

Algoritmi e Complessità

Massimo Perego

Contents

1	Notazione	4
2	Algoritmi 101	7
2.1	Cos'è un problema?	7
2.2	Cos'è un algoritmo?	9
2.3	Cos'è un algoritmo per un problema π ?	10
2.4	Teoria della complessità	11
2.4.1	Algoritmica	11
2.4.2	Strutturale	13
2.4.3	Classi di complessità:	14
2.4.4	Riduzione in tempo polinomiale	15
	Teorema	15
	Teorema di Cook	16
3	Problemi di ottimizzazione	17
3.1	Problema di decisione associato	18
3.2	Classe di complessità \mathcal{PO}	19
	Teorema	19
3.3	Rapporto di prestazione	20
3.4	Altre classi	21
3.5	Problema di MaxMatching	22
	Visite di grafi	24
3.5.1	Trovare un cammino aumentante	26
	Teorema	26
4	Tecniche greedy	28
4.1	Problema di LoadBalancing	28
	Teorema	28

4.1.1	Algoritmo greedy per LoadBalancing	29
	Teorema	29
	Teorema	31
4.1.2	SortedGreedyBalance	32
	Teorema	32
4.2	Problema della selezione dei centri CenterSelection	34
4.2.1	Algoritmo CenterSelectionPlus	35
	Teorema 1	36
	Teorema 2	36
4.2.2	GreedyCenterSelection	38
	Teorema	39
	Teorema	40
	Funzione Armonica	42
4.3	Problema di SetCover	43
4.3.1	GreedySetCover	43
	Teorema	46
	Tightness	47
5	Tecnica di Pricing	49
5.1	Problema VertexCover	49
5.1.1	PricingVertexCover	51
	Teorema	52
5.2	Problema DisjointPaths	53
5.2.1	GreedyPaths	54
	Teorema	58
6	Tecniche basate sull'arrotondamento	59
	Programmazione Lineare	59
	PL Intera	59
6.1	VertexCover via Rounding	60
7	Altri esempi	63
	Nascita della teoria dei grafi	63
	Teorema di Eulero	63
	Handshaking Lemma	64
	Teorema	64
7.1	Problema del Traveling Salesman (TSP)	65
7.1.1	Algoritmo di Cristophides	67
	Lemma 1	68
	Lemma 2	68

	Teorema	68
	Tightness	69
7.1.2	Inapprossimabilità del TSP	71
	Teorema	71
8	Schemi di approssimazione PTAS e FPTAS	73
8.1	PTAS per 2-LoadBalancing	73
	Teorema	74
	Teorema	76
8.2	Knapsack Problem	77
8.2.1	Soluzione di Programmazione Dinamica: Versione A .	79
8.2.2	Soluzione di Programmazione Dinamica: Versione B .	81
8.2.3	Algoritmo DP Approssimato	82
	Teorema	85

1 Notazione

Definire una notazione che verrà usata da qui in avanti.

Verranno usati tanto gli **insiemi**: \mathbb{N} , \mathbb{Z} , \mathbb{Q} , \mathbb{R} ed anche le versioni con **solo i positivi** di ognuno di questi \mathbb{N}^+ , ...

Un **magma** è un **insieme più un'operazione binaria** definita su di esso.

Esempio

$$(A, \otimes)$$

Se il magma ha la **proprietà associativa** sull'operazione è un **semigrupp**

$$(y \otimes z) \otimes x = y \otimes (z \otimes x)$$

Se **esiste elemento neutro** rispetto all'operazione si parla di **monoide**.

Se esiste elemento neutro è unico

$$x \otimes \bar{e} = \bar{e} \otimes x = x$$

Se l'**operazione è anche commutativa** allora **monoide abelliano**. L'esempio più semplice sono i naturali con l'operazione somma.

Se per ogni elemento c'è **anche un inverso** allora è un **gruppo**.

Definendo un **alfabeto come un insieme finito non vuoto** di elementi.

Monoide libero: prendendo un alfabeto Σ e considerando tutte le sequenze finite costituite da elementi dell'alfabeto, si nota con

$$w \in \Sigma^* \quad w = w_1, w_2, w_3, \dots, w_n \quad n \geq 0$$

$$w_i \in \Sigma \quad |w| = n$$

La concatenazione è un'operazione interna a Σ^* , due parole se le concateno le metto una dopo l'altra

$$w = w_1, w_2, w_3, \dots, w_n$$

$$w' = w'_1, w'_2, w'_3, \dots, w'_m$$

$$w \cdot w' = w_1, w_2, w_3, \dots, w_n, w'_1, w'_2, w'_3, \dots, w'_m$$

(Σ^*, \cdot) è un monoide, chiamato **monoide libero** su Σ (abelliano solo quando nell'insieme è presente solo una lettera).

In generale un monoide libero è un monoide con una base, ovvero un insieme di elementi che generano il monoide e che non possono essere generati da altri elementi della base.

Se A e B sono due insiemi allora B^A è l'**insieme delle funzioni da A a B**

$$B^A = \{f | f : A \rightarrow B\}$$

La cardinalità finale è quella di $|B|^{|A|}$, per questo la notazione è in quest'ordine.

Con l'**insieme** K si intende

$$K = \{0, 1, \dots, k-1\}$$

Con 0 che significa l'insieme vuoto.

Sostanzialmente l'insieme di tutti i valori inferiori a K .

L'insieme dei valori dei bit si chiama $2 = \{0, 1\}$.

Considerando

$$2^A = \{f | f : A \rightarrow 2\} = \{f | f : A \rightarrow \{0, 1\}\}$$

Questa è una funzione che **mappa ogni elemento di A in 0 o in 1**, in biezione con $\cong \{x | x \text{ è sottoinsieme di } A\} = P(A)$ (insieme delle parti di A).

A^2 , **insieme delle coppie di elementi di A** , ma sarebbe

$$A^2 = \{f|f : 2 \rightarrow A\} = \{f|f : \{0, 1\} \rightarrow A\}$$

che è in biezione con

$$A \times A$$

Quindi con il prodotto cartesiano di A con A .

2^* è un monoide libero sull'alfabeto binario, **tutte le combinazioni binarie**. Quindi una qualsiasi stringa in binario sarà parte di questo monoide.

Mentre 2^{2^*} rappresenta la **famiglia di tutti i linguaggi binari** (su $\{0, 1\}$). Ogni linguaggio che consiste di numeri binari è membro di 2^{2^*} .

2 Algoritmi 101

2.1 Cos'è un problema?

Un **problema** π è **costituito da**:

- Un **insieme di input** possibili I_π
- Un **insieme di output** possibili O_π
- Funzione $Sol_\pi : I_\pi \rightarrow 2^{O_\pi} \setminus \{\emptyset\}$ **mappa ogni input ad un insieme di output corretti possibili**, non è detto che dobbiamo avere un solo output corretto, possono esserci più output validi. Molti problemi considerati potranno avere più output corretti.

Esempio di problema: Descrizione informale: decidere se un numero è primo.

Supponendo che si tratta, in input, di numeri naturali scritti in base 2 (non specificato, da specificare, questo sono assunzioni).

Mentre in output si può avere solo s o no , si tratta di un problema di decisione

$$O_\pi = \{no, yes\}$$

Anche se non esplicitamente detti, **input e output dovrebbero essere specificati come su un linguaggio binario**, sono **sequenze di bit**.

Altro esempio: emettere il MCD tra due interi positivi x e y . Non è più un problema di decisione.

Come distinguo i due valori in input?

Soluzioni: raddoppio tutti i bit e 01 diventa il separatore.

Altra soluzione: Elias γ_x sapendo la lunghezza l in bit dell'input x , scrivo l in unario, quindi l bit ed uno 0 prima dell'input, così in base a quanti 1 leggo so quanti bit leggere dopo.

I problemi generalmente saranno scritti “approssimativamente”, quindi sono necessarie diverse assunzioni per specificarli formalmente.

Nella pratica quindi i dettagli di implementazione dell'algoritmo cambieranno, ma tra n^2 e $n^2 \log n$ ci cambia poco, sempre dominato da un polinomio ed un polinomio vale l'altro (assunzione solo teorica, nella realtà non è proprio così ma sticazzi).

2.2 Cos'è un algoritmo?

Senza definirlo esattamente in modo formale, si presume che a questo punto tu ne abbia un'idea.

Ma se proprio dobbiamo: tutti gli algoritmi sono Macchine di Turing (programmi eseguibili da una MdT).

Tesi di Church-Turing: storicamente, hanno preso tutte le definizioni di algoritmi presenti, hanno provato che sono tutte equivalenti ed ogni definizione riporta alla stessa famiglia di problemi risolvibili: problemi che con una MdT sono risolvibili allora sono risolvibili anche secondo qualsiasi altra definizione.

Al giorno d'oggi è importante perché si traspone in “ogni linguaggio di programmazione (Turing-completo) può risolvere gli stessi problemi risolvibili da tutti gli altri”.

In passato è stato discusso come sia impossibile concepire una definizione realistica di algoritmo non equivalente alla MdT.

2.3 Cos'è un algoritmo per un problema π ?

Dando $x \in I_\pi$ ad un algoritmo A eseguito su una MdT restituirà $y \in O_\pi$ tale che $y \in Sol_\pi(x)$, ovvero **dato un input restituisce una soluzione tra quelle ammesse**.

Molti problemi non sono risolvibili. Considerando i problemi di decisione, data una stringa binaria bisogna rispondere *s* o *no*.

Se ne può avere uno per ogni possibile insieme di stringhe binarie, gli elementi di 2^{2^*} sono tutti problemi di decisione, e possiamo vedere che

$$|2^{2^*}| \cong |2^{\mathbb{N}}| \cong |\mathbb{R}|$$

ha la potenza del continuo, quindi il numero possibile di problemi di decisione è un infinito non numerabile.

Ma i programmi (algoritmi per risolvere questi problemi) sono infiniti? Sì, ma hanno cardinalità $|\mathbb{N}|$, essendo questa la dimensione dell'insieme di risposte possibili ($|2^*|$).

Quindi i problemi sono di più dei possibili programmi per risolverli. Per forza di cose la maggior parte dei problemi non sono risolvibili.

2.4 Teoria della complessità

Tutti i problemi che verranno considerati sono risolvibili, ma **quanto efficientemente**? La teoria della complessità stabilisce la **quantità di risorse consumate da un problema**.

Si può parlare di teoria della complessità in due termini:

- **Algoritmica**
- **Strutturale**

2.4.1 Algoritmica

Dato π

1. Stabilire **se è risolvibile** (per noi sempre sì); c'è un algoritmo che lo risolve?
2. **Risolvibile con che costo?** Ci sono più **misure** possibili:
 - Tempo (solitamente sarà usata questa)
 - Spazio
 - Numero di processori utilizzati/somma dei tempi di tutte le CPU
 - Energia dissipata

Considerando il tempo, vogliamo stabilire **dato un input quanto tempo impiega**, ovvero **quanti passi** deve svolgere

$$T_A : I_\pi \rightarrow \mathbb{N}$$

Invece di calcolare la **complessità** input per input si considera “per taglia”, **per ogni dimensione**, prendendo l'input che ci impiega il tempo massimo per taglia

$$t_a : \mathbb{N} \rightarrow \mathbb{N}$$

$$t_a(n) = \max\{T_A(x) \mid x \in I_\pi \text{ t.c. } |x| = n\}$$

Si tratta di un'assunzione worst-case, dati input di stessa dimensione prendo il peggiore. Potrebbe essere utile fare una media al posto di un'assunzione pessimistica, ma noi teniamo così perché è comodo, inoltre nella realtà possiamo solo metterci di meno.

Se dobbiamo valutare quale algoritmo usare il **confronto va fatto asintoticamente**, i casi piccoli sono “facili” (di solito, ma andrebbero considerati i casi d’uso effettivi? In estremo potrei fare una tabella). Quindi consideriamo come le performance dell’algoritmo scalano con la dimensione dell’input.

Quindi due assunzioni: worst-case e asintotico.

Generalmente quindi tra

$$t_a = O(n^2) \quad t_a = O(n^7)$$

meglio il primo tendenzialmente.

Bisogna anche provare che non ci possono essere cose migliori/peggio, quindi upper and lower bound.

2.4.2 Strutturale

Fissato un problema π , qual'è la sua complessità? Non vogliamo cercare la complessità di un algoritmo, ma la **complessità del miglior algoritmo che risolve un problema** π . Questo permette di eliminare facilmente problemi che non hanno soluzioni efficienti.

Con la notazione worst case asintotica sopracitata possiamo avere un **upper bound** per il **tempo impiegato** per la risoluzione di un problema. Quindi nel caso peggiore avrà asintoticamente quel comportamento. $O(n^2)$ vuol dire che nel caso peggiore ci mette tempo quadratico rispetto all'input.

Esistono anche **lower bounds**, ovvero limiti inferiori per il **tempo impiegato** da un algoritmo e dimostrarlo implica **dimostrare che per come è fatto un problema, questo non può impiegare meno di un certo tempo**. $\Omega(n)$ vuol dire che non ci può mettere meno di tempo lineare.

Un problema viene “chiuso” quando upper e lower bound coincidono, vuol dire che si ha l'algoritmo ottimo e si conosce esattamente la complessità del problema. Per esempio, l'ordinamento di array basato su confronti e scambi ci mette esattamente $n \log n$, quindi si dice $\Theta(n \log n)$.

Le casistiche problematiche sono quando gli **upper bound sono esponenziali ed i lower bound polinomiali**, quindi le possibilità sono:

- esiste un upper bound da migliorare, quindi esistono algoritmi polinomiale non ancora scoperti
- esiste un lower bound esponenziale, quindi il problema non è risolvibile in tempo polinomiale

In entrambi i casi saremmo contenti, il non avere informazioni a riguardo è il problema.

2.4.3 Classi di complessità:

Per la complessità strutturale sono state inventate delle classi di problemi per **dividerli in base alla difficoltà di risoluzione**:

- \mathcal{P} : problemi di decisione risolvibili in tempo polinomiale; la domanda diventa “il mio π è in \mathcal{P} ?”
- \mathcal{NP} : problemi di decisione risolvibili in tempo polinomiale **su macchine non deterministiche**

Sostanzialmente macchine non deterministiche vuol dire che: una macchina su cui il programma può essere eseguito più volte in parallelo con determinate variabili settate a 0 e 1, permettendo di eseguire in parallelo tante “versioni” ed ottenere una soluzione per ogni esecuzione parallela. La soluzione alla fine è *yes* se esiste almeno una esecuzione che ha dato *yes* e *no* se tutte le esecuzioni hanno riportato *no*.

CNF-Sat problem: l’input è una espressione logica in forma CNF (Conjunctive Normal Form), quindi un **and** tra clausole, ognuna delle quali contiene **or** di letterali, i quali possono essere variabili o negazioni di variabili.

L’output è se l’espressione logica φ è soddisfacibile, ovvero, esiste \exists assegnazione che rende φ *true*?

Attualmente questo problema si pensa sia in \mathcal{NP} , non sappiamo se possa stare in \mathcal{P} .

Sappiamo rientra in \mathcal{NP} perché è facile pensare ad un algoritmo non deterministico per questo problema, basta avere una esecuzione parallela per ogni possibile valore di ognuna delle variabili, e se almeno un output è *sì* allora è soddisfacibile.

Relazione tra \mathcal{P} e \mathcal{NP} : \mathcal{P} è ovviamente **incluso in \mathcal{NP}** , dato che se posso risolvere un problema in modo polinomiale, posso risolverlo anche con il parallelismo aggiunto delle macchine non deterministiche.

La vera domanda è se $\mathcal{P} = \mathcal{NP}$, quindi **se le due classi coincidono**, le possibilità sono:

- le due **classi sono uguali** $\mathcal{P} = \mathcal{NP}$: sarebbe molto bello, vorrebbe dire che **tutti i problemi sono risolvibili polinomialmente**
- \mathcal{P} **contenuta in \mathcal{NP}** , $\mathcal{P} \subset \mathcal{NP}$: meno emozionante ma vorrebbe dire che **esistono dei problemi non risolvibili in tempo polinomiale**

Ma è uno dei millennium problem, quindi penso rimarrà lì.

2.4.4 Riduzione in tempo polinomiale

Un problema π_1 è polinomialmente riducibile a π_2 iff:

$$\pi_1 \leq_p \pi_2 \text{ iff:}$$

$$\exists f : I_{\pi_1} \rightarrow I_{\pi_2} \text{ t.c. } x \in I_{\pi_1} \quad Sol_{\pi_1}(x) = yes \quad Sol_{\pi_2}(f(x)) = yes$$

Esiste una funzione che a partire da un input di π_1 arriva ad un input di π_2 in tempo polinomiale e se x è un input di π_1 possiamo avere lo stesso output che x fornisce su π_1 per π_2 con input $f(x)$.

Sostanzialmente vuol dire che possiamo **trasformare input di π_1 in input di π_2 in tempo polinomiale**, allora **se π_2 è risolvibile in tempo polinomiale lo è anche π_1** .

Teorema: Se $\pi_1 \leq_p \pi_2$ e $\pi_2 \in \mathcal{P}$ allora $\pi_1 \in \mathcal{P}$. Se un problema può essere polinomialmente ridotto ad un altro all'interno di \mathcal{P} , allora anch'esso è in \mathcal{P} .

Sostanzialmente \leq_p può essere letto come “non è più difficile di”.

Teorema di Cook: dato che $\text{CNF-Sat} \in \mathcal{NP}$ e $\forall \pi \in \mathcal{NP}$ allora $\pi \leq_p \text{CNF-Sat}$. Esistono **problemi all'interno di \mathcal{NP} a cui tutti gli altri problemi si possono ridurre** (provato per la prima volta da Cook con il CNF-Sat).

Questi problemi sono **\mathcal{NP} -completi** (\mathcal{NP}_c)

$$(\mathcal{NP}_c) = \{\pi \in \mathcal{NP} \mid \forall \pi' \in \mathcal{NP} \pi' \leq_p \pi\}$$

\mathcal{NP} completi vuol dire **tutti i problemi in \mathcal{NP} si possono ridurre ad un altro problema in \mathcal{NP}** , quindi alla complessità di questi altri. Vuol dire sostanzialmente che sono tutti circa “difficili uguali” e che “cambia poco” (secondo la riduzione polinomiale).

Se noi sapessimo che un qualunque problema \mathcal{NP}_c è risolvibile in tempo polinomiale, l'implicazione di questo teorema è che tutti i problemi in \mathcal{NP} sarebbero risolvibili in tempo polinomiale, facendo collassare \mathcal{P} e \mathcal{NP} .

Corollario: se $\pi \in \mathcal{NP}_c$ e $\pi \in \mathcal{P}$ allora $\mathcal{P} = \mathcal{NP}$.

Nello scenario in cui $\mathcal{P} \neq \mathcal{NP}$ i problemi \mathcal{NP}_c sono sottoinsieme di \mathcal{NP} , ma completamente staccati dall'insieme di \mathcal{P} .

Dire che un **problema è \mathcal{NP}_c** vuol dire che è **risolvibile esattamente solo in tempo esponenziale**.

3 Problemi di ottimizzazione

Sono un caso particolare dei problemi come precedentemente definiti.

Un **problema di ottimizzazione** π :

1. **Input** $I_\pi \subseteq 2^*$
2. Una **funzione** $Amm_\pi : I_\pi \rightarrow 2^{2^*} \setminus \{\emptyset\}$ **mappa** ogni **input** in un **insieme di soluzioni ammissibili**
3. $c_\pi : 2^* \times 2^* \rightarrow \mathbb{N}$, $\forall x \in I_\pi$, $\forall y \in Amm_\pi(x)$, $c_\pi(x, y)$, ovvero un costo/guadagno, **funzione obiettivo**
4. $T_\pi \in \{\max, \min\}$: “tipo del problema”, voglio massimizzare o minimizzare il mio costo/guadagno

Esempio: Max-Sat

- input: formule logiche in forma CNF, esempio:

$$(x_7 \vee x_2) \wedge (x_4 \vee \neg x_5 \vee \neg x_7) \wedge (x_9 \neg x_2)$$

- Le soluzioni ammissibili per una data formula φ sono tutti gli assegnamenti delle variabili di φ , quindi per l'esempio tutte le 2^5 possibili combinazioni di assegnamenti delle 5 variabili
- $c_\pi(\varphi, ass)$: il costo per ogni formula con un dato assegnamento è il numero di clausole rese vere dall'assegnamento
- $T_\pi = \max$: problema di massimizzazione, voglio un assegnamento che massimizzi il numero di clausole soddisfatte

Ovviamente non si può avere un algoritmo polinomiale che risolve questo problema in quanto sarebbe equivalente al CNF-SAT, avere il massimo numero di clausole soddisfacibili vorrebbe dire anche sapere se la formula è soddisfacibile.

3.1 Problema di decisione associato

Ad ogni problema di ottimizzazione π si può associare un problema di decisione $\hat{\pi}$:

- $I_{\hat{\pi}} = (x, k)$, $x \in I_{\pi}$, $k \in \mathbb{N}$.
- La risposta per l'input (x, k) è sì iff $\exists y \in \text{Amm}_{\pi}(x)$ esiste una soluzione ammissibile tale che $c_{\pi}(x, y) \geq k$ per i problemi di massimizzazione, oppure $c_{\pi}(x, y) \leq k$ per problemi di minimizzazione (rientro nel costo k ?).

Sostanzialmente, “trovare il minimo” diventa “esiste una soluzione con costo minore di k ?”.

Esempio: per Max-Sat

- $I_{\text{Max-Sat}} = \{(\varphi, k) \mid \varphi \text{ formula CNF e } k \in \mathbb{N}\}$.
- (φ, k) è sì iff \exists assegnamento che rende vere almeno k clausole di φ .

3.2 Classe di complessità \mathcal{PO}

Si definisce \mathcal{PO} l'insieme dei problemi di ottimizzazione π tali che π è risolvibile in tempo polinomiale

$$\mathcal{PO} = \{ \pi \text{ problemi di ott. } \mid \pi \text{ risolvibile in t polinomiale} \}$$

Teorema: Se $\pi \in \mathcal{PO}$, allora il suo problema di decisione associato sta in \mathcal{P} , i.e. $\hat{\pi} \in \mathcal{P}$.

Corollario: Al contrario, se il problema di decisione associato $\hat{\pi}$ è \mathcal{NP}_c , allora il problema non può essere risolto in tempo polinomiale $\pi \notin \mathcal{PO}$ (assumendo $\mathcal{P} \neq \mathcal{NP}$).

Esempio di problemi in \mathcal{PO} : problemi di programmazione lineare. Mentre restringere il campo a solamente a soluzioni intere fa diventare il problema \mathcal{NP}_c .

3.3 Rapporto di prestazione

Verranno trattati principalmente problemi che rientrano in $\mathcal{N}\mathcal{P}\mathcal{C}$, quindi non saranno risolvibili in modo esatto, ma solamente in modo approssimato. Ma come si quantifica l'approssimazione?

Dato un p.o. π , chiamiamo $\text{opt}_\pi(x)$ il **valore ottimo della funzione obiettivo su input x** .

Dato un **algoritmo approssimato** A (in input prende un possibile input ed in output fornisce una soluzione ammissibile, ma non necessariamente ottima) per π , definisco il **rapporto di approssimazione**

$$R_\pi(x) = \max \left\{ \frac{c_\pi(x, y)}{\text{opt}_\pi(x)}, \frac{\text{opt}_\pi(x)}{c_\pi(x, y)} \right\}$$

Il max serve a prendere un valore sempre ≥ 1 , sia per problemi di minimizzazione che massimizzazione, in modo da avere una definizione unica.

Si dice che A è una α -**approssimazione** per π iff

$$\forall x \in I_\pi : R_\pi(x) \leq \alpha$$

Per ogni input $R_\pi(x)$ è minore o uguale di α , sostanzialmente comunque vada quell'algoritmo non può essere peggio di α sul problema π . Ovviamente con $\alpha = 1$ l'algoritmo è esatto ed il problema sta in \mathcal{PO} .

3.4 Altre classi

Si hanno altre possibili classi di problemi:

- Fuori da \mathcal{PO} ($\mathcal{PO} \subset$), ci sono le classi n -**APX**, quindi 2-APX ad esempio sono gli algoritmi che risolvono con α fino a 2 volte, 3 fino a 3 volte, ...; sostanzialmente classi sempre più grandi fino ad avere approssimazioni arbitrariamente grandi.
- **APX**: problemi risolvibili a meno di una costante di approssimazione di qualche tipo arbitrariamente grande, ma costante. Sempre algoritmi polinomiali, ma approssimati.
- Poi $\log n$ -**APX**: approssimati ma l'approssimazione dipende anche dalla dimensione dell'input, fuori da APX ($\text{APX} \subset$).
- Tutte queste classi sono contenute in \mathcal{NPO} , ovvero l'equivalente di ottimizzazione di \mathcal{NP} .
- esistono anche problemi \mathcal{NPO} -completi \mathcal{NPO}_c , ed interseca tutti i sottoinsiemi fin'ora (tranne \mathcal{PO} ovviamente).
- **PTAS** (polynomial time approximation scheme), (appena fuori \mathcal{PO}): problemi approssimabili “quanto voglio”, quindi posso decidere il tasso di approssimazione ma il tempo peggiora di conseguenza, anche in modo anche esponenziale.
- **FPTAS** (fully PTAS): dentro PTAS, fuori \mathcal{PO} , come PTAS, ma il tempo peggiora esclusivamente in modo polinomiale.

Dovrei mettere un disegno con i cerchietti ma non ho assolutamente voglia, si capisce anche così dai (spero).

3.5 Problema di MaxMatching

Caratteristiche del problema:

- **Input:** un **grafo non orientato** $G = (V, E)$ e bipartito (due gruppi di nodi, i nodi all'interno di un gruppo non hanno collegamenti tra loro)
- **Soluzioni ammissibili:** chiamate **matching**, insieme M , una selezione di lati per cui nessun vertice incide su > 1 lato, scegliere lati in modo da coprire il maggior numero di vertici possibili senza coprire nessun vertice più di una volta
- **Funzione obiettivo:** numero di lati scelti, ovvero **cardinalità di M**
- **Tipo:** problema di **massimizzazione** = max, massimizzare il numero di vertici coperti

In questo caso la **soluzione ottima** si può trovare in **tempo polinomiale**, questo problema rientra in \mathcal{PO} .

Algoritmo del cammino aumentante: un cammino aumentante viene definito su un grafo su cui è già presente un matching, anche parziale (solo in realtà, altrimenti non aumento nulla). Un **vertice** viene definito “**esposto**” quando su quel vertice **non incide nessun lato del matching**.

Un **augmenting path** è un **cammino** che **parte e arriva su un vertice esposto** ed **alterna lati “liberi” (non nel matching) e lati “presi” (nel matching)**. Sequenza di lati alternata che inizia e termina su un vertice esposto. Partendo da vertici esposti il primo lato è per forza non preso.

Sapendo che c'è un cammino aumentante, è possibile scambiare i lati presi e non presi, si può fare uno **switch** sul cammino aumentante trovato. Partendo ed arrivando su nodi esposti per il cammino, **facendo lo switch** ci sarà sempre un **lato in più “preso” nel matching** e di conseguenza due nodi in più.

Se ho un cammino aumentante posso migliorare il matching, quindi quest'ultimo non è ottimo. Per migliorare il matching basta trovare un cammino aumentante, di conseguenza se lo trovo posso migliorare il matching, altrimenti il matching è massimo.

Lemma: se esiste un cammino aumentante per il matching M , allora M non è massimo.

Lemma 2: se M non è massimo allora esiste un cammino aumentante.

Proof. Sia M' un altro matching con $|M'| > |M|$ (M non è massimo, ne esiste uno con cardinalità maggiore). Di conseguenza, considerando gli insiemi dei lati presi da ognuno dei matching, abbiamo tre zone tra i due insiemi:

- lati in comune $M \cap M'$
- lati solo nell'insieme più piccolo $M \setminus M'$
- lati solo nell'insieme più grande $M' \setminus M$

Prendendo la differenza simmetrica tra i due

$$M \Delta M' = (M \setminus M') \cup (M' \setminus M)$$

Considerazione: Nessun vertice può avere più di 2 lati coincidenti in $M \Delta M'$; se in ogni insieme possono avere un solo lato incidente ed ho due insiemi non possono essere più di 2 lati totali, il grado di ogni vertice sarà ≤ 2 , inoltre se ne ha 2 uno viene da un insieme ed uno dall'altro.

Di conseguenza sul grafo generato dai lati in $M \Delta M'$ esiste \exists almeno un cammino, composto da lati presi alternativamente uno da un insieme ed uno dall'altro, non possono essere di fila altrimenti vorrebbe dire che un vertice è stato considerato due volte all'interno dello stesso matching.

Sapendo che $|M'| > |M|$ ci deve essere almeno un cammino che inizia e finisce con un lato di $M' \setminus M$ dato che sono di più e non può essere un ciclo (altrimenti sarebbero due lati dello stesso matching sullo stesso vertice).

Essendo questo cammino alternanza di lati presenti in $M \setminus M'$ e lati in $M' \setminus M$ questo vuol dire che è un cammino aumentante per M .

□

Visite di grafi

Un po' estemporaneo ma serve spiegarlo.

Con “visite di grafi” si intende **metodi sistematici per “scoprire” un grafo**. Durante una visita i **nodi** possono rientrare in **3 categorie**:

- sconosciuti, bianchi W
- conosciuti ma non visitati, grigi G , detti anche frontiera di visita
- visitati, neri B

La visita funziona iniziando **mettendo nella frontiera un solo nodo**, chiamato seed della visita.

$$F \leftarrow \{x_{seed}\}$$

Quindi in questo momento il nodo seme è grigio, tutti gli altri bianchi.

Algorithm 1 VisitaGrafo ($Graph$)

```
 $\forall x \ c(x) \leftarrow W$ 
 $F \leftarrow \{x_{seed}\}$ 
 $c_{(seed)} \leftarrow G$ 
while  $F \neq \emptyset$  do
   $x \leftarrow \text{pick}(F)$ 
  visit  $x$ 
   $c(x) \leftarrow B$ 
  for  $i \in \text{neighbor}(x)$  do
    if  $c(y) == W$  then
       $F \leftarrow F \cup \{y\}$ 
       $c(y) = G$ 
    end if
  end for
end while
```

Quando la frontiera è non-vuota, prendo un nodo dalla frontiera e visito questo nodo (qualunque cosa significhi). Coloro questo nodo di nero. Guardo tutti i vicini di questo nodo e:

- se sono neri, già visitati, oppure grigi, già conosciuti, li lascio lì
- se sono bianchi li metto nella frontiera e li coloro di grigio

Per un grafo orientato l'unica cosa che cambia è scegliere se visitare i vicini uscenti o entranti (collegati con archi uscenti o entranti).

Se il **grafo è connesso** questo **garantisce di visitare ogni vertice una volta sola**. Se non è connesso verrà visitata solo la parte connessa del seme, per continuare serve un altro seed.

In base a **come viene scelto il nodo da prendere nella frontiera cambia il tipo di visita**, ad esempio con uno stack viene una DFS, con una coda viene una BFS, ma in generale, in base a come viene implementata la scelta del nodo si ha un tipo di visita diversa.

3.5.1 Trovare un cammino aumentante

Un cammino aumentante deve **iniziare e terminare su un nodo esposto**, partendo da un grafico ed un matching.

Come **trovo nodi esposti**? faccio, partendo da ogni nodo **una BFS che alterna lati all'interno di M e lati al di fuori di M** .

Se **una BFS** ad un certo punto **trova altri nodi esposti** allora abbiamo un **cammino aumentante**, altrimenti si passa alle BFS sul prossimo nodo, se **non viene trovato nessun altro nodo esposto** in nessun nodo allora **non ci sono cammini aumentanti**.

Algorithm 2 FindAugmenting(G, M)

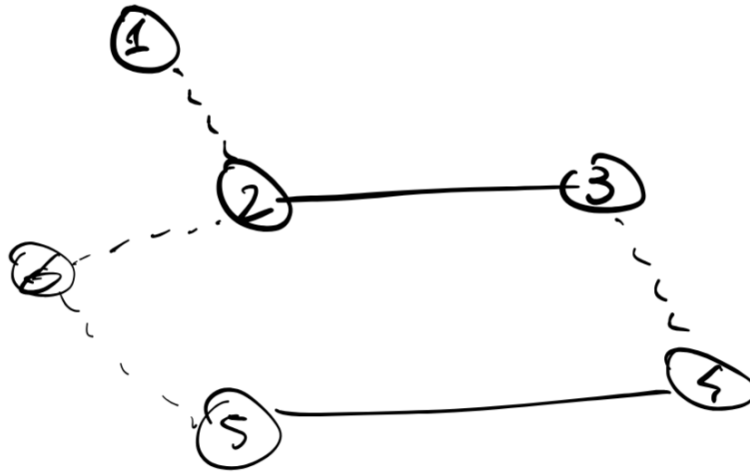
```
 $X \leftarrow$  vertici esposti in  $M$ 
for all  $x \in X$  do
    BFS( $x$ ) con alternanza di lati  $\notin M$  e  $\in M$ 
end for
```

Tempo: la BFS richiede tempo proporzionale al numero di lati $O(N \cdot M)$, una BFS per ogni vertice esposto, quindi al massimo $O(n^2)$. Quante volte dobbiamo eseguire FindAugmenting? al limite $n/2$, quindi $O(n/2 \cdot n^2) \subset O(n^3)$.

Teorema: BipartiteMaxMatching $\in \mathcal{PO}$.

Corollario: il problema di decisione PerfectMatching (il grafo ammette un matching massimale) è $\in \mathcal{P}$.

Perché la BFS non funziona su grafi non bipartiti?



Partendo da qualsiasi nodo esposto in questo grafo non si possono trovare cammini aumentanti (prova a farci la BFS alternata se non mi credi, stronzo (non ho voglia di scriverla)).

4 Tecniche greedy

4.1 Problema di LoadBalancing

Non ammette una soluzione polinomiale ottima $\in \mathcal{NPO}$.

Problema in cui si ha **un numero di task** di una certa **durata** e bisogna **dividerli su varie macchine** in modo da avere la massima **durata di ogni macchina al minimo possibile**:

- **Input:** $m > 0$ numero di macchine, $n > 0$ numero di task, $(t_i)_{i \in n}$, durate dei task, $t_i > 0$
- **Output:** $\alpha : n \rightarrow m$, assegnare ogni task ad una macchina. Il carico della macchina j è definito come

$$L_j := \sum_{i: \alpha(i)=j} t_i$$

La somma di tutte le durate dei task assegnati a quella macchina. Il carico massimo è

$$L := \max_j \{L_j\}$$

il valore più alto di una singola macchina

- **Funzione obiettivo:** L
- **Tipo:** problema di minimizzazione min, devo minimizzare il costo della macchina con il costo più alto

Teorema: LoadBalancing $\in \mathcal{NPOc}$.

Definizione di \mathcal{NPOc} : un problema di ottimizzazione $\pi \in \mathcal{NPOc}$ iff $\pi \in \mathcal{NPO}$ e $\hat{\pi} \in \mathcal{NPC}$, il problema di decisione associato è \mathcal{NPC} .

4.1.1 Algoritmo greedy per LoadBalancing

Fornisce una **soluzione approssimata**. Sostanzialmente l'algoritmo consiste nel **prendere i task** ed associarli sempre alla macchina più “scarica”, con la attuale durata complessiva delle task minore.

Chiamando A_i l'insieme dei **task associati alla macchina i** .

All'inizio sia task associati che carichi totali partono a zero. Si itera su tutti i task, si sceglie la macchina con il carico minore e si aggiunge alla sua lista di task il task j considerato (ed il carico relativo).

Ovviamente è polinomiale: $O(n)$ per cercare tra i task e $O(m)$ per cercare le macchine, quindi il totale $O(nm)$

Algorithm 3 GreedyLoadBalancing(n, m)

```
 $A_i \leftarrow \emptyset \forall i \in n$   
 $L_i \leftarrow \emptyset \forall i \in n$   
for  $j = 0, \dots, m - 1$  do  
   $\bar{i} \leftarrow \arg \min_i L_i$   
   $A_{\bar{i}} \leftarrow A_{\bar{i}} \cup \{j\}$   
   $L_{\bar{i}} \rightarrow L_{\bar{i}} + t_j$   
end for
```

Teorema: GreedyBalance è una 2-Approssimazione per Loadbalancing

Proof. Chiamando L^* il valore della funzione obiettivo della soluzione ottima.

Osservazione 1: posso dire che

$$L^* \geq \frac{1}{m} \sum_j t_j$$

Ovviamente se potessi dividere il totale di tutti i task in modo perfetto su m macchine otterrei una soluzione ideale ($1/m$ del tempo totale per ogni macchina), quindi la soluzione ottima deve essere per forza \geq di questa divisione.

Osservazione 2

$$L^* \geq \max_j t_j$$

Il compito più grande qualcuno lo deve fare, quindi il tempo massimo non può essere inferiore al tempo del task più lungo.

Sia \hat{i} tale che $L_{\hat{i}} = L$ e sia \hat{j} l'ultimo compito che le è stato assegnato.

Prima di assegnare alla macchina \hat{i} l'ultimo carico, questa doveva essere quella con il carico minore di tutte

$$\begin{aligned} L_{\hat{i}} - t_{\hat{j}} &\leq L'_i \quad \forall i \in m \\ m(L_{\hat{i}} - t_{\hat{j}}) &\leq \sum_i L_i = \sum_j t_j \\ \implies L_{\hat{i}} - t_{\hat{j}} &\leq \frac{1}{m} \sum_j t_j \leq L^* \end{aligned}$$

L'ultima parte della disequazione possiamo aggiungerla essendo la stessa cosa della osservazione 1.

$$L = L_{\hat{i}} = L_{\hat{i}} - t_{\hat{j}} + t_{\hat{j}} \geq 2 \cdot L^*$$

Per la osservazione 2, posso far vedere che $t_{\hat{j}} \leq L^*$, quindi il totale è tutto minore di $2 \cdot L^*$. Di conseguenza possiamo evincere il rapporto di approssimazione

$$\frac{L}{L^*} \leq 2$$

L'algoritmo quindi è 2-Approssimato. □

Dimostrazione di tightness: questa dimostrazione serve a stabilire che L/L^* è davvero 2, ovvero che **il bound non può diminuire ulteriormente** e quel fattore di approssimazione dipende intrinsecamente dall'algoritmo e non da qualche errore nella dimostrazione.

Teorema: per ogni $\epsilon > 0$, esiste un input per il problema LoadBalancing tale che GreedyLoadBalancing produce un output con

$$2 - \epsilon \leq \frac{L}{L^*} \leq 2$$

(ovvero schiaccio quanto voglio il bound).

Dobbiamo **costruire un input** che fa “andare male” l'algoritmo:

- numero di macchine $m > 1/\epsilon$
- numero di task $n = m(m - 1) + 1$
- ci sono $m(m-1)$ task lunghi 1 e 1 task lungo m , presentati in quest'ordine

I primi m verranno dati ciascuno ad una macchina, e si ripete $m - 1$ volte. Alla fine di questi tutte le macchine hanno lo stesso carico. Poi arriva big task che verrà messo sulla macchina meno carica (sono tutte uguali, quindi una a caso praticamente).

Di conseguenza la funzione obiettivo finale sarà:

$$L = m - 1 + m = 2m - 1$$

Non sapendo che alla fine arriva il taskone non possiamo prepararci al suo arrivo. La soluzione ottima sarebbe dare il Titanic ad una macchina e dividere gli altri $m(m - 1)$ alle altre macchine; alla fine tutte le macchine avranno lo stesso carico, quindi è sicuramente la soluzione ottima, con valore:

$$L = m$$

Il rapporto di prestazione nel primo caso diventa:

$$\frac{L}{L^*} = \frac{2m - 1}{m} = 2 - \frac{1}{m}$$

Essendo un **algoritmo greedy**, l'ordine dei task è **importante** ed in questo caso è la differenza tra un input ottimale ed il peggiore.

Comunque, magari gli input reali non sono così male, questo algoritmo su input reali potrebbe avere un tasso di approssimazione più basso, potrebbero essere input più favorevoli rispetto alle casistiche considerate in questo modo.

4.1.2 SortedGreedyBalance

Sempre sul problema di LoadBalancing, sempre stesso input e output dell'algoritmo precedente.

L'algoritmo funziona **ordinando** in ordine **decrescente i task** e poi **chiamando l'algoritmo greedy**:

Algorithm 4 SortedGreedyBalance(n, m)

Sort t_i in Non-increasing order
Run GreedyLoadBalancing(n, m)

Ma è davvero meglio?

Teorema: SortedGreedyBalance produce una $\frac{3}{2}$ -Approssimazione

Proof. **Osservazione 1:** se $n < m$, la soluzione prodotta è ottima (tutti i task a macchine distinte, meglio di così non si può).

Da qui in poi viene considerato $n > m$ (almeno una macchina deve avere 2 task assegnati).

Osservazione 2: $L^* \geq 2 \cdot t_m$. Considerando i task

$$t_0 \geq t_1 \geq \dots \geq t_m \geq t_{m+1} \geq \dots \geq t_{n-1}$$

Almeno una macchina dovrà avere 2 task, uno dai primi m task ed uno dai restanti.

Sia \hat{i} tale che $L_{\hat{i}} = L$:

- Se \hat{i} ha un task solo, la soluzione è ottima
- Se \hat{i} ha più di un compito. Sia \hat{j} l'ultimo compito (ovvero l'indice del task) assegnato a \hat{i}
- Sicuramente $\hat{j} \geq m$ in quanto i primi m task sono sicuramente dati a macchine distinte (non può esserci una macchina meno carica di una vuota)

Quindi

$$L = L_{\hat{i}} = (L_{\hat{i}} - t_{\hat{j}}) + t_{\hat{j}}$$

Di conseguenza, sapendo che

$$t_{\hat{j}} \leq t_m \leq \frac{1}{2}L^*$$

Possiamo dire

$$\begin{array}{ccc} \begin{array}{c} (L_{\hat{i}} - t_{\hat{j}}) \\ t_{\hat{j}} \end{array} & \begin{array}{c} \leq L^* \\ \leq t_{\hat{j}} \end{array} & \implies (L_{\hat{i}} - t_{\hat{j}}) + t_{\hat{j}} \leq L^* + \frac{1}{2}L^* = \frac{3}{2}L^* \end{array}$$

In conclusione

$$L = L_{\hat{i}} \leq \frac{3}{2}L^*$$

□

Ci vorrebbe una dimostrazione di tightness, ma questa analisi non è tight, si può dimostrare che l'algoritmo in realtà è una $\frac{4}{3}$ -Approssimazione (dimostrazione di Graham 1969).

Hochbaum-Shmoys 1988 hanno dimostrato che in realtà LoadBalancing appartiene a *PTAS*, ovvero si può approssimare con una precisione ϵ qualunque, e scala esponenzialmente il tempo necessario per l'approssimazione, quindi si sa anche che \notin *FPTAS*.

4.2 Problema della selezione dei centri CenterSelection

Avendo vari posti dove smistare pacchi e n soldi per costruire un numero limitato di magazzini, bisogna decidere dove piazzarli. Una volta piazzati tutti gli altri centri di smistamento si rivolgono al magazzino più vicino. Si creano delle celle di Varanoi (Check nome) attorno ai magazzini sostanzialmente. Di conseguenza c'è una distanza massima che un centro di smistamento deve coprire per arrivare al magazzino, ovvero un raggio di copertura, tra tutte le soluzioni si vuole minimizzare questo raggio.

Hai dei **punti in uno spazio metrico** e si vogliono **scegliere k punti** in questo spazio in modo che il **raggio di copertura** per tutti sia il **minore possibile**.

Spazio metrico: un'insieme con una funzione distanza (Ω, d)

$$d : \Omega \times \Omega \rightarrow \mathbb{R}^+$$

Con le seguenti proprietà:

- $d(x, y) = d(y, x)$ (simmetria)
- $d(x, y) = 0$ iff $x = y$
- $d(x, y) \leq d(x, z) + d(z, y)$ (disuguaglianza triangolare)

L'input per la selezione dei centri sarà immerso in uno spazio metrico.

Definizioni: nel mio insieme di punti, qual'è il punto scelto (magazzino) più vicino?

$$\text{dist}(x, A) := \min_{y \in A} d(x, y)$$

$$\forall s \in S \quad \delta_c(s) := \text{dist}(s, C)$$

$$\rho_c := \max_{s \in S} \delta_c(s)$$

Definizione del problema: fissiamo uno spazio metrico (Ω, d)

- **Input:** un insieme $S \subseteq \Omega$ di n punti immersi in uno spazio metrico fissato ed un **budget** $k > 0$
- **Soluzione accettabile:** è un sottoinsieme $C \subseteq S$ con $|C| \leq k$ ($\leq k$ punti nello spazio)
- **Funzione obiettivo:** ρ_c
- **Tipo:** min

Vogliamo minimizzare il massimo dei percorsi.

4.2.1 Algoritmo CenterSelectionPlus

Facciamo finta di avere in input un parametro che non avremo. Stesso input del problema, ma aggiungiamo

$$r \in \mathbb{R}^+$$

Algorithm 5 CenterSelectionPlus(S, k)

```
 $C \leftarrow \emptyset$ 
while  $S \neq \emptyset$  do
  take any  $\bar{s} \in S$ 
   $C \leftarrow C \cup \{\bar{s}\}$ 
  Remove from  $S$  all  $s$  t.c.  $d(s, \bar{s}) \leq 2r$ 
end while
if  $|C| \leq k$  then
  Output  $C$ 
else
  Output “Impossibile”
end if
```

Insieme di centri vuoto, poi finché l’insieme di punto non è vuoto:

- prende un punto in in S
- toglie da S tutti i punti con una distanza dal punto preso minore di $2r$

Se la cardinalità di C è $\leq k$ torna C , altrimenti dice che è impossibile.

Teorema 1: Se CenterSelectionPlus **emette un output**, esso è una $\frac{2r}{\rho^*}$ -**Approssimazione**.

Proof. $\forall s \in S$ con s cancellato da S e sia \bar{s} il centro scelto quando abbiamo cancellato s , di conseguenza $d(s, \bar{s}) \leq 2r$:

$$\rho_c \leq \delta_c(s) \leq d(s, \bar{s}) \leq 2r$$

Di conseguenza

$$\frac{\rho_c}{\rho^*} \leq \frac{2r}{\rho^*}$$

□

Teorema 2: se $r \geq \rho^*$, l'algoritmo **emette un output**.

Proof. Sia C^* una soluzione ottima, una distribuzione di centri che ha come raggio di copertura ρ^* .

Sia $\bar{s} \in C$. Chiamiamo $\bar{c}^* \in C^*$ un centro tale che nella soluzione ottima \bar{s} si rivolge a \bar{c}^* .

Sia X l'insieme dei punti che nella soluzione ottima C^* si rivolgono a \bar{c}^*

$$\forall s \in X \quad d(s, \bar{s}) \leq d(s, \bar{c}^*) + d(\bar{c}^*, \bar{s}) \leq 2\rho^* \leq 2r$$

\implies Quando seleziono s , cancello tutti i punti di X (come minimo), perché distano da s meno di $2r$.

\implies Elimino da S un'intera cella di Voronoi di C^* .

C^* ha al massimo k celle di Voronoi (avendo k centri al suo interno).

\implies Dopo $\leq k$ passi ho cancellato tutto.

□

Il rapporto di approssimazione abbiamo detto essere $\frac{2r}{\rho^*}$, di conseguenza l'insieme delle soluzioni ammissibili è quello dei valori $\geq \rho^*$ e la scelta di r **migliora all'avvicinarsi a ρ^*** , fino ad essere una 2-Approssimazione con $r = \rho^*$.

Con $\frac{1}{2}\rho^* < r < \rho^*$ non è ben definito che output si può ottenere (e se si ottiene un output), ma sicuramente con un $r \leq \frac{1}{2}\rho^*$ non si può avere output perché altrimenti porterebbe, secondo l'approssimazione detta sopra, ad un risultato migliore dell'ottimo.

Quindi voglio avere un r **pari a ρ^*** , ma è un dato che non conosciamo.

Al posto di scegliere un punto a caso potrei scegliere punti \bar{s} che sono almeno a distanza $> 2r$ da C , facendo terminare l'algoritmo quando non posso sceglierne altri. Equivalente all'algoritmo detto, al posto di cancellare i punti quegli stessi non potranno essere presi in considerazione ma è la stessa cosa.

4.2.2 GreedyCenterSelection

Stesso input di CenterSelection ma NON abbiamo più il parametro r .

Algorithm 6 GreedyCenterSelection(S, k)

```
if  $|S| \leq k$  then
  output  $S$ 
end if
choose any  $\bar{s} \in S$ 
 $C \leftarrow \{\bar{s}\}$ 
while  $|C| \leq k$  do
  select  $\bar{s}$  maximizing  $d(\bar{s}, C)$ 
   $C \leftarrow C \cup \{\bar{s}\}$ 
end while
Output  $C$ 
```

Se l'insieme dei punti ha cardinalità $\leq k$, torna S (basta prendere tutti i punti). Altrimenti sceglie un qualsiasi $\bar{s} \in S$ e lo sceglie come centro.

Finché la cardinalità di C è $< k$ (quindi sceglierà esattamente k punti) **sceglie il punto che massimizza la distanza** $d(\bar{s}, C)$ e lo aggiunge a C . Alla fine restituisce C .

Cancellare i punti o scegliere il punto più lontano sono equivalenti, non cambia nulla.

Teorema: GreedyCenterSelection è una **2-Approssimazione** per CenterSelection.

Proof. Supponendo che, per assurdo, l'algoritmo GreedyCenterSelection emetta una soluzione con $\rho > 2\rho^*$ (soluzione fuori da 2 volte il raggio di copertura ottimale). Questo vuol dire che esiste un elemento dell'insieme S , $\exists \bar{s} \in S$, con distanza $d(\bar{s}, C) > 2\rho^*$.

Sia \bar{s}_i l' i -esimo centro aggiunto e sia \bar{C}_i l'insieme dei centri in quel momento. Il punto va scelto in modo che massimizzi la distanza dai centri attuali:

$$d(\bar{s}_i, \bar{C}_i) \geq d(\hat{s}_i, C_i) \geq d(\hat{s}, C) > 2\rho^*$$

Quindi sarà maggiore della distanza di uno dei centri precedenti, che sarà maggiore della distanza finale, a sua volta maggiore di $2\rho^*$.

Alla fine, uno dei punti deve essere distante $> 2\rho^*$ da tutti i punti scelti.

Ma allora l'esecuzione è una delle esecuzioni possibili di CenterSelectionPlus quando $r = \rho^*$, ovvero la scelta (prendere quello a distanza massima) è compatibile quella che farebbe l'altro algoritmo (sostanzialmente sta prendendo uno dei punti lontani più di $2r$, non importa quale). GreedyCenterSelection è una particolare esecuzione di CenterSelectionPlus con $r = \rho^*$.

Ma CenterSelectionPlus con $r = \rho^*$ produce un output:

\implies termina entro k iterazioni

\implies tutti gli $s \in S$ sono tali che $d(s, C) \leq 2\rho^*$

(per i teoremi dimostrati per CenterSelectionPlus).

Ma non è vero, dato che: $d(\hat{s}, C) > 2\rho^*$. Assurdo.

□

Si potrebbe anche dimostrare che l'algoritmo è tight, ma sappiamo anche che c'è un lower bound per l'approssimazione in tempo polinomiale, quindi possiamo avere una dimostrazione di inapprossimabilità per mostrare che non si può approssimare con un fattore di approssimazione migliore di 2.

Teorema: Se $\mathcal{P} \neq \mathcal{NP}$ non esiste un algoritmo polinomiale che α -**approssimi** CenterSelection per **qualche** $\alpha < 2$ (quindi l'algoritmo greedy è polinomialmente ottimo).

Proof. La dimostrazione si basa su un problema di decisione chiamato DominatingSet:

- Input: un grafo $G = (V, E)$ ed un $k > 0$
- Output: $\exists D \subseteq V$, sapere se esiste un insieme di vertici di cardinalità al massimo k , $|D| \leq k$, tale che $\forall x \in V$ allora $\exists y \in D$ con $xy \in E$. Sostanzialmente metto sui vertici k guardie ed ogni arco deve essere collegato ad almeno una guardia

Si sa che DominatingSet $\in \mathcal{NPC}$.

Dati G, k , input di DominatorSet, dobbiamo costruire un'istanza di CenterSelection, quindi innanzitutto uno spazio metrico:

$$\Omega = V = S$$

$$d(x, y) = \begin{cases} 0 & \text{se } x = y \\ 1 & \text{se } xy \in E \\ 2 & \text{se } xy \notin E \end{cases}$$

Una cosa non ovvia da dimostrare è la disuguaglianza triangolare, quindi

$$d(x, y) \leq d(x, z) + d(z, y)$$

Quindi il valore di $d(x, y)$ può essere solo 1 o 2, mentre dall'altro lato della disuguaglianza può valere 2, 3 o 4. Non devo dirtelo io che 2 è sempre ≤ 2 .

Abbiamo la nostra istanza di CenterSelection(S, k), quindi come input insieme S e budget k .

Quanto vale il raggio di copertura?

$$\rho^*(S, k) \in \{1, 2\}$$

Tutte le distanze sono 1 o 2, quindi il raggio sarà 1 o 2.

Quando succede che la distanza è 1?

$$\rho^*(S, k) = 1$$

iff $\exists C^* \subseteq S$, con cardinalità $|C^*| \leq k$ e centri tali che $\forall s, d(s, C^*) \leq 1$

iff $\exists C^* \subseteq S$ con cardinalità $|C^*| \leq k$ tale che $\min_{c \in C^*} d(s, c) = 1$.

iff $\exists C^* \subseteq S$ con cardinalità $|C^*| \leq k$ tale che $\forall s, \exists c \in C^*$ con $sc \in E$.

iff C^* è un dominating set, in quanto è esattamente la definizione di DominatingSet.

Un algoritmo α -approssimante per CenterSelection fornirà

$$\rho^* \leq \rho(S, k) \leq \alpha \rho^*(s, k)$$

Ed i casi possibili sono:

- $\rho^* = 1$, quindi

$$1 \leq \rho(S, k) \leq \alpha$$

Se la soluzione ottima è 1 allora mi produce α al limite.

- $\rho^* = 2$

$$2 \leq \rho(S, k) \leq 2\alpha$$

Sostanzialmente, se potessi risolvere ottimamente CenterSelection in tempo polinomiale, avrei anche una risposta in tempo polinomiale per DominatingSet, dato che, con CenterSelection, se viene 1 la risposta è “sì”, mentre se viene 2 la risposta è “no”.

□

Comunque si parla di un’istanza di CenterSelection abbastanza particolare, quindi questo non significa che non esistano altre versioni ristrette di CenterSelection con migliori fattori di approssimazione. Limitare le istanze solitamente permette di approssimare meglio. Sapere come sono fatti gli input permette di fare meglio.

Funzione Armonica

Serve per dopo, la lascio qui.

Una funzione che **mappa numeri naturali positivi in reali**:

$$H : \mathbb{N}^+ \rightarrow \mathbb{R}$$

Secondo la funzione:

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

Proprietà 1: può essere approssimata per eccesso

$$H(n) \leq 1 + \int_1^n \frac{1}{x} dx$$

$$\implies H(n) \leq 1 + \ln(n)$$

Proprietà 2: possiamo dare anche un limite inferiore

$$\int_t^{t+1} \frac{1}{x} dx \leq \int_t^{t+1} \frac{1}{t} dx = \frac{1}{t}$$

$$\begin{aligned} H(n) &= \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{n} \geq \int_1^2 \frac{1}{x} dx + \int_2^3 \frac{1}{x} dx + \dots + \int_n^{n+1} \frac{1}{x} dx = \\ &= \int_1^{n+1} \frac{1}{x} dx = \ln(n+1) \end{aligned}$$

Quindi

$$H(n) \geq \ln(n+1)$$

Unendo le due cose:

$$\ln(n+1) \leq H(n) \leq 1 + \ln(n)$$

4.3 Problema di SetCover

Definizione del problema:

- **Input:** degli **insiemi** S_1, \dots, S_m e la loro unione $\bigcup_{i \in m} S_i = U$ (universo), ognuno con dei pesi w_0, \dots, w_{m-1}
- **Soluzioni ammissibili:** scegliere un certo numero di insiemi in modo che siano coperti tutti i punti. $I \subseteq m$ tale che

$$\bigcup_{i \in I} S_i = U$$

- **Funzione obiettivo:** la somma dei pesi per gli insiemi scelti

$$w = \sum_{i \in I} w_i$$

- **Tipo:** min, problema di minimizzazione

4.3.1 GreedySetCover

Algorithm 7 GreedySetCover()

```
 $R \leftarrow U$   
 $I \leftarrow \emptyset$   
while  $R \neq \emptyset$  do  
  choose  $S_i$  minimizing  $\frac{w_i}{|S_i \cap R|}$   
   $I \leftarrow I \cup \{i\}$   
   $R \leftarrow R \setminus S_i$   
end while  
output  $I$ 
```

Metto da parte l'insieme universo e prendo un insieme vuoto per i punti. Finché rimangono punti in R scelgo l'insieme che minimizza il rapporto tra costo e punti ancora da coprire presenti in quell'insieme. Una volta scelto, aggiungo l'insieme ed i punti che copre. Alla fine restituisco I .

Questo algoritmo è come se assegnasse ad ogni vertice da coprire un costo

$$\forall s \in S_i \cap R \quad c(s) := \frac{2_i}{|S_i \cap R|}$$

Lemma 1: Alla fine dell'esecuzione $w = \sum_{s \in U} c(s)$, il **costo totale** è la **somma di tutti i costi singoli pagati**.

Proof. $w = \sum_{i \in I} w_i$, ma il suo costo è ripartito in tutti gli elementi al momento in cui li ho scelti, quindi:

$$w = \sum_{i \in I} w_i = \sum_{i \in I} \sum_{s \in S_i \cap R} c(s) = \sum_{s \in U} c(s)$$

Insomma, la somma dei costi è il costo complessivo, tutto abbastanza ovvio. \square

Lemma 2: per ogni k

$$\sum_{s \in S_k} c(s) \leq H(|S_k|)w_k$$

Se sommiamo i costi attribuiti all'interno di ogni insieme, quello che si ottiene è minore uguale della funzione armonica calcolata nel punto pari alla cardinalità dell'insieme S_k , moltiplicato per il peso w_k . Si insomma, si capisce meglio dalla disequazione.

Proof. Prendendo $S_k = \{s_1, \dots, s_d\}$, con i vertici enumerati in ordine di copertura (mettendo per primi i vertici che sono stati coperti per primi, quindi a blocchi di elementi, un "blocco" per ogni volta che viene scelto un'insieme).

Consideriamo l'iterazione che copre s_j e di conseguenza il momento prima che s_j venga coperto tutti i vertici dopo sicuramente non sono coperti

$$\begin{aligned} \{s_j, s_{j+1}, \dots, s_d\} &\subseteq R \\ \implies |S_k \cap R| &\geq d - j + 1 \end{aligned}$$

Per la definizione di costo sappiamo che al passo h prima che venga scelto il vertice s_j :

$$\implies c(s_j) = \frac{w_h}{|S_h \cap R|} \leq \frac{w_k}{|S_k \cap R|} \leq \frac{w_k}{d - j + 1}$$

Ed il costo è sempre minore di quelli successivi essendo che l'algoritmo cerca di minimizzarlo.

Di conseguenza:

$$\begin{aligned} \sum_{s \in S_k} c(s) &\leq \sum_{j=1}^d c(s_j) \leq \sum_{j=1}^d \frac{w_k}{d - j + 1} = \\ &= \frac{w_k}{d} + \frac{w_k}{d-1} + \dots + \frac{w_k}{1} = w_k \left(\frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{d} \right) = \\ &= w_k H(d) = w_k H(|S_k|) \end{aligned}$$

□

Dai due lemmi segue il teorema.

Teorema: Sia $M = \max_i |S_i|$ (massima cardinalità di un insieme). Allora GreedySetCover è **una $H(M)$ -approssimazione** per SetCover (l'**approssimazione è in funzione dell'input**, al crescere di quest'ultimo cresce il fattore di approssimazione).

Proof. Sia I^* una soluzione ottima (come insieme di indici) e di conseguenza

$$w^* = \sum_{i \in I^*} w_i$$

Sarà il minimo possibile.

Per il Lemma 2

$$w_i \geq \frac{\sum_{s \in S_i} c(s)}{H(|S_i|)} \geq \frac{\sum_{s \in S_i} c(s)}{H(M)} \implies$$

H è una funzione monotona, quindi con un valore maggiore risultato maggiore, di conseguenza al denominatore diventa più piccolo.

Quindi

$$\implies w^* = \sum_{i \in I^*} w_i = \sum_{i \in I^*} \sum_{s \in S_i} c(s)$$

La somma dei pesi degli insiemi è pari alla somma del costo di tutti gli elementi di tutti gli insiemi. In questa sommatoria stiamo sommando tutti gli elementi dell'universo, quindi (per il Lemma 1)

$$\geq \sum_{s \in U} c(s) = w$$

Ed inoltre possiamo vedere che, applicando, in ordine, la prima e seconda uguaglianza vista sopra

$$\begin{aligned} w^* = \sum_{i \in I^*} w_i &\geq \sum_{i \in I^*} \frac{\sum_{s \in S_i} c(s)}{H(M)} \geq \frac{w}{H(M)} \\ \implies \frac{w}{w^*} &\leq H(M) \end{aligned}$$

□

Osservazione: M è sicuramente minore della dimensione dell'input, $M \leq |U| = n$, quindi $H(M) \leq H(n) = O(\log n)$.

Corollario: GreedySetCover è una $O(\log n)$ -**approssimazione**. L'approssimazione peggiora con $O(\log n)$. Non eccezionale come cosa ma ci accontentiamo.

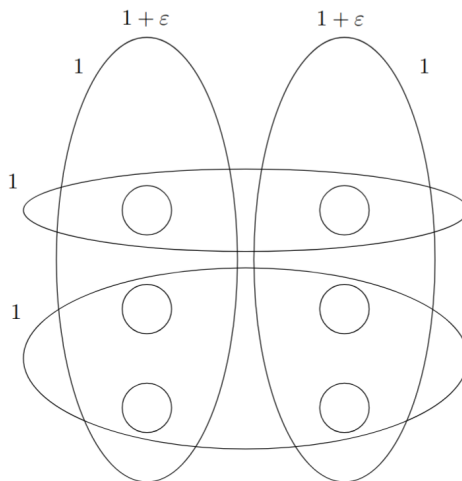
Tightness: dobbiamo costruire un **input** su cui l'algoritmo va abbastanza male da **verificare** il nostro **upper bound** per l'approssimazione.

L'input (universo) di dimensione n è composto da due insiemi di costo $1 + \epsilon$, ognuno con al suo interno $n/2$ punti.

Poi ci sono altri insiemi, tutti di costo 1:

- uno che contiene $n/4$ elementi da ognuno dei due insiemi iniziali (cardinalità totale $n/2$)
- uno che contiene $n/8$ elementi da ognuno dei due insiemi iniziali (cardinalità totale $n/4$)
- Puoi indovinare come continua

Esempio:



La soluzione ottima ovviamente è prendere i due insiemi più grandi per un costo di $2(1 + \epsilon)$.

Ma GreedySetCover su questo input sceglierà per primo l'insieme che copre $n/2$ punti con costo 1 in quanto

$$\frac{1}{n/2} < \frac{1+\epsilon}{n/2}$$

Al passaggio dopo i due insiemi grossi conterranno $n/4$ nuovi elementi e si ripeterà:

$$\frac{1}{n/4} < \frac{1+\epsilon}{n/4}$$

Quindi non sceglierà mai i due insiemi grandi ma sempre le intersezioni che coprono sempre meno elementi. Di conseguenza:

$$\begin{aligned} w &= \log n \\ w^* &= 2 + 2\epsilon \\ \implies \frac{w}{w^*} &= \frac{\log n}{2 + 2\epsilon} = \Omega(\log n) \end{aligned}$$

Quindi questo algoritmo fa le scelte peggiori possibili, ed è l'input peggiore possibile.

Fun fact: Se $P \neq NP$ non esiste un algoritmo che approssimi SetCover meglio di $(1 - O(1)) \log n$. Questo pone SetCover nell'insieme dei problemi $\log n$ -APX.

5 Tecnica di Pricing

5.1 Problema VertexCover

Definizione del problema:

- **Input:** un grafo non orientato $G = (V, E)$ con dei pesi/costi per ogni nodo $w_i \in \mathbb{R}^+$, $i \in V$
- **Soluzione ammissibile:** $X \subseteq V$ tale che $\forall e \in E$, $e \cap X \neq \emptyset$, ogni lato deve avere una delle due estremità dentro la soluzione
- **Funzione obiettivo:** somma dei pesi $\sum_{i \in X} w_i$
- **Tipo:** minimizzazione min

Pricing-based Solution: I problemi basati sul pricing funzionano basandosi sull'idea che ogni lato offre una cifra per farsi coprire, ed i vertici devono decidere cosa fare, ovvero se entrare nella soluzione o meno, ottenendo la cifra di tutti i lati che copre. I vertici sono una gang che vuole massimizzare il costo dei lati. Ad ogni lato si associa un prezzo per “farsi coprire”.

Quindi si ha un **pricing** $(P_e)_{e \in E}$ **per ogni lato**.

Un pricing $(P_e)_{e \in E}$ è **equo** se **iff** $\forall i$

$$\sum_{e \in E, i \in e} P_e \leq w_i$$

Ovvero, i **lati**, in totale, **non offriranno più** di quanto “richiede” il vertice, ovvero **del costo del vertice**. Zero su tutti i lati soddisfa banalmente questa proprietà. Verranno considerati solo di pricing equi.

Lemma 1: Se $(P_e)_{e \in E}$ è un **pricing equo** allora se **sommo tutti i prezzi** di tutti i lati **ottengo un valore minore uguale** della **soluzione ottima**

$$\sum_{e \in E} P_e \leq w^*$$

Proof. Considerando $X^* \subseteq V$ soluzione ottima e il peso relativo

$$w^* = \sum_{i \in X^*} w_i$$

Per la definizione di equità: $\forall i$

$$w_i \geq \sum_{e \in E, i \in e} P_e$$

il peso di ogni nodo è maggiore della somma dei costi dei lati su di esso.

Quindi

$$w^* = \sum_{i \in X^*} w_i \geq \sum_{i \in X^*} \sum_{e \in E, i \in e} P_e \geq \sum_{e \in E} P_e$$

La soluzione ottima è per forza maggiore della somma di tutti i lati, quindi maggiore della somma del prezzo di tutti i lati (per ammissibilità). \square

Sostanzialmente, se il pricing è equo il costo del nodo deve essere \geq della somma dei pricing dei suoi lati, quindi la soluzione ottima, in quanto composta dal peso dei nodi, non potrà essere minore della somma totale dei pricing.

Pricing stretto: Il pricing $(P_e)_{e \in E}$ è **stretto** su \hat{i} (un vertice specifico) iff

$$\sum_{e \in E, \hat{i} \in e} P_e < w_{\hat{i}}$$

In breve, “**pricing stretto**” vuol dire che **non soddisfa quanto chiede il vertice**, somma dei pricing di quel vertice minore del costo del vertice.

5.1.1 PricingVertexCover

L'input è sempre $G = (V, E)$, $w_i \in \mathbb{R}^+$, $i \in V$.

Algorithm 8 PricingVertexCover

$P_e \leftarrow 0, \forall e \in E$
while $\exists \hat{e} = \{\hat{i}, \hat{j}\}$ t.c. (P_e) è stretto su \hat{i} e \hat{j} **do**
 Sia $\hat{e} = \{\hat{i}, \hat{j}\}$ quello che minimizza

$$\Delta \leftarrow \min \left(w_{\hat{i}} - \sum_{e \in E, \hat{i} \in e} P_e, w_{\hat{j}} - \sum_{e \in E, \hat{j} \in e} P_e \right)$$

$P_{\hat{e}} \leftarrow P_{\hat{e}} + \Delta$
end while

Inizialmente attribuisce a tutti i lati un prezzo 0.

Finché **esiste un lato** tale che (P_e) è stretto su \hat{i} e su \hat{j} , $\exists \hat{e} = \{\hat{i}, \hat{j}\}$, ovvero il **pricing è stretto su entrambe le estremità**, quindi per entrambi il costo del vertice è maggiore della somma dei pricing offerti, finché vale questa condizione si **viene scelto il lato** $\hat{e} = \{\hat{i}, \hat{j}\}$ tale che **minimizza** di quanto dovrebbe **aumentare la sua offerta** per **rendere non stretto** il costo di uno dei vertici adiacenti.

Alla fine **restituisce l'insieme dei vertici** i tali per cui P_e su i **non è stretto**.

Partono tutti a zero, guardo vertice per vertice quanto stanno ricevendo dai lati adiacenti, ed all'inizio ovviamente non sarà stretto su nessun vertice. Quanto dovrebbe offrire in più ogni lato per rendere stretto il pricing su uno dei due vertici a cui è collegato? Calcolo il valore per tutti e prendo il minimo.

Ripeto sui lati non ancora stretti per entrambe le estremità, finché sono presenti lati con questa caratteristica.

Alla fine restituisce i vertici “contenti”, quindi i vertici che prendono quanto chiedono, ovvero i vertici con pricing non stretto.

Ricordando il lemma sul pricing equo.

Lemma: alla fine di PricingVertexCover il costo $w \leq 2 \sum_{e \in E} P_e$ è minore uguale al doppio della somma dei prezzi.

Proof.

$$w = \sum_{i \in \text{output}} w_i = \sum_{i: P_e \text{ non stretto su } i} w_i =$$

Emettiamo in output i vertici tali per cui il pricing su quel vertice non è stretto. Per la definizione di “stretto”, possiamo continuare dicendo che:

$$= \sum_{i \in \text{output}} \sum_{e \in E, i \in e} P_e$$

Qua stiamo sommando per tutti i vertici nell’output i prezzi di alcuni lati, ma **ognuno dei lati può comparire al massimo due volte**, se entrambe le estremità di quel lato stanno nell’output. Quindi:

$$\sum_{i \in \text{output}} \sum_{e \in E, i \in e} P_e \leq 2 \sum_{e \in E} P_e$$

□

Teorema: PricingVertexCover è una 2-Approssimazione.

Proof.

$$\frac{w}{w^*} \leq \frac{2 \sum_{e \in E} P_e}{w^*} \leq \frac{2 \sum_{e \in E} P_e}{\sum_{e \in E} P_e} = 2$$

Per il Lemma 2, w è minore di 2 volte la somma totale dei prezzi, mentre la soluzione ottima w^* è per forza maggiore della somma totale dei prezzi.

□

Non si sa se esiste un’approssimazione migliore, non si conosce una γ -approssimazione con $\gamma < 2$. Si sa che non esiste un PTAS, quindi c’è per forza un minimo di approssimazione ma non si sa per certo quale sia.

5.2 Problema DisjointPaths

Problema dei **cammini disgiunti**. L'idea è, su un grafo orientato, ci sono k **sorgenti** ed altrettante **destinazioni**, una lista di sorgenti e destinazioni. Vogliamo **collegare il maggior numero di sorgenti e destinazioni tramite un cammino**, con il vincolo di non usare ogni lato più di una volta. Verrà considerata una versione con un parametro c che determina la congestione, ovvero **ogni lato non può essere usato più di c volte**.

Definizione:

- **Input:** un grafo orientato $G = (N, A)$, la lista delle sorgenti $s_0, \dots, s_k \in N$, la lista delle destinazioni $t_0, \dots, t_k \in N$ ed un parametro $c \in \mathbb{N}^+$
- **Output:** $I \subseteq k$ e $\forall i \in I$ un cammino $\pi_i : s_i \rightarrow t_i$ tale che nessun arco di G sia usato da più di c cammini
- **Funzione obiettivo:** la cardinalità $|I|$, ovvero il numero di cammini trovati
- Tipo: massimizzazione, max

Da notare che con i grafi orientati si parla di nodi (da qui N), non vertici, e archi (da qui A), non lati.

Per l'esecuzione dell'algoritmo **assoceremo ad ogni arco un funzione lunghezza:**

$$\ell : A \rightarrow \mathbb{R}^+$$

Che può **variare nel tempo**, quindi va **specificato il momento** in cui viene considerata. Di conseguenza si avrà una lunghezza dei cammini: con $\pi = \langle x_1, \dots, x_i \rangle$

$$\ell(\pi) = \ell(x_1, x_2) + \ell(x_2, x_3) + \dots + \ell(x_{i-1}, x_i)$$

5.2.1 GreedyPaths

Input come sopra, grafo, sorgenti, destinazioni, capacità. Si aggiunge un parametro $\beta > 1$.

Algorithm 9 GreedyPaths

```
 $I \leftarrow \emptyset$ 
 $P \leftarrow \emptyset$ 
 $\ell(a) = 1, \forall a \in A$ 
while true do
  find shortest path  $\pi_i : s_i \rightarrow t_i$  with  $i \notin I$ 
  if such path does not exist then
    break
  end if
   $I \leftarrow I \cup \{i\}$ 
   $P \leftarrow P \cup \{\pi_i\}$ 
  for all  $a \in \pi_i$  do
     $\ell(a) \leftarrow \ell(a) \cdot \beta$ 
    if  $\ell(a) = \beta^c$  then
      remove  $a$ 
    end if
  end for
end while
Output  $I, P$ 
```

L'insieme I è l'insieme delle **coppie già collegate**, all'inizio vuoto, così come l'insieme di **cammini** P . La funzione **lunghezza all'inizio vale 1 per tutti** gli archi. Poi si ha un ciclo infinito in cui:

- trova il **percorso più breve** (secondo la lunghezza ℓ) **tra una coppia sorgente e destinazione non ancora collegata**
- **se tale percorso non esiste** per nessuna delle coppie rimaste, **esce dal ciclo**
- se ho trovato un percorso, **aggiungo** la coppia di **nodi collegati ad** I ed il relativo path in P
- per tutti gli archi nel cammino, **moltiplico il lunghezza per** β , rendendoli più costosi e meno appetibili per i prossimi path. Inoltre se un **arco è stato usato** c volte **viene rimosso**

Alla fine **restituisce l'insieme delle coppie collegate ed i relativi path**.

Quando si dice “trovare il cammino minimo”, bisogna usare più iterazioni di Dijkstra per trovarli tutti.

Definizione: In un certo istante dell'esecuzione un **cammino** π è definito **corto** iff la sua **lunghezza** è **minore di** β^c , quindi $\ell(\pi) < \beta^c$.

Definizione: In un certo istante, un **cammino** π è **utile** se **collega una coppia nuova** $i \notin I$. L'algoritmo considera solo cammini utili ed il più corto tra quelli esistenti.

L'algoritmo avrà più **fasi**, una in cui seleziona solo cammini corti utili, poi una in cui sceglie cammini lunghi utili, e termina quando sono finiti tutti i cammini utili.

Considerando il programma nella fase in cui sono **finiti cammini corti utili**, chiameremo $\bar{\ell}, \bar{I}, \bar{P}$ rispettivamente lunghezza, insieme di coppie collegate ed i relativi path in quel momento, ovvero subito dopo che è stato aggiunto l'ultimo cammino utile.

Alla fine dell'esecuzione chiameremo i parametri $\ell_{out}, I_{out}, P_{out}$.

Lemma 1: Sia $i \in I^* \setminus I_{out}$, una coppia **collegata dalla soluzione ottima ma non dalla soluzione dell'algoritmo**, allora

$$\bar{\ell}(\pi_i^*) \geq \beta^c$$

Ovvero il costo del path è superiore a β^c .

Proof. $i \in I^* \setminus I_{out}$, se fosse vera la condizione $\bar{\ell}(\pi_i^*) < \beta^c$ selezioneremmo quel path, ma noi stiamo guardando il momento in cui non sono più presenti cammini corti, quindi non è possibile.

Quindi se $\bar{\ell}(\pi_i^*) < \beta^c$ per costruzione dell'algoritmo verrebbe incluso, ma il presupposto è esattamente il contrario.

□

Lemma 2: se sommo la lunghezza di tutti gli archi ottengo:

$$\sum_{a \in A} \bar{\ell}(a) \leq \beta^{c+1} |\bar{I}| + m$$

La somma della lunghezza è minore uguale di β^{c+1} per il numero di cammini corti $+m$, con m il numero totale di archi.

Proof. Per induzione. All'inizio:

$$\sum_{a \in A} \ell(a) = m \leq \beta^{c+1} |\bar{I}| + m$$

Supponendo che

$$\ell_1 \xrightarrow{i, \pi_i} \ell_2 \rightarrow \dots \rightarrow \bar{\ell}$$

Quindi nel momento di transizione tra ℓ_1 e ℓ_2 vengono aggiunti i e π_i . Possiamo vedere che ℓ_2 diventa:

$$\ell_2(a) = \begin{cases} \ell_1(a) & a \notin \pi_i \\ \ell_1(a) \cdot \beta & a \in \pi_i \end{cases}$$

Quindi domandandoci di quanto sia aumentata la lunghezza degli archi possiamo vedere che:

$$\sum_{a \in A} \ell_2(a) - \sum_{a \in A} \ell_1(a) =$$

Ma possiamo considerare solo gli archi nel path π in quanto l'aumento per gli altri è 0:

$$= \sum_{a \in \pi} (\ell_2(a) - \ell_1(a)) =$$

Ma per passare da ℓ_1 a ℓ_2 abbiamo moltiplicato ℓ_1 per β , quindi

$$= \sum_{a \in \pi} (\beta \ell_1(a) - \ell_1(a)) =$$

Quindi possiamo raccogliere e dire che, per definizione di tutti i path scelti negli istanti precedenti $\bar{\ell}$

$$= (\beta - 1) \sum_{a \in A} \ell_1(a) = (\beta - 1) \ell_1(\pi) < (\beta - 1) \beta^c \leq \beta^{c+1}$$

Ad ognuno dei $|\bar{I}|$ passi (quindi per ognuno dei cammini trovati) possiamo dire che aumento la somma di al più β^{c+1} . Si aggiunge m in quanto valore iniziale di tutti gli archi, ovviamente presente.

□

Osservazione 1: Tutti i **cammini** nella **soluzione ottima** ma **non in quella dell'algoritmo** sono **maggiori di**

$$\sum_{i \in I^* \setminus I_{out}} \bar{\ell}(\pi_i^*) \geq \beta^c |I^* \setminus I_{out}|$$

Per il Lemma 1.

Osservazione 2: Nella **soluzione ottima** nessun arco è usato più di c volte, comunque è vincolata, quindi

$$\sum_{i \in I^* \setminus I_{out}} \bar{\ell}(\pi_i^*) \leq c \sum_{a \in A} \bar{\ell}(a) \leq$$

e per il lemma 2

$$\leq c (\beta^{c+1} |\bar{I}| + m)$$

Quindi, **usando le osservazioni:**

$$\beta^c |I^*| \leq \beta^c |I^* \setminus I_{out}| + \beta^c |I^* \cap I_{out}| \leq$$

La prima parte, per l'osservazione 1, è maggiorata da

$$\leq \sum_{i \in I^* \setminus I_{out}} \bar{\ell}(\pi_i^*) + \beta^c |I_{out}| \leq$$

Ma questo corrisponde all'osservazione 2:

$$\leq c (\beta^{c+1} |\bar{I}| + m) + \beta^c |I_{out}| \leq c (\beta^{c+1} |I_{out}| + m) + \beta^c |I_{out}|$$

Dividendo per β^c :

$$|I^*| \leq c \cdot \beta \cdot |I_{out}| + \frac{cm}{\beta^c} + |I_{out}| \leq$$

Aggiungendo un $|I_{out}|$ sicuramente il valore rimane maggiore, lo aggiungiamo per raccogliere:

$$\leq c \cdot \beta \cdot |I_{out}| + \frac{cm}{\beta^c} |I_{out}| + |I_{out}| = |I_{out}| \left(c\beta + \frac{cm}{\beta^c} + 1 \right)$$

Quindi

$$\frac{|I^*|}{|I_{out}|} \leq c(\beta + m\beta^{-c}) + 1$$

Ma bisogna trovare un β adeguato, e considereremo

$$\beta = m^{\frac{1}{c+1}}$$

Con questa scelta il rapporto visto può essere maggiorato da:

$$\frac{|I^*|}{|I_{out}|} \leq c \left(m^{\frac{1}{c+1}} + m^{1-\frac{1}{c+1}} \right) + 1 = 2cm^{\frac{1}{c+1}} + 1$$

Teorema: GreedyPath fornisce una $\left(2c \cdot m^{\frac{1}{c+1}} + 1\right)$ -**approssimazione** per Disjoint paths.

Esempio di valori:

c	
1	$2\sqrt{m} + 1$
2	$4\sqrt[3]{m} + 1$
3	$6\sqrt[4]{m} + 1$

Dipende da c e dal numero di archi m . Si ha c che determina un parametro lineare ed una radice che riduce il valore di m . Cresce al numero di archi, ma decresce con c .

L'approssimazione migliora nel caso di coppie ripetute.

6 Tecniche basate sull'arrotondamento

Programmazione Lineare

Serve per il prossimo algoritmo, bear with me.

La programmazione lineare (LP) può essere **descritta come problema di ottimizzazione**:

- **Input:** una **matrice di razionali** $A \in \mathbb{Q}^{m \times m}$, un **vettore** $b \in \mathbb{Q}^m$, **vettore** $c \in \mathbb{Q}^n$
- **Soluzioni ammissibili:** tutti i **vettori** $x \in \mathbb{Q}^n$ **tali che** $Ax \geq b$
- **Funzione obiettivo:** $c^T x$
- **Tipo:** minimizzazione min

Si sa che $LP \in \mathcal{PO}$ (il più famoso è algoritmo di Karmarkar).

PL Intera

Programmazione lineare intera (ILP):

- **Input:** matrice $A \in \mathbb{Q}^{m \times m}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$
- **Soluzioni ammissibili:** tutti i **vettori** $x \in \mathbb{Z}^n$ **tali che** $Ax \geq b$
- **Funzione obiettivo:** $c^T x$
- **Tipo:** minimizzazione min

Cercando una **soluzione intera al posto che razionale** il problema diventa $ILP \in \mathcal{NPO}$.

6.1 VertexCover via Rounding

Soluzione per VertexCover **basata sull'arrotondamento**.

Programmazione lineare per VertexCover: Per il problema Vertex-Cover π : Input: $G = (V, E)$, con pesi $w_i \in \mathbb{Q}^+$, $\forall i \in V$ (già definito prima, non riscritto).

Il problema di programmazione lineare intera associato a π diventa $ILP(\pi)$: $x \in \mathbb{Z}^n$, ed i vincoli sono:

$$\begin{cases} x_i + x_j \geq 1 & \forall \{i, j\} \in E \\ x_i \geq 0 & \forall i \in V \\ x_i \leq 1 & \forall i \in V \end{cases}$$

La funzione **obiettivo da minimizzare** è:

$$\sum_{i \in V} w_i x_i$$

I vincoli sostanzialmente dicono che il nodo è o 0 o 1, possono essere solo numeri interi, quindi il nodo è preso oppure no, ed il vincolo ≥ 1 stabilisce che ogni arco è preso da almeno un nodo.

Ma questo rimane un problema $\in \mathcal{NPOc}$.

La versione “rilassata” di π è uguale ma **risolto con** $x \in \mathbb{Q}^n$ (LP non ILP). Stessi vincoli, stessa funzione obiettivo.

Quindi, partiamo dal **problema di VertexCover** π , lo abbiamo trasformato in un problema di **programmazione lineare intera** $ILP(\pi)$, ci dimentichiamo della parte intera, **diventa** $LP(\pi)$, ed **arrotondiamo** a partire dalla soluzione ottima x^* di $LP(\pi)$:

$$\pi = (V, E), w_i \rightarrow \begin{matrix} ILP(\pi) \\ w_{ILP}^* \end{matrix} \rightarrow \begin{matrix} LP(\pi) \\ w_{LP}^* \end{matrix} \xrightarrow{x^*} \text{Rounding} \rightarrow \hat{x}$$
$$\hat{x} = \begin{cases} 0 & \text{se } x_i^* < \frac{1}{2} \\ 1 & \text{se } x_i^* \geq \frac{1}{2} \end{cases}$$

Chiamando w_{ILP}^* la funzione obiettivo di $ILP(\pi)$ e w_{LP}^* la funzione obiettivo di $LP(\pi)$.

Lemma 1: $w_{LP}^* \leq w_{ILP}^*$.

Proof. Il problema rilassato ha un superinsieme di soluzioni ammissibili, quindi la soluzione ottima sicuramente non può peggiorare, può solo essere migliore. □

Lemma 2: \hat{x} è una **soluzione ammissibile** di $ILP(\pi)$.

Proof. Ricordando i vincoli

$$\begin{cases} \hat{x}_i + \hat{x}_j \leq 1 & \forall \{i, j\} \in E \\ 0 \leq \hat{x}_i \leq 1 & \forall i \in V \end{cases}$$

E la definizione di \hat{x}_i (sopra).

Inoltre sappiamo che x_i^* rispetterà, in quanto soluzione ammissibile, gli stessi vincoli che sono presenti per \hat{x} .

L'unico vincolo importante da verificare è che la somma sia ≥ 1 e l'unico caso in cui può non succedere è quando entrambi sono $= 0$

$$\hat{x}_i + \hat{x}_j \not\geq 1 \implies \hat{x}_i + \hat{x}_j = 0 \implies \hat{x}_i = 0, \hat{x}_j = 0$$

Ma questo succede solo quando gli x_i^* erano entrambi $< 1/2$

$$x_i^* = 0, x_j^* = 0$$

Ma questo è impossibile in quanto vorrebbe dire che:

$$x_i^* + x_j^* < 1$$

Il che è impossibile per i vincoli del problema $LP(\pi)$, risolto in maniera ottima. □

Lemma 3: $\forall i, \hat{x}_i \leq 2x_i^*$.

Proof. Guardando i possibili casi:

$$\hat{x}_i = 0 \implies x_i^* < 1/2 \implies 0 \leq 2 \cdot 1/2$$

$$\hat{x}_i = 1 \implies x_i^* \geq 1/2 \implies 2x_i^* \geq 1 = \hat{x}_i$$

□

Lemma 4:

$$w = \sum_i w_i \hat{x}_i \leq 2 \sum_i w_i x_i^* = 2w_{LP}^*$$

Questo dice che la soluzione ottenuta dall'algoritmo di arrotondamento è \leq (per il Lemma 3) di 2 volte la soluzione ottima del problema $LP(\pi)$.

Per il Lemma 1 per il Lemma 4, rispettivamente, possiamo vedere che:

$$\frac{w}{w_{ILP}^*} \leq \frac{w}{w_{LP}^*} \leq \frac{2w_{LP}^*}{w_{LP}^*} = 2$$

Quindi l'algoritmo di arrotondamento è una **2-Approssimazione di VertexCover**.

Sperimentalmente, l'approssimazione peggiora, sia per rounding che pricing VertexCover, con grafi sparsi. In qualsiasi caso sempre abbastanza meglio della 2-Approssimazione, generare input pessimi è difficile, solitamente sono casi estremi.

7 Altri esempi

Nascita della teoria dei grafi

La città di Königsberg è attraversata dal fiume Pregel, nel quale sono presenti due isole collegate da 7 ponti in totale.

La domanda posta ad Eulero fu: si può passare da tutti i ponti e tornare all'inizio? Ovvero, esiste un cammino che passa per tutti gli archi (ponti) del grafo per poi tornare al nodo iniziale?

Le due isole diventano 2 vertici e le sponde altri 2. Al giorno d'oggi sarebbe chiamato “multigrafo”, avendo più lati incidenti sugli stessi due vertici, ma il problema rimane lo stesso.

Se volessimo formalizzare il problema nella terminologia moderna: dato un multigrafo, esiste un circuito che passa per tutti i lati? Si chiama “circuito Euleriano” (indovina perché).

Teorema di Eulero: Esiste un circuito Euleriano iff il grafo è connesso (ovviamente) e **tutti i vertici hanno grado pari**.

Come si **costruisce un circuito**, a partire da un grafo con vertici di **grado pari**?

Partendo da un qualsiasi vertice x_0 e si percorre un qualunque lato, arrivando ad x_1 , il quale avrà almeno un altro lato uscente (avendo esso grado pari), che arriva ad x_2 , e la cosa si ripete.

Nel caso si crei un ciclo prima di toccare tutti i lati vuol dire che questo ha almeno 4 lati incidenti e quindi si può continuare.

Se si torna ad x_0 prima di toccare tutti i lati, si ricomincia senza considerare i lati già visti.

Handshaking Lemma

“Lemma delle strette di mano”, se un gruppo di persone si stringono la mano, il numero di persone che stringono la mano ad un numero dispari di persone sono in quantità pari.

Teorema: In un grafo $G = (V, E)$, il numero di vertici di grado ($d()$) dispari è pari.

Proof. Sommando i gradi di tutti i vertici

$$\sum_{x \in V} d(x) = 2m$$

In questo modo ogni lato viene contato 2 volte, quindi è pari 2 volte il numero di lati m . $2m$ è ovviamente pari, ed i nodi con grado pari possono non essere considerati ai fini della parità (se tolgo o aggiungo un numero pari la parità non cambia), quindi i nodi con grado dispari devono essere pari, altrimenti la somma potrebbe non essere pari.

□

7.1 Problema del Traveling Salesman (TSP)

Avendo un insieme di città collegate da strade bidirezionali (grafo pesato) un commesso vuole passare per tutte le città una ed una sola volta, facendo un percorso di lunghezza minima per poi tornare a casa (vuole fare il giro).

Definizione:

- **Input:** un **grafo non orientato** $G = (V, E)$ con delle **lunghezze** $[\delta_e]_{e \in E}$ per ogni **lato**
- **Soluzioni ammissibili:** circuiti che passano per ogni vertice esattamente una volta (**circuito hamiltoniano**)
- **Funzione obiettivo:** la **lunghezza** del circuito
- **Tipo:** minimizzazione, min

Non è detto che il circuito hamiltoniano esista.

Questo problema è **equivalente al TSP su clique** (quindi su un grafo fully connected). Per usare un algoritmo che funziona solo su clique per grafi generici basta aggiungere i lati mancanti con un peso molto alto (più di ogni possibile cammino tra lati tendenzialmente), in modo che non vengano mai scelti, se la soluzione passa per uno dei lati “finti” nel grafo originale non c’è un circuito hamiltoniano.

Per questo da qui in poi verrà **presupposto che i grafi siano clique** $G = K_n$ (notazione per una clique di n nodi).

TSP Metrico: Un TSP Metrico ha come proprietà:

- G è una clique
- $\forall x, y, z, \delta_{xy} \leq \delta_{xz} + \delta_{zy}$ (disuguaglianza triangolare)

Ingredienti mancanti: Due concetti che serviranno per l'algoritmo vero e proprio:

- **Minimum spanning tree:** dato un grafo pesato connesso $G = (V, E)$ trovare un albero di copertura (che tocca tutti i vertici) di peso totale minimo. Risolvibile in tempo polinomiale (es. Algoritmo di Kruskal).
- **Minimum weight perfect matching:** Data una clique pesata un numero pari di vertici trovare un matching perfetto (completo) di peso minimo. Risolvibile esattamente in tempo polinomiale (Blossom algorithm)

Ora possiamo cucinare.

7.1.1 Algoritmo di Cristophides

Algoritmo per il TSP metrico:

- Input: $G = (V, E)$ clique, con pesi $[\delta_e]_{e \in E}$ che sono una metrica

Passaggi dell'algoritmo:

1. **Troviamo un minimum spanning tree T .**
2. Sia D l'insieme dei **vertici di grado dispari in T** . Per Handshaking Lemma sappiamo che D ha **cardinalità pari**.
3. Scegliamo un **minimum weight perfect match M sui nodi presenti in D** (sicuramente presente, dato che quei nodi fanno parte di una clique e sono pari).
4. Il perfect matching potrebbe riusare lati dell'albero, se il lato compare sia nel MST che nel matching bisognerà andare a considerare un multigrafo.
Sia $H = T \dot{\cup} M$ (unione tenendo le ripetizioni). Tutti i vertici di H avranno grado pari, in quanto tutti quelli che avevano grado dispari hanno ricevuto un lato bonus dal matching.
5. Avendo tutti grado pari si può usare il teorema di Eulero, quindi **troviamo un circuito euleriano π** .
6. Ma questo può passare per lo stesso vertice più di una volta, si tratta di un circuito euleriano, noi lo vogliamo hamiltoniano. Quindi **trasformo π in un circuito hamiltoniano $\tilde{\pi}$** . Per farlo bisogna “strozzare i cappi”, quando π passa per un nodo già preso, lo salto, tanto siamo in una clique, quindi al posto di fare il giro x, y, z , se ho già preso prima y , posso fare direttamente x, z .
7. Insomma, seguo il circuito euleriano ma salto i nodi che visita due volte, diventa un circuito hamiltoniano. Questo è l'output.

Lemma 1: $\delta(T) \leq \delta^*$, la somma dei pesi presenti nel MST è minore il peso del circuito hamiltoniano ottimo.

Proof. Sia π^* un TSP ottimo. Se da π^* togliamo un lato otteniamo un albero di copertura minimo (da un circuito, se tolgo un lato non ho più cicli, quindi ho un albero), quindi

$$\delta(T) \leq \delta^* - \delta_e \leq \delta^*$$

□

Lemma 2: $\delta(M) \leq \frac{1}{2}\delta^*$.

Proof. Sia π^* is TSP ottimo. Dentro questo ci saranno anche i vertici di grado dispari presenti in D che deve coprire il matching M .

Prendendo un circuito tra elementi di D (vertici di grado dispari) e chiamandolo $\bar{\pi}^*$, ma questo ovviamente

$$\bar{\pi}^* \leq \delta(\pi^*)$$

Sto saltando dei pezzi dal circuito originale, sicuramente è meno.

Dividiamo i lati di $\bar{\pi}^*$ in due insiemi M_1 e M_2 alternati. Ognuno di questi due gruppi di lati è un perfect matching su D , ma sappiamo che M è il minimum perfect matching, quindi:

$$\delta(M) \leq \delta(M_1)$$

$$\delta(M) \leq \delta(M_2)$$

Di conseguenza

$$2\delta(M) \leq \delta(M_1) + \delta(M_2) = \delta(\bar{\pi}^*) \leq \delta(\pi^*)$$

$$\implies \delta(M) \leq \frac{1}{2}\delta^*$$

□

Teorema: L'algoritmo di Cristophides fornisce una $3/2$ -**approssimazione per il TSP metrico**.

Proof.

$$\delta(\pi) = \delta(T) + \delta(M)$$

Usando i due lemmi:

$$\begin{aligned} \delta(T) + \delta(M) &\leq \delta^* + \frac{1}{2}\delta^* \\ \implies \delta(\pi) &\leq \frac{3}{2}\delta^* \end{aligned}$$

Per la triangolare (durante la trasformazione prendo solo lati in meno, mai in più):

$$\delta(\tilde{\pi}) \leq \delta(\pi) \leq \frac{3}{2}\delta^*$$

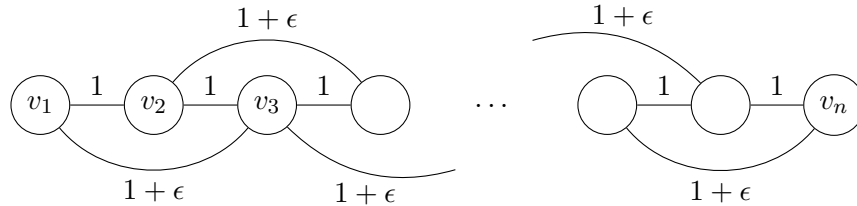
□

Tightness: l'analisi è stretta, esistono input per cui l'algoritmo restituisce $3/2\delta^*$.

Proof. $n \in \mathbb{N}$ pari e $\epsilon \in (0,1)$. Prendiamo n nodi, li colleghiamo in un cammino in fila e tutti questi lati hanno lunghezza 1.

Aggiungo lati da 1 a 3, da 2 a 4, da 3 a 5, da 5 a 7, tra tutti i pari e tra tutti i dispari, ognuno di lunghezza $1 + \epsilon$.

Sia $K_{n,\epsilon}$ la clique ottenuta aggiungendo tutti i lati mancanti, ognuno dei quali con lunghezza pari al cammino minimo tra i due vertici collegati. Ovviamente è metrica in quanto per costruzione rispetta la triangolare.



Cristophides sceglie il MST, che in questo caso saranno tutti i lati di lunghezza 1. T prende tutti gli $n - 1$ lati di lunghezza 1, quindi

$$\delta(T) = n - 1$$

Gli unici vertici di grado dispari sono il primo e l'ultimo, quindi il minimum perfect matching sarà il lato tra v_1 e v_n , il quale sarà pari al cammino minimo tra i due ovvero tutti gli $1 + \epsilon$ possibili, più l'ultimo lato:

$$\delta(M) = (1 + \epsilon)\frac{n}{2} + 1$$

Il cammino hamiltoniano coinciderà con quello euleriano quindi:

$$\delta = \delta(T) + \delta(M) = n - 1 + (1 + \epsilon)\frac{n}{2} + 1 = \frac{3}{2}n + \epsilon\frac{n}{2}$$

Ma il circuito ottimo è prendere due lati da 1 affianco a primo ed ultimo nodo, poi tutti i lati da $1 + \epsilon$, quindi il totale:

$$\delta^* = (1 + \epsilon)n + 2$$

Quindi il rapporto di approssimazione diventa:

$$\frac{\delta}{\delta^*} = \frac{\frac{3}{2}n + \epsilon\frac{n}{2}}{(1 + \epsilon)n + 2} \rightarrow \frac{3}{2}$$

Tende a $3/2$ per $n \rightarrow +\infty$ e $\epsilon \rightarrow 0$.

□

7.1.2 Inapprossimabilità del TSP

TSP Metrico è approssimabile, **quello generale no**.

Teorema: Non esiste $\alpha > 1$ tale che TSP sia α -approssimabile (se $\mathcal{P} \neq \mathcal{NP}$). Il problema sta fuori da APX.

Proof. Fact: Il problema di **decidere se un grafo ammetta un circuito hamiltoniano** è \mathcal{NP} -Completo (problema di decisione).

Supponiamo per assurdo di **avere un algoritmo α -approssimante per TSP**.

Supponendo di avere un grafo $G = (V, E)$ e voler sapere se questo grafo ha un circuito hamiltoniano. Per **trasformarlo in un'istanza di TSP**, dobbiamo **aggiungere pesi e trasformarlo in una clique**

$$G' = \left(V, \binom{V}{2}, d \right)$$

Bisogna aggiungere le **distanze**, dove

$$d(x, y) = \begin{cases} 1 & \text{se } \{x, y\} \in E \\ \lceil \alpha n \rceil + 1 & \text{altrimenti} \end{cases}$$

Si tratta di una clique, quindi per forza c'è un circuito hamiltoniano su G' , ma se G **ammetteva** un circuito hamiltoniano allora G' ne ha $\leq n$.

Se G **non ha un circuito hamiltoniano**, qualunque circuito hamiltoniano viene trovato su G' ha **lunghezza** $\geq \lceil \alpha n \rceil + 1$.

Diamo G' come **input dell'algoritmo** α -approssimante per TSP, questo **emetterà** nei casi:

- G ha un c.H.: l'algoritmo emette un output $\leq \alpha n$ (dato che si tratta di una α -approssimazione)
- G non ha un c.H.: l'algoritmo emette un output $\geq \lceil \alpha n \rceil + 1$

Quindi possiamo avere output $\leq \alpha n$ e $\geq \lceil \alpha n \rceil + 1$, ma *è possibile che questi due intervalli si sovrappongano?*

$$\begin{aligned}\alpha n &\geq \lceil \alpha n \rceil + 1 \\ \implies \alpha &\geq \frac{\lceil \alpha n \rceil + 1}{n} \geq \frac{\alpha n + 1}{n} = \alpha + \frac{1}{n} \\ \alpha &\geq \alpha + \frac{1}{n}\end{aligned}$$

Quindi sono **intervalli sempre disgiunti** e di conseguenza le soluzioni sono ben distinte.

Questo vorrebbe dire risolvere il problema di decidere se un grafo ammetta un circuito hamiltoniano efficientemente, che sappiamo non essere possibile essendo in \mathcal{NP} .

□

8 Schemi di approssimazione PTAS e FPTAS

8.1 PTAS per 2-LoadBalancing

Ricordando cos'è un PTAS: problemi approssimabile a meno di una costante desiderata (arriva al tasso di approssimazione desiderato).

LoadBalancing già visto, quello delle macchine e task, bla bla bla. 2-LoadBalancing è la stessa cosa ma ha **solo 2 macchine**.

Definizione del problema:

- **Input:** sequenza di n task t_0, \dots, t_{n-1}
- **Soluzioni ammissibili:** assegnamenti dei task su 2 macchine, una funzione $\alpha : n \rightarrow 2$
- **Funzione obiettivo:** massimo carico tra le due macchine

$$\max \left(\sum_{i:\alpha(i)=0} t_i, \sum_{i:\alpha(i)=1} t_i \right)$$

- **Tipo:** minimizzazione, min

Per l'**algoritmo PTAS**:

- **Input:** i task t_0, \dots, t_{n-1} ed un **costante di approssimazione** (circa) $\epsilon > 0$ (per avere una $1 + \epsilon$ -approssimazione)

Passaggi:

- If $\epsilon \geq 1$:
 - Assegna tutti i task alla prima macchina
 - Termina
- Altrimenti, ordina i t_i in **ordine decrescente**.
- **Fase 1:**
 - **Calcola** $k \leftarrow \lceil \frac{1}{\epsilon} - 1 \rceil$
 - Si **cerca esaustivamente** l'assegnamento ottimo dei task t_0, \dots, t_{k-1} (primi k task, 2^k assegnamenti possibili)
- **Fase 2:**
 - I **restanti task** t_k, \dots, t_{n-1} sono **assegnati** in modo **greedy** (sempre alla macchina più scarica)

L'assegnamento dei task “grossi” (i primi) viene risolto a forza bruta, il resto greedy.

Teorema: l'algoritmo è una $(1 + \epsilon)$ -approssimazione di 2-LoadBalancing.

Proof. Chiamando la **somma dei task**

$$T := \sum_{i \in n} t_i$$

Se $\epsilon \geq 1$, otteniamo una 2-Approssimazione (butto tutto su una, di sicuro non può essere peggio di 2 volte la soluzione ottima).

$$L^* \geq \frac{T}{2}, \quad L = T$$

$$\implies \frac{L}{L^*} \leq \frac{T}{T/2}$$

Nel caso in cui $\epsilon < 1$: alla **fine della prima fase** abbiamo l'**assegnamento ottimo** per i primi k task ed un carico sulle macchine 0 e 1, rispettivamente y_0 e y_1 .

Rimangono **da assegnare** i task t_k, \dots, t_{n-1} per arrivare agli **assegnamenti finali** L_0 e L_1 .

Assumiamo (senza perdita di generalità) che $L_0 \geq L_1$ (**alla fine** la macchina con il **carico maggiore è la prima**).

Due casi:

1. $L_0 = y_0$, dopo la prima fase tutti i restanti task sono stati attribuiti alla seconda macchina.
 \implies abbiamo **ottenuto l'assegnamento ottimo**, per costruzione dell'algoritmo.
2. Uno o più task di t_k, \dots, t_{n-1} sono stati assegnati alla prima macchina, sia t_h l'**ultimo assegnato** a suddetta macchina (il più piccolo). Di conseguenza

$$L_0 - t_h \leq L'_1 \leq L_1$$

(dove L'_1 è il carico della seconda macchina prima dell'assegnamento di t_h). Ovviamente prima di assegnare t_h il carico sulla prima macchina era minore del carico sulla seconda macchina

$$\implies 2L_0 - t_h \leq L_0 + L_1$$

$$T = L_0 + L_1$$

$$\implies L_0 - \frac{t_h}{2} \leq \frac{T}{2}$$

$$L_0 \leq \frac{T}{2} + \frac{t_h}{2}$$

Ricordiamo che i **primi** k **task sono** $\geq (k+1)t_h$, in quanto ordinati in ordine decrescente

$$T \geq (k+1)t_h$$

I task dopo t_h li minoriamo con 0.

$$\frac{T}{2} \geq (k+1)\frac{t_h}{2}$$

Quindi

$$\frac{L_0}{L^*} \leq \frac{L_0}{T/2} \leq \frac{\frac{T}{2} + \frac{t_h}{2}}{T/2} = 1 + \frac{t_h}{T} \leq 1 + \frac{t_h}{(k+1)t_h} = 1 + \frac{1}{(k+1)}$$

Ma **per definizione di k**

$$\begin{aligned} \frac{L_0}{L^*} &\leq 1 + \frac{1}{(k+1)} \leq 1 + \frac{1}{\frac{1}{\epsilon} - 1 + 1} = 1 + \epsilon \\ \implies \frac{L_0}{L^*} &\leq 1 + \epsilon \end{aligned}$$

□

Teorema: Il PTAS richiede tempo $O\left(n \log n + 2^{\min(\frac{1}{\epsilon}, n)}\right)$.

Proof. $n \log n$ è il **tempo** per l'**ordinamento**.

L'**esponenziale** è la **fase di assegnamento esatto**.

Abbiamo 2^k assegnamenti ma

$$2^k = 2^{\lceil \frac{1}{\epsilon} - 1 \rceil} \sim 2^{\frac{1}{\epsilon}}$$

Ci sarebbe anche la parte di assegnamento greedy ma talmente breve che non conta.

□

Quindi abbiamo un algoritmo PTAS, con un **tasso di approssimazione arbitrariamente preciso** ma con **tempo che peggiora esponenzialmente** in base a quest'ultimo.

8.2 Knapsack Problem

I pirati arrivano in una grotta con degli oggetti, ognuno di questi n oggetti ha un valore

$$v_0, \dots, v_{n-1}$$

ed i pirati vogliono portarsi a casa il maggior valore possibile, ma per portarli hanno solo uno zaino con una capacità W ed ogni oggetto ha un peso

$$w_0, \dots, w_{n-1}$$

I pirati vogliono portare a casa il maggior valore possibile con un peso che rientra nello zaino.

Definizione del problema:

- **Input:** $n > 0$ numero di oggetti, $v_i, w_i \in \mathbb{N}^+$ valori e pesi associati agli oggetti, $i < n$, capacità dello zaino $W > 0$
- **Soluzioni ammissibili:** $X \subseteq n$ tale che

$$\sum_{i \in X} w_i \leq W$$

- **Funzione obiettivo:** la somma dei valori

$$v = \sum_{i \in X} v_i$$

- **Tipo:** massimizzazione, max

Side quest: Programmazione Dinamica: vogliamo risolvere un certo problema π , nel quale sono presenti dei parametri $\pi[\bar{a}, \bar{b}]$, ma il problema si può presentare in varianti con valori diversi da \bar{a} e \bar{b} .

Voglio poter **risolvere** una certa **istanza di un problema a partire da istanze precedenti**/più semplici/con valori più semplici risolte in precedenza.

Devo quindi risolvere soluzioni più semplici per arrivare a quelle più grosse, le soluzioni dipendono da quelle prima.

Il numero di problemi da risolvere prima della soluzione che cerco è pari a tutte le possibili combinazioni di valori di parametri fino a \bar{a} e \bar{b} . Si possono pensare le diverse istanze da risolvere come se fossero in una tabella con da un lato tutti i valori possibili del parametro \bar{a} e dell'altro i possibili valori di \bar{b} .

	0	1	...	\bar{a}
0				
1				
...				
\bar{b}				

In ognuna delle celle la soluzione dell'istanza di π con i relativi valori dei parametri.

Fondamentale i parametri siano valori interi, altrimenti si può risolvere tutto in tempo polinomiale.

Side-side-quest: pseudopolinomialità: Un algoritmo è pseudopolinomiale quando il suo **tempo di esecuzione dipende dal valore numerico degli input** considerati, al posto che dalla grandezza della rappresentazione. Un valore n in binario richiede $b = \lceil \log_2 n \rceil$ bit, ma il valore considerato è $n \leq 2^b$, quindi il tempo d'esecuzione è esponenziale rispetto alla dimensione dell'input.

8.2.1 Soluzione di Programmazione Dinamica: Versione A

Partendo dalle n coppie di valori-pesi.

I parametri secondo cui iterare con la programmazione dinamica (righe e colonne della tabella) sono:

- **W , capienza dello zaino**, risolvo il problema per tutti i valori da 0 a W
- **Numero di oggetti**, considero solo i primi i oggetti, variando il valore di i tra 0 e n

Posso costruire una tabella con in **ogni riga tutte le soluzioni per un possibile valore per la capienza dello zaino W** e la colonna i contiene le soluzioni considerando solo i primi i oggetti. Quindi la cella 3,2 contiene il valore massimo ottenibile con i primi 3 oggetti e capienza 2.

Ogni cella della tabella contiene il valore massimo che i pirati si portano a casa, i.e., il risultato del problema, i.e., il **valore della funzione obiettivo**.

La prima riga, non importa quanti oggetti considero, se lo zaino ha capienza 0 non posso metterci nulla.

La prima colonna anch'essa è tutti zero, in quanto non ho oggetti, anche con un camion al posto dello zaino nella grotta non trovo nulla.

Quindi come posso trovare il **valore $v(w, i)$ per una qualsiasi cella?** Pre-supponendo di riempire la tabella da sinistra verso destra, dall'alto verso il basso. La w rappresenta la capienza considerata, la funzione v rappresenta il valore portato a casa e la i indica che abbiamo considerato i primi i oggetti, quelli che hanno da indice 0 a $i - 1$.

Guardo i **valori precedenti**, partendo da $v(w, i - 1)$:

- Se decido di non aggiungere l' i -esimo elemento il risultato non cambia, in quanto ho già trovato il risultato migliore per quello stesso spazio
- Altrimenti, se voglio aggiungere l' i -esimo elemento con valore v_i , aggiungo il suo valore, ma lo zaino deve essere abbastanza capiente

Formalmente:

$$v(w, i) = \begin{cases} v(w, i-1) & \text{se } w < w_{i-1} \\ \max(v(w, i-1), v_{i-1} + v(w - w_{i-1}, i-1)) & \text{altrimenti} \end{cases}$$

Quindi, se non c'è abbastanza spazio ($w < w_{i-1}$) per aggiungere l' i -esimo elemento il valore rimane lo stesso.

Altrimenti prendo il massimo tra il valore della cella subito a sinistra e il valore della migliore cella la cui capienza associata mi permette di aggiungere l' i -esimo elemento.

Il numero di colonne è n , ma tutti i possibili valori di W sono pseudopolinomiali \implies complessità esponenziale.

8.2.2 Soluzione di Programmazione Dinamica: Versione B

Tengo la nostra tabella per la programmazione dinamica, in riga sempre il numero di oggetti considerati, ma **in colonna metto il valore che i pirati vogliono portarsi a casa**, di conseguenza **nelle celle calcolo la minima capacità di uno zaino per ottenere quel valore**. Stesso problema, stiamo solo cambiando la rappresentazione.

Se il valore obiettivo è 0, la dimensione necessaria è sempre 0, si tratta di un pirata molto umile.

Se voglio portare a casa un valore > 0 , ma sto considerando 0 elementi risulta difficile \implies la prima colonna è tutta $+\infty$ (tranne la prima cella a zero).

Come ottengo **una cella qualunque** $w(v, i)$?

Se decido di non portare a casa anche l' i -esimo oggetto rimane tutto come $w(v, i - 1)$, altrimenti diventa $w_{i-1} + w(v - v_{i-1}, i - 1)$, quindi

$$w(v, i) = \begin{cases} \min(w(v, i - 1), w_{i-1}) & \text{se } v < v_{i-1} \\ \min(w(v, i - 1), w_{i-1} + w(v - v_{i-1}, i - 1)) & \text{altrimenti} \end{cases}$$

Se il valore cercato è minore del valore dell'oggetto che ho appena introdotto posso scegliere di prendere solo quest'ultimo.

Altrimenti prendo il minimo tra il valore subito a sinistra e la somma tra il peso dell' i -esimo oggetto e il peso della cella contenente il valore minimo necessario per arrivare al nostro obiettivo una volta sommato il valore dell'oggetto appena inserito.

Quindi, compilo la tabella, ma **avrò delle celle in cui il peso minimo è too much**, $> W$, l'**output** per il nostro problema lo **ottengo risalendo** dalla cella in basso a destra verso l'alto **finché non trovo un valore di capienza minima ammissibile** $w < W$, i.e., prendo il valore ammissibile più alto possibile.

8.2.3 Algoritmo DP Approssimato

Problema: gli algoritmi precedenti sono esatti, ma sono entrambi esponenziali nel numero di righe.

Per **approssimare il problema** possiamo “*comprimere le righe*”, ovvero **dividiamo tutti i valori per un certo numero, diminuendo** così il **numero di righe** necessarie, fino a far diventare l’algoritmo polinomiale.

Si può fare **solo con la seconda versione**, in cui le righe rappresentano la funzione obiettivo (che possiamo approssimare, andando ad ottenere un sub-ottimo), mentre con la prima versione non possiamo comprimere la capacità dello zaino, in quanto parte dei constraint del problema, rischieremmo di andare verso soluzioni non ammissibili.

Quindi, applichiamo il secondo algoritmo, ma con **scaling**.

Al problema $\pi(v_i, w_i, W)$ aggiungiamo $\epsilon \in (0, 1]$ e vogliamo una $(1 + \epsilon)$ -**approssimazione** di π

$$\theta = \frac{\epsilon \cdot v_{max}}{2n}$$

Aggiungiamo **due ulteriori versioni del problema**:

$$\bar{\pi} \left(\bar{v}_i = \left\lceil \frac{v_i}{\theta} \right\rceil, \theta, w_i, W \right)$$

Quindi arrotondo un po’ i valori per eccesso, mentre in

$$\hat{\pi} = \left(\hat{v}_i = \left\lceil \frac{v_i}{\theta} \right\rceil, w_i, W \right)$$

approssimo ma lascio compresso, quindi noi vogliamo risolvere $\hat{\pi}$ in modo esatto tramite il secondo algoritmo di programmazione dinamica visto precedentemente.

Dobbiamo capire **come sono correlate le tre soluzioni**:

$$\hat{X}^* \subseteq n, \quad \overline{X}^* \subseteq n \quad X^* \subseteq n$$

Osservazione: l'ammissibilità è la stessa per tutti e 3 i problemi, tutti i problemi hanno le stesse soluzioni ammissibili.

Osservazione: Le soluzioni

$$\hat{X}^* = \hat{X}^*$$

in quanto sono problemi uguali, i valori sono solo moltiplicati per una costante. **Queste soluzioni sono uguali.**

Lemma: Sia X una soluzione ammissibile,

$$(1 + \epsilon) \sum_{i \in \hat{X}^*} v_i \geq \sum_{i \in X} v_i$$

La soluzione ottima dei nostri problemi modificati, moltiplicata per la costante di approssimazione, è sempre \geq di qualsiasi soluzione ammissibile.

Proof. Dato che $\hat{\pi}$ ha degli arrotondamenti per eccesso, tutti i valori saranno un po' più grandi, quindi

$$\sum_{i \in X} v_i \leq \sum_{i \in X} \hat{v}_i \leq \sum_{i \in \overline{X}^*} \hat{v}_i$$

E ovviamente, la soluzione ottima per $\hat{\pi}$ ha un valore maggiore di una qualsiasi soluzione ammissibile.

Per come sono definiti

$$\hat{v}_i - v_i \leq \theta$$

Quindi

$$\sum_{i \in \hat{X}^*} \hat{v}_i \leq \sum_{i \in \overline{X}^*} (v_i + \theta) = \sum_{i \in \overline{X}^*} v_i + |\overline{X}^*| \theta \leq$$

$$\leq \sum_{i \in \bar{X}^*} v_i + n\theta = \sum_{i \in \bar{X}^*} v_i + n \frac{\epsilon \cdot v_{max}}{2n}$$

Quindi

$$\sum_{i \in X} v_i \leq \sum_{i \in \bar{X}^*} v_i + \frac{\epsilon \cdot v_{max}}{2}$$

è vera per ogni soluzione ammissibile X .

Prendo solo l'indice dell'elemento del valore massimo:

$$X = \{i_{max}\}$$

(ovviamente escludendo gli elementi con $w_i > W$).

Questa è una soluzione ammissibile, quindi

$$v_{max} \leq \sum_{i \in \bar{X}^*} v_i + \frac{\epsilon \cdot v_{max}}{2} \leq$$

Essendo $\epsilon \leq 1$

$$\begin{aligned} &\leq \sum_{i \in \bar{X}^*} v_i + \frac{v_{max}}{2} \\ \implies \frac{v_{max}}{2} &\leq \sum_{i \in \bar{X}^*} v_i \end{aligned}$$

Quindi la disuguaglianza precedente

$$\sum_{i \in X} v_i \leq \sum_{i \in \bar{X}^*} v_i + \frac{\epsilon \cdot v_{max}}{2}$$

diventa

$$\sum_{i \in X} v_i \leq (1 + \epsilon) \sum_{i \in \bar{X}^*} v_i$$

Ma ricordando che $\hat{X}^* = \bar{X}^*$

$$\sum_{i \in X} v_i \leq (1 + \epsilon) \sum_{i \in \hat{X}^*} v_i$$

□

Teorema: Il lemma vale per tutte le soluzioni, anche quella originale

$$(1 + \epsilon)\hat{v}^*\theta \geq v^*$$

Proof. Il Lemma dice che

$$(1 + \epsilon) \sum_{i \in \hat{X}^*} v_i \geq \sum_{i \in X} v_i$$

Ma la prima parte è $\leq \hat{v}^*\theta$ e la seconda è una qualsiasi soluzione ammissibile, quindi anche X^* e di conseguenza v^* .

□

Non sono un cazzo sicuro di questo teorema ma trust me bro.

Comprimo i valori, faccio i miei calcoli, ma poi alla fine devo decomprimerli per sapere il mio risultato, per questo ho il θ in giro, la discrepanza è data dall'approssimazione per eccesso effettuata.

Corollario: L'algoritmo è una $(1 + \epsilon)$ -**approssimazione**.

Complessità: devo guardare le celle della tabella, ovvero il numero di colonne per il numero di righe, quindi

$$\begin{aligned} n \cdot \sum \hat{v}_i &\leq n^2 \hat{v}_{max} = n^2 \left\lceil \frac{v_{max}}{\theta} \right\rceil = \\ &= n^2 \left\lceil \frac{2v_{max}n}{\epsilon \cdot v_{max}} \right\rceil = O\left(\frac{n^3}{\epsilon}\right) \end{aligned}$$

la disuguaglianza è data dal fatto che ognuno degli n elementi è per forza minore del massimo v_{max}

In totale quindi diventa

$$O\left(\frac{n^3}{\epsilon}\right)$$

Ovvero, polinomiale in n e in ϵ , quindi è **un FPTAS**.