Teoria dei grafi e applicazioni

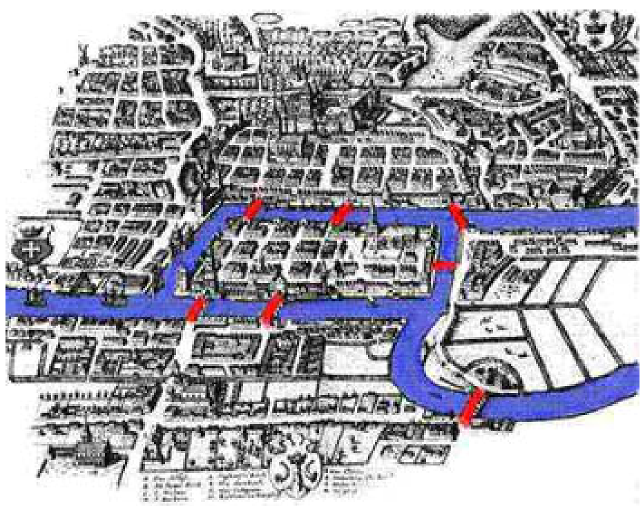
S. Bonaccorsi

Corso di *Mathematical Models for the Physical, Natural and Social Sciences*

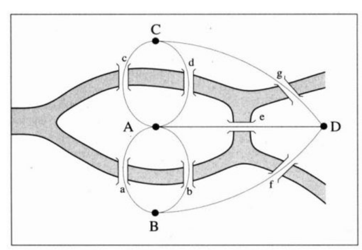
1 Teoria dei grafi e applicazioni

1.1 I sette ponti di Königsberg

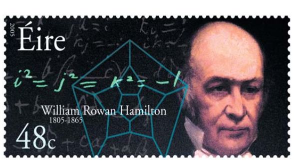
La teoria dei grafi ha una data di nascita precisa: il 1736. In quella data, il matematico svizzero Leonhard Euler risolse il problema noto come i sette ponti di Königsberg.



Ci si chiedeva se fosse possibile fare una passeggiata in città, che partisse e arrivasse allo stesso punto, in modo da attraversare tutti i ponti esattamente una volta.

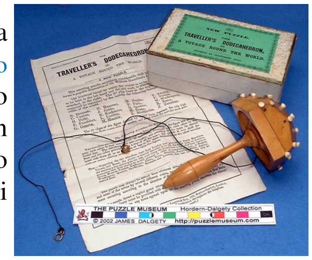


1.2 Il gioco dell'icosaedro

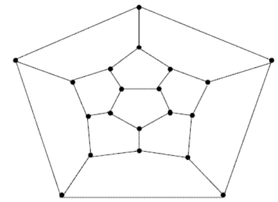


Sir William Rowan Hamilton
William Hamilton nacque a Dublino nel 1805. Fu astronomo, fisico e matematico: a lui si deve la definizione dei quaternioni.

Nel 1859 Hamilton propose un gioco che, per diversi aspetti, era legato alla sua teoria dei quaternioni: lo cedette per 25 sterline a un commerciante (quello fu l'unico denaro che ricevette direttamente per una scoperta o pubblicazione, ci dice il suo biografo). Il gioco dell'icosaedro (che in realtà si giocava su un dodecaedro) era il seguente: Hamilton aveva assegnato a ogni vertice il nome di una città e richiedeva di trovare un percorso che facesse il giro del mondo, ossia visitasse tutte le città una sola volta, per poi tornare al vertice di partenza.



Se immaginiamo che la superficie di un dodecaedro sia fatta di gomma, possiamo bucare una delle sue facce e stirarla su una superficie: allora gli spigoli formeranno la rete mostrata in figura. Su un dodecaedro con vertici non segnati sono possibili solo due circuiti di Hamilton, immagini speculari l'uno dell'altro. Se però i vertici vengono marchiati e consideriamo ogni percorso differente se passa per i 20 vertici in ordine diverso, allora i circuiti diversi diventano 30 (non contando le sequenze uguali percorse in senso inverso).



Rif.: *M. Gardner*, Enigmi e giochi matematici, 1967

1.3 Il problema del commesso viaggiatore

Una variante del gioco dell'icosaedro è il problema del commesso viaggiatore (TSP). Si tratta di trovare il percorso chiuso più breve in un grafo completo *pesato*, ossia i cui lati hanno *lunghezze* diverse.

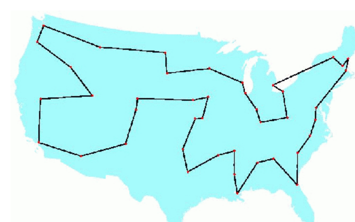
- Questo è **il** problema per eccellenza nella ottimizzazione combinatoria.

Non è un problema trovare un circuito chiuso: in un grafo completo a n nodi esistono $\frac{1}{2}(n-1)!$ circuiti chiusi. Il problema è trovare il migliore. Questo problema è contenuto nella classe dei problemi $\mathcal{NP}$-hard.

Trovare un algoritmo che possa risolvere ogni esempio di TSP sarebbe un cambio di orizzonte importante in matematica: usando questo metodo, saremmo in grado di risolvere efficientemente ogni problema computazionale per cui la risposta sia facilmente verificabile. Molti lo ritengono impossibile.

Le origini del problema sono oscure. Negli anni '20 del XX secolo era stato proposto, a Vienna, dal matematico e economista Karl Menger e riapparve, dieci anni più tardi, nel circolo matematico di Princeton.

Il primo passo importante fu compiuto da George Dantzig nel 1954 che pubblicò la descrizione di un algoritmo per risolvere il TSP e lo applicò al problema di trovare il cammino minimo tra 49 città.



Gli algoritmi sono diventati sempre più sofisticati negli anni; dalle 49 città del '54 alle quasi 25mila di 50 anni più tardi:

anno	1954	1971	1975	1980
team-leader	G. Dantzig	M. Held	P.M. Camerini	H. Crowder
# città	49	64	67	318
	1987	1987	1994	2004
	M. Grötschel	M. Padberg	D. Applegate	
	666	2392	7397	24978

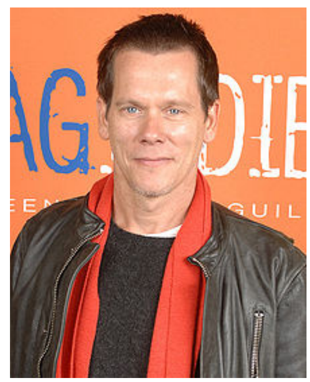
Oggi, il problema aperto è

creare un percorso ottimale tra tutte le 1.904.711 città del mondo, che comprendono anche numerose basi di ricerca in Antartide usando la distanza (in metri) calcolata sugli archi di cerchio massimo.

Il limite inferiore calcolato per questo circuito è pari a 7.512.218.268 metri ed è stato trovato nel 2007.

Il migliore circuito ad oggi è stato trovato da Keld Helsguan ed è lungo **7.515.772.212** metri, ossia al più lo 0,0474% in più del circuito ottimale.
http://www.math.uwaterloo.ca/tsp/world/images/anim10.htmlLink

1.4 Sei gradi di Kevin Bacon



In una intervista del gennaio 1994, l'attore Kevin Bacon affermò di avere lavorato con tutti gli attori di Hollywood o, almeno, con qualcuno che ci aveva recitato insieme.
Ispirato da questa notizia, il 7 aprile 1994 uscì un lunghissimo articolo su un newsgroup intitolato *Kevin Bacon is the Center of the Universe*.
Poco dopo, tre studenti di college crearono il gioco Six Degrees of Kevin Bacon, seguito a breve da un libro e da un sito web, The Oracle of Bacon, che calcola il numero di Bacon di ogni attore presente nel database IMD (Internet Movie database).

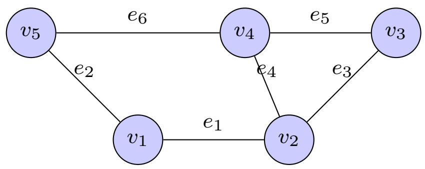
Trivia:

- Il calcolo del numero di Bacon per un attore si basa sulla ricerca di un *cammino minimo*:
  - Kevin Bacon ha numero di Bacon 0;
  - gli attori che hanno collaborato direttamente con lui hanno numero di Bacon 1;
  - se tra i collaboratori di un attore X il minimo numero di Bacon è n, allora il numero di Bacon di X è $n + 1$
- nel febbraio 2013, il numero di Bacon finito più alto era 11;
- allo stesso tempo, circa il 12% delle persone citate nel database non erano connessi con Kevin Bacon in nessun modo;

Kevin Bacon *non* è il centro dell'universo di Hollywood, se per centro dell'universo intendiamo l'attore con la distanza media minima da tutte le altre persone a lui connesse. Il numero di persone connesse con ogni attore e le loro distanze variano con l'uscita di ogni film e il centro dell'universo cambia di conseguenza.
Secondo il sito web The Oracle of Bacon, ad Aprile 2013 il centro dell'universo era Harvey Keitel.

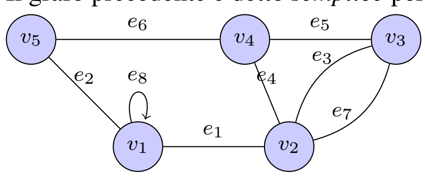
2 La teoria dei grafi

Un grafo $G = (V, E)$ è un insieme di *vertici* (o nodi) V e un insieme di *lati* E, che per semplicità supporremo finiti, quindi $|V| = n$ e $|E| = m$.



In questo esempio, $V(G) = \{v_1, \dots, v_5\}$ e $E(G) = \{e_1, \dots, e_6\}$. Il lato $e_k = (v_i, v_j)$ è detto *incidente* ai vertici v_i e v_j.

Il grafo precedente è detto *semplice* perché non contiene *self-loop* né lati multipli (come nell'esempio sotto).

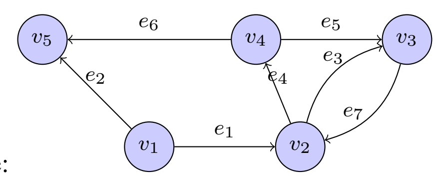


Il *grado* di un vertice $v \in V$ è il numero $d(v)$ dei lati incidenti.
Un *self-loop* conta 2 nel calcolo del grado di un vertice.
Un vertice *isolato* ha grado 0.

Teorema (*lemma delle strette di mano*)
La somma dei gradi di un grafo $G = (V, E)$ vale $\sum d(v) = 2|E| = 2m$. (banale)

Corollario: ***Il numero di vertici di grado dispari è pari.***
Questo corollario è certamente degno di nota, perché rappresenta un vincolo a cui deve sottostare ogni rete: ad esempio, non è possibile mettere insieme nove persone in modo tale che ognuna di loro ne conosca esattamente altre cinque del gruppo.

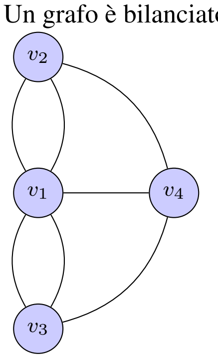
In un grafo diretto, ogni lato ha una direzione:



Diremo che nel lato $e = (v_s, v_t)$, v_s è la partenza e v_t è l'arrivo.

Ogni nodo ha un *grado entrante* $d_{in}(v)$ e un *grado uscente* $d_{out}(v)$.

Un grafo è bilanciato se $d_{in}(v) = d_{out}(v)$ per ogni nodo v.

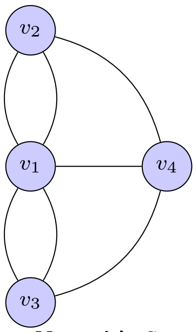


La *distanza* tra due vertici u e v è la lunghezza del minimo cammino che si deve attraversare per andare da u a v. Ad esempio, nel grafo di Königsberg, il vertice v_1 ha distanza 1 da ogni altro vertice, ma la distanza tra v_2 e v_3 è pari a 2.

In generale, diremo che due vertici sono *adiacenti* se hanno distanza 1 (se sono connessi da un lato).

Il *diametro* di un grafo è il massimo delle distanze tra due nodi del grafo. Così, il grafo di Königsberg ha diametro 2; un grafo completo ha sempre diametro 1.

2.1 Circuiti euleriani

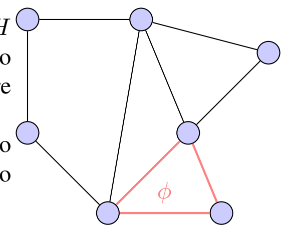


Un *circuito* in un grafo è un cammino che inizia e finisce dallo stesso vertice. Un *circuito euleriano* è un circuito che passa per ogni lato del grafo esattamente una volta.

Teorema di Eulero
Un grafo ha un circuito euleriano se e solo se ogni vertice ha grado pari.

Necessità. Se esiste un circuito Euleriano, allora questo visita tutti i nodi. Ogni volta che raggiungiamo un nodo v, lo dobbiamo lasciare *con un altro lato*, quindi il suo grado è pari.

Sufficienza. In G deve esistere un ciclo ϕ. Consideriamo il sottografo H ottenuto rimuovendo da G i lati in ϕ. I nodi in H che non sono isolati hanno ancora grado pari. Allora da ogni componente connessa di H posso ottenere un ciclo ϕ_i. Continuo finché non riduco tutto il grafo a nodi isolati.
Per costruire il circuito euleriano, partiamo da ϕ; ogni volta che incontriamo un nodo che sta in un altro ciclo ϕ_i, aggiungiamo tutto il ciclo ϕ_i al nostro cammino. Alla fine, si ottiene un circuito euleriano.



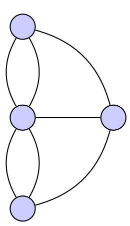
3 Il problema del postino cinese

Il problema del postino cinese è una generalizzazione del problema dei ponti di Königsberg. Abbiamo visto che un circuito esiste solo in grafi che hanno tutti i nodi di grado pari. Il problema di Königsberg non verifica questa condizione, però possiamo domandarci: è possibile costruire nuovi ponti (aggiungere nuovi lati) affinché un circuito euleriano esista?

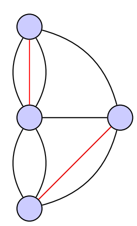
Il problema del postino cinese (così chiamato dal matematico cinese Meigu Guan che lo ha introdotto per primo) è il problema di trovare un circuito che passi su tutti i lati di un grafo *almeno* una volta e che sia di lunghezza minima.

Non è un problema solo astratto: pensate ad esempio alla pulizia delle strade, oppure alla consegna della posta: tutte le strade devono essere visitate, il servizio deve ritornare al punto di partenza e, soprattutto, si devono minimizzare i costi del servizio stesso.

Nella versione semplificata, tutti i lati hanno lunghezza unitaria. Allora il problema si riduce a trovare il numero minimo di lati da duplicare per rendere tutti i nodi di grado pari.

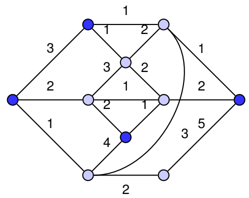


Nel grafo, tutti i vertici hanno grado dispari. Aggiungendo un lato, possiamo cambiare la parità di due vertici. Allora, per portare tutti i vertici ad avere grado pari, dobbiamo aggiungere almeno due lati.

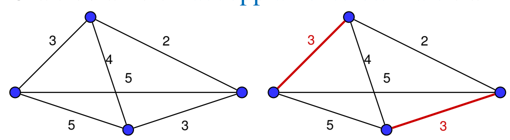


È sempre possibile risolvere questo problema? Si! Un modo, che non è forse troppo efficiente, è quello di raddoppiare il numero di lati. In questo modo, tutti i vertici hanno grado pari.

Consideriamo ora il caso in cui ogni lato abbia un peso; il problema del postino cinese richiede di trovare il circuito euleriano di lunghezza minima nel grafo.

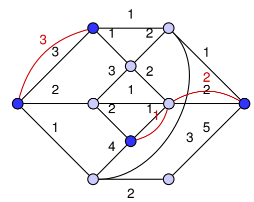


Iniziamo con l'isolare tutti i nodi di grado dispari; creiamo un sottografo in cui consideriamo i percorsi più brevi tra di loro
Ora troviamo un accoppiamento ottimale tra i nodi di questo grafo



Un accoppiamento ottimale è un sottografo per cui ogni vertice ha grado 1 e la lunghezza totale dei lati è minima

e otteniamo infine la soluzione del problema, come grafo aumentato che ammette un circuito euleriano



4 Small world

Negli anni '60 dello scorso secolo, lo psicologo sperimentale Stanley Milgram concepì alcuni fondamentali studi sul fenomeno detto *small world* nelle reti sociali dell'uomo. Milgram cercò di stimare la distanza tipica tra due nodi qualunque di una rete sociale, come ad esempio quella degli attori; il suo scopo era dimostrare che, in generale, queste distanze dovevano essere piccole.

I suoi esperimenti avevano lo scopo di mostrare che la globalizzazione rendeva il mondo più piccolo.

6 gradi di separazione
Nel suo esperimento più famoso, Milgram spedì 96 pacchi a altrettante persone che vivevano a Omaha, NE, USA, che aveva scelto a caso dall'elenco telefonico.

Ogni pacchetto conteneva un messaggio, scritto sulla carta intestata dell'Università di Harvard (di cui Milgram era membro), che invitava il destinatario a partecipare all'esperimento e le istruzioni per lo stesso.

Si chiedeva alle persone di cercare di fare avere il pacchetto a una persona (un amico di Milgram, che viveva a Boston, MA, USA). Di questa persona veniva dato solo il nome (quindi il sesso), l'indirizzo e l'occupazione (agente di borsa). Alle persone di Omaha si chiedeva di mandare il messaggio a qualcuno che conoscevano personalmente e che ritenevano (per similitudine di lavoro, oppure per vicinanza geografica, o altro) potesse essere più vicino al bersaglio.
Ogni persona che riceveva il pacchetto avrebbe dovuto inoltrarlo, con le stesse regole, semplicemente aggiungendo i dettagli del passaggio.

6 gradi di separazione
L'amico di Milgram ricevette 18 dei 96 pacchetti. Questo tasso di successo è molto alto, e quando si è tentato di ripetere l'esperimento usando la posta elettronica (Dodds, 2003) non si è assolutamente ripetuto.

Il numero medio di passaggi di mano che aveva avuto ogni pacchetto era di circa 5,9. Questo numero venne poi popolarizzato nella locuzione sei gradi di separazione.

Ancora più particolare, poi, notare come la velocità di trasmissione dell'informazione era creata da una conoscenza in massima parte locale della rete sociale, era analizzata con informazioni soprattutto locali.

Abbiamo già visto come questa idea funzioni nel caso del cinema; sappiamo che vale anche in matematica. Il sito MathSciNet della American Mathematical Society, oltre a contenere un importante catalogo di pubblicazioni di matematica e materie affini, fornisce per ogni autore un numero di Erdös che misura la distanza di un matematico dal matematico Paul Erdös, che fu uno dei fondatori della teoria dei grafi.

Con una opportuna ricerca, si ottiene:

MR Erdos Number = 4

Stefano Bonaccorsi	coauthored with	Lorenzo Zambotti	(2004)
Lorenzo Zambotti	coauthored with	Marc Yor	(2006)
Marc Yor	coauthored with	Endre Csáki	(2005)
Endre Csáki	coauthored with	Paul Erdös	(1985)

5 Flusso sui network

Un tipo di problema che viene spesso analizzato utilizzando le tecniche della teoria dei grafi è il problema dei flussi sui grafi. La rete può rappresentare un sistema stradale e il flusso rappresenta il movimento dei veicoli da un punto all'altro sulle strade. Oppure il network può essere una rete di computer, dove chi si muove sono i pacchetti dati da un computer all'altro. Oppure una rete idrica, postale o altro.

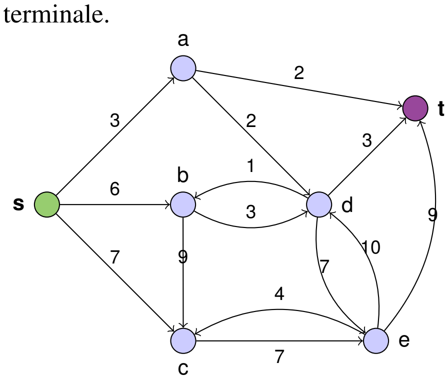
Il modello

Il network che consideriamo è dato da un grafo diretto $G = (V, E)$ con un nodo sorgente s (con grado uscente $d_{out}(s) > 0$) e un nodo terminale t con grado entrante $d_{in}(t) > 0$). Inoltre, ad ogni lato associamo una capacità positiva $c(e) > 0$.

La capacità di un lato determina quanto materiale (veicoli, pacchetti, energia, acqua) può passare lungo il lato in una data unità di tempo. Così, strade piccole o piccole condotte avranno bassa capacità, mentre condotte o strade larghe hanno alta capacità.

Per semplicità, supporremo che le capacità siano numeri interi.

In queste note, saremo interessati in un problema piuttosto semplice, quello di massimizzare il trasporto di materiale dalla sorgente al terminale. Per semplicità, supporremo che non esistano lati entranti nella sorgente e lati uscenti dal terminale.



Questo è un esempio di sistema di distribuzione dell'acqua, dove i lati sono le condotte, i vertici sono stazioni di pompaggio e la direzione di un lato indica in che direzione le nostre pompe spingono l'acqua. La capacità di ogni lato rappresenta il massimo numero di metri cubi al secondo di acqua che possiamo immettere in ogni condotta.

Il nostro scopo sarà quello di muovere la maggiore quantità possibile di acqua dalla sorgente al terminale, sottostando alle seguenti condizioni:

1. possiamo muovere l'acqua sempre nella direzione della condotta;
2. non possiamo immettere in una condotta più acqua di quanta sia la sua capacità;
3. in una stazione intermedia tra sorgente e terminale, il flusso totale entrante deve essere uguale al flusso totale uscente.

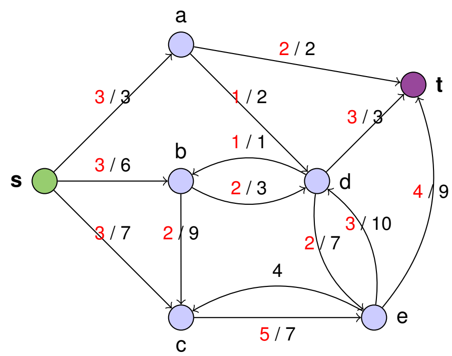
Flusso

Un flusso ammissibile f è una funzione sui lati $e = (v_i, v_j) \in E$ tale che per ogni lato e si ha $0 \le f(e) \le c(e)$;

per ogni vertice $v \in V \setminus \{s, t\}$ vale $\displaystyle\sum_{u \in V \mid (u,v) \in E} f(u, v) = \sum_{z \in V \mid (v,z) \in E} f(v, z)$.

Il flusso totale del network è allora pari al flusso uscente dalla sorgente e al flusso entrante al terminale: $F(G) = \displaystyle\sum_{u \in V \mid (s,u) \in E} f(s, u) = \sum_{z \in V \mid (z,t) \in E} f(z, t)$.

Diamo un esempio di flusso sulla rete vista precedentemente, dove indichiamo in rosso il flusso assegnato ad ogni nodo.



Questo flusso è ammissibile, in quanto ogni nodo è in pareggio e nessuna condotta è utilizzata oltre la sua capacità. Il flusso totale del network è pari a $3 + 3 + 3 = 4 + 3 + 2 = 9$. Ci domandiamo allora se è possibile fare di meglio, ossia aumentare il flusso totale del network.

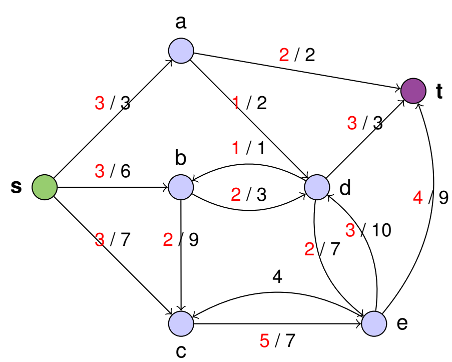
Parleremo di *cammini migliorabili* se, in un network con un flusso ammissibile,

1. considerando il network come grafo non diretto, esiste un cammino $e_1, \ldots, e_n$ che congiunge s a t tale che
2. ogni lato e_k che punta in avanti (verso t) ha un flusso associato inferiore alla capacità massima, e
3. ogni lato e_j che punta indietro (verso s) ha un flusso associato positivo.

Se abbiamo a disposizione un *cammino migliorabile* possiamo aumentare il flusso totale del network:

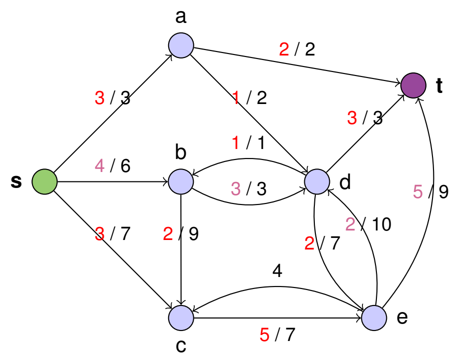
1. si aumenta di una unità il flusso su tutti i lati che puntano in avanti
2. si diminuisce di una unità il flusso su tutti i lati che puntano indietro.

Diamo un esempio di cammino migliorabile sulla rete vista precedentemente.

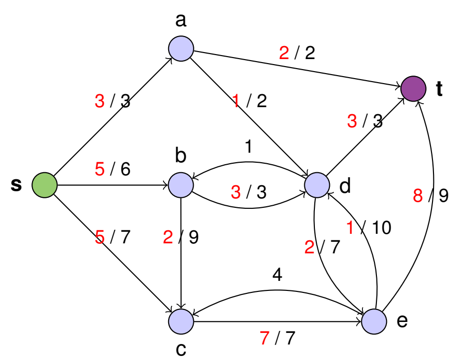


Si prenda il cammino $\overrightarrow{(s,b)}, \overrightarrow{(b,d)}, \overrightarrow{(a,d)}, \overleftarrow{(d,e)}, \overrightarrow{(e,t)}$. È un cammino migliorabile perché tutti i flussi in avanti sono aumentabili (ad esempio, il primo vale 3 / 6 e tutti i flussi indietro sono positivi (il flusso su $\overleftarrow{(d,e)}$ vale 3 / 10.

Usando l'algoritmo descritto in precedenza possiamo allora costruire un flusso ammissibile che ha flusso totale 10.



Ripetendo l'algoritmo, otteniamo ad un certo punto questo flusso, che non ammette altri cammini migliorabili.



Questo ci dà un flusso totale pari a 13. Come possiamo sapere di non poter fare di meglio? ad esempio, ci sono ancora condotte non completamente utilizzate in partenza dalla sorgente s e anche tra quelle in arrivo al terminale t: chi mi garantisce che non possa arrivare ad avere un flusso totale pari a 14?

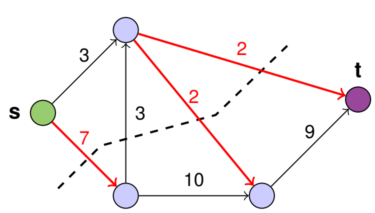
Sezioni

Una sezione di un grafo è una partizione dell'insieme dei vertici $V = P \cup \bar{P}$ tale che P contiene la sorgente s, $\bar{P}$ contiene il terminale t.

Sezioni

La capacità di una sezione è la somma delle capacità dei lati che uniscono i punti a monte con i punti a valle della sezione:

$$\kappa(P, \bar{P}) = \sum_{(u,v) \in E | u \in P, v \in \bar{P}} c(u, v).$$



La sezione in figura ha capacità pari a $7 + 2 + 2 = 11$.

È abbastanza evidente che il flusso totale di un grafo può essere calcolato anche analizzando il flusso sui lati che attraversano una sezione:

$$F(G) = \sum_{(u,v) \in E | u \in P, v \in \bar{P}} f(u, v) - \sum_{(v,u) \in E | v \in \bar{P}, u \in P} f(v, u)$$

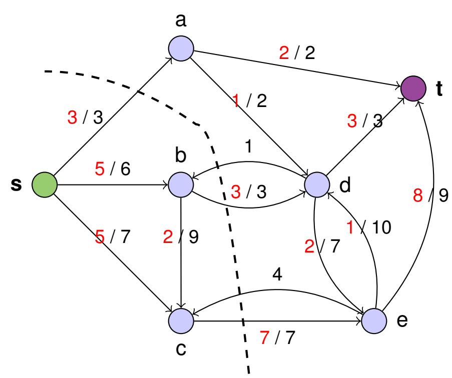
Siccome questo è vero per ogni sezione, otteniamo che:

Il flusso totale in un grafo è limitato dalla capacità di una sezione minimale, ossia $F(G) \leq \kappa(P_m, \bar{P_m}) = \min \kappa(P, \bar{P})$.

Possiamo raggiungere questo limite? In effetti, sì. Lo garantisce questo teorema (che non dimostriamo, provate a casa!).

Teorema
Il flusso totale in un grafo è uguale alla capacità di una sezione minimale, ossia $F(G) = \kappa(P_m, \bar{P_m}) = \min \kappa(P, \bar{P})$.

Ritorniamo al nostro esempio. Abbiamo costruito un flusso ammissibile sul grafo che ha flusso totale 13.



Si prenda la sezione $\{s, b, c\} \cup \{a, d, e, t\}$: la sua capacità vale 13. Allora il flusso che abbiamo costruito è massimale.