \text{ t.c. } |U| = k, \\
&\quad \text{indipendente in } G) \\
&= P\left(\bigcup_{\substack{U \subseteq V \\ |U|=k}} \{U \text{ è indipendente in } G\}\right) \\
&\leq \sum_{\substack{U \subseteq V \\ |U|=k}} P(U \text{ è indipendente in } G) \\
&= \sum_{\substack{U \subseteq V \\ |U|=k}} 2^{-\binom{k}{2}} = \binom{n}{k} 2^{-\binom{k}{2}}
\end{aligned}$$

Usiamo ora il fatto che

$$\binom{n}{k} \leq \left(\frac{n}{2}\right)^k$$

per $4 \leq k \leq n$. Quindi:

$$\binom{n}{k} 2^{-\binom{k}{2}} \leq \left(\frac{n}{2}\right)^k 2^{-\frac{k(k-1)}{2}}$$

Sfruttando le proprietà dei logaritmi scriviamo

$$\begin{aligned}
2^{\log_2 \left(\frac{n}{2}\right)^k} 2^{-\frac{k(k-1)}{2}} &= 2^{\log_2 \left(\frac{n}{2}\right)^k - \frac{k(k-1)}{2}} \\
&= 2^{k(\log_2 n - 1) - \frac{k(k-1)}{2}}
\end{aligned}$$

Con la condizione che $k \geq 2 \log_2 n$

$$\geq 2^{\frac{k^2}{2} - k - \frac{k(k-1)}{2}} = 2^{-\frac{k}{2}}$$

Siccome $k \geq 4$, allora

$$2^{-\frac{k}{2}} < \frac{1}{2}$$

Dato $U \subseteq V$ con $|U| = k$ la probabilità che U formi una clique in G è $2^{-\frac{k}{2}}$. Rifacendo la stessa derivazione di prima arriviamo a dire che $P(\omega(G) \geq k) < \frac{1}{2}$. Quindi possiamo affermare

$$\begin{aligned}
& P(\alpha(G) < k, \omega(G) < k) \\
&= 1 - P(\alpha(G) \geq k \vee \omega(G) \geq k) \\
&\geq 1 - P(\alpha(G) \geq k) - P(\omega(G) \geq k) \\
&> 1 - \frac{1}{2} - \frac{1}{2} = 0
\end{aligned}$$

Abbiamo dimostrato che $P(\alpha(G) < k, \omega(G) < k) > 0$, questo implica che debba esistere un grafo G di ordine n tale che $\alpha(G) < k$ e $\omega(G) < k$.

Discutiamo un'altra proprietà che, similmente alla precedente, sembra essere contrastante.

Teorema 7. Teorema di Erdős.
 $\forall k \exists G$ tale che

$$g(G) > k \wedge \chi(G) > k$$

Prima di dimostrare il teorema introduciamo un lemma utile alla sua dimostrazione.

Lemma 1. Il valore atteso del numero di cicli di lunghezza k in $G \sim \mathcal{G}(k, n)$ è

$$\frac{n(n-1) \dots (n-k+1)}{2k} p^k$$

Dimostrazione 16. Dimostriamo il lemma precedente. Sia C_k l'insieme con tutti i cicli possibili di lunghezza k su n vertici. Ci chiediamo quanti siano. Un ciclo è determinato dai vertici che ne fanno parte. Il numero di modi per scegliere k vertici è

$$n(n-1) \dots (n-k+1)$$

Dobbiamo però considerare che in questo modo contiamo k volte ogni ciclo per il fatto che cambiamo solamente l'ordine dei nodi, e un addizionale 2 volte per ogni verso. Quindi

$$|C_k| = \frac{n(n-1) \dots (n-k+1)}{2k}$$

Dove dividiamo per $2k$ perché ci sono esattamente $2k$ sequenze che corrispondono allo stesso ciclo. Sia $C \in C_k$, ci chiediamo qual è la probabilità che G contenga C ? p^k , perché ogni arco tra i k nodi deve essere estratto, gli altri non ci interessa cosa succede. Indichiamo con $N_k(G)$ il numero di cicli di lunghezza k in G , allora

$$\mathbb{E}[N_k(G)] = \sum_{C \in C_k} P(G \text{ contiene } C)$$

$$= \sum_{C \in C_k} p^k$$

La sommatoria diventa indipendente da C

$$= |C_k| \cdot p^k \\ = \frac{n(n-1) \dots (n-k+1)}{2k} p^k$$

Capitolo 8

Ottava lezione

Dimostriamo ora il teorema di Erdős.

Dimostrazione 17. Cominciamo con il porre $p_n = n^{\varepsilon-1}$ dove $0 < \varepsilon < \frac{1}{k}$. Sia $N_{\leq k}(G)$ la variabile casuale che rappresenta il numero di cicli di lunghezza al più k in G . Guardiamo il valore atteso di questa variabile casuale.

$$\mathbb{E}[N_{\leq k}(G)] = \sum_{i=3}^k \mathbb{E}[N_i(G)]$$

Dove la sommatoria parte da 3 perché non esistono cicli di lunghezza minore, per definizione stessa. $N_i(G)$ è il numero di cicli di lunghezza i in G .

$$= \sum_{i=3}^k \frac{n(n-1) \dots (n-i+1)}{2i} p^i$$

A denominatore abbiamo i fattori, ciascuno è al più n .

$$\leq \sum_{i=3}^k \frac{n^i}{2i} p^i$$

per l'indicizzazione stessa della sommatoria, $i > 1$

$$\leq \frac{1}{2} \sum_{i=3}^k (np)^i$$

Non è una disuguaglianza stretta perché ambo i lati potrebbero valere 0.

$$\leq \frac{k-2}{2} (np)^k$$

dove $k-2$ sono il numero di termini della sommatoria. Notiamo che la sostituzione di $(np)^k$ con $(np)^i$ vale perché la base è maggiore di 1. Infatti:

$$np = n \cdot n^{\varepsilon-1} = n^{\varepsilon} > n^{\frac{1}{k}} = \sqrt[k]{n} > 1$$

Concentriamoci ora sul valore atteso di $N_{\leq k}(G) \geq \frac{n}{2}$. A cosa serve vi chiederete voi? Le strade del Signore sono infinite.

$$\mathbb{E}[N_{\leq k}(G) \geq \frac{n}{2}] \leq \frac{\mathbb{E}[N_{\leq k}(G)]}{\frac{n}{2}}$$

Che è l'applicazione della disuguaglianza di Markov. Usiamo il risultato trovato prima,

$$\begin{aligned} &\leq \frac{(k-2)(np)^k}{n} = (k-2)n^{k-1}p^k \\ &= (k-2)n^{k-1} \cdot n^{k\varepsilon-k} = (k-2)n^{k\varepsilon-1} \\ &= (k-2)n^{-(1-k\varepsilon)} \end{aligned}$$

Notiamo che $k\varepsilon - 1 < \frac{k}{k} - 1 < 0$. Quindi se $n \rightarrow 0$ possiamo affermare

$$\mathbb{E}[N_{\leq k}(G) \geq \frac{n}{2}] < \frac{1}{2}$$

Analizziamo ora $P(\alpha(G) \geq \frac{n}{2k})$. Dove ci porteranno questi conti? Lo scopriremo solo vivendo. Notiamo che una probabilità simile l'abbiamo già analizzata nelle precedenti lezioni. E allora risparmiamo inchiostro

$$\begin{aligned} P(\alpha(G) \geq \frac{n}{2k}) &\leq \binom{n}{\frac{n}{2k}} (1-p)^{\binom{\frac{n}{2k}}{2}} \\ &= \binom{n}{r} (1-p)^{\binom{r}{2}} \end{aligned}$$

dove $r = \frac{n}{2k}$. Introduciamo due importanti disequazioni note

$$\binom{n}{r} < 2^n$$

e

$$1-p \leq e^{-p}$$

Andiamo a utilizzarle

$$\leq 2^n e^{-p \binom{r}{2}}$$

Per $r \geq 2$ sappiamo che $\binom{r}{2} \geq \frac{r^2}{4}$ (infatti $\frac{r(r-1)}{2} \geq \frac{r^2}{4}$).

$$\leq 2^n e^{-p \frac{r^2}{4}} = 2^n e^{-p \frac{n}{16k^2}}$$

Vale che $p \cdot n = n^\varepsilon$. Scegliamo n abbastanza grande, ovvero $n^\varepsilon \geq 16k^2$.

$$\leq 2^n e^{-n} = \left(\frac{2}{e}\right)^n$$

Sempre per n abbastanza grande possiamo affermare

$$\left(\frac{2}{e}\right)^n < \frac{1}{2}$$

Utilizziamo tutto quello che abbiamo appreso.

$$\begin{aligned} &P(N_{\leq k}(G) < \frac{n}{2} \wedge \alpha(G) < \frac{n}{2k}) \\ &= 1 - P(N_{\leq k}(G) \geq \frac{n}{2} \vee \alpha(G) \geq \frac{n}{2k}) \\ &\geq 1 - P(N_{\leq k}(G) \geq \frac{n}{2}) - P(\alpha(G) \geq \frac{n}{2k}) \\ &> 1 - \frac{1}{2} - \frac{1}{2} \geq 0 \end{aligned}$$

Quindi

$$\exists G \ N_{\leq k} < \frac{n}{2} \wedge \alpha(G) < \frac{n}{2k}$$

G ha al più $\frac{n}{2}$ cicli di lunghezza al più k . Togliamo un vertice da ogni ciclo di lunghezza al più k in G . Sia

$H = (V_H, E_H)$ il grafo ottenuto. Allora $|V_H| \geq \frac{n}{2}$, perché al peggio leviamo un nodo per ogni ciclo. Inoltre, $g(H) > k$. Quanto vale $\chi(H)$? Sappiamo che

$$\chi(H)\alpha(H) \geq |V_H|$$

Quindi

$$\chi(H) \geq \frac{|V_H|}{\alpha(H)} \geq \frac{\frac{n}{2}}{\alpha(G)}$$

Dove l'ultima disequazione deriva dal fatto che $\alpha(G) \geq \alpha(H)$ (non dimostrato).

$$> \frac{\frac{n}{2}}{\frac{n}{2}} k = k$$

Abbiamo così dimostrato il teorema.

8.1 Proprietà asintotiche

Dati $p_1, p_2, \dots \in (0, 1)$, $G_n \sim \mathcal{G}(n, p_n)$ per $n = 1, 2, \dots$. Consideriamo una qualche proprietà, per esempio

$$\omega(G_n) \leq f(n)$$

Dove f è una qualsiasi funzione. Ora introduciamo il predicato

$$\pi_n(G_n) \in \{0, 1\}$$

E quindi, nel nostro esempio, potrebbe essere

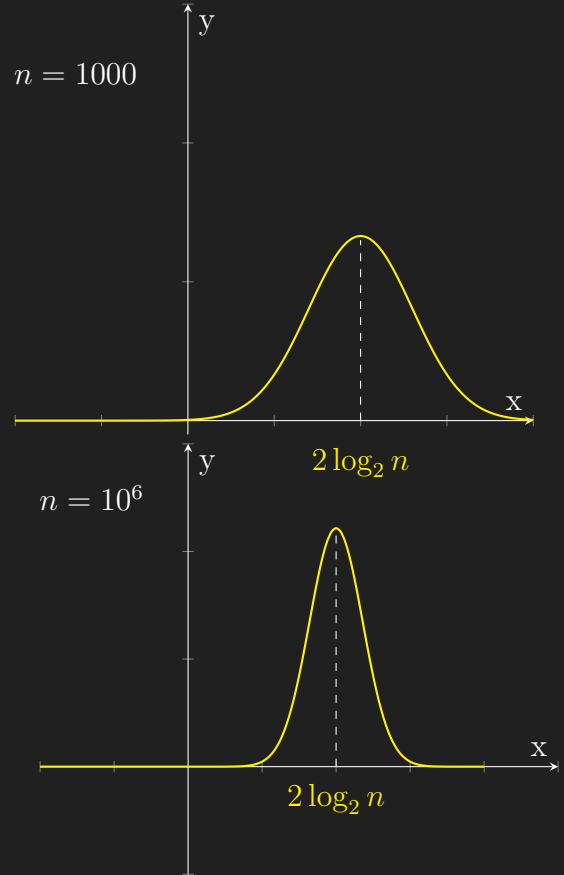
$$\pi_n^{G_n} = \mathbb{I}\{\omega(G_n) \leq f(n)\}$$

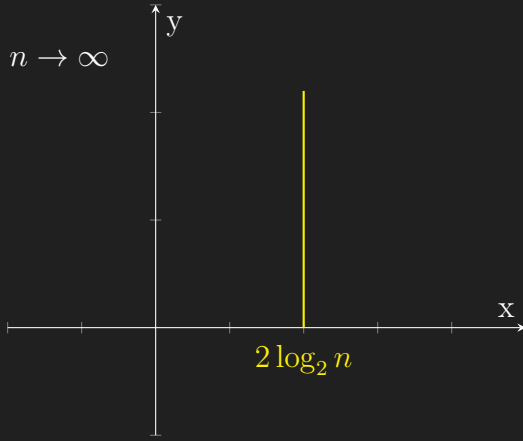
Avremo una sequenza di predicati π_1, π_2, \dots , e ci chiediamo se

$$\lim_{n \rightarrow \infty} P(\pi_n(G_n) = 1)$$

Cioè se il grafo G_n ha la proprietà $\omega(G_n) \leq f(n)$ con n molto grande. Se $= 1$ allora la proprietà **vale asintoticamente** per quasi tutti i grafi G_n .

Esempio 13. Dato $p_n = \frac{1}{2}$ e $\pi_n(G_n) = \mathbb{I}\{\omega(G_n) = 2 \log_2 n\}$ (stesso discorso può essere fatto per $\pi_n(G_n) = \mathbb{I}\{\alpha(G_n) = 2 \log_2 n\}$). Vediamo cosa succedere con il crescere di n .





Nella precedente lezione abbiamo dimostrato che

$$P(\alpha(G) \geq k) \leq 2^{-\frac{k}{2}}$$

con $p_n = \frac{1}{2}$ e $k \geq 2 \log_2 n$. Poniamo $k = 2 \log_2 n$, allora

$$P(\alpha(G) \geq k) \leq 2^{-\log_2 n} = \frac{1}{n}$$

se $n \rightarrow \infty$ tende a 0. Similmente possiamo dire che

$$P(\omega(G_n) \geq 2 \log_2 n) \leq \frac{1}{n} \rightarrow 0$$

Per un fatto che non dimostriamo vale che

$$P(\omega(G_n) < 2 \log_2 n) \leq 2^{-n^2+o(1)} \rightarrow 0$$

Quindi la clique massima non è né più grande né più piccola, quindi è esattamente $2 \log_2 n$ se $n \rightarrow \infty$. Lo

stesso risultato vale per il numero di indipendenza. Concentriamoci ora sul numero cromatico. Sappiamo già che

$$\chi(G_n) \geq \frac{n}{\alpha(G_n)}$$

Sappiamo che con n abbastanza grande il numero di indipendenza è $2 \log_2 n$

$$\geq \frac{n}{2 \log_2 n}$$

Fatto 13. Vale la seguente disuguaglianza

$$\chi(G_n) \leq \frac{n}{2 \log_2 n}$$

Dimostrazione 18. Prendiamo $G = (V, E)$ e un sottoinsieme dei suoi vertici, $S \subseteq V$. Sia $G[S]$ il grafo indotto con $|S| = m$. Notiamo che $G \sim \mathcal{G}(n, \frac{1}{2})$ implica $G[S] \sim \mathcal{G}(m, \frac{1}{2})$. Infatti basta estrarre gli archi relativi agli m vertici per primi. Concentriamoci sulla probabilità che il grafo $G[S]$ abbia un numero di indipendenza minore di $2 \log_2 m$.

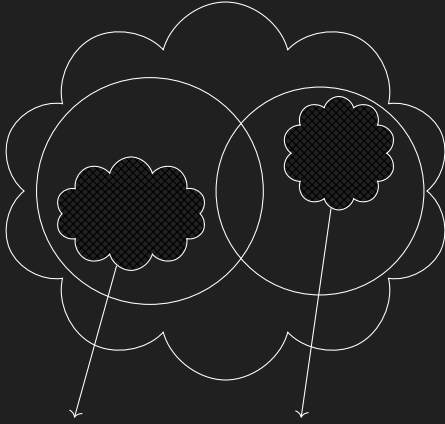
$$P(\exists S \subseteq V \mid S| = m \text{ t.c. } \alpha(G[S]) \leq 2 \log_2 m)$$

$$\begin{aligned} &\leq \sum_{\substack{S \subseteq V \\ |S|=m}} P(\alpha(G[S]) < 2 \log_2 m) \\ &\leq \binom{n}{m} 2^{-m^2+o(1)} \leq 2^{n-m^2+o(1)} \end{aligned}$$

Dove abbiamo usato il fatto che $\binom{n}{m} \leq 2^n$. Scegliamo $m = \frac{n}{(\log_2 n)^2}$

$$\begin{aligned}
&= 2^{n - (\frac{n}{(\log_2 n)^2})^{2+o(1)}} \\
&= 2^{n - \frac{n^{2+o(1)}}{(\log_2 n)^{4+o(1)}}} \rightarrow 0
\end{aligned}$$

Allora $\forall S \subseteq V$ con $|S| = m$ sono tali che $\alpha(G[S]) \geq 2 \log_2 m$. Finché ci sono sottoinsiemi di taglia m , possiamo isolare un insieme indipendente di taglia almeno $2 \log_2 m$



In figura, le nuvolette colorate sono gli insiemi indipendenti rimossi, i cerchi indicano gli insiemi S . A ogni insieme di vertici indipendenti assegniamo lo stesso colore. Quindi se abbiamo N insiemi indipendenti, avremo N colori diversi. Il numero di insiemi indipendenti trovati è $\frac{n}{k}$, dove n è il numero di vertici totali e k la dimensione degli insiemi indipendenti. A questo punto ci avvanzeranno al più m nodi, a cui assegniamo un colore diverso per

ognuno. Quindi possiamo affermare che

$$\begin{aligned}
\chi(G_n) &\leq \frac{n}{k} + m = \frac{n}{2 \log_2 n} + \frac{n}{(\log_2 n)^2} \\
&= \frac{n}{2 \log_2 n} \left(1 + \frac{2}{\log_2 n}\right) = (1 + o(1)) \frac{n}{2 \log_2 n}
\end{aligned}$$

In questo modo concludiamo la dimostrazione.

Alla luce del fatto, possiamo affermare che per $n \rightarrow \infty$

$$\chi(G_n) = \frac{n}{2 \log_2 n}$$

vale asintoticamente per quasi tutti i G_n .

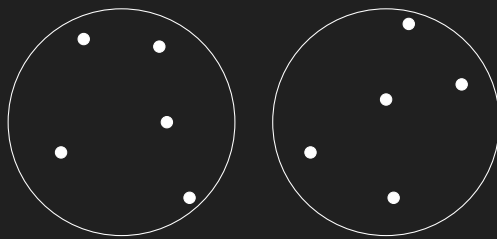
Capitolo 9

Nona lezione

9.1 Introduzione

Iniziamo la seconda parte del corso e ci concentreremo sulla **clusterizzazione spettrale**. Per clusterizzazione si intende il partizionamento di un insieme di dati in modo che dati simili siano nello stesso elemento della partizione.

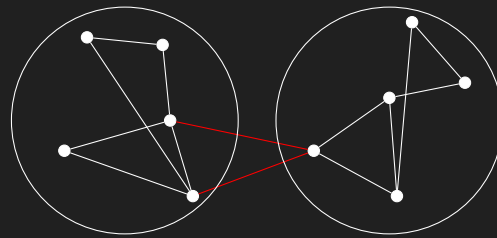
Esempio 14. Dati dei punti possiamo decidere di raggrupparli per vicinanza



In un grafo i dati sono i vertici e gli archi rappresentano la relazione di similarità. Se abbiamo un grafo con $G = (V, E)$ tale che $|V| = n$ e abbiamo due cricche ciascuna da $\frac{n}{2}$ è facile trovare un modo ottimo per parti-

zionarlo. Ma questo non è sempre il caso.

Esempio 15. Dati dei punti possiamo decidere di raggrupparli per vicinanza



Con questa scelta andiamo a tagliare 2 archi, ovvero violiamo 2 similarità.

Come partizionare i vertici di un grafo in modo da ottenere due o più cluster il più possibile densi e ben separati (con pochi archi tra un cluster e l'altro)? In altre parole ci stiamo chiedendo se esista, e quale sia, l'algoritmo per partizionare in modo migliore i vertici. Per rispondere a

questa domanda utilizzeremo l'algebra lineare applicata ai grafi. Questo approccio non è l'unico ma l'algebra lineare si sposa in modo elegante con la teoria dei grafi. Infatti, possiamo rappresentare un grafo $G = (V, E)$, $|V| = n$, con una **matrice di adiacenza** $A \in \{0, 1\}^{n \times n}$. Questa è

- Binaria: $A_{ij} = 1 \iff (i, j) \in E$ altrimenti $A_{ij} = 0$.
- Simmetrica: perché ci troviamo nel contesto di grafi non orientati.
- Diagonale composta da 0: perché non abbiamo self-loop (capi).

9.2 Ripasso di algebra lineare

Consideriamo matrici $M \in \mathbb{R}^{n \times n}$, ovvero matrici quadrate a coefficienti lineari. Se

$$\exists \lambda \in \mathbb{R}, u \in \mathbb{R}^n \setminus \{0\} \rightarrow u \neq (0, \dots, 0) \text{ t.c. } Mu = \lambda u$$

u viene chiamato **autovettore** di M con **autovalore** λ . Possiamo assumere che tutti i vettori u abbiano norma 1. Infatti, se $\|u\| \neq 1$ dividiamo ambo i lati per $\|u\|$

$$M \frac{u}{\|u\|} = \lambda \frac{u}{\|u\|}$$

e consideriamo come vettore $u' = \frac{u}{\|u\|}$, che ha norma 1. Quindi ogni autovettore ha lunghezza unitaria. Notiamo che $\lambda \in \mathbb{R}$ è un autovalore di M se e solo se

$$\exists x \neq (0, \dots, 0) \text{ t.c. } (M - \lambda I)x = (0, \dots, 0)$$

dove I è la matrice identità di dimensioni $n \times n$

$$I = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}$$

con la proprietà che $Ix = x \forall x \in \mathbb{R}^n$. L'equazione $(M - \lambda I)x = (0, \dots, 0)$ vale per $x \neq (0, \dots, 0)$ se e solo se $\det(M - \lambda I) = 0$. Non ci interessa in questo momento come si calcola il determinante di una matrice, sottolineiamo soltanto che $\det(M - \lambda I)$ è un polinomio di grado λ . Il **teorema fondamentale dell'algebra** ci dice che l'equazione $\det(M - \lambda I)$ ha n radici, non necessariamente distinte e non tutte necessariamente reali. Quindi ogni $A \in \mathbb{R}^{n \times n}$ ha n autovalori.

Fatto 14. Se $M \in \mathbb{R}^{n \times n}$ è simmetrica allora $\exists \lambda \in \mathbb{R}$ e $u \in \mathbb{R}^n \setminus \{0\}$ tale che $Mu = \lambda u$, ovvero ha almeno un autovalore reale.

Fatto 15. Se $M \in \mathbb{R}^{n \times n}$ è simmetrica allora se $Mu = \lambda u$ e $Mu' = \lambda' u'$ con $\lambda \neq \lambda'$, vale che $u^T u' = 0$, cioè u, u' sono **ortogonali**.

Dimostrazione 19. Siccome M è simmetrica deve valere

$$(Mu')^T u' = u'^T Mu'$$

infatti $M^T = M$. Ma $Mu' = \lambda'u'$, quindi

$$(Mu')^T u' = \lambda u'^T u' = \lambda' u'^T u'$$

Siccome $\lambda \neq \lambda'$, l'unico modo per cui l'equazione $\lambda u'^T u' = \lambda' u'^T u'$ possa valere è che

$$u'^T u' = 0$$

Introduciamo un teorema importante che utilizzeremo estensivamente.

Teorema 8. Teorema spettrale.

Sia $M \in R^{n \times n}$ simmetrica allora esistono $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, non necessariamente distinti, e $u_1, \dots, u_n \in R^n \setminus \{0\}$ ortonormali (ovvero $\|u_i\| = 1$ per $i = 1, \dots, n$ e $u_i^T u_j = 0$ con $i \neq j$) tali che $Mu_i = \lambda_i u_i$ per $i = 1, \dots, n$.

Dimostrazione 20. Dimostrazione per induzione.

CASO BASE: Se $n = 1$, allora $M \in R^{1 \times 1} \equiv \mathbb{R}$ e $\forall x \in \mathbb{R} \setminus \{0\}$ è un autovettore con autovalore M perché $Mx = Mx$.

IPOTESI INDUTTIVA: Il teorema vale per $n - 1$. Per il primo fatto che abbiamo dimostrato in questa lezione vale

$$\exists \lambda_n \in \mathbb{R}, \exists x_n \in R^n \text{ t.c.}$$

$$Mx_n = \lambda_n x_n, \quad x_n \neq (0, \dots, 0)$$

Claim 1. Per $y \in R^n \setminus \{0\}$ vale che $y^T x_n = 0 \rightarrow (My)^T x_n = 0$. Infatti

$$x_n^T (My) = (Mx_n)^T y = \lambda_n x_n^T y = 0$$

Il claim ci dice che se y è ortogonale a x_n allora My è ortogonale a x_n .

Sia $V \subset R^n$ un sottospazio che contiene tutti e soli i vettori ortogonali a x_n . Quindi $\dim(V) = n - 1$ perché togliamo la dimensione della direzione del vettore usato per definire l'ortogonalità. Sia $\{u_1, \dots, u_{n-1}\}$ una base ortonormale di V e sia $B = \{u_1, \dots, u_{n-1}\} \in R^{(n-1) \times (n-1)}$. Quindi,

$$B : R^{n-1} \rightarrow V \subset R^n$$

B è una matrice che proietta i vettori di dimensione n nello spazio V . La matrice BB^T è tale che

$$BB^T : R^n \rightarrow V \subset R^n \quad \forall z \in V$$

inoltre $BB^T z \in V$. Sia $n' = n - 1$. Sia $n' = B^T M B \in R^{(n-1) \times (n-1)}$, simmetrica. Appliciamo l'ipotesi induttiva a x' e troviamo $\lambda_1, \dots, \lambda_{n-1} \in \mathbb{R}$ autovalori e $y_1, \dots, y_{n-1} \in R^{n-1}$ autovettori. Per $i = 1, \dots, n - 1$ vale

$$M'y_i = B^T M B y_i = \lambda_i y_i$$

Moltiplichiamo per B ambo i lati

$$BB^T M B y_i = \lambda_i B y_i \in V$$

la posizione di λ_i è irrilevante dato che è uno scalare. Siccome By_i è ortogonale a x_n (quindi $x_n^T(MBy_i) = 0$, e dunque $MBy_i \in V$) e per il claim vale che

$$\lambda_i By_i = BB^T MBy_i = MBy_i$$

con $By_i = x_i$. L'equazione sopra implica

$$Mx_i = \lambda_i x_i \quad \forall i = 1, \dots, n-1$$

Notiamo che BB^T proietta vettori in U , ma MBy_i è già in U , e quindi lo proietta su stesso. Concludiamo dicendo che, per costruzione,

$$x_i^T x_j = 0 \quad \forall i \neq j \quad 1 \leq i, j \leq n-1$$

Infatti,

$$x_i^T x_j = (By_i)^T By_j = y_i^T B^T By_j$$

Vale che $B^T B = I$, infatti $U_i U_j = 0$ se $i \neq j$, dato che sono ortogonali, altrimenti è 1 se $i = j$.

$$= y_i^T I y_j = y_i^T y_j = 0$$

per $i \neq j$.

Corollario 1. Esponiamo un corollario comodo. Se $M \in R^{n \times n}$ simmetrica allora

$$M = U^T \Lambda U = \sum_{i=1}^n \lambda_i U_i U_i^T$$

Dove $U = [U_1, \dots, U_n]$ e

$$\Lambda = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$$

Dimostrazione 21. Dato $MU_i = \lambda_i U_i \quad \forall i = 1, \dots, n$ vale che $MU[\lambda_1 U_1, \dots, \lambda_n U_n] = U\Lambda$. Siccome $U = [U_1, \dots, U_n]$ è una matrice ortogonale (ovvero $U^{-1} = U^T$) abbiamo

$$UU^T = UU^{-1} = I$$

Concludiamo che

$$M = MUU^T = U\Lambda U^T$$

Capitolo 10

Decima lezione

Riprendiamo il teorema spettrale. Ogni matrice simmetrica $M \in \mathbb{R}^{n \times n}$ può essere scritta come

$$M = U \Lambda U^T = \sum_{i=1}^n \lambda_i U_i U_i^T$$

dove $U = [U_1, \dots, U_n]$ e $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, con (u_i, λ_i) autovettori e corrispondenti autovalori di M per $i = 1, \dots, n$. Notiamo che dati $U, V \in \mathbb{R}^n$ vale che

$$(UV^T)_{ij} = u_i v_j$$

Enunciamo ora un teorema che ci permette di definire gli autovettori e autovalori in un altro modo.

Teorema 9. Caratterizzazione variazionale degli autovalori. Sia $M \in \mathbb{R}^{n \times n}$ simmetrica e siano $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ i suoi autovalori. Per $m < n$ u_1, \dots, u_m siano autovettori tali che $M u_i = \lambda_i u_i$ per $i = 1, \dots, m$. Allora

$$\lambda_{m+1} = \min_{\substack{u \in \mathbb{R}^n \setminus \{0\} \\ u \perp \{u_1, \dots, u_m\}}} \frac{u^T M u}{u^T u}$$

dove $u \perp \{u_1, \dots, u_m\}$ indica che u è ortogonale ai vettori $\{u_1, \dots, u_m\}$

La dimostrazione del teorema è omessa, ci limitiamo a fare delle osservazioni. Prima di tutto, il coefficiente

$$\frac{u^T M u}{u^T u}$$

è detto **quoziente di Rayleigh**. Sappiamo che $u^T u = \|u\|^2$, quindi possiamo riscriverlo come

$$\frac{u^T}{\|u\|} M \frac{u}{\|u\|}$$

Senza perdita di generalità possiamo assumere che il vettore u abbia norma 1 (se così non fosse, possiamo porre $u' = \frac{u}{\|u\|}$, u' ha norma 1). Allora

$$\lambda_{m+1} = \min_{\substack{u \in \mathbb{R}^n \setminus \{0\} \\ u \perp \{u_1, \dots, u_m\}}} u^T M u$$

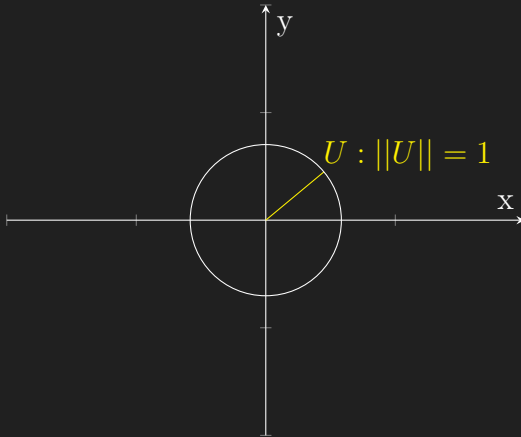
Concentriamoci sul caso λ_1 .

$$\lambda_1 = \min_{u \in \mathbb{R}^n \setminus \{0\}} \frac{u^T M u}{u^T u} = \min_{u \in \mathbb{R}^n \setminus \{0\}} u^T M u$$

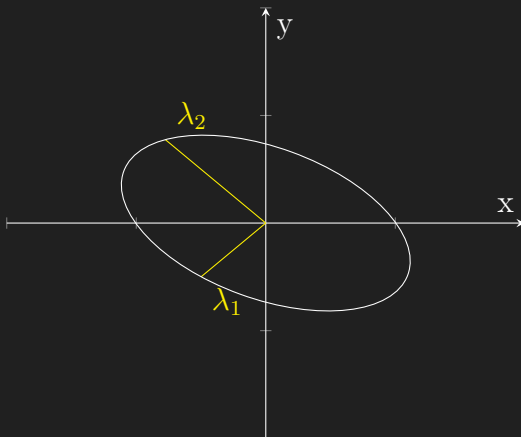
Sia u_1 l'autovettore associato a λ_1

$$u_1^T M u_1 = u_1^T \lambda_1 u_1 = \lambda_1 \|u_1\|^2 = \lambda_1$$

Andiamo a vedere cosa vuol dire a livello geometrico, per ovvi motivi $n = 2$. Inizialmente la situazione è



Se applichiamo a u la trasformazione $u^T M$ otteniamo



La lunghezza degli assi dell'ellisse corrisponde al valore degli autovalori. L'asse più corto è λ_1 , il più lungo è λ_2 . La rappresentazione geometrica riesce a catturare lo spettro della matrice. Sappiamo che λ_1 è il minimo degli autovalori, vediamo ora che λ_n (nell'esempio a 2 dimensioni λ_2) è il maggiore. Consideriamo la matrice $-M$ che ha autovalori $-\lambda_n \leq -\lambda_{n-1} \leq \dots \leq -\lambda_1$.

$$-\lambda_n = \min_{u \in \mathbb{R}^n \setminus \{0\}} -\frac{u^T M u}{u^T u} = -\max_{u \in \mathbb{R}^n \setminus \{0\}} \frac{u^T M u}{u^T u}$$

moltiplichiamo entrambi per -1

$$\lambda_n = \max_{u \in \mathbb{R}^n \setminus \{0\}} \frac{u^T M u}{u^T u}$$

Ricapitolando, l'autovalore 1 corrisponde all'autovettore u che minimizza il coefficiente di Rayleigh, mentre l'autovalore n a quello massimo. Quelli in mezzo (autovalore $1 < i < n$) corrispondono agli autovettori che minimizzano lo stesso coefficiente ma con il vincolo di ortogonalità. Concludiamo questa parte di algebra lineare con la seguente definizione

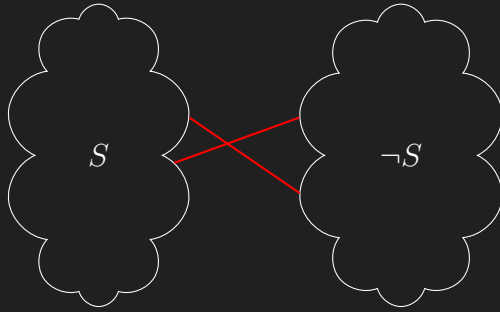
Definizione 7. Una matrice simmetrica M è **positiva semidefinita** se $x^T M x > 0 \forall x \in \mathbb{R}^n$

Gli autovalori di una matrice positiva semidefinita (se è positiva semidefinita è anche simmetrica) sono

tutti non negativi. La dimostrazione è ovvia guardando il coefficiente di Rayleigh.

10.1 Clustering Spettrale

Sia $G = V, E$ vogliamo riuscire a partizionare il grafo così che il numero archi tra le partizioni sia piccolo.



Sia $\neg S \equiv V \setminus S$. Andiamo a tagliare gli archi

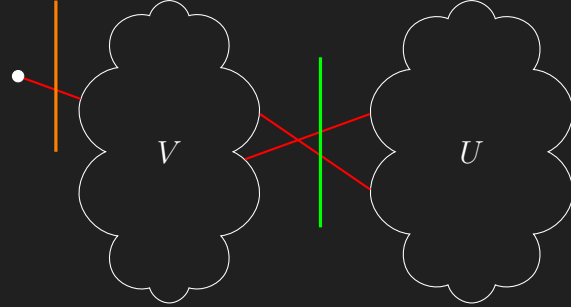
$$E(S, \neg S) = \{(i, j) : i \in S, j \in \neg S\}$$

Definizione 8. La sparsità di un taglio $(S, \neg S)$ è definita come

$$\sigma(S) = \frac{|E(S, \neg S)|}{|S||\neg S|}$$

dove $|E(S, \neg S)|$ è il numero di archi tra S e $\neg S$ mentre $|S||\neg S|$ è il numero di archi possibili.

Il denominatore serve per bilanciare. Senza di esso potremmo trovarci con partizioni senza senso che però sono privilegiate dalla sparsità.



Senza l'effetto mitigatore del denominatore, la partizione ottenuta usando il taglio arancione sarebbe migliore di quella ottenuta usando il taglio verde, ma è evidente che non sia vero. Nel caso in cui usiamo il taglio arancione abbiamo $|S| = 1$ e $|\neg S| = n - 1$ e quindi $|S||\neg S| = n - 1$. Nel caso in cui $|S| = \frac{n}{2} = |\neg S|$ quindi $|S||\neg S| = \frac{n^2}{4}$, se n abbastanza grande $\frac{n^2}{4} \gg n - 1$. Quindi la sparsità privilegia partizioni bilanciate.

Definizione 9. La sparsità di un grafo $G = (V, E)$ è

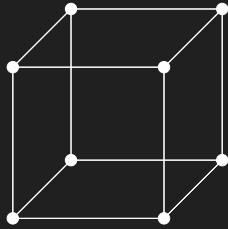
$$\sigma(G) = \min_{\substack{S \subset V \\ |S| \neq 0}} \sigma(S)$$

Da ora in poi facciamo un'assunzione per semplificare i conti (ma non troppo, abbastanza per guadagnare i sei crediti del corso) consideriamo grafi **d-regolari**.

Definizione 10. Un grafo $G = (V, E)$ si dice **d-regolare** se

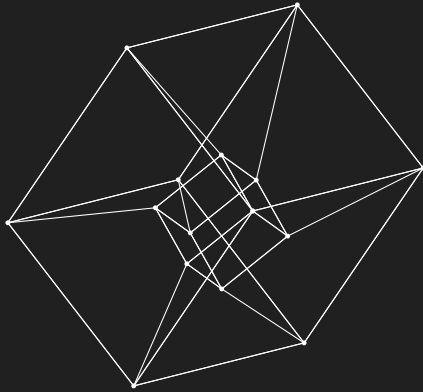
$$d(v) = d \quad \forall v \in V$$

Una cricca K_n è $n - 1$ -regolare. Invece, un ciclo su n vertici C_n è 2-regolare. Un cubo è 3-regolare.



In generale un ipercubo (cubo a più di 3 dimensioni) è $\log_2 n$ -regolare, dove n è l'ordine del grafo.

Esempio 16. Un tesseracto (ipercubo a 4 dimensioni) è 4-regolare.



Definiamo un parente stretto della sparsità.

Definizione 11. Sia G un grafo d -regolare e $S \subset V$ $|S| \neq 0$, definiamo l'espansione come

$$xpn(S) = \frac{|E(S, \neg S)|}{d \cdot |S|}$$

Notiamo che la definizione non è simmetrica! Infatti, $xpn(S) \neq xpn(\neg S)$, ovvero scambiando S con $\neg S$ otteniamo due risultati diversi. L'espansione è simile alla sparsità, con la differenza che la prima ha come denominatore il numero di archi massimi che possono uscire da S .

Siccome xpn non è simmetrica definiamo una simmetrizzazione, detta **conduttanza**.

Definizione 12. La **conduttanza** è definita come

$$\begin{aligned} \phi(S) &= \max \{xpn(S), xpn(\neg S)\} \\ &= \frac{|E(S, \neg S)|}{d \cdot \min \{|S|, |\neg S|\}} \end{aligned}$$

La conduttanza di un grafo G è

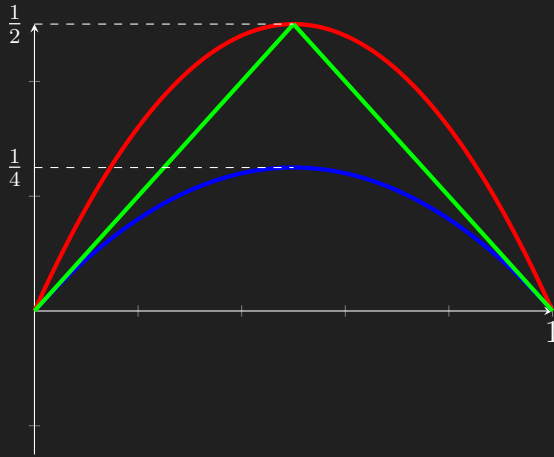
$$\phi(G) = \min_{S \subset V} \phi(S)$$

Ovvero al conduttanza minima possibile per un sottoinsieme del grafo. Che relazione c'è tra la conduttanza e la sparsità? Cominciamo con il dire che vale

$$\frac{1}{n} \min \{|S|, |\neg S|\} = \min \left\{ \frac{|S|}{n}, \frac{|\neg S|}{n} \right\}$$

Inoltre, data la partizione $S, \neg S$ per definizione sono tali che $|S| + |\neg S| = n$. Poniamo $\frac{|S|}{n} = \alpha$ e $\frac{|\neg S|}{n} = 1 - \alpha$. Sicuramente α è tale che $0 \leq \alpha \leq 1$. Vale la seguente disuguaglianza

$$\alpha(1-\alpha) \leq \min\{\alpha, 1-\alpha\} \leq 2\alpha(1-\alpha)$$



1. La parabola più esterna rappresenta $2\alpha(1-\alpha)$
2. La parabola più interna rappresenta $\alpha(1-\alpha)$
3. In mezzo tra le due $\min\{\alpha, 1-\alpha\}$

Quindi

$$\begin{aligned} \frac{|S||\neg S|}{n} &\leq \min\left\{\frac{|S|}{n}, \frac{|\neg S|}{n}\right\} \\ &\leq 2 \frac{|S||\neg S|}{n} \end{aligned}$$

Ricaviamo che

$$\phi(S) \leq \frac{n}{d} \sigma(S) \leq 2\phi(S)$$

per tutti i tagli possibili. Dunque

$$\phi(G) \leq \frac{n}{d} \sigma(G) \leq 2\phi(G)$$

possiamo concludere che minimizzare la sparsità è equivalente a minimizzare la conduttanza.

10.2 Matrice laplaciana

Definiamo la **matrice laplaciana** di un grafo d -regolare. Innanzitutto, sia A la matrice di adiacenza di $G = (V, E)$ dove $A_{ij} = \mathbb{I}\{(i, j) \in E\} \in \{0, 1\}^{n \times n}$. La matrice laplaciana L è definita come

$$L = I - \frac{A}{d}$$

dove

$$\left(\frac{A}{d}\right)_{ij} = \frac{1}{d} \mathbb{I}\{(i, j) \in E\}$$

Per ogni $x \in \mathbb{R}^n$ vale che

$$x^T L x = x^T I x - \frac{1}{d} x^T A x$$

sappiamo che $x^T I x = x^T x = \|x\|^2 = \sum_{i \in V} x_i^2$

$$\sum_{i \in V} x_i^2 - \frac{1}{d} \sum_{i \in V} \sum_{j \in V} x_i A_{ij} x_j$$

Sappiamo che $A_{ij} = 0$ se $(i, j) \notin E$ e attraverso qualche manipolazione algebrica otteniamo

$$\begin{aligned}
&= \frac{1}{d} \sum_{i \in V} \sum_{j: (i,j) \in E} x_i^2 - \frac{1}{d} \sum_{i \in V} \sum_{j: (i,j) \in E} x_i x_j \\
&= \frac{1}{d} \sum_{i \in V} \sum_{j: (i,j) \in E} (x_i^2 - x_i x_j)
\end{aligned}$$

Notiamo che stiamo prendendo due volte ogni arco i, j . La prima volta con i e la seconda con j . Quindi eliminiamo la prima sommatoria, aggiungiamo x_j^2 (dato che ci spostiamo solo sugli archi adesso) e raddoppiamo $-x_i x_j$ per contarla due volte come prima:

$$\begin{aligned}
&= \frac{1}{d} \sum_{j: (i,j) \in E} (x_i^2 + x_j^2 - 2x_i x_j) \\
&= \frac{1}{d} \sum_{j: (i,j) \in E} (x_i - x_j)^2
\end{aligned}$$

Che è sempre non negativo. Questo vuol dire che la matrice laplaciana è positiva semidefinita.

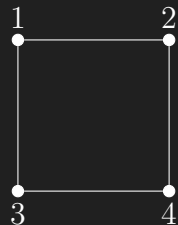
Capitolo 11

Undicesima lezione

11.1 Legame matrice laplaciana e autovalori

Nella scorsa lezione abbiamo scoperto che la matrice laplaciana L è positiva semidefinita. Questo implica che ha n autovalori tali che $0 \leq \lambda_1 \leq \dots \leq \lambda_n$. Prima di andare a lavorare sulla matrice laplaciana vediamo almeno una volta.

Esempio 17. Sia $G = (V, E)$ un grafo 2-regolare:



La matrice d'adiacenza è A è

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{matrix}$$

Da questa otteniamo la matrice laplaciana L

$$L = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \end{matrix}$$

In generale, una matrice laplaciana è tale che la somma sulle righe è $= 0$ e la somma sulle colonne è $= 0$. Prendiamo il primo autovalore

$$\lambda_1 = \min_{U \in \mathbb{R}^n \setminus 0} \frac{U^T L U}{U^T U}$$

Sappiamo che $\lambda_1 \geq 0$ perché L è positiva semidefinita. Definiamo il vettore $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ ovvero il vettore contenente tutti 1. Allora

$$L \cdot \mathbf{1} = \begin{bmatrix} \sum_{j=1}^n L_{1j} \\ \vdots \\ \sum_{j=1}^n L_{nj} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Quindi $\mathbf{1}^T L \mathbf{1} = 0$. Quindi $\lambda_1 = 0$ dato che 0 è il minimo valore pos-

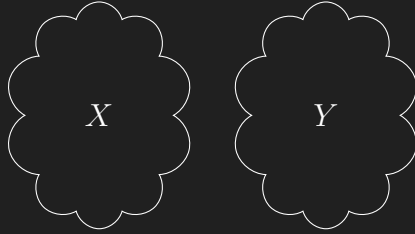
sibile ottenibile ed esiste un vettore, ovvero $\mathbf{1}$ tale per cui il coefficiente di Rayleigh è 0. Per il teorema spettrale $U_1 = \frac{\mathbf{1}}{\sqrt{n}}$ (dove \sqrt{n} serve a normalizzarlo) è un autovettore con autovalore $\lambda_1 = 0$. Ogni altro autovettore U è tale che

$$U^T \frac{\mathbf{1}}{n} = 0 \Leftrightarrow U^T \mathbf{1} = 0 \Leftrightarrow \sum_{i=1}^n u_i = 0$$

Concentriamoci sul secondo autovalore, λ_2 . Vale che

$$\begin{aligned} \lambda_2 &= \min_{\substack{U \in \mathbb{R}^n \setminus \{0\} \\ U^T \mathbf{1} = 0}} \frac{U^T L U}{U^T U} \\ &= \min_{\substack{U \in \mathbb{R}^n \setminus \{0\} \\ U^T \mathbf{1} = 0}} \frac{\sum_{(i,j) \in E} (u_i - u_j)^2}{d \sum_{i \in V} u_i^2} \end{aligned}$$

Se G ha due componenti connesse



Scegliamo $U \in \mathbb{R}^n$ tale che

$$u_i = \begin{cases} \frac{1}{|X|} & \text{se } i \in X \\ \frac{1}{|Y|} & \text{se } i \in Y \end{cases}$$

Vale che $U^T \mathbf{1} = 0$. Infatti, per $\sum_x u_i = 1$ e $\sum_y u_i = -1$, quindi $\sum_x u_i + \sum_y u_i = 0$. Inoltre $U^T L U = 0$ perché:

□ Se $i, j \in X$ allora $u_i - u_j = 0$.

□ Se $i, j \in Y$ allora $u_i - u_j = 0$.

□ Non può esistere il caso in cui $i \in X, j \in Y$ o viceversa.

Quindi $\frac{U}{\|U\|}$ è un autovettore di G con autovalore 0. Più in generale possiamo affermare che $\lambda_k = 0$ se G ha k componenti connesse. Questo ci permette di ricavare il numero di componenti connesse di G guardando il numero di autovalori pari a 0. Vediamo cosa succede con λ_n .

$$\begin{aligned} \lambda_n &= \max_{U \in \mathbb{R}^n \setminus \{0\}} \frac{U^T L U}{U^T U} = \max \frac{\sum_{(i,j) \in E} (u_i - u_j)^2}{d \sum_{i \in V} u_i^2} \\ &= \max \frac{1}{d \sum u_i^2} \sum_{(i,j) \in E} (u_i^2 + u_j^2 - 2u_i u_j) \end{aligned}$$

Sapendo che ogni nodo appare come estremo in d archi distinti

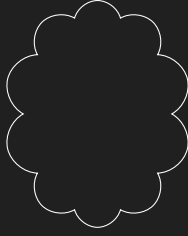
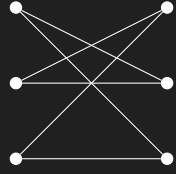
$$= \max \frac{1}{d \sum u_i^2} (d \sum_i u_i^2 - \sum_{(i,j) \in E} (2u_i u_j))$$

sommiamo e sottraiamo $d \sum_i u_i^2$

$$\begin{aligned} &= \max \frac{1}{d \sum u_i^2} (2d \sum_i u_i^2 - (d \sum_i u_i^2 + \sum_{(i,j) \in E} (2u_i u_j))) \\ &= \max 2 - \frac{\sum_{(i,j) \in E} (u_i + u_j)^2}{d \sum_i u_i^2} \end{aligned}$$

$$= 2 - \min \frac{\sum_{(i,j) \in E} (u_i + u_j)^2}{d \sum_i u_i^2}$$

Quindi $0 \leq \lambda_n \leq 2$. Notiamo che λ_n vale 2 nel caso in cui nel grafo sia presente una componente bipartita.



Costruiamo l'autovettore $U \in \mathbb{R}^n$

$$u_i = \begin{cases} 1 & \text{se } i \in X \\ -1 & \text{se } i \in Y \\ 0 & \text{altrimenti} \end{cases}$$

Quindi $\sum_{(i,j) \in E} (u_i + u_j)^2 = 0$ e allora $\lambda_n = 0$.

11.2 Disuguaglianza di Cheeger

Come si lega la sparsità, o la conduttanza, allo spettro?

Definizione 13. La disuguaglianza di **Cheeger** ci dà un'approssimazione di $\phi(G)$:

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda}$$

La disuguaglianza di Cheeger può essere anche scritta come

$$\frac{2\lambda_2}{2n} \leq \sigma(G) \leq \frac{2d}{n} \sqrt{2\lambda_2}$$

Dimostrazione 22. Cominciamo a dimostrare la parte sinistra, ovvero

$$\frac{\lambda_2}{2} \leq \phi(G)$$

Consideriamo $U \in \mathbb{R}^d$ tale che $U^T \mathbf{1} = 0$.

$$\sum_i \sum_j (u_i - u_j)^2 = \sum_i \sum_j (u_i^2 + u_j^2 - 2u_i u_j)$$

$$= n \sum_i u_i^2 + n \sum_j u_j^2 - 2 \sum_i \sum_j u_i u_j$$

Sapendo che $(\sum_i u_i)^2 = (\sum_i u_i)(\sum_j u_j) = \sum_i \sum_j u_i u_j$ scriviamo

$$= 2n \sum_i u_i^2 - 2(\sum_i u_i)^2 = 2n \sum_i u_i^2 - 2(U^T \mathbf{1})^2$$

dato che $U^T \mathbf{1} = 0$

$$= 2n \sum_i u_i^2$$

Quindi abbiamo che

$$\lambda_2 = \min_{\substack{U \in \mathbb{R}^n \setminus \{0\} \\ U^T \mathbf{1} = 0}} \frac{\sum_{(i,j) \in E} (u_i - u_j)^2}{d \sum_i u_i^2}$$

Se $U \neq 0$ e $U^T \mathbf{1} = 0$ allora $U \neq 0$ e $U \neq \mathbf{1}$. Inoltre usiamo il fatto che $\sum_i u_i^2 = \frac{1}{2n} \sum_{i,j} (u_i - u_j)^2$

$$= \min_{U \in \mathbb{R}^n \setminus \{0, \mathbf{1}\}} \frac{\sum_{(i,j) \in E} (u_i - u_j)^2}{\frac{d}{2n} \sum_{i,j} (u_i - u_j)^2}$$

Sia $\forall S \subseteq V$, $U \in \{0, 1\}^n$ tale che $u_i = 1 \leftrightarrow i \in S$ (o in altre parole $u_i = \mathbb{I}\{i \in S\}$). Allora

$$|E(S, \neg S)| = \sum_{(i,j) \in E} (u_i - u_j)^2$$

e sapendo che $u_i = u_i^2$ (dato che $1^2 = 1$ e $0^2 = 0$) possiamo dire

$$|S| |\neg S| = \left(\sum_i u_i^2 \right) (n - \sum_i u_i^2)$$

$$= n \sum_i u_i^2 - \sum_{i,j} u_i u_j$$

In una dimostrazione precedente avevamo trovato che $\sum_i \sum_j (u_i - u_j)^2 = 2n \sum_i u_i^2 - 2(\sum_i u_i)^2$. Il lato destro è esattamente quello che abbiamo noi moltiplicato per un fattore 2, quindi

$$= \frac{1}{2} \sum_{i,j} (u_i - u_j)^2$$

Allora

$$\sigma(G) = \min_{(S, \neg S) \text{ cut}} \frac{|E(S, \neg S)|}{|S| |\neg S|}$$

$$= \min_{U \in \mathbb{R}^n \setminus \{0, \mathbf{1}\}} \frac{\sum_{(i,j) \in E} (u_i - u_j)^2}{\frac{d}{2n} \sum_{i,j} (u_i - u_j)^2} \geq \frac{d}{n} \lambda_2$$

Che implica che

$$\frac{d}{n} \lambda_2 \leq \sigma(G) \leq \frac{2d}{n} \phi(G)$$

Quindi abbiamo dimostrato che

$$\frac{\lambda_2}{2} \leq \phi(G)$$

Capitolo 12

Dodicesima lezione

Dimostriamo l'altra parte della disuguaglianza di Cheeger.

$$\phi(G) \leq \sqrt{2\lambda_2}$$

Introduciamo il seguente algoritmo:

Algorithm 3 (Fiedler)

Input: $G = (V, E)$, vettore $x \in \mathbb{R}^n \setminus \{0\}$

- 1: Ordiniamo le componenti di x dalla più piccola alla più grande, ovvero (v_1, \dots, v_n) tali che $v_1 \leq \dots \leq v_n$.
- 2: Troviamo $k \in \{1, \dots, n-1\}$ che minimizza la conduttanza ϕ del taglio $\phi\{v_1, \dots, v_k\}$

Output: taglio $\{v_1, \dots, v_k\}$

L'algoritmo di **Fiedler** ordina le componenti di x e poi costruisce vari tagli, dividendo in due gruppi le componenti, e trova il migliore tra questi. L'ordinamento ha una complessità di $O(|V| \log |V|)$ mentre trovare il minimo taglio costa $O(|E|)$ perché

dobbiamo passare tutta la lista, infatti inizialmente avremo il taglio $\{v_1\}$ e calcoleremo la relativa conduttanza, poi $\{v_1, v_2\}$ fino a $\{v_1, \dots, v_{n-1}\}$.

Fiedler trova un cut $(S_F, \neg S_F)$ tale che

$$\phi(S_f) \leq \sqrt{2\lambda_2}$$

quando il vettore x in input è autovettore relativo a λ_2 . Il problema di trovare il $\phi(G)$ minimo è NP-completo, questo algoritmo ci permette di trovarne un'approssimazione (abbastanza buona) in tempo lineare. Teniamo a mente che per trovare l'autovettore di λ_2 serve tempo n^3 , che comunque è sempre migliore di un tempo esponenziale. Definiamo

$$R_L(x) = \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{d\|x\|^2}$$

ovvero il coefficiente di Rayleigh con argomento x e dove utilizziamo della

matrice laplaciana L . Vale che

$$\lambda_2 = \min_{\substack{x \in \mathbb{R}^n \setminus \{0\} \\ x^T \mathbf{1}}} R_L(x)$$

Teorema 10. Sia $x \in \mathbb{R}^n \setminus \{0\}$ tale che $x^T \mathbf{1} = 0$ e sia $S_f \subset V$ il cut trovato da Fiedler con input x . Allora

$$\phi(S_f) \leq \sqrt{2R_L(x)}$$

Se x è autovettore di λ_2 allora

$$\begin{aligned} \phi(G) &= \min_{S \subset V} \phi(S) \leq \phi(S_f) \\ &\leq \sqrt{2R_L(x)} \leq \sqrt{2\lambda_2} \end{aligned}$$

Per provare questo teorema abbiamo bisogno del seguente lemma.

Lemma 2. Sia $x \in \mathbb{R}^n \setminus \{0\}$ tale che $x^T \mathbf{1} = 0$, allora $\exists y \in \mathbb{R}^n$ con $y_i \geq 0 \ \forall i = 1, \dots, n$ tale che $R_L(y) \leq R_L(x)$. Inoltre, per ogni $0 < t \leq \max_i y_i$ il taglio $(\{v \in V : y_v \geq t\}, \{v \in V : y_v < t\})$ è uno di quelli considerati da Fiedler su input x .

Dimostrazione 23. Proviamo il lemma precedente. Innanzitutto prendiamo $x \in \mathbb{R}^n$, sia m la mediana e costruiamo $x^+, x^- \in \mathbb{R}^n$ definiti, componente per componente, come

$$\square \quad x_v^+ = [x_v - m]_+, \text{ cioè se } x_v > m \text{ allora } x_v^+ = x_v - m \text{ altrimenti } x_v^+ = 0.$$

$$\square \quad x_v^- = [m - x_v]_+, \text{ cioè se } x_v < m \text{ allora } x_v^- = m - x_v \text{ altrimenti } x_v^- = 0.$$

Dove $[z]_+ = z\mathbb{I}\{z > 0\}$.

Esempio 18. Per capire meglio come sono costruiti x^+, x^- consideriamo un esempio in cui $x = (3, -1, 5, 2)$, quindi la mediana è $m = 2.5$. Costruiamo x^+ :

1. 3 è maggiore di 2.5, quindi $x_1^+ = 3 - 2.5 = 0.5$
2. -1 è minore di 2.5, quindi $x_2^+ = 0$
3. 5 è maggiore di 2.5, quindi $x_3^+ = 5 - 2.5 = 2.5$
4. 2 è minore di 2.5, quindi $x_4^+ = 0$

Quindi $x^+ = (0.5, 0, 2.5, 0)$. In modo abbastanza simile possiamo costruire x^- :

1. 3 è maggiore di 2.5, quindi $x_1^- = 0$
2. -1 è minore di 2.5, quindi $x_2^- = 2.5 - (-1) = 3.5$
3. 5 è maggiore di 2.5, quindi $x_3^- = 0$
4. 2 è minore di 2.5, quindi $x_4^- = 2.5 - 2 = 0.5$

Quindi $x^- = (0, 3.5, 0, 0.5)$. Notiamo che

$$x = x^+ - x^- + m\mathbf{1}$$

infatti

$$(0.5, 0, 2.5, 0) - (0, 3.5, 0, 0.5) = \\ = (0.5, -3.5, 2.5, -0.5)$$

$$(0.5, -3.5, 2.5, -0.5) + \\ + (2.5, 2.5, 2.5, 2.5) = (3.5, -1, 2.5, 2)$$

Notiamo che $x_v^+, x_v^- \geq 0 \ \forall v \in V$.
Per ogni $t > 0, t \leq \max_v y_v$

$$\{v \in V : x_v^+ \geq t\} \\ \equiv \{v \in V : [x_v - m]_+ \geq t\}$$

Notiamo che siccome $t > 0$, sicuramente $x_v > m$, quindi $[x_v - m]_+ = x_v - m$.

$$\equiv \{v \in V : x_v \geq m + t\}$$

La stessa cosa vale per x^-

$$\{v \in V : x_v^- \geq t\} \\ \equiv \{v \in V : [m - x_v]_+ \geq t\} \\ \equiv \{v \in V : x_v \leq m - t\}$$

Ci rimane da dimostrare che $\exists y \geq 0 : R_L(y) \leq R_L(x)$. Poniamo

$$y = \operatorname{argmin}_{z \in \{x^-, x^+\}} R_L(z)$$

e anche $x' = x - m\mathbf{1} = x^+ - x^-$.
Prima di continuare osserviamo che $\forall c \ R_L(x + c\mathbf{1}) = R_L(x)$. Il numeratore non cambia, infatti su $(x_i - x_j)$ l'aggiunta di una costante non ha

effetto, dato che si eliminano a vicenda. Il denominatore diventa più grande (o al limite uguale)

$$\|x + c_1\|^2 = \|x\|^2 + c^2\|\mathbf{1}\|^2 + 2cx^T\mathbf{1}$$

siccome $x^T\mathbf{1} = 0$

$$> \|x\|^2$$

Quindi $R_L(x') \leq R_L(x)$. Consideriamo $R_L(y)$, vogliamo dimostrare che $R_L(y) \leq R_L(x')$.

$$R_L(y) = \min\{R_L(x^+), R_L(x^-)\}$$

Usiamo il fatto che $\min\{A, B\} \leq \alpha A + (1 - \alpha)B$ dove $0 \leq \alpha \leq 1$. Ovvero il minimo è minore o uguale della combinazione lineare di A e B . Nel nostro caso $\alpha = \frac{\|x^+\|^2}{\|x^+\|^2 + \|x^-\|^2}$

$$\leq \frac{\|x^+\|^2 R_L(x^+) + \|x^-\|^2 R_L(x^-)}{\|x^+\|^2 + \|x^-\|^2} \\ = \frac{\sum_{(i,j) \in E} (x_i^+ - x_j^+)^2 + \sum_{(i,j) \in E} (x_i^- - x_j^-)^2}{d(\|x^+\|^2 + \|x^-\|^2)}$$

Facciamo un passaggio (al numeratore) che dimostriamo dopo

$$\leq \frac{\sum_{(i,j) \in E} ((x_i^+ - x_j^+) - (x_i^- - x_j^-))^2}{d(\|x^+\|^2 + \|x^-\|^2)}$$

dove il denominatore è uguale perché, dato che $(x^+)^T x^- = 0$ (sicché se $x_i^+ = 0$ allora $x_i^- > 0$ e viceversa) abbiamo che

$$\|x^+\|^2 + \|x^-\|^2$$

$$\begin{aligned}
&= \|x^+\|^2 + \|x^-\|^2 - 2(x^+)^T x^- \\
&= \|x^+ - x^-\|^2 = \|x'\|^2
\end{aligned}$$

possiamo quindi concludere che

$$\frac{\sum_{(i,j) \in E} ((x_i^+ - x_j^+) - (x_i^- - x_j^-))^2}{d(\|x^+\|^2 + \|x^-\|^2)}$$

è uguale a

$$= \frac{\sum_{(i,j) \in E} (x'_i - x'_j)^2}{d\|x'\|^2}$$

Dimostriamo ora quel passaggio che abbiamo lasciato in sospeso, ovvero che $\forall (i, j) \in E$ vale

$$\begin{aligned}
&(x_i^+ - x_j^+)^2 + (x_i^- - x_j^-)^2 \leq \\
&\leq ((x_i^+ - x_j^+) - (x_i^- - x_j^-))^2
\end{aligned}$$

Abbiamo due casi:

1. $(x_i^+ - x_j^+)(x_i^- - x_j^-) = 0$. Allora o $x_i^+ - x_j^+$ è uguale a 0, oppure $x_i^- - x_j^-$ è uguale a 0. Quindi vale la disuguaglianza. Inoltre questo implica che o $x'_i, x'_j \geq 0$ oppure $x'_i, x'_j \leq 0$.
2. $(x_i^+ - x_j^+)(x_i^- - x_j^-) \neq 0$. Senza perdita di generalità possiamo dire che $x'_i < 0 < x'_j$, altrimenti saremmo nel primo caso. Allora

$$0 > x'_i = x_i^+ - x_i^- \leftrightarrow x_i^+ = 0$$

$$0 < x'_j = x_j^+ - x_j^- \leftrightarrow x_j^- = 0$$

Quindi

$$\begin{aligned}
&(x_i^+ - x_j^+)^2 + (x_i^- - x_j^-)^2 \\
&= (-x_j^+)^2 + (x_i^-)^2 \\
&= (x_j^+)^2 + (-x_i^-)^2
\end{aligned}$$

ricordando che x_j^+ e x_i^- hanno segno discorde:

$$\leq (x_j^+ - x_i^-)^2$$

Capitolo 13

Tredicesima lezione

Concludiamo la dimostrazione della disuguaglianza $\phi(G) \leq \sqrt{2\lambda_2}$. Per fare ciò dimostriamo un fatto e poi un lemma.

Fatto 16. $\forall X, Y$ tali che $y > 0$ e $\mathbb{E}[X], \mathbb{E}[Y] \leq \infty$ vale che

$$P\left(\frac{X}{Y} \leq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}\right) > 0$$

Dimostrazione 24. Poniamo

$$r = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$$

e notiamo che

$$\mathbb{E}[X] - r \cdot \mathbb{E}[Y] = \mathbb{E}[X - rY] = 0$$

siccome il valore atteso è 0 deve per forza valere

$$P(X - rY \leq 0) > 0$$

altrimenti il valore atteso sarebbe positivo. Se dividiamo per Y

$$P\left(\frac{X}{Y} - r \leq 0\right) > 0$$

e abbiamo concluso la dimostrazione.

Lemma 3. $\forall y \in \mathbb{R}_{\geq 0}^n \exists t$ con $0 < t < \max_v y_v$ tale che

$$xpn(\{v \in V : y_v \geq t\}) \leq \sqrt{2R_L(y)}$$

Dimostrazione 25. Dato che $R_L(y) = R_L(c \cdot y) \forall c > 0$ possiamo assumere che $\max_v y_v = 1$, ovvero scaliamo senza perdita di generalità. Sia $t \in [0, 1]$ una variabile casuale tale che

$$P(t \leq \sqrt{a}) = a \quad \forall a \in [0, 1]$$

quindi t^2 è distribuita uniformemente in $[0, 1]$. Definiamo

$$S_t = \{v \in V : y_v \geq t\}$$

vale che

$$|S_t| > 0 \quad \forall t \in (0, 1]$$

dato che y_t è non negativo, se t è minimo allora prendo tutti i vertici, se $t = 1$ prendo solo il massimo. Qualsiasi sia la scelta non è mai vuoto. Prendiamo l'espansione

$$xpn(S_t) = \frac{|E(S_t, \neg S_t)|}{d|S_t|}$$

Usando il fatto di prima possiamo dire che

$$\leq \frac{\mathbb{E}[|E(S_t, \neg S_t)|]}{d\mathbb{E}[|S_t|]}$$

con probabilità maggiore di 0 rispetto a estrazione di t . Allora esiste t tale che la disequazione vale.

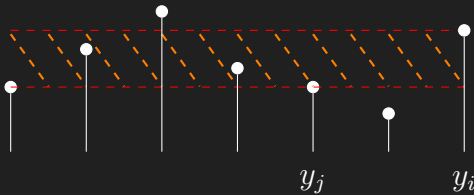
Analizziamo il denominatore

$$\begin{aligned} \mathbb{E}[|S_t|] &= \mathbb{E}\left[\sum_{v \in V} \mathbb{I}\{v \in S_t\}\right] \\ &= \sum_{v \in V} P(v \in S_t) = \sum_{v \in V} P(t \leq y_v) \\ &= \sum_{v \in V} y_v^2 \end{aligned}$$

Scegliamo qualsiasi $(i, j) \in E$ e assumiamo che $y_j \leq y_i$. Allora,

$$P(i \in S_t, j \in \neg S_t) = P(y_j < t \leq y_i)$$

Si consideri il disegno sottostante



la parte tratteggiata è la parte che "teniamo". Il resto viene cancellato, quindi:

$$= P(t \leq y_i) - P(t \leq y_j) = y_i^2 - y_j^2$$

Ci concentriamo ora sul numeratore:

$$\mathbb{E}[|E(S_t, \neg S_t)|] = \mathbb{E}\left[\sum_{(i,j) \in E} \mathbb{I}\{i \in S_t, j \in \neg S_t\}\right]$$

$$\begin{aligned} &= \sum_{(i,j) \in E} P(i \in S_t, j \in \neg S_t) \\ &= \sum_{(i,j) \in E} ((y_i^2 - y_j^2) \mathbb{I}\{y_i \leq y_j\} + (y_j^2 - y_i^2) \mathbb{I}\{y_j \leq y_i\}) \\ &= \sum_{(i,j) \in E} |y_i^2 - y_j^2| = \sum_{(i,j) \in E} |y_i - y_j| (y_i + y_j) \end{aligned}$$

Prendiamo $V, U \in \mathbb{R}^n$, per la **disuguaglianza di Cauchy Schwarz** vale che

$$U^T V \leq \|U\| \cdot \|V\|$$

nel nostro caso

$$\leq \sqrt{\sum_{(i,j) \in E} (y_i - y_j)^2} \sqrt{\sum_{(i,j) \in E} (y_i + y_j)^2}$$

Usando la disuguaglianza $(a + b)^2 \leq 2(a^2 + b^2)$ la sommatoria interna può essere trasformata in

$$\sum_{(i,j) \in E} (y_i + y_j)^2 \leq 2 \sum_{(i,j) \in E} (y_i^2 + y_j^2) = 2d \sum_{v \in V} y_v^2$$

quindi

$$xpn(S_t) \leq \frac{\sqrt{2d\|Y\|^2 \sum_{(i,j) \in E} (y_i - y_j)^2}}{d\|Y\|^2}$$

portiamo tutto dentro radice

$$= \sqrt{\frac{2 \sum_{(i,j) \in E} (y_i - y_j)^2}{d\|Y\|^2}} = \sqrt{2R_L(Y)}$$

concludendo così la dimostrazione.

Possiamo terminare la dimostrazione del teorema. Per il lemma

$\exists Y \in \mathbb{R}_{\geq 0}^n$ $S_t = \{v \in V : y_v \geq t\}$ possiamo quindi scrivere

$$\phi(S_f) = \min_t \phi(S_t)$$

Ricordiamo che

$$\phi(S) = \frac{|E(S, \neg S)|}{d \min\{|S|, |\neg S|\}}$$

e

$$xpn(S) = \frac{|E(S, \neg S)|}{d|S|}$$

ma Y , per costruzione, ha al più $\frac{n}{2}$ componenti non zero ($Y \in \{x^-, x^+\}$), ricordiamo che Y lo avevamo definito come $Y = \arg \min_{Z \in \{x^-, x^+\}} R_L(Z)$. Questo implica che $|S_t| \leq \frac{n}{2} \forall t$, invece $|\neg S_t| \geq \frac{n}{2}$. E' quindi ovvio che $\min(|S_t|, |\neg S_t|) = |S_t|$, quindi $\phi(S_t) = xpn(S_t)$.

$$\phi(S_f) = \min_t \phi(S_t) = \min_t xpn(S_t)$$

$$\leq \sqrt{2R_L(Y)}$$

per il lemma $R_L(Y) \leq R_L(X)$

$$\leq \sqrt{2R_L(X)}$$

che conclude la dimostrazione! Quindi, per grafi d -regolari, vale che

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

13.1 Grafi non regolari

Analizziamo brevemente cosa succede se il grafo non è d -regolare. Guardiamo alla generalizzazione del coefficiente di Rayleigh R_L .

$$R_L = \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{v \in V} d(v)x_v^2}$$

$$= \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{v \in V} (\sqrt{d(v)}x_v)(\sqrt{d(v)}x_v)}$$

Costruiamo la matrice diagonale D

$$D = \begin{bmatrix} d(v_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d(v_n) \end{bmatrix}$$

e siccome è diagonale ogni operazione sulla matrice si applica sui singoli elementi

$$\sqrt{D} = \begin{bmatrix} \sqrt{d(v_1)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{d(v_n)} \end{bmatrix}$$

stessa cosa vale per l'inversione

$$D^{-1} = \begin{bmatrix} d(v_1)^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d(v_n)^{-1} \end{bmatrix}$$

Ricordiamoci che

$$x^T Lx = \frac{1}{d} \sum_{(i,j) \in E} (x_i - x_j)^2$$

possiamo ottenere un risultato simile nel caso generalizzato. Prendiamo $x^T L D x$, dove $L = I - \frac{A}{D}$, e analizziamo cosa succede.

$$x^T L D x = x^T \left(I - \frac{A}{D} \right) D x$$

$$= x^T (D - A) x = x^T D x - x^T A x$$

$$= \sum_i \sum_j x_i D_{ij} x_j - \sum_i \sum_j x_i A_{ij} x_j$$

notiamo che $D_{ij} \neq 0$ solo quando $i = j$.

$$= \sum_i x_i^2 D_{ii} + \sum_i \sum_j x_i A_{ij} x_j$$

come avevamo fatto introducendo la Laplaciana possiamo scrivere

$$= \sum_{i \in V} \sum_{j: (i,j) \in E} \frac{x_i^2}{D_{ii}} D_{ii} + \sum_{i \in V} \sum_{j: (i,j) \in E} x_i x_j$$

da qui facilmente, come fatto nelle lezioni precedenti, otteniamo

$$= \sum_{(i,j) \in E} (x_i - x_j)^2$$

abbiamo dimostrato che

$$x^T (D - A) x = \sum_{(i,j) \in E} (x_i - x_j)^2$$

Allora riprendiamo il coefficiente di Rayleigh e scriviamo

$$R_L = \frac{x^T (D - A) x}{\sum_{v \in V} (\sqrt{d(v)} x_v) (\sqrt{d(v)} x_v)}$$

il denominatore può essere riscritto come

$$= \frac{x^T (D - A) x}{(D^{\frac{1}{2}} x)^T (D^{\frac{1}{2}} x)}$$

poniamo $U = D^{\frac{1}{2}} x$ e quindi ne deriva che $x = D^{-\frac{1}{2}} U$

$$= \frac{(D^{-\frac{1}{2}} U)^T (D - A) D^{-\frac{1}{2}} U}{U^T U}$$

sappiamo che $U^T U = \|U\|^2$ e anche che $D^T = D$ perché diagonale

$$\frac{U^T D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} U}{\|U\|^2}$$

espandiamo il prodotto

$$= \frac{U^T D^{-\frac{1}{2}} D D^{-\frac{1}{2}} U - U^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} U}{\|U\|^2}$$

notiamo che

$$D^{-\frac{1}{2}} D D^{-\frac{1}{2}} = D^{-\frac{1}{2}} D^{\frac{1}{2}} D^{\frac{1}{2}} D^{-\frac{1}{2}} = I$$

quindi

$$R_L = \frac{U^T (I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) U}{\|U\|^2}$$

Sia $G = (V, E)$ qualunque il Laplaciano normalizzato di G è

$$L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

e le disuguaglianze di Cheeger continuano a valere. Notiamo che se siamo nel caso d -regolare, ovvero

$$D^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{d}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{d}} \end{bmatrix}$$

allora

$$D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = \frac{1}{d} A$$

e quindi

$$L = I - \frac{1}{d} A$$

13.2 Introduzione clique pianate

Vogliamo usare lo spettro per trovare comunità in reti sociali. Dato un grafo $G = (V, E)$ i nodi rappresentano le persone e i, j sono amici (o hanno interessi in comune) se e solo se $(i, j) \in E$. Una comunità è un sotto-grafo denso in un grafo, o alternativamente possiamo dire che è un insieme di utenti densamente connessi. Prendiamo un grafo $G_n \sim \mathcal{G}(n, \frac{1}{2})$ (dove scegliamo $\frac{1}{2}$ per semplicità) e consideriamo una clique $S \subseteq V$ tale che $|S| = o(|V|)$ costruita nel modo seguente:

1. Scelgo $S \subseteq V$, tale che $|S| = k$
2. Estraggo $G_n \sim \mathcal{G}(n, \frac{1}{2})$
3. Completo la clique su S aggiungendo gli archi mancanti.

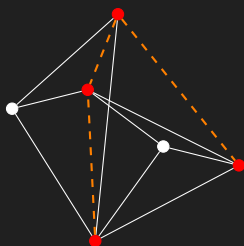
Sappiamo che $\omega(G_n) \sim 2 \log_2 n$, quindi ha senso cercare una clique solamente se $k \gg 2 \log_2 n$, perché altrimenti non è una comunità ma una clique che si forma casualmente.

Capitolo 14

Quattordicesima lezione

Sia $|V| = n$ e $S \subset V$ tale che $|S| = k$. Sia $G_n \sim \mathcal{G}(n, \frac{1}{2})$ un grafo casuale. Definiamo G come il grafo G_n a cui vengono aggiunti gli archi mancanti per ottenere una clique su S .

Esempio 19. In figura, il grafo generato G_n . In rosso i nodi appartenenti a S e in arancione gli archi aggiunti per ottenere il grafo G dove S è una cricca.



Come spiegato la scorsa lezione, ci interessa solo il caso in cui $k \gg 2 \log_2 n$. Un primo algoritmo banale per trovare una cricca in un grafo è il seguente

Algorithm 4

- 1: Enumera tutti i sottoinsiemi di taglia k in V
 - 2: Verifica che c'è una clique
-

Questo algoritmo ha tempo

$$\binom{n}{k} \binom{k}{2} \simeq k^2 \left(\frac{n}{k}\right)^k$$

ovvero esponenziale in k , dove k è funzione di n . Diamo adesso un algoritmo più furbo

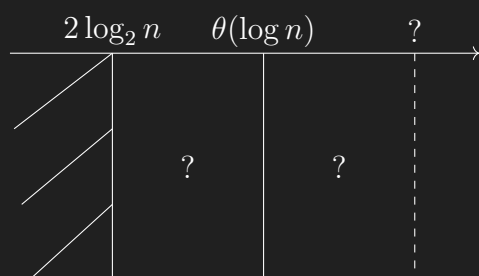
Algorithm 5

- 1: Enumera tutti i sottoinsiemi di taglia $3 \log_2 n < k$.
 - 2: Verifica se c'è una clique sul sottoinsieme S' considerato.
 - 3: Cerca $T \subset V \setminus S'$ tale che ogni vertice in T è collegato con tutti i vertici di S' .
 - 4: Verifico che $T \cup S'$ sia una clique.
-

Il tempo dell'algoritmo è

$$\binom{n}{3 \log_2 n} \simeq k^2 n^{3 \log_2 n}$$

ovvero è quasi polinomiale. E' un problema aperto determinare se esista un algoritmo polinomiale per trovare una clique di taglia $\log n$. In dipendenza della grandezza della clique che cerchiamo riusciamo a trovare algoritmi più o meno ottimi.



Non ha senso andare a cercare clique più piccole di $2 \log_2 n$. Infatti, in questo caso, sarebbero naturali all'interno del grafo. Tra $2 \log_2 n$ e $\theta(\log n)$ non sappiamo se esista un algoritmo efficiente. Sappiamo che più una clique è grande e più è facile trovarla. Ci chiediamo qual è quel valore tale per cui, da quel valore in su, esista un algoritmo efficiente per trovare una clique? Si consideri il seguente algoritmo

Algorithm 6

- 1: Ordina i vertici per grado decrescente.
 - 2: Trova il più grande k tale che i primi k vertici in questo ordine formano una clique.
-

E' ovvio che più un vertice abbia grado alto più è facile che compaia all'interno di una clique. Il tempo che ci impiega questo algoritmo è $n \log n + k^2$. Analizziamo l'algoritmo appena presentato. Sia $G_n \sim \mathcal{G}(n, \frac{1}{2})$ e sia $d_{G_n}(v)$ una variabile aleatoria binomiale di parametri $(n - 1, \frac{1}{2})$, per ogni v , che rappresenta il numero di vicini di v . Poniamo, per semplicità, $d_{G_n}(v) = N_v$. Possiamo affermare quanto segue:

$$\forall v \ N_v \leq \frac{n-1}{2} + \sqrt{\frac{n-1}{2} \ln \frac{1}{\delta}}$$

con probabilità $1 - \delta$. Questo vale per il lemma di Chernoff-Hoffding.

Sempre per questo lemma possiamo affermare che

$$\forall v \ N_v \geq \frac{n-1}{2} - \sqrt{\frac{n-1}{2} \ln \frac{1}{\delta}}$$

con probabilità $1 - \delta$. Andiamo a considerare il grafo G , definito all'inizio della lezione come il grafo G_n a cui vengono aggiunti gli archi per completare la clique S . Qual è il numero minimo di vicini di un qualsiasi $v \in S$? Cominciamo con il dire che sicuramente

$$\min_{v \in S} d_G(v) \geq \frac{n-1}{2} - \sqrt{\frac{n-1}{2} \ln \frac{2}{\delta^*}}$$

dove $\delta^* = \frac{\delta}{2}$, e quindi la disequazione vale con probabilità $1 - \frac{\delta}{2}$. Questa

disequazione segue dalla disequazione di prima, ed è il numero di archi random che vengono generati. Possiamo ottenere un bound più stretto. Cominciamo con il notare che $\forall v \in S$ il numero di archi è tale che

$$|N_{G_n}(v) \cap S| \leq \frac{k-1}{2} + \sqrt{\frac{k-1}{2} \ln \frac{2}{\delta^*}}$$

Ci chiediamo, qual è il massimo valore che può assumere $|N_{G_n}(v) \cap S|$, con $v \in S$? Utilizziamo la regola dell'union bound:

$$P(\max_{v \in S} |N_{G_n}(v) \cap S| \geq \varepsilon)$$

$$\text{dove } \varepsilon = \frac{k-1}{2} + \sqrt{\frac{k-1}{2} \ln \frac{2}{\delta^*}}.$$

$$= P(\exists v \in S : |N_{G_n}(v) \cap S| > \varepsilon)$$

$$= P(\bigcup_{v \in S} |N_{G_n}(v) \cap S| > \varepsilon)$$

$$\leq \sum_{v \in S} P(|N_{G_n}(v) \cap S| > \varepsilon)$$

Sappiamo che $|N_{G_n}(v) \cap S| \leq \varepsilon$ con probabilità $1 - \delta^*$, quindi $|N_{G_n}(v) \cap S| > \varepsilon$ con probabilità $1 - (1 - \delta^*) = \delta^*$. Siccome la sommatoria ha k elementi scriviamo

$$\leq k\delta^*$$

vogliamo rendere questa probabilità minore di $\frac{\delta'}{2}$, ovvero

$$k\delta^* \leq \frac{\delta'}{2}$$

Quindi

$$\delta^* \leq \frac{\delta'}{2k}$$

Questo vuol dire che con probabilità $1 - \frac{\delta'}{2}$ vale che

$$\max_{v \in S} |N_{G_n}(v) \cap S| \leq \frac{k-1}{2} + \sqrt{\frac{k-1}{2} \ln \frac{2k}{\delta'}}$$

Notiamo che quanto valga δ' non ci interessa, quello ci serve è la relazione tra l'argomento del logaritmo e la probabilità. Quindi, possiamo porre $\delta' = \delta$, dato che non ha alcun significato numerico, e non perdere alcuna informazione. Per arrivare ad avere una clique completa, dovrò aggiungere, almeno

$$\frac{k-1}{2} + \sqrt{\frac{k-1}{2} \ln \frac{2k}{\delta'}}$$

infatti

$$\frac{k-1}{2} + \sqrt{\frac{k-1}{2} \ln \frac{2k}{\delta'}} - \sqrt{\frac{k-1}{2} \ln \frac{2k}{\delta'}} = \frac{k-1}{2}$$

che è il numero di vicini per ogni vertice all'interno di una clique. Quindi il vertice $v \in S$ con il minimo numero di vicini in G ha almeno i vertici che ha ottenuto randomicamente in G_n , più tutti quelli che servono a completare la clique.

$$\min_{v \in S} d_G(v) \geq \frac{n-1}{2} - \sqrt{\frac{n-1}{2} \ln \frac{2}{\delta}} + \frac{k-1}{2} - \sqrt{\frac{k-1}{2} \ln \frac{2k}{\delta}}$$

siccome $-\sqrt{\frac{n-1}{2} \ln \frac{2}{\delta}} \geq -\sqrt{\frac{k-1}{2} \ln \frac{2k}{\delta}}$ possiamo scrivere

$$\geq \frac{n-1}{2} + \frac{k-1}{2} - 2\sqrt{\frac{n-1}{2} \ln \frac{2}{\delta}}$$

con probabilità $1 - \delta$ e $n \geq k$. Ricordando che non abbiamo aggiunto archi a vertici non in S possiamo affermare

$$\max_{v \in V \setminus S} d_G(v) \leq \frac{n-1}{2} + \sqrt{\frac{n-1}{2} \ln \frac{n}{\delta}}$$

Supponiamo di prendere


$$\frac{k-1}{2} \geq 4\sqrt{\frac{n-1}{2} \ln \frac{n}{\delta}}$$

che vuol dire che $k = \Omega(\sqrt{n \log n})$. Allora, a parte un piccolo fattore 2, possiamo dire che

$$\min_{v \in S} d_G(v) \geq \frac{n-1}{2} + 2\sqrt{\frac{n-1}{2} \ln \frac{n}{\delta}}$$

Notiamo che, il vertice v con il grado minore tra quelli in S ha un grado che è maggiore del vertice con il grado maggiore tra quelli non in S .

Quindi l'algoritmo ha un'alta probabilità di trovare la clique considerando per i primi i vertici con grado maggiore.

$2 \log_2 n$	$\theta(\log n)$?	$\Omega(\sqrt{n \log n})$	
	?	?	?	✓

Ci interessa ora capire cosa succeda tra $\theta(\log n)$ e $\Omega(\sqrt{n \log n})$. Consideriamo il caso $k = \Omega(\sqrt{n})$ e cerchiamo un algoritmo efficiente. E' ovvio che il caso di prima non funziona, quindi usiamo tecniche spettrali. Definiamo la matrice J_n di dimensioni $n \times n$ come

$$J_n = \mathbf{1}^T \mathbf{1} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

Data M simmetrica definiamo $\lambda_{\max}(M)$ come l'autovalore massimo. Notiamo che

$$\mathbf{1}^T M \mathbf{1} = \sum_{i,j} M_{ij}$$

Enunciamo un lemma che non dimostriamo.

Lemma 4. Sia $G_n \sim \mathcal{G}(n, \frac{1}{2})$ con matrice di adiacenza A_n . Allora

$$\lambda_{max}(A - \frac{1}{2}J_n) \leq 2\sqrt{n}$$

con alta probabilità.

La matrice $-\frac{1}{2}J_n$ è tale che

$$\begin{bmatrix} -\frac{1}{2} & \cdots & -\frac{1}{2} \\ \vdots & \ddots & \vdots \\ -\frac{1}{2} & \cdots & -\frac{1}{2} \end{bmatrix}$$

La matrice A avrà su ogni riga circa $\frac{n-1}{2}$ uni, perché è il grado medio dei nodi in G_n . Ci saranno, quindi, alcuni zeri, che diventeranno $-\frac{1}{2}$, e ci saranno alcuni uni che diventeranno $\frac{1}{2}$. La matrice $(A - \frac{1}{2}J_n)$ sarà formata da $\frac{1}{2}$ e $-\frac{1}{2}$. Se la matrice fossero tutti $\frac{1}{2}$ (la diagonale sarebbe $-\frac{1}{2}$ perché sono degli zeri in A , ma non ha alcuna importanza) e la moltiplicassimo per il vettore di tutti 1 otteniamo

$$\begin{bmatrix} \frac{1}{2} & \cdots & \frac{1}{2} \\ \vdots & \ddots & \vdots \\ \frac{1}{2} & \cdots & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{n}{2} \\ \vdots \\ \frac{n}{2} \end{bmatrix}$$

E' abbastanza evidente che $\lambda_{max} = \frac{n}{2}$, e cioè sarebbe una clique l'intero grafo.

Capitolo 15

Quindicesima lezione

15.1 Caso $k = \Omega(\sqrt{n})$

Nella scorsa lezione abbiamo visto che dato un grafo $G_n \sim \mathcal{G}(n, \frac{1}{2})$, con $|V| = n$ e $S \subseteq V$ $|S| = k$, la clique completata su S viene trovata in modo efficiente se $k = \Omega(\sqrt{n} \log n)$. In questa lezione ci concentriamo sul caso in cui $k = \Omega(\sqrt{n})$ e andremo a utilizzare il fatto introdotto la scorsa lezione. Cominciamo a definire

□ G il grafo perturbato, ovvero il grafo in cui è presente la clique nascosta.

□ A la matrice di adiacenza di G

ci chiediamo come si comporta l'autovalore massimo λ_{max} . Cominciamo a esprimerlo attraverso il coefficiente di Rayleigh

$$\lambda_{max}(A - \frac{J_n}{2}) = \max_{x: x \neq 0} \frac{x^T (A - \frac{J_n}{2}) x}{x^T x}$$

Costruiamo il vettore $\mathbf{1}_s \in \{0, 1\}^n$ che è definito come

$$(\mathbf{1}_s)_i = \begin{cases} 1 & \text{se } i \in S \\ 0 & \text{altrimenti} \end{cases}$$

siccome $\mathbf{1}_s$ è un qualunque vettore diverso da 0, vale che

$$\begin{aligned} &\geq \frac{\mathbf{1}_s^T (A - \frac{J_n}{2}) \mathbf{1}_s}{\mathbf{1}_s^T \mathbf{1}_s} \\ &= \frac{\mathbf{1}_s^T A \mathbf{1}_s - \frac{1}{2} \mathbf{1}_s^T J_n \mathbf{1}_s}{\mathbf{1}_s^T \mathbf{1}_s} \end{aligned}$$

Facciamo le seguenti considerazioni:

□ Il vettore $\mathbf{1}_s$ ha esattamente k posizioni poste a 1, mentre le altre sono tutte poste a 0. Per questo motivo il denominatore è tale che

$$\mathbf{1}_s^T \mathbf{1}_s = k$$

□ L'espressione $\frac{1}{2} \mathbf{1}_s^T J_n \mathbf{1}_s$ può essere riscritta come segue

$$\sum_{i=1}^n \sum_{j=1}^n (\mathbf{1}_s)_i (\mathbf{1}_s)_j = k^2$$

dove l'1 compare perché J_n è una matrice di soli 1.

□ L'espressione $\mathbf{1}_s^T A \mathbf{1}_s$ è invece esprimibile come

$$\sum_{i \in S} \sum_{j \in S} A_{ij} = k(k-1)$$

e intuitivamente indica che ci sono k elementi connessi con $k-1$ elementi, cioè formano una clique.

Fatte queste considerazioni possiamo riscrivere l'espressione precedente come

$$= \frac{k(k-1) - \frac{k^2}{2}}{k} = \frac{k}{2} - 1$$

Ricapitolando abbiamo che

$$\lambda_{\max}(A_n - \frac{J_n}{2}) \leq 2\sqrt{n}$$

$$\lambda_{\max}(A - \frac{J_n}{2}) \geq \frac{k}{2} - 1$$

Se $k = \Omega(\sqrt{n})$ possiamo distinguere G_n da G con una clique nascosta. Non abbiamo un modo per capire quale sia questa clique nascosta. Consideriamo la matrice $A = A_n + A_s$ dove A_n è la matrice di adiacenza di G_n e A_s è la matrice che contiene gli archi aggiunti. Potremmo vedere A_s come la matrice di adiacenza di un grafo $G_k \sim \mathcal{G}_{\parallel}(k, \frac{1}{2})$, anche se

non sarebbe propriamente corretto dato che A_s ha dimensioni $n \times n$. Per il lemma sappiamo che con alta probabilità vale

$$\lambda_{\max}(A_n - \frac{J_n}{2}) \leq 2\sqrt{n}$$

$$\lambda_{\max}(A_s - \frac{J_s}{2}) \leq 2\sqrt{k}$$

dove $J_s = \mathbf{1}_s \mathbf{1}_s^T$. Sia $x \in R^n$ un autovettore con autovalore $\lambda_{\max}(A - \frac{J_n}{2})$. Assumiamo che x abbia norma 1 per semplicità. Sappiamo che

$$\frac{k}{2} - 1 \leq \frac{x^T(A - \frac{J_n}{2})x}{x^T x}$$

siccome $\|x\| = 1$ allora $x^T x = 1$. Utilizziamo il fatto che $A = A_n + A_s$

$$= x^T A_s x + x^T(A_n - \frac{J_n}{2})x$$

$$\leq x^T A_s x + 2\sqrt{n}$$

Sistemando i membri

$$\frac{k}{2} - 1 - 2\sqrt{n} \leq x^T A_s x$$

sommiamo e sottraiamo $\frac{J_s}{2}$

$$= x^T(A_s - \frac{J_s}{2} + \frac{J_s}{2})x$$

espandiamo la parentesi e usiamo il fatto che $J_s = \mathbf{1}_s \mathbf{1}_s^T$

$$= x^T(A_s - \frac{J_s}{2})x + x^T \frac{\mathbf{1}_s \mathbf{1}_s^T}{2} x$$

ricordando la definizione di autovalore massimo e che $(x^T y)^2 = x^T y y^T x$ possiamo scrivere

$$\leq \lambda_{\max}(A_s - \frac{J_s}{2}) + \frac{1}{2}(x^T \mathbf{1}_s)^2$$

allora con alta probabilità

$$\leq 2\sqrt{k} + \frac{1}{2}(x^T \mathbf{1}_s)^2$$

Definiamo il vettore $y_i = |x_i| \forall i = 1, \dots, n$. Notiamo che

$$\begin{aligned} y^T \mathbf{1}_s &= \sum_{i \in S} y_i = \sum_{i \in S} |x_i| \\ &\geq |\sum x_i| = |x^T \mathbf{1}_s| \end{aligned}$$

Riprendiamo il risultato che abbiamo trovato poco fa e risolviamo per x

$$(x^T \mathbf{1}_s)^2 \geq k - 2 - 4\sqrt{n} - 4\sqrt{k}$$

se $k = \Omega(\sqrt{n})$ e per n sufficientemente grande

$$\geq (1 - \alpha)^2 k$$

Per comprendere come mai questa affermazione valga, ricordiamo che $k = \Omega(\sqrt{n})$ e quindi $k = c\sqrt{n}$. Quindi,

$$\begin{aligned} k - 2 - 4\sqrt{n} - 4\sqrt{k} \\ = c\sqrt{n} - 2 - 4\sqrt{n} - 4\sqrt{c\sqrt{n}} \end{aligned}$$

ignoriamo il fattore $4\sqrt{\sqrt{n}}$ e -2 dato che influiscono poco

$$= (c - 4)\sqrt{n}$$

vogliamo che valga

$$(c - 4)\sqrt{n} \geq (1 - \alpha)^2 \sqrt{n}$$

ovvero

$$(c - 4) \geq (1 - \alpha)^2$$

Risolviamo per c e otteniamo

$$c \geq \frac{(1 - \alpha)^2}{4}$$

Quindi per un qualsiasi α posso trovare k dell'ordine di \sqrt{n} tale per cui vale la disequazione

$$(x^T \mathbf{1}_s)^2 \geq (1 - \alpha)^2 k$$

Prendiamo la radice

$$||x^T \mathbf{1}_s|| \geq (1 - \alpha)\sqrt{k}$$

Fatto 17. Il vettore $\sqrt{k}y$ è vicino a $\mathbf{1}_s$.

Dimostrazione 26. Cominciamo con l'osservare che $\mathbf{1}_s$ è il vettore che descrive la clique, se conosciamo $\mathbf{1}_s$ conosciamo la clique, sappiamo dove sta. Quindi il vettore $\sqrt{k}y$ è una buona proxy per trovare la clique. Per dimostrare che sono vicini andiamo a vedere la loro distanza al quadrato.

$$||\mathbf{1}_s - \sqrt{k}y||^2 = ||\mathbf{1}_s||^2 + k||y||^2 - 2\sqrt{k}(y^T \mathbf{1}_s)$$

sappiamo che $\|\mathbf{1}_s\|^2 = k$ e che $k\|y\|^2 = k$ quindi

$$\leq 2k - 2(1 - \alpha)k$$

dove la disuguaglianza segue dal fatto che

$$y^T \mathbf{1}_s \geq |x^T \mathbf{1}_s| \geq (1 - \alpha)\sqrt{k}$$

allora concludiamo con il dire

$$= 2\alpha k$$

Dato che possiamo scegliere α come ci pare sono vicini tanto quanto vogliamo.

15.2 Algoritmo Improved Clique Finder

Diamo adesso un algoritmo per trovare la clique

Algorithm 7

Input: $G, k, \epsilon > 0$

- 1: Calcola A per G
- 2: Sia x l'autovettore di $\lambda_{\max}(A - \frac{J_n}{2})$
- 3: Sia L l'insieme contenente i k vertici corrispondenti alle componenti più grandi di y

Output: Genero i vertici di y con almeno $(\frac{3}{4} - \epsilon)k$ vicini in L

Chiariamo il contenuto del vettore L . Se y fosse il vettore $[0, 1, 2, 3, -3]$ e $k = 3$, il vettore L sarebbe $[1, 2, 3]$ ovvero i 3 elementi più grandi.

Cominciamo con il considerare un valore soglia t tale che

$$i \in L \Leftrightarrow \sqrt{k}y_i \geq t$$

Definiamo

$$m = |S \setminus L| = |L \setminus S|$$

dove la seconda uguaglianza vale perché L e S hanno la stessa cardinalità. Quindi

$$\|\mathbf{1}_s - \sqrt{k}y\|^2 = \sum_{i \in S} (1 - \sqrt{k}y_i)^2 + \sum_{j \notin S} ky_j^2$$

dove abbiamo diviso gli elementi che stanno in S (e quindi $(\mathbf{1}_s)_i = 1$) da quelli che non stanno in S (e quindi $(\mathbf{1}_s)_i = 0$). Possiamo minorare l'espressione precedente con

$$\geq \sum_{i \in S \setminus L} (1 - \sqrt{k}y_i)^2 + \sum_{j \in L \setminus S} ky_j^2$$

è facile convincersi che questo valga osservando che $S \setminus L \subseteq S$, $L \setminus S \subseteq L$ e anche che tutti gli elementi della sommatoria sono positivi dato che sono dei quadrati. Osserviamo che se $j \in L$ allora $y_j \sqrt{k} \geq t$ per costruzione. Quindi, $\forall j \in L$

$$y_l \geq \frac{t}{\sqrt{k}}$$

Siccome y è un vettore con norma 1, se poniamo $y_j = 0$ $j > k$ allora

$$y_1, \dots, y_k = \frac{1}{\sqrt{k}}$$

e dunque t è al massimo 1. Possiamo dire che

$$\begin{aligned} &\geq \sum_{i \in S \setminus L} (1 - \sqrt{k}y_i)^2 + \sum_{j \in L \setminus S} ky_j^2 \\ &\geq \sum_{i \in S \setminus L} (1 - t)^2 + \sum_{j \in L \setminus S} t^2 \end{aligned}$$

dato che:

□ Se $i \notin L$ allora $\sqrt{k}y_i < t$, per definizione.

□ Se $i \in L$ allora $y_i \geq \frac{t}{\sqrt{k}}$

Sappiamo che la cardinalità dei due insiemi $S \setminus L$ e $L \setminus S$ è esattamente m

$$= m(1 - t)^2 + mt^2$$

con qualche manipolazione algebrica arriviamo a

$$= m(1 - 2t + 2t^2)$$

cerchiamo un minorante per t . Quando $1 - 2t + 2t^2$ assume un valore minimo? Prendiamo la derivata e poniamola = 0.

$$-4t + 2 = 0 \Leftrightarrow t = \frac{1}{2}$$

Allora

$$m(1 - 2t + 2t^2) \geq \frac{m}{2}$$

Quindi

$$\frac{m}{2} \leq \|\mathbf{1}_S - \sqrt{k}y\|^2 \leq 2\alpha k$$

Scegliamo $\alpha = \frac{1}{16}$

$$m \leq \frac{k}{4}$$

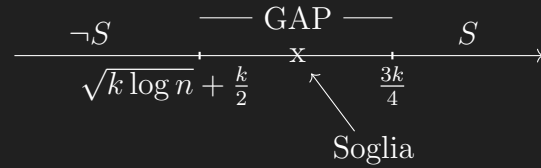
Questo vuol dire che L contiene almeno $\frac{3}{4}k$ elementi di S . Ogni elemento di S ha almeno $\frac{3k}{4}$ vicini in L

$$\forall v \in S \quad |N_v \cap L| \geq \frac{3}{4}k$$

con alta probabilità. I nodi che non stanno in S hanno un vicinato che è generato in modo casuale e quindi

$$\forall v \notin S \quad |N_v \cap L| \leq \frac{k}{2} + \sqrt{k \log n}$$

con alta probabilità.



Riusciamo a discriminare elementi di S , ovvero parte della clique, da elementi che non fanno parte di S , ovvero non della clique, con alta probabilità.

[12pt]report mioStile

Teoria dei grafi Andrea Cosentino
13 agosto 2024

Capitolo 16

Sedicesima lezione

16.1 Introduzione camminate casuali su grafi

Le camminate casuali su grafi (random walks) sono dei processi stocastici su una struttura discreta (come il grafo). Trovano applicazioni in molti contesti, tra cui:

- PageRank: ranking di pagine web. Veniva usato da google per identificare le pagine più importanti. Una pagina è tanto importante tanto quanto sono le connessioni ad altre pagine web (definizione ricorsiva).
- Diffusione di informazione su una rete sociale online. Modellizza processo di diffusione virale, come notizie, meme, etc... Chiamato anche "cascading process".
- Stima di proprietà di una rete grande: n° di nodi, diametro, distribuzione dei gradi.

- Estrazione casuale di elementi da un insieme combinatoriale: alberi di copertura di un grafo, cicli hamiltoniani, le permutazioni di un insieme finito che soddisfano certi vincoli.

16.2 Relazione tra autovalori e grado del grafo

Consideriamo un grafo (semplice) $G = (V, E)$, la sua matrice di adiacenza A con autovalori $\alpha_1 \geq \dots \geq \alpha_n$ e la matrice laplaciana L con autovalori $\lambda_1 \leq \dots \leq \lambda_n$. Notiamo che la numerazione degli autovalori di A e L è invertita. Infatti, se G è d -regolare, abbiamo che

$$L = I - \frac{1}{d} A \Rightarrow \lambda_i = 1 - \frac{\alpha_i}{d}$$

il segno $-$ fa ribaltare l'ordine degli autovalori. Inoltre, $\alpha_i \in [-d, d]$ poiché $\lambda_i \in [0, 2]$ (già dimostrato in passato), infatti:

$$\lambda_i = 1 - \frac{\alpha_i}{d} \Rightarrow \alpha_i = d(1 - \lambda_i)$$

Ricordiamo che con $D(G)$ indichiamo il grado medio dei nodi in G e con $\Delta(G)$ il grado massimo. E' facile vedere che $\forall G$ (anche non regolare) vale che

$$D(G) \leq \alpha_1 \leq \Delta(G)$$

Dimostrazione 27. Cominciamo con il dimostrare la parte sinistra della relazione.

$$\begin{aligned} \alpha_i &= \max_{x: x \neq 0} \frac{x^T A x}{\|x\|^2} \geq \frac{\mathbf{1}^T A \mathbf{1}}{\|\mathbf{1}\|^2} \\ &= \frac{\sum_i \sum_j A_{ij}}{n} \end{aligned}$$

fissiamo una riga e sommiamo su tutte le colonne, otteniamo così il grado del nodo

$$\frac{1}{n} \sum_i d(i) = D(G)$$

Dimostriamo la seconda parte della relazione. Sia u autovettore tale che $Au = \alpha_1 u$ e sia $u_i = \max_j u_j > 0$. Se u_i (componente massima di u) fosse ≤ 0 , consideriamo $-u$ (ribaltiamo il vettore), in questo modo vale la relazione e u resta autovettore.

$$\alpha_1 = \frac{(Au)_i}{u_i} = \frac{1}{u_i} \sum_j A_{i,j} u_j$$

$$= \sum_j A_{ij} \frac{u_j}{u_i} \leq \sum_i A_{ij} = d(i) \leq \Delta(G)$$

dove siccome u_i è la componente più grande, allora $\frac{u_j}{u_i} \leq 1$.

16.3 Risultati utili su autovettori e autovalori

Fatto 18. Se A_n è una matrice simmetrica $n \times n$ allora

$$\sum_{i=1}^n A_{ii} = \sum_{i=1}^n \lambda_i$$

ovvero la somma degli elementi sulla diagonale è uguale alla somma degli autovalori.

Se A è la matrice di adiacenza di G (connesso) allora $A_i i = 0 \forall i = 1, \dots, n$, dato che non ci sono cappi. Allora per il fatto di prima

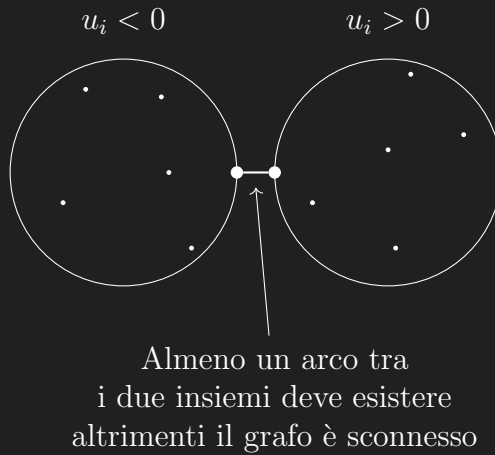
$$0 = \sum_{i=1}^n A_{i,i} = \sum_{i=1}^n \alpha_i$$

Dato che $\alpha_i \geq d(G) > 0$ allora $\alpha_n < 0$.

Lemma 5. Sia $G = (V, E)$ connesso. Sia M simmetrica e non negativa tale che $m_{ij} > 0$ se e solo se $(i, j) \in E$. Se un vettore u non negativo è un autovettore per M , allora u è strettamente positivo.

Dimostrazione 28. Se u non è strettamente positivo allora c'è almeno

una componente pari a zero, $u_r = 0$. Dato che G è un grafo connesso esiste $(r, s) \in E$ tale che $u_s > 0$. Questo perché un autovettore non può essere nullo quindi c'è una componente, almeno, strettamente positiva. Inoltre, la componente nulla (u_r) deve essere connessa a una non nulla (u_s) perché il grafo è connesso.



Definiamo μ tale che $Mu = \mu u$, ovvero μ è autovalore di u . Deve valere

$$0 = (Mu)_r = \sum_{i=1}^n M_{ri} u_i \geq M_{rs} u_s > 0$$

ottenendo quindi una contraddizione.

Enunciamo ora un teorema utile nell'analisi di camminate casuali su grafi.

Teorema 11. Perron-Frobenius per matrici simmetriche

Sia $G = (V, E)$ connesso e sia M una matrice simmetrica non negativa tale che

$$M_{ij} > 0 \Leftrightarrow (i, j) \in E$$

allora gli autovalori $\mu_1 \geq \dots \geq \mu_n$ di M soddisfano

1. μ_1 ha un autovettore strettamente positivo
2. $\mu_1 \geq -\mu_n$
3. $\mu_1 > \mu_2$ (unico autovalore distinto per u_1 , μ_1 ha molteplicità 1)

Dimostrazione 29. Definiamo $Mu_1 = \mu_1 u_1$ e $x_i = |(u_1)_i| \forall i = 1, \dots, n$. Notiamo che

$$\|x\| = \|u_1\| = 1$$

Inoltre

$$\mu_1 = u_1^T M u_1$$

questo perché

$$Mu_1 = \mu_1 u_1 \Leftrightarrow u^T M u_1 = u^T \mu u_1$$

dove abbiamo moltiplicato per u_1 ambo i membri. Siccome μ è uno scalare possiamo raccoglierlo a fattore comune

$$\Leftrightarrow u_1^T M u_1 = \mu (u_1^T u_1)$$

sapendo che $u_1^T u_1 = 1$ troviamo che

$$\begin{aligned}\mu_1 &= u_1^T M u_1 \\ &= \sum_i \sum_j M_{ij} (u_1)_i (u_1)_j\end{aligned}$$

sappiamo che M_{ij} è sempre ≥ 0 quindi sostituendo u con x

$$\leq \sum_i \sum_j M_{ij} x_i x_j = x^T M x$$

Dato che μ_1 è autovalore massimo, x deve essere autovettore di μ_1 . Siccome x è non negativo allora ha componenti strettamente positive per il lemma precedente. Abbiamo dimostrato la prima asserzione, dimostriamo ora la seconda.

Sia u_n autovalore minimo tale che $M u_n = \mu_n u_n$. Definiamo $x_i = |(u_n)_i|$ $\forall i = 1, \dots, n$. Allora

$$\begin{aligned}|\mu_n| &= |u_n^T M u_n| \\ &= \sum_i \sum_j M_{ij} |(u_n)_i| |(u_n)_j| = x^T M x\end{aligned}$$

per definizione di autovalore massimo

$$x^T M x \leq \mu_1$$

Dimostriamo ora la parte terza parte. Sia μ_2 tale che $M u_2 = \mu_2 u_2$. Notiamo che $u_1 > 0$ e che $u_2^T u_1 = 0$ per definizione. Allora u_2 deve contenere sia componenti positiva che negative. Infatti, se contenesse solo componenti positive il prodotto

$u_2^T u_1$ sarebbe positivo, se contenesse solo componenti negative sarebbe negativo.

$$\min_i (u_2)_i < 0 < \max_i (u_2)_i$$

Definiamo $x_i = |(u_2)_i|$ per $i = 1, \dots, n$.

Ancora una volta notiamo che

$$\mu_2 = u_2^T M u_2 \leq x^T M x \leq \mu_1$$

Dato che G è connesso (la spiegazione è stata fatta con disegno sopra)

$$\exists (r, s) \in E : (u_2)_r < 0 < (u_2)_s$$

Tenendo presente che

□ la sommatoria

$$u_2^T M u_2 = \sum_{i,j} M_{ij} (u_2)_i (u_2)_j$$

contiene il termine $M_{rs} (u_2)_s (u_2)_r$ che è minore di 0

□ la sommatoria

$$x_2^T M x_2 = \sum_{i,j} M_{ij} x_i x_j$$

contiene il termine $M_{rs} x_s x_r$ che è maggiore di 0

possiamo dire che

$$M_{ij} (u_2)_i (u_2)_j \leq M_{ij} x_i x_j$$

per ogni i, j e che

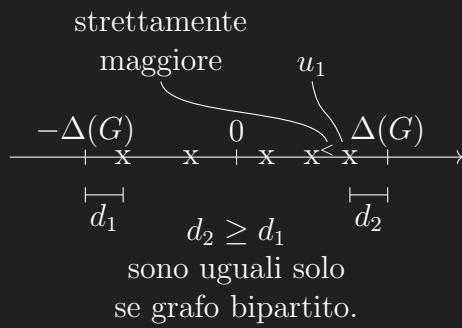
$$M_{rs} (u_2)_r (u_2)_s \leq M_{rs} x_r x_s$$

e quindi

$$u_2^T M u_2 < x^T M x$$

possiamo concludere che $\mu_2 < \mu_1$.

Fatto 19. $\mu_n = -\mu_1$ se e solo se il grafo G è bipartito.



Capitolo 17

Diciassettesima lezione

17.1 Definizione camminata casuale

Dato $G = (V, E)$ connesso con $V = \{1, \dots, n\}$, una **camminata casuale** è definita come

$$V_0 \rightarrow V_1 \rightarrow V_2 \rightarrow V_3 \rightarrow \dots$$

dove V_0 è il vertice di partenza e a ogni step la camminata passa da v_t a v_{t+1} . Noi consideriamo solo camminature uniformi. I termini V_i sono delle variabili casuali su \mathbb{V} con distribuzione p_i . Possiamo assumere che l'inizio (ovvero V_0) sia deterministico, ovvero lo scegliamo. Definiamo la distribuzione p_t come

$$\begin{aligned} p_t(i) &= \mathbb{P}(V_t = i) \\ &= \sum_{\substack{j: \\ (i,j) \in E}} \mathbb{P}(V_t = i | V_{t-1} = j) \mathbb{P}(V_{t-1} = j) \end{aligned}$$

dove

$$\mathbb{P}(V_t = i | V_{t-1} = j) = \frac{A_{ij}}{d(j)}$$

ovvero V_t è scelto uniformemente a caso in $N_{V_{t-1}}$. Possiamo dedurre che ci occupiamo di camminature stazionarie poiché la probabilità di transizione è indipendente dal tempo. Introduciamo i vettori e_i per ogni vertice. Contengono 1 solo nella posizione i – *esima*

$$e_i = (0, \dots, 0, 1, 0, \dots, 0)$$

Possiamo esprimere la distribuzione di probabilità del primo vertice come $p_0 = e_0$. Cerchiamo di esprimere $p_t(i)$ in modo più compatto. Introduciamo la matrice diagonale D tale che

$$D = \begin{bmatrix} d(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d(n) \end{bmatrix}$$

Siccome è una matrice diagonale ha inverso

$$D_{ij}^{-1} = \frac{\mathbb{I}\{i = j\}}{d(j)}$$

Notiamo che

$$\begin{aligned}(AD^{-1})_{ij} &= \sum_{k=1}^n A_{ik} D_{kj}^{-1} \\ &= \sum_{k=1}^n A_{ik} \frac{\mathbb{I}\{k=j\}}{d(j)} = \frac{A_{ij}}{d(j)}\end{aligned}$$

dove l'ultimo passaggio vale perché tutti i termini, tranne il j -esimo, della sommatoria sono pari a 0, per definizione stessa della funzione indicatrice. A questo punto possiamo riscrivere p_t come

$$p_t = AD^{-1}p_{t-1}$$

Poniamo $W = AD^{-1}$ e quindi $\forall t \geq 1$

$$p_t = Wp_{t-1} = W^t p_0$$

dove abbiamo srotolato p_{t-1} . Siccome $W_{ij} = \frac{A_{ij}}{d(j)}$, la matrice W non è simmetrica ed è non negativa. Notiamo che

$$(\mathbb{1}^T W)_j = \sum_{i=1}^n \frac{A_{ij}}{d(j)}$$

sommiamo le colonne di A e otteniamo il grado del nodo

$$= \frac{d(j)}{d(j)} = 1$$

Questo vuol dire che la matrice $\mathbb{1}^T W$ è una matrice stocastica sulle colonne, ovvero la somma delle colonne (che sono non negative) è 1.

17.2 Distribuzione stazionaria

Diciamo che una distribuzione π su \mathbb{V} è una distribuzione stazionaria per W se $W\pi = \pi$, dove π è un autovettore non normalizzato di W con autovalore 1. Sia $d = (d_1, \dots, d_n)$, definiamo $\pi(i)$ come

$$\pi(i) = \frac{d(i)}{\mathbb{1}^T d}$$

Intuitivamente il tempo trascorso su un nodo è proporzionale al suo grado. Siccome vale

$$(W\pi)_i = \sum_{j=1}^n W_{ij} \pi(j)$$

$$\begin{aligned}&= \sum_{j=1}^n \frac{A_{ij}}{d(j)} \frac{d(j)}{\mathbb{1}^T d} = \frac{1}{\mathbb{1}^T d} \sum_{j=1}^n A_{ij} \\ &= \frac{d(i)}{\mathbb{1}^T d} = \pi(i)\end{aligned}$$

quindi $W\pi = \pi$ e allora π è stazionaria per W .

17.3 Convergenza della camminata casuale alla distribuzione stazionaria

Definiamo

$$A_{norm} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

che è il modo matriciale per dire che dividiamo A per D . Siccome $W = AD^{-1}$

$$A_{norm} = D^{-\frac{1}{2}} W D^{\frac{1}{2}}$$

Definiamo anche

$$L_{norm} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = I - A_{norm}$$

con $\lambda_1 \leq \dots \leq \lambda_N \in [0, 1]$ gli autovalori di L_{norm} e $\omega_1 \geq \dots \geq \omega_n \in [-1, 1]$ gli autovalori di A_{norm} . Notiamo che $\lambda_i = 1 - \omega_i$.

Fatto 20. ψ è un autovettore di A_{norm} con autovalore ω se e solo se $D^{\frac{1}{2}}\psi$ è un autovettore di W con autovalore ω .

Dimostrazione 30. Siccome

$$A_{norm} = D^{-\frac{1}{2}} W D^{\frac{1}{2}}$$

allora moltiplicando per $D^{\frac{1}{2}}$ ambo i membri

$$D^{\frac{1}{2}} A_{norm} = W D^{\frac{1}{2}}$$

Se $A_{norm}\psi = \omega\psi$ allora

$$\begin{aligned} W(D^{\frac{1}{2}}\psi) &= D^{\frac{1}{2}} A_{norm}\psi = D^{\frac{1}{2}} \omega\psi \\ &= (D^{\frac{1}{2}}\psi)\omega \end{aligned}$$

quindi $(D^{\frac{1}{2}}\psi)$ è un autovettore di W con autovalore ω . Analogamente $Wu = \omega u$ implica che (non lo dimostriamo ma è facile da fare)

$$A_{norm}(D^{-\frac{1}{2}}u) = \omega(D^{-\frac{1}{2}}u)$$

Possiamo applicare Perron-Frobenius ad A_{norm} . L'autovalore massimo di W (ω_1) ha un unico autovettore le cui componenti sono strettamente positive e $\omega_n = \omega_1 \Leftrightarrow G$ è bipartito. Inoltre, se ψ_1 è l'autovettore di A_{norm} con autovalore 1 allora $\pi = D^{\frac{1}{2}}\psi_1$. Definiamo delle proprietà spettrali che ci torneranno utili più avanti.

Fatto 21. Sia A una matrice simmetrica $n \times n$ con spettro $\lambda_1, \dots, \lambda_n$ e u_1, \dots, u_n . Allora $\forall t \in \mathbb{N}$

$$A^t = \sum_{i=1}^n \lambda_i^t u_i u_i^T = U \Lambda U^T$$

Dimostrazione 31. Dimostrazione per induzione.

□ Caso base, $t = 1$, vero per il teorema spettrale.

□ Passo induttivo. Scriviamo A^t come

$$A^t = A A^{t-1}$$

e applichiamo l'ipotesi induttiva

$$\begin{aligned} &= \left(\sum_i \lambda_i u_i u_i^T \right) \left(\sum_j \lambda_j^{t-1} u_j u_j^T \right) \\ &= \sum_i \sum_j \lambda_i \lambda_j^{t-1} u_i u_i^T u_j u_j^T \end{aligned}$$

siccome u_i e u_j sono ortogonali vale che $u_i^T u_j = \mathbb{I}\{i = j\}$. Allora

$$= \sum_i \lambda_i \lambda_i^{t-1} u_i u_i^T = \sum_i \lambda_i^t u_i u_i^T$$

Enunciamo un teorema, che dimostreremo nella prossima lezione, che prova la convergenza della camminata casuale alla distribuzione stazionaria.

Teorema 12. $\forall G$ connesso, non bipartito, vale che

$$\lim_{t \rightarrow \infty} W^t p_0 = \pi = \frac{d}{1^T d}$$

$\forall p_0$.

Cerchiamo di capire intuitivamente perché il grafo non possa essere bipartito. Consideriamo un caso limite di grafo bipartito



Se $p_0 = e_1$ allora

$$V_t = \begin{cases} 1 & \text{se } t \text{ è pari} \\ 2 & \text{altrimenti} \end{cases}$$

se invece $p_0 = e_2$

$$V_t = \begin{cases} 1 & \text{se } t \text{ è dispari} \\ 2 & \text{altrimenti} \end{cases}$$

ovvero le distribuzioni non convergono! E' proprio il fatto che $\omega_n =$

$-\omega_1$ che ci blocca. Ovviamente esistono distribuzioni anche sui grafi bipartiti che convergono, per esempio

$$p_0 = \left(\frac{1}{2}, \frac{1}{2}\right)$$

in questo caso $W p_0 = p_0$.

Capitolo 18

Diciottesima lezione

18.1 Dimostrazione

Andiamo a dimostrare il teorema della lezione precedente.

Dimostrazione 32. Cominciamo con il considerare ψ_1, \dots, ψ_n autovettori di A_{norm} . Questi formano una base ortonormale di \mathbb{R}^n . Andiamo a a proiettare $D^{-\frac{1}{2}}p_0$ sulla base ortonormale

$$D^{-\frac{1}{2}}p_0 = \sum_{i=1}^n (\psi_i^T D^{-\frac{1}{2}}p_0) \psi_i =$$

poniamo $\psi_i^T D^{-\frac{1}{2}}p_0 = c_i$

$$= \sum_{i=1}^n c_i \psi_i$$

Andiamo a concentrarci su p_t

$$p_t = W^t p_0 = (D^{\frac{1}{2}} A_{norm} D^{-\frac{1}{2}})^t p_0$$

che abbiamo già dimostrato in precedenza. Espandiamo il prodotto

$$\overbrace{D^{\frac{1}{2}} A_{norm} D^{-\frac{1}{2}} \cdot D^{\frac{1}{2}} A_{norm} D^{-\frac{1}{2}} \dots}^I \dots D^{\frac{1}{2}} A_{norm} D^{-\frac{1}{2}} p_0$$

siccome $D^{-\frac{1}{2}} \cdot D^{\frac{1}{2}} = I$ e A_{norm} si ripete t volte

$$= D^{\frac{1}{2}} A_{norm}^t D^{-\frac{1}{2}} p_0$$

$$= D^{\frac{1}{2}} A_{norm}^t \sum_{i=1}^n c_i \psi_i$$

che vale per quello che abbiamo scritto prima. Dal teorema spettrale sappiamo che $A_{norm} = \sum_{i=1}^n \omega_i \psi_i \psi_i^T$. Inoltre, la scorsa lezione abbiamo dimostrato che è robusto rispetto alle potenze, quindi

$$= D^{\frac{1}{2}} \left(\sum_{i=1}^n \omega_i^t \psi_i \psi_i^T \right) \sum_{i=1}^n c_i \psi_i$$

uniamo le sommatorie

$$= D^{\frac{1}{2}} \sum_{i=1}^n \sum_{j=1}^n \omega_i^t c_j \psi_i \psi_i^T \psi_j$$

Siccome sono ortogonali tra loro, $\psi_i^T \psi_i$ è uguale a 1 solo se $i = j$, altrimenti è 0. Possiamo sostituire $\psi_i^T \psi_i$ con $\mathbb{I}\{i = j\}$ e quindi eliminare tutti i casi in cui $i \neq j$:

$$= D^{\frac{1}{2}} \sum_{i=1}^n \omega_i^t c_i \psi_i$$

Ricordiamo che l'autovalore massimo è tale che $\omega_i = 1$. Se G fosse bipartito (ed è importante che non lo sia per questa dimostrazione) $\omega_n = -\omega$. Il teorema però specifica che il grafo non può essere bipartito, quindi $|\omega_n| < \omega_1$. Inoltre, siccome $\omega_2 < \omega_1$ possiamo dire che $\omega_2, \dots, \omega_n \in (-1, 1)$. Questo è interessante dato che $\lim_{t \rightarrow \infty} \omega^t = 0$ se $\omega \in (-1, 1)$. Stacciamo $i = 1$ dalla sommatoria e esprimiamo $\omega_1 = 1$

$$= D^{\frac{1}{2}} c_1 \psi_1 + D^{\frac{1}{2}} \sum_{i=2}^n \omega_i^t c_i \psi_i$$

Consideriamo

$$\lim_{t \rightarrow \infty} p_t$$

per le considerazioni fatte prima sappiamo che

$$= D^{\frac{1}{2}} c_1 \psi_1$$

Consideriamo π autovettore di W associato a $\omega_1 = 1$. W e A_{norm} hanno gli stessi autovettori ma autovalori diversi. Possiamo scrivere

$$\frac{d}{\mathbf{1}^T d} = \pi = D^{\frac{1}{2}} \psi_1 c$$

dove c è una costante di scalatura. Questo perché π è normalizzato con norma 1, mentre ψ con norma 2. Ribaltiamo l'equazione ed esprimiamo c come la norma due

$$\psi_1 = \frac{D^{-\frac{1}{2}} \pi}{\|D^{-\frac{1}{2}} \pi\|} = \frac{D^{-\frac{1}{2}} d}{\|D^{-\frac{1}{2}} d\|}$$

Allora riprendiamo $D^{\frac{1}{2}} c_1 \psi_1$ e lavoriamo su c_1 . Utilizziamo la definizione di c_1 e scriviamo

$$c_1 = \psi_1^T D^{-\frac{1}{2}} p_0$$

usiamo la definizione appena trovata di ψ_1 (ricordandoci che è trasposto)

$$= \frac{d^T D^{-\frac{1}{2}}}{\|D^{-\frac{1}{2}} d\|} D^{-\frac{1}{2}} p_0$$

dato che $D^{-\frac{1}{2}} D^{\frac{1}{2}} = D^{-1}$

$$= \frac{d^T D^{-1} p_0}{\|D^{-\frac{1}{2}} d\|}$$

d è il vettore dei gradi mentre D^{-1} è la matrice con i reciproci dei gradi. Quindi

$$= \frac{\mathbf{1}^T p_0}{\|D^{-\frac{1}{2}} d\|}$$

inoltre $\mathbf{1}^T p_0 = 1$ dato che p_0 è una distribuzione di probabilità, e la somma delle probabilità è uguale a 1

$$= \frac{1}{\|D^{-\frac{1}{2}} d\|}$$

Allora

$$D^{\frac{1}{2}} c_1 \psi_1 = \frac{D^{\frac{1}{2}} \psi_1}{\|D^{-\frac{1}{2}} d\|}$$

esprimiamo ψ_1

$$\begin{aligned} &= \frac{D^{\frac{1}{2}} D^{-\frac{1}{2}} d}{\|D^{-\frac{1}{2}} d\|^2} = \frac{d}{\|D^{-\frac{1}{2}} d\|^2} \\ &= \frac{d}{d^T D^{-1} d} = \frac{d}{\mathbf{1}^T d} = \pi \end{aligned}$$

dove il penultimo passaggio vale perché

$$\begin{aligned} d^T D^{-1} d &= \sum_i \sum_j d(i) \frac{\mathbb{I}\{i=j\}}{d(j)} d(j) \\ &= \sum_i \frac{d(i)^2}{d(i)} = \mathbf{1}^T d \end{aligned}$$

Abbiamo quindi dimostrato che

$$\lim_{t \rightarrow \infty} W^T p_0 = \lim_{t \rightarrow \infty} p_t = \pi$$

18.2 Velocità di convergenza

Sappiamo che $p_t = \pi$ se t tende a infinito. Cosa possiamo dire per t finito? Ovvero, $\forall v \in V$

$$|p_t(v) - \pi(v)| \leq ?$$

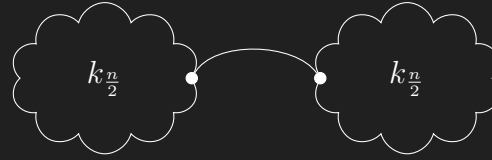
Dopo un t quanto sono lontano da π ? Dopo quanto tempo posso fermarmi essendo sicuro di avere una buona approssimazione?

Teorema 13. Per $u, v \in V$ e $t \in \mathbb{N}$, se $p_0 = e_u$ allora

$$|p_t(v) - \pi(v)| \leq \left(\sqrt{\frac{d(v)}{d(u)}} \right) k^t$$

con $k = \max\{|\omega_n|, |\omega_2|\}$.

Notiamo che k è minore di 1 (almeno che il grafo non sia bipartito). Quindi per $t \rightarrow \infty$ k tende a 0. Questo teorema è interessante però il problema è che dobbiamo curarci sia di ω_n che di ω_2 . Ci piacerebbe se dipendesse solo da ω_2 . Infatti, se avessimo un grafo del tipo



allora

$$\lambda_2 \ll 1$$

e potremmo affermare che la convergenza è lenta (dato che $\omega_2 = 1 - \lambda_2$). Però il problema è che non stiamo tenendo in conto ω_n e quindi non lo sappiamo. Cominciamo con il dimostrare il teorema sopra citato, poi ci occuperemo di questo problema.

Dimostrazione 33. Sappiamo che

$$p_t(v) = e_v^T p_t = e_v^T \pi + e_v^T D^{\frac{1}{2}} \sum_{i=2}^n c_i \omega_i^t \psi_i$$

dove abbiamo utilizzato la definizione di p_t che abbiamo trovato nella dimostrazione del teorema precedente. Ricordiamoci che

□ $c_i = \psi_i^T D^{-\frac{1}{2}} e_u = \psi_i^T \frac{e_u}{\sqrt{d(u)}}$ che abbiamo ottenuto semplicemente moltiplicando $D^{-\frac{1}{2}}$ con e_u .

□ $e_v^T D^{\frac{1}{2}} = \sqrt{d(v)} e_v^T$ per la stessa ragione di sopra.

$$\begin{aligned} &= \pi(v) + e_v^T D^{\frac{1}{2}} \sum_{i=2}^n \omega_i^t \psi_i \psi_i^T D^{-\frac{1}{2}} e_u \\ &= \pi(v) + \sqrt{\frac{d(v)}{d(u)}} e_v^T \sum_{i=2}^n \omega_i^t \psi_i e_u \psi_i^T \end{aligned}$$

dove abbiamo utilizzato i due fatti sopra citati. Portiamo dentro e_v^T

$$= \pi(v) + \sqrt{\frac{d(v)}{d(u)}} \sum_{i=2}^n \omega_i^t (e_v^T \psi_i) (e_u \psi_i^T)$$

consideriamo la sommatoria

$$\sum_{i=2}^n \omega_i^t (e_v^T \psi_i) (e_u \psi_i^T)$$

prendiamo i valori assoluti

$$\leq \sum_{i=2}^n |\omega_i^t| |e_v^T \psi_i| |e_u \psi_i^T|$$

Sappiamo che k è il massimo tra il valore assoluto di tutti gli autovalori tali che $i = 2, \dots, n$, quindi

$$\leq k^t \sum_{i=1}^n |e_v^T \psi_i| |e_u \psi_i^T|$$

per la disuguaglianza di Cauchy-Schwarz scriviamo

$$\leq k^t \sqrt{\sum_{i=1}^n (e_v^T \psi_i)^2} \sqrt{\sum_{i=1}^n (e_u^T \psi_i)^2}$$

Siccome ψ_1, \dots, ψ_n è una base ortonormale vale che $\|v\| = \sqrt{\sum_{i=1}^n (v_i^T \psi_i)^2}$

$$= k^t \|e_v\| \cdot \|e_u\| = k^t$$

dato che la norma di $e_i = 1$. Concludiamo la dimostrazione mettendo tutto insieme

$$p_t(v) \leq \pi(v) + \sqrt{\frac{d(v)}{d(u)}} k^t$$

18.3 Camminata pigra

Prima abbiamo accennato al fatto che ci piacerebbe se

$$|\pi(v) - p_t(v)| \leq (\dots)(1 - \lambda_2)^t$$

dove con ... indichiamo un qualche coefficiente. Definiamo

$$W' = \frac{1}{2}I + \frac{1}{2}W = \frac{1}{2}(I + W)$$

W' è la camminata casuale tale che

$$\begin{aligned} \mathbb{P}(v_{t+1} = i | v_t = j) &= \frac{1}{2} \mathbb{I}\{i = j\} + \\ &+ \frac{A_{ij}}{2d(j)} \mathbb{I}\{i \neq j\} \end{aligned}$$

ovvero abbiamo probabilità $\frac{1}{2}$ di restare sul nodo in cui siamo. In pratica usiamo il seguente "algoritmo"

1. Tiro una moneta
2. Se testa resto altrimenti mi sposto sul vicino scelto a caso

$\omega_1 \geq \dots \geq \omega_n$ sono gli autovalori di W mentre $\omega'_1 \geq \dots \geq \omega'_n$ sono gli autovalori di W' tali che

$$\omega'_i = \frac{1}{2}(1 + \omega_i)$$

quindi $\omega'_i \in [0, 1]$, abbiamo scalato ω_i . La distribuzione stazionaria π per W lo è anche per W' . Applicando la definizione di W possiamo scrivere

$$W' = I - \frac{1}{2}D^{\frac{1}{2}}L_{norm}D^{-\frac{1}{2}}$$

e $\omega'_i = \frac{1}{2}(1 + 1 - \lambda_i) = 1 - \frac{\lambda_i}{2}$, dove abbiamo usato $\omega_i = 1 - \lambda_i$. Andiamo a considerare

$$p_t = (W')^t p_0$$

possiamo scrivere

$$= \pi + D^{\frac{1}{2}} \sum_{i=2}^n c_i \left(1 - \frac{\lambda_i}{2}\right)^t \psi_i$$

che abbiamo già dimostrato, l'unica differenza è che stiamo usando $\omega'_i = (1 - \frac{\lambda_i}{2})^t$ invece di ω_i . Per la dimostrazione che abbiamo fatto nella sezione precedente, possiamo scrivere

$$p_t(v) \leq \pi(v) + \sqrt{\frac{d(v)}{d(u)}} \left(1 - \frac{\lambda_2}{2}\right)^t$$

perché il massimo tra $|\omega'_n|$ e $|\omega'_2|$ è $|\omega'_2|$, in virtù di quella riscrittura che abbiamo fatto. Infatti ora $\omega' \geq 0 \ \forall i$ e $\omega'_2 \geq \omega'_3 \geq \dots \geq \omega'_n > 0$.

Usiamo la nota disuguaglianza $e^{-x} \geq 1 - x$ per scrivere

$$|p_t(v) - \pi(v)| \leq \sqrt{\frac{d(v)}{d(u)}} e^{-\frac{\lambda_2}{2}}$$

Possiamo concludere che la velocità di convergenza è esponenziale.

Capitolo 19

Diciannovesima lezione

19.1 Tempo di Mixing

Il **tempo di mixing** è definito come il minimo tempo prima del quale non ha senso considerare la distribuzione corrente della nostra camminata casuale con un'approssimazione ragionevole della distribuzione stazionaria. In altre parole, dopo il tempo di mixing possiamo iniziare a campionare la distribuzione corrente. Formalmente definiamo il tempo di mixing su W' come

$$\min\{t : |p_t(v) - \pi(v)| \leq \frac{\pi(v)}{2} \forall v \in V\}$$

dove $p_t = (W')^t p_0$. Quindi se abbiamo una camminata

$$V_0 \rightarrow V_1 \rightarrow \dots \rightarrow V_t$$

se $t > \text{mix}$ per W' allora $V_t \sim \pi_t$, a meno di un'approssimazione (noi abbiamo usato $\frac{1}{2}$ nella definizione). Come calcoliamo il tempo di mixing? Sappiamo che

$$|p_t(v) - \pi(v)| \leq \sqrt{\frac{d(v)}{d(u)}} e^{\frac{-\lambda_2 t}{2}}$$

e vogliamo che

$$\leq \frac{\pi(v)}{2} = \frac{d(v)}{2\mathbf{1}^T D}$$

osserviamo che $\mathbf{1}^T D = \sum_i d(i) = 2|E|$, fatto che abbiamo dimostrato in una delle prime lezioni. Risolviamo la disequazione per t

$$t \geq \frac{2}{\lambda_2} \ln \frac{4|E|}{\sqrt{d(v)d(u)}}$$

siccome $d(v), d(u) \geq \delta$ riscriviamo come

$$t \geq \frac{2}{\lambda_2} \ln \frac{4|E|}{\delta}$$

Allora possiamo dire che

$$t = \Omega\left(\frac{1}{\lambda_2} \ln |E|\right)$$

19.1.1 Caratterizzazione autovalori

Prima di andare ad analizzare il tempo di mixing su alcuni grafi mettiamo in relazione lo spettro di $L = D - A$ (con autovalori $\lambda'_1, \dots, \lambda'_n$) con lo spettro di $L_{norm} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ (con autovalori $\lambda_1, \dots, \lambda_n$ e autovettori v_1, \dots, v_n). Questo perché lo spettro del laplaciano (L) è più facile da calcolare. Enunciamo due fatti.

Fatto 22. $x^T L_{norm} x = y^T L y$ dove scegliamo y tale che $y = D^{-\frac{1}{2}} x$.

Dimostrazione 34. Appliciamo la definizione di L_{norm}

$$\begin{aligned} x^T L_{norm} x &= x^T (I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) x \\ &= x^T I x - (x^T D^{-\frac{1}{2}}) A (D^{-\frac{1}{2}} x) \\ &= x^T I x - y^T A y \end{aligned}$$

Concentriamoci sul primo termine. Sappiamo che $D^{-\frac{1}{2}} D^{\frac{1}{2}} = I$ quindi

$$\begin{aligned} &= x^T I I x = (x^T D^{-\frac{1}{2}}) (D^{\frac{1}{2}} D^{\frac{1}{2}}) (D^{-\frac{1}{2}} x) \\ &= y^T D y \end{aligned}$$

mettendo insieme le due parti

$$= y^T D y - y^T A y = y^T (D - A) y = y^T L y$$

che conclude la dimostrazione.

Notiamo che la relazione appena esposta ($y = D^{-\frac{1}{2}} x$) è biettiva (ovvero va da R^n a R^n) e che la matrice D è a rango pieno.

Teorema 14. Sia L la matrice laplaciana di un grafo con autovalori $\lambda'_1, \dots, \lambda'_n$ e sia L_{norm} la sua laplaciana normalizzata con autovalori $\lambda_1, \dots, \lambda_n$. Allora, $\forall i = 1, \dots, n$ vale che

$$\frac{\lambda'_i}{\Delta} \leq \lambda_i \leq \frac{\lambda'_i}{\delta}$$

Dimostrazione 35. Usando il teorema di Courant-Fischer sulla caratterizzazione degli autovalori scriviamo

$$\begin{aligned} \lambda_i &= \min_{\substack{x \in R^n \setminus \{0\} \\ x \perp \{u_1, \dots, u_{i-1}\}}} \frac{x^T L_{norm} x}{x^T x} \\ &= \min_{S: \dim(S)=i} \max_{x \in S} \frac{x^T L_{norm} x}{x^T x} \end{aligned}$$

ruotiamo il sottospazio, traslazioni lineari biettive di sottospazi sono sempre sottospazi quindi

$$= \min_{S: \dim(S)=i} \max_{y \in S} \frac{y^T L_{norm} y}{y^T y}$$

Facciamo due osservazioni:

1.

$$\begin{aligned} y^T D y &= \sum_{i=1}^n d(i) y_i^2 \\ &\geq \delta \sum_{i=1}^n y_i^2 = \delta y^T y \end{aligned}$$

2.

$$y^T D y = \sum_{i=1}^n d(i) y_i^2$$

$$\leq \Delta \sum_{i=1}^n y_i^2 = \Delta y^T y$$

Quindi possiamo dire che λ_i è

$$\geq 1 \frac{1}{\Delta} \min \max \frac{y^T L y}{y^T y} = \frac{\lambda'_i}{\Delta}$$

e per la stessa ragione anche che è

$$\leq \frac{\lambda'_i}{\delta}$$

Concludiamo che

$$\frac{\lambda'_i}{\Delta} \leq \lambda_i \leq \frac{\lambda'_i}{\delta}$$

19.1.2 Tempo di Mixing su clique

Andiamo ad analizzare il tempo di mixing su una clique K_n . Cominciamo con il notare che

$$\begin{aligned} L &= D - A = nI - \mathbf{1}\mathbf{1}^T = \\ &= \begin{bmatrix} n-1 & \dots & -1 \\ \vdots & \ddots & \vdots \\ -1 & \dots & n-1 \end{bmatrix} \end{aligned}$$

Possiamo dire che

$$u_1 = \frac{1}{\sqrt{n}}$$

questo è vero per qualsiasi laplaciano. Inoltre, siccome la somma delle righe e delle colonne è 0, l'autovalore di u_1 è 0. L'autovettore u_2 deve essere perpendicolare a u_1

$$u_2^T u_1 = 0 \Leftrightarrow \sum_{i=1}^n u_{2,i} = 0$$

$$\Leftrightarrow \mathbf{1}^T u_2 = 0$$

Possiamo generalizzare per un vettore qualunque $x \neq 0$.

$$\mathbf{1}x = 0$$

troviamo l'autovalore

$$Lx = (n\mathbb{I} - \mathbf{1}\mathbf{1}^T)x = (nx) - \mathbf{1}(\mathbf{1}^T x)$$

siccome $\mathbf{1}^T x = 0$

$$= nx$$

quindi qualsiasi vettore ortogonale a u_1 è autovettore di L con autovalore n . Siccome $\lambda'_2, \lambda'_3, \dots, \lambda'_n = n$ e anche che $\Delta = \delta = n - 1$ allora

$$\lambda_i = \frac{\lambda'_i}{n-1} = \frac{n}{n-1}$$

Allora il mixing time per k_n è

$$t \sim \frac{1}{\lambda_2} \ln \frac{|E|}{n-1} \sim \left(1 - \frac{1}{n}\right) \ln n$$

che tende a 1 per n che tende a infinito.

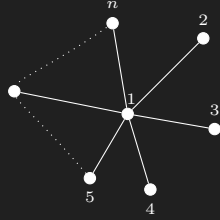
19.1.3 Tempo di Mixing su grafo a stella

Cominciamo con il calcolare la componente i -esima della matrice Lx .

$$(Lx)_i = ((D - A)x)_i$$

$$\begin{aligned}
&= (Dx)_i - \sum_{j=1}^n A_{ij}x_j = d(i)x_i - \sum_{j \in N_i} x_j \\
&= \sum_{j \in N_i} (x_i - x_j)
\end{aligned}$$

Consideriamo ora il grafo a stella S_n



Per ogni vertice che non sia quello centrale, quindi $\forall a, b > 1$ consideriamo il vettore $e_a - e_b$ e calcoliamo

$$(L(e_a - e_b))_i$$

per quello che abbiamo scritto prima vale che

$$\begin{aligned}
&= \sum_{j \in N_i} \mathbb{I}\{i = a\} - \mathbb{I}\{i = b\} - \\
&\quad - \mathbb{I}\{j = a\} + \mathbb{I}\{j = b\}
\end{aligned}$$

siccome stiamo considerando nodi diversi da quello centrale, tutti i nodi hanno un solo vicino, e quindi possiamo scrivere

$$= \mathbb{I}\{i = a\} - \mathbb{I}\{i = b\} = e_a - e_b$$

In generale, $\forall a, b > 1$ vale che

$$L(e_a - e_b) = e_a - e_b$$

Andiamo ora a considerare $n - 2$ autovettori linearmente indipendenti e

non ortogonali. Li costruiamo in questo modo

$$e_i - e_{i+1} = v_i$$

$\forall i = 2, \dots, n-1$. Prendiamo il caso $n = 5$ e consideriamo la matrice 5×5 formata da questi vettori

$$\begin{bmatrix}
0 & 1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Vediamo che i primi tre vettori sono indipendenti, infatti non è possibile ottenere uno applicando trasformazioni lineari agli altri. Consideriamo i vettori U_1, \dots, U_{n-2} di $\text{span}\{e_2 - e_3, \dots, e_{n-1} - e_n\}$ tali che

$$U_i = \sum_{j=2}^{n-1} c_j v_j$$

allora

$$\begin{aligned}
LU_i &= L \sum_{j=2}^{n-1} c_j v_j = \sum_j c_j L v_j \\
&= \sum_j c_j v_j = U_i
\end{aligned}$$

Siamo partiti da dei vettori non ortogonali e abbiamo ottenuto dei vettori ortogonali. Questi sono autovettori di L con autovalori tali che $\lambda'_1 = 0$, come al solito perché autovalore del autovettore più grande del

laplaciano, e $\lambda'_2 = \dots = \lambda'_{n-1} = 1$ perché $e_a - e_b$ ha autovalore 1 (infatti $L(e_a - e_b) = 1(e_a - e_b)$). Ci resta da capire l'autovalore di λ'_n . Consideriamo la sommatoria della diagonale

$$\sum_i L_{ii} = n - 1 + n - 1 = 2n - 2$$

perché è la somma del vertice centrale $(n-1)$ con tutti gli altri vertici $(1(n-1))$. Per definizione è uguale a

$$\sum_i \lambda'_i$$

quindi

$$\sum_i \lambda'_i = 2n - 2$$

possiamo riscrivere $\sum_i \lambda'_i$ come

$$\sum_{i=2}^{n-1} (\lambda'_i) + \lambda'_n = n + \lambda'_n$$

allora

$$2n - 2 = n + \lambda'_n \Leftrightarrow \lambda'_n = n - 2$$

Troviamo λ_2 per capire come si comporta il mixing time.

$$\frac{\lambda'_2}{n-1} \leq \lambda_2 \leq \frac{\lambda'_2}{1}$$

dove $n-1$ è il grado massimo e 1 il grado minimo. Siccome $\lambda'_2 = 1$

$$\frac{1}{n-1} \leq \lambda_2 \leq 1$$

Quindi se λ_2 fosse vicino a $\frac{1}{n-1}$ avremmo

$$t \sim n - 1 \ln n$$

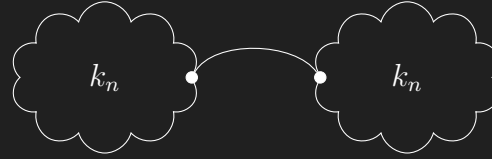
che è peggio della clique. Mentre se fosse vicino a 1

$$t \sim \ln n$$

Non lo dimostriamo ma è noto che λ_2 è molto vicino a 1.

19.1.4 Tempo di mixing grafo a manubrio

Consideriamo il grafo a manubrio (dumbbell in inglese)



Siccome

$$\sum_i \lambda'_i = |E| \sim n^2$$

nel caso migliore

$$\lambda'_2 = O(n)$$

e dato che $\Delta = n$ allora

$$\lambda_2 \geq 1$$

In generale però

$$\lambda'_2 = \min_{\substack{x \neq 0 \\ 1^T x = 0}} \frac{x^T L x}{x^T x} \leq \frac{v^T L v}{v^T v}$$

dove v è un vettore tale che

$$v = (\overbrace{-1, \dots, -1}^n, \underbrace{1, \dots, 1}_n)$$

dove -1 in posizione i indica che il vertice è nella parte sinistra del grafo, mentre 1 che è nella parte destra. Il numeratore è pari a

$$v^T L v = \sum_{(i,j) \in E} (v_i - v_j)^2 = 4$$

perché tutti i vertici tranne 2 sono connessi solo ad altri vertici nella stessa parte di grafo, e quindi $v_i - v_j = 0$. L'unico caso che non è 0 è per i vertici di "frontiera", infatti che vale $v_i - v_j = 1 - (-1) = 2$. Consideriamo il denominatore

$$v^T v = 2n$$

perché è la somma dei quadrati dei coefficienti. Quindi

$$\lambda'_2 \leq \frac{4}{2n} = \frac{2}{n}$$

A questo punto sappiamo che

$$\lambda_2 \leq \frac{\lambda'_2}{\delta} = \frac{2}{n(n-1)}$$

e allora

$$t \sim \frac{1}{\lambda_2} \ln \frac{|E|}{\delta} \sim n^2 \ln n$$

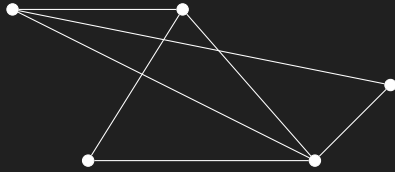
vuol dire che dobbiamo aspettare un tempo n^2 . Questo ha senso se pensiamo che la probabilità di cambiare clique è $\frac{1}{n}$ e quindi per esplorare il grafo abbiamo bisogno di ordine n^2 step.

Capitolo 20

Ventesima lezione

20.1 Algoritmi distribuiti

Vediamo un'applicazione della teoria spettrale dei grafi. Parliamo di algoritmi distribuiti e nello specifico di unanimità distribuita. Abbiamo una rete di comunicazione $G = (V, E)$



Ogni nodo ha un processore. Questi nodi vogliono calcolare qualcosa in modo distribuito. Ogni nodo non conosce G e può solo scambiare messaggi con i propri vicini. A ciascun nodo assumiamo un numero

$$x_0 = (\underbrace{x_0(1), \dots, x_0(n)}_{v_1})$$

Per esempio vogliamo calcolare la media di questi numeri

$$\mu = \frac{1}{n} \sum_{i=1}^n x_0(n)$$

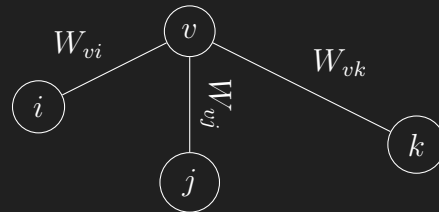
definiamo $\mu_t(i)$ come lo stato del nodo i dopo t passi e $x_0(i)$ è lo stato iniziale. Creiamo un processo che converge per ciascun nodo, quindi tale che

$$\lim_{t \rightarrow \infty} x_t(i) = x$$

$\forall i = 1, \dots, n$. Costruiamo $x_{t+1}(i)$ a partire da $x_t(i)$ in questo modo (cambiamo notazione, indichiamo nodo con v)

$$x_{t+1}(v) = \sum_{u:(v,u) \in E} W_{v,u} x_t(u)$$

Per capire cos'è W consideriamo il grafo seguente



Ogni arco (e quindi ogni vicino) ha un peso. Il nodo v chiede lo stato

dei nodi vicini e li somma in modo pesato. La richiesta degli stati viene fatta da tutti i nodi simultaneamente (non ci sono problemi di sincronizzazione perché viene chiesto il valore al tempo t). Possiamo scrivere il vettore x_{t+1} come

$$x_{t+1} = Wx_t$$

la matrice W è detta **matrice di gossip** ed è

1. Non negativa e simmetrica
2. Doppia stocastica, quindi $W\mathbf{1} = \mathbf{1}$ e $\mathbf{1}^T W = \mathbf{1}^T$
3. Le sue componenti sono $W_{ij} > 0 \Leftrightarrow (i, j) \in E$ altrimenti sono nulle.

Fatto 23. Il più grande autovalore di una matrice stocastica sulle righe ($W\mathbf{1} = \mathbf{1}$) è 1

Dimostrazione 36. Chiaramente, 1 è autovalore di W per definizione stessa di matrice stocastica. Per assurdo esista $\mu > 1$ e $x \neq 0$ tale che $Wx = \mu x$. Sia quindi $x_k = \max_i(x_i)$. Senza perdita di generalità assumiamo $x_k > 0$ (infatti se x è autovettore anche $-x$ lo è). Gli elementi di W sono non negativi

$$\forall i \ W_{ij} \geq 0 \ j = 1, \dots, n$$

e ogni riga somma a 1

$$\sum_{j=1}^n W_{ij} = 1$$

Vale che

$$(Wx)_i = \sum_{j=1}^n W_{ij}x_j \leq \max_j x_j = x_k$$

Possiamo anche dire che $\forall i$

$$(\mu x)_i = (Wx)_i \leq x_k$$

Siccome μ è maggiore di 1

$$x_k < \mu x_k \leq x_k$$

ed è una contraddizione.

Assumiamo $\omega_n \leq \dots \leq \omega_1 = 1$ autovalori di W . Ricordiamoci che vogliamo calcolare

$$\frac{1}{n} \sum_{i=1}^n x_0(i) = \frac{1}{n} \mathbf{1}^T x_0 = \mu$$

Definiamo il vettore (μ_1, \dots, μ_n) , che è il vettore a cui vorremmo convergesse il vettore degli stati dei nodi. Chiamiamo questo vettore x . Vale che

$$x = \mu \mathbf{1} = \frac{1}{n} \mathbf{1} \mathbf{1}^T x_0$$

Verifichiamo che x sia una distribuzione stazionaria per W , vogliamo quindi che $Wx = x$.

$$Wx = \frac{1}{n} W \mathbf{1} \mathbf{1}^T x_0$$

siccome W è stocastica allora $W\mathbf{1} = \mathbf{1}$, dunque

$$= x$$

Fatto 24. Vale che

$$\frac{1}{n} \mathbf{1} \mathbf{1}^T x_t = \frac{1}{n} \mathbf{1} \mathbf{1}^T W x_{t-1} = \frac{1}{n} \mathbf{1} \mathbf{1}^T x_{t-1}$$

iterando il processo otteniamo

$$\frac{1}{n} \mathbf{1} \mathbf{1}^T x_t = \frac{1}{n} \mathbf{1} \mathbf{1}^T x_0 = x \quad \forall t$$

Ci interessa mostrare che

$$\|x_t - x\| \rightarrow 0$$

per $t \rightarrow \infty$. Questo implica che $x_t \rightarrow x$, cioè che con tempo abbastanza grande ci avviciniamo al risultato giusto. Sappiamo che $\forall W, \forall x$ (per ogni matrice quadrata e per qualsiasi vettore) vale

$$\|Wx\| \leq \|W\| \cdot \|x\|$$

dove

$$\|W\|$$

è l'autovalore massimo, cioè la direzione in cui la matrice "stiraccia" di più il vettore x . Se x è l'autovettore corrispondente all'autovalore massimo allora

$$\|Wx\| = \|W\| \cdot \|x\|$$

Andiamo a vedere come progredisce la differenza (guardiamo il quadrato, è più facile)

$$\|x_{t+1} - x\|^2 = \|Wx_t - Wx\|^2 =$$

raccogliamo W e aggiungiamo un fattore che è uguale a 0

$$= \|W(x_t - x) - \frac{\mathbf{1} \mathbf{1}^T}{n}(x_t - x)\|$$

dove $\frac{\mathbf{1} \mathbf{1}^T}{n}(x_t - x) = 0$ per il fatto che

$$\frac{\mathbf{1} \mathbf{1}^T}{n} x_t = \frac{\mathbf{1} \mathbf{1}^T}{n} x$$

Raccogliamo a fattore comune e otteniamo

$$= \|(W - \frac{\mathbf{1} \mathbf{1}^T}{n})(x_t - x)\|^2$$

che sappiamo essere

$$\leq \|W - \frac{\mathbf{1} \mathbf{1}^T}{n}\|^2 \cdot \|x_t - x\|^2$$

Per far sì che $x_t \rightarrow x$ abbiamo bisogno che

$$\|W - \frac{\mathbf{1} \mathbf{1}^T}{n}\|^2 \ll 1$$

Fatto 25. Se G non è bipartito allora

$$\|W - \frac{\mathbf{1} \mathbf{1}^T}{n}\| = \max\{|\omega_2|, |\omega_n|\}$$

Dimostrazione 37. Sappiamo che $\omega_1 = 1$, $u_1 = \frac{1}{\sqrt{n}}$. Consideriamo la matrice $W = U \Lambda U^T$ tale che

$$\Lambda = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \omega_n \end{bmatrix}$$

ovvero $\Lambda = \text{diag}(1, \omega_2, \dots, \omega_n)$, mentre $U = [u_1, \dots, u_n]$. Definiamo la matrice maschera M

$$M = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

che è formata da tutti 0 tranne che nella posizione $(1, 1)$. Vale che

$$UMU^T = u_1 u_1^T = \frac{\mathbb{1}\mathbb{1}^T}{n}$$

Allora

$$\begin{aligned} \|W - \frac{\mathbb{1}\mathbb{1}^T}{n}\| &= \|U\Lambda U^T - UMU^T\| \\ &= \|U(\Lambda - M)U^T\| \\ &= \|U \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \omega_2 & \vdots \\ 0 & \dots & \omega_n \end{bmatrix} U^T\| \end{aligned}$$

concludiamo la dimostrazione

$$= \max\{|\omega_2|, |\omega_n|\} = \kappa$$

Possiamo quindi dire che

$$\|x_t - x\| \leq \kappa^t \|x_0 - x\|$$

dove $0 < \kappa < 1$. Come scegliamo W in pratica? Consideriamo $L = D - A$ e $0 < \alpha \leq \frac{1}{\Delta(G)}$. Allora W è tale che

$$W = I - \alpha L$$

Notiamo che $\mathbb{1}^T I = 1 = I \mathbb{1}$, cioè I è bi-stocastica. Inoltre $L \mathbb{1} = 0 = \mathbb{1}^T L$, cioè la somma delle righe e delle colonne di L è 0. Anche se perturbo L con $-\alpha$ resta vero. Quindi noi stiamo sommando una matrice bi-stocastica(I) a una matrice $(-\alpha L)$ tale che la somma delle colonne e delle righe è 0. La matrice risultante non può che essere bi-stocastica. Per capire se è una matrice di gossip dobbiamo vedere se $\omega_{ij} \geq 0$. Possiamo riscrivere W come

$$W = (I - \alpha D) + \alpha A$$

dove

$$D = \begin{bmatrix} d(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d(n) \end{bmatrix} \cdot \alpha$$

siccome $\alpha \leq \frac{1}{\Delta(G)}$

$$= \begin{bmatrix} < 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & < 1 \end{bmatrix}$$

Quindi $I - \alpha D > 0$, e la somma di due quantità positive è > 0 . Adesso andiamo a prendere il grafo G e il laplaciano L con i suoi autovalori $0 \leq \lambda_1 \leq \dots \leq \lambda_n$. Abbiamo detto che $W = I - \alpha L$. Quindi gli autovalori di W sono tali che

$$\omega_i = 1 - \alpha \lambda_i > 0$$

In particolare

$$\kappa = \max\{|\omega_n|, |\omega_2|\} = \omega_2 = 1 - \alpha\lambda_2$$

Analizziamo ω_2

$$\omega_2^t = (1 - \alpha\lambda_2)^t$$

usiamo $\alpha = \frac{1}{\Delta(G)}$

$$= (1 - \frac{\lambda_2}{\Delta(G)}) \leq e^{-\frac{\lambda_2 t}{\Delta(G)}}$$

Siccome $\omega_2 = \kappa$ e κ è il fattore di convergenza, possiamo dire che abbiamo una convergenza esponenziale di x_t , limitata da $\frac{\lambda_2}{\Delta(G)}$. Per questa scelta di W possiamo definire x_{t+1} , cosa che ci eravamo preposti a inizio lezione,

$$x_{t+1} = Wx_t \Leftrightarrow$$

$$x_{t+1}(v) = \alpha + x_t(v) + \sum_{\substack{u: \\ (u,v) \in E}} (x_t(u) - x_t(v))$$

dove

$$\sum_{\substack{u: \\ (u,v) \in E}} (x_t(u) - x_t(v))$$

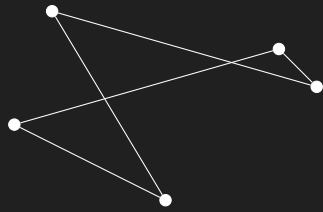
è la somma della differenza tra i vecchi stati.

Capitolo 21

Ventunesima lezione

21.1 Decisioni sequenziali

Parliamo della teoria delle decisioni sequenziali usando la teoria spettrale. Vogliamo raccomandare dei film a degli utenti. Quindi abbiamo una sequenza di utenti e a ogni utente ne raccomandiamo uno. Consideriamo un grafo $G = (V, E)$ noto



ogni nodo corrisponde a un film e ogni arco è una relazione di similarità. Se tra due film c'è un arco, i film sono simili. Definiamo $l_t(i)$ come quanto dispiace il film i -esimo al t -esimo. Questa è una funzione di loss, quindi più piccolo è il valore meglio è. Non conosciamo $l_t(i) \forall t = 1, 2, \dots, i \in V$, sappiamo solamente che $l_t(i) \in [0, 1]$. Definiamo un algoritmo per la scelta delle raccomandazioni.

Algorithm 8

- 1: **for** $t = 1, 2, \dots$ **do**
 - 2: Raccomandiamo $I_t \in V$ al t -esimo utente
 - 3: Rileviamo $l_t(I_t) \in [0, 1]$, ovvero il gradimento dell'utente
 - 4: Osserviamo $l_t(i) \forall i \in N_{I_t}$, ovvero vediamo il gradimento se avessimo scelto uno dei vicini di I_t
-

Il nostro obiettivo è minimizzare il dispiacere, quindi

$$\sum_t l_t(I_t)$$

il problema è che questo valore è dipendente dal numero di film e dal numero di vicini di un nodo. Consideriamo allora il regret

$$R_T = \sum_{i=1}^T l_t(I_t) - \min_{i \in V} \sum_{t=1}^T l_t(i)$$

ovvero quanto perdiamo rispetto alla migliore scelta, cioè il singolo film

che **globalmente** (non per singola scelta) piaceva di più. Notiamo che nel peggiore dei casi R_t è lineare rispetto al numero di scelte, $R_t = O(T)$. Noi vorremmo un regret sub-lineare, ovvero

$$\frac{R_T}{T} \rightarrow 0$$

per $T \rightarrow \infty$. Dimostriamo che $\forall G \forall l_1, \dots, l_T \in [0, 1]^k$ vale

$$R_T \leq c\sqrt{\alpha(G)T \ln k}$$

dove $k = |V|$, c è una costante, $\alpha(G)$ il numero di indipendenza e T il numero di decisioni che prendo. Notiamo che se il grafo è sparso allora $\alpha(G) \rightarrow k$, mentre se è denso $\alpha(G) \rightarrow 1$. Cioè, più è denso più ottengo informazioni rapidamente. In che modo raccomando un film? Consideriamo la raccomandazione al tempo t , $I_t \sim p_t$ su V . p_t è tale che

$$\begin{aligned} p_t(i) &= \mathbb{P}(I_t = i | I_1, \dots, I_{t-1}) = \\ &= \frac{\exp(-\eta \sum_{s=1}^{t-1} \hat{l}_s(i))}{\sum_{j \in V} \exp(-\eta \sum_{s=1}^{t-1} \hat{l}_s(j))} \end{aligned}$$

dove \exp indica semplicemente l'esponenziale (e^x) mentre μ è un fattore > 0 . Il denominatore è semplicemente un fattore di normalizzazione. Usiamo l'esponenziale perché funziona bene, dilata i valori. Cosa sono $\hat{l}_s(i)$? Noi non conosciamo $l_s(i)$ per tutti i film, infatti lo conosciamo solo per il film scelto e per i

suoi vicini. Quindi costruiamo degli stimatori, $\hat{l}_s(i)$ è una stima di $l_s(i)$. Ci chiediamo qual è la probabilità di osservare $l_t(i)$?

$$\mathbb{P}(\text{osservare } l_t(i) | I_1, \dots, I_t) =$$

$$\sum_{j \in N_i} p_t(j) + p_t(i) = q_t(i)$$

Questo perché, come già detto prima, conosciamo $l_t(i)$ se lo scegliamo direttamente oppure se scegliamo uno dei suoi vicini. Definiamo la stima $\hat{l}_t(i)$ come

$$\hat{l}_t(i) = \frac{l_t(i)}{q_t(i)} \mathbb{I}_{\{\text{osservo } l_t(i)\}}$$

dove dividiamo per $q_t(i)$ per rendere più grande il valore. Lo stimatore può assumere solo due valori

$$\hat{l}_t(i) \in \left\{ \frac{l_t(i)}{q_t(i)}, 0 \right\}$$

notiamo che è uno stimatore stocastico, perché dipende da scelte stocastiche. Il problema è che all'inizio non abbiamo ancora fatto alcuna scelta, quindi definiamo

$$P_1(i) = \frac{1}{k} \quad i = 1, \dots, k$$

Da ora indicheremo il numeratore di $p_t(i)$ con $w_t(i)$ e il denominatore con W_t , quindi

$$p_t(i) = \frac{w_t(i)}{W_t} \geq 0$$

Fissiamo I_1, \dots, I_n e andiamo a considerare il rapporto

$$\frac{W_{t+1}}{W_t}$$

non è chiaro perché partiamo da questo rapporto, però ci servirà a trovare un bound per R_T .

$$= \sum_{i \in V} \frac{w_{t+1}(i)}{W_t}$$

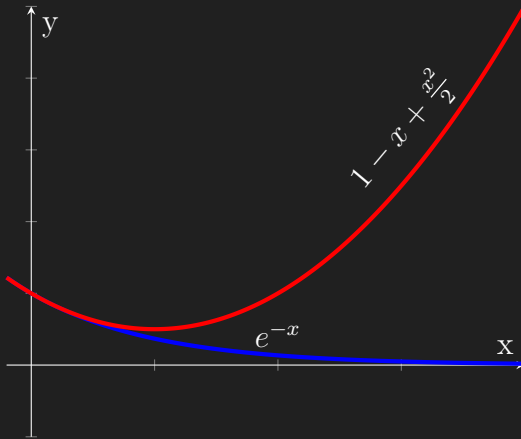
notiamo che

$$w_{t+1}(i) = e^{-\eta \sum_{s=1}^t \hat{l}_s(i)} = w_t(i) e^{-\eta \hat{l}_t(i)}$$

quindi il rapporto può essere riscritto come

$$= \sum_{i \in V} \frac{w_t(i)}{W_t} e^{-\eta \hat{l}_t(i)} = \sum_{i \in V} p_t(i) e^{-\eta \hat{l}_t(i)}$$

sappiamo che $e^{-x} \leq 1 - x + \frac{x^2}{2} \forall x \geq 0$

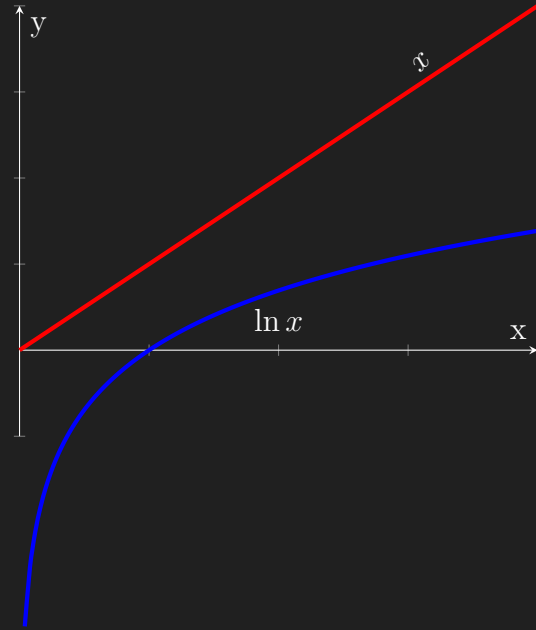


quindi

$$\leq \sum_{i \in V} p_t(i) \left(1 - \eta \hat{l}_t(i) + \frac{\eta^2}{2} \hat{l}_t(i)^2\right)$$

$$= 1 - \sum_{i \in V} p_t(i) \eta \hat{l}_t(i) + \sum_{i \in V} p_t(i) \frac{\eta^2}{2} \hat{l}_t(i)^2$$

Lasciamo questo risultato da parte per ora. Ora consideriamo prima di tutto la nota disuguaglianza $\ln(1 - x) \leq x \forall x > 0$



Prendiamo lo stesso rapporto di prima e mettiamolo in logaritmo

$$\ln \frac{W_{t+1}}{W_1} = \sum_{t=1}^T \ln \frac{W_{t+1}}{W_t}$$

questo vale perché la somma è telescopica, svolgendola ci restano solo il primo e l'ultimo elemento. Usiamo ora la disuguaglianza $\ln(1 - x) \leq x$

$$\leq \sum_{t=1}^T \ln \left(1 - \sum_{i \in V} p_t(i) \eta \hat{l}_t(i) + \sum_{i \in V} p_t(i) \frac{\eta^2}{2} \hat{l}_t(i)^2\right)$$

$$= - \sum_{t=1}^T \sum_i p_t(i) \hat{l}_t(i) + \frac{\eta^2}{2} \sum_t \sum_i p_t(i) \hat{l}_t(i)^2$$

Abbiamo trovato un maggiorante per $\ln \frac{W_{t+1}}{W_1}$, troviamo un minorante.

$$\ln \frac{W_{t+1}}{W_1} \geq \ln \frac{e^{-\eta \sum_{s=1}^T \hat{l}_s(k)}}{\sum e^{-\eta \cdot 0}}$$

dove abbiamo scelto un termine qualsiasi della somma. E' ovvio che $\sum_{i=1}^n x \geq x_i$, dato che tutti gli elementi della sommatoria sono positivi. Inoltre il denominatore è e^0 poiché al tempo 1 le stime sono = 0. Concludiamo scrivendo

$$= -\eta \sum_{t=1}^T \hat{l}_t(k) - \ln k$$

Abbiamo quindi ottenuto un minorante e un maggiorante. Li mettiamo insieme (minorante < maggiorante), spostiamo i termini a sinistra e a destra in modo che tutti siano positivi e dividiamo per η . In questo modo otteniamo $\forall I_1, \dots, I_t \in V$

$$\sum_t \sum_i p_t(i) \hat{l}_t(i) \leq \sum_t \hat{l}_t(k) + \frac{\ln k}{\eta} + \frac{\eta}{2} \sum_t \sum_i p_t(i) \hat{l}_t(i)^2$$

Ci interessa guardare il valore atteso rispetto all'estrazione di I_1, \dots, I_n . Prima però ricordiamo che

$$\mathbb{E}[x] = \mathbb{E}[\mathbb{E}[x|y]]$$

che ci tornerà utile più avanti. Prendiamo il valore atteso della disequazione di prima

$$\mathbb{E}\left[\sum_t \sum_i p_t(i) \hat{l}_t(i)\right] \leq \mathbb{E}\left[\sum_t \hat{l}_t(k)\right] + \frac{\ln k}{\eta} + \frac{\eta}{2} \mathbb{E}\left[\sum_t \sum_i p_t(i) \hat{l}_t(i)^2\right]$$

Usiamo il fatto sul valore atteso che abbiamo presentato prima e scriviamo (formula un po' lunga)

$$\begin{aligned} \mathbb{E}\left[\sum_t \sum_i p_t(i) \mathbb{E}[\hat{l}_t(i) | I_1, \dots, I_{t-1}]\right] &\leq \\ \mathbb{E}\left[\sum_t \mathbb{E}[\hat{l}_t(k) | I_1, \dots, I_{t-1}]\right] + \frac{\ln k}{\eta} &+ \\ + \frac{\eta}{2} \mathbb{E}\left[\sum_t \sum_i p_t(i) \mathbb{E}[\hat{l}_t(i)^2 | I_1, \dots, I_{t-1}]\right] \end{aligned}$$

Notiamo due fatti

$$\begin{aligned} \mathbb{E}[\hat{l}_t(i) | I_1, \dots, I_n] &= \frac{l_t(i)}{q_t(i)} \mathbb{P}(\text{osservo } l_t(i)) \\ &+ 0 \cdot \mathbb{P}(\text{non osservo } l_t(i)) \end{aligned}$$

dove le probabilità sono condizionate a I_1, \dots, I_{t-1} . Siccome la probabilità di osservare $l_t(i)$ è $q_t(i)$ scriviamo

$$= l_t(i)$$

Questo vuol dire che $\hat{l}_t(i)$ è uno stimatore non distorto. Consideriamo ora il valore atteso di $\hat{l}_t(i)^2$

$$\begin{aligned} & \mathbb{E}[\hat{l}_t(i)^2 | I_1, \dots, I_{t-1}] \\ &= \frac{l_t(i)^2}{q_t(i)} q_t(i) + 0(1 - q_t(i)) \leq \frac{1}{q_t(i)} \end{aligned}$$

dove la disuguaglianza vale perché $l_t(i)^2$ al massimo vale 1. Possiamo usare queste informazioni nella disuguaglianza di prima, e inoltre portiamo a primo membro $\sum_t l_t(k)$

$$\begin{aligned} & \mathbb{E}[\sum_t \sum_i p_t(i) l_t(i)] - \sum_t l_t(k) \\ & \leq \frac{\ln k}{\eta} + \frac{\eta}{2} \mathbb{E}[\sum_t \sum_i \frac{p_t(i)}{q_t(i)}] \end{aligned}$$

Analizziamo il lato sinistro, possiamo dire che

$$\mathbb{E}[\sum_t \sum_i p_t(i) l_t(i)] = \mathbb{E}[l_t(I_t)]$$

Inoltre, siccome k è un elemento qualsiasi possiamo sceglierlo tale che

$$k = \operatorname{argmin}_{i \in V} \sum_{t=1}^T l_t(i)$$

La parte sinistra della disequazione è R_T , il regret. Concentriamoci sulla parte destra, in particolare su

$$\sum_i \frac{p_t(i)}{q_t(i)}$$

Sappiamo che il teorema di Turàn ci dice

$$Q(G) = \sum_{i \in V} \frac{1}{1 + d(i)} \leq \alpha(G)$$

moltiplicando e dividendo sopra e sotto per $\frac{1}{k}$, dove $k = |V|$, otteniamo

$$Q(G) = \sum_{i \in V} \frac{\frac{1}{k}}{\frac{1}{k} + \frac{d(i)}{k}}$$

e scegliendo $p(i) = \frac{1}{k}$

$$= \sum_{i \in V} \frac{p(i)}{p(i) + \sum_{j \in N_i} p(j)}$$

Questo risultato è molto interessante, il problema è che non possiamo applicarlo al nostro caso perché la probabilità non è uniforme. In realtà il teorema di Turàn si può estendere per qualunque $p(i)$ e la dimostrazione è uguale a quella che abbiamo fatto. Quindi, per il teorema di Turàn, scriviamo

$$\begin{aligned} \sum_{i \in V} \frac{p_t(i)}{q_t(i)} &= \sum_{i \in V} \frac{p_t(i)}{p_t(i) + \sum_{j \in N_i} p_t(j)} \\ &\leq \alpha(G) \end{aligned}$$

Concludiamo mettendo tutto assieme

$$\mathbb{E}[R_T] \leq \frac{\ln k}{\eta} + \frac{\eta}{2} \alpha(G) T$$

scegliamo η in modo che i due addendi siano bilanciati, quindi vogliamo che

$$\frac{\ln k}{\eta} = \frac{\eta}{2} \alpha(G) T$$

risolviamo per η

$$\eta = \sqrt{2 \frac{\ln k}{\alpha T}}$$

quindi

$$\mathbb{E}[R_T] \leq \frac{3}{2} \sqrt{\alpha T \ln k}$$

e questo risultato non è migliorabile. Notiamo che il risultato ottenuto dipende dal grado dei nodi. Infatti la varianza dello stimatore dipende dal grado e quindi l'algoritmo, che usa gli stimatori, ne dipende in modo indiretto.