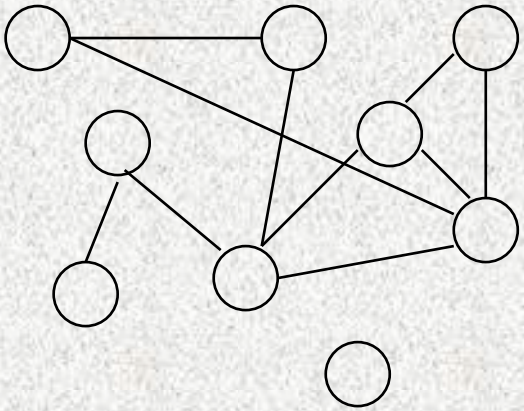
Lezione 6
Grafi

Grafi

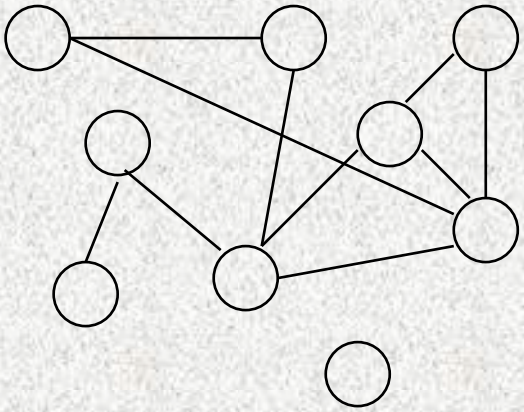
- Estensione di alberi e liste
  - Il collegamento fra due nodi in un grafo rappresenta una relazione di adiacenza o di vicinanza tra essi
- Sono importanti, perché innumerevoli situazioni possono essere modellati tramite grafi
- Un grafo è definito come una coppia G=(V,E)
  - V = vertici o nodi
  - E = archi, che collegano i nodi
- La dimensione di G è data da n+m, con n=$|V|$ e m=$|E|$
- Qual è il massimo numero di archi possibile?
- Il grafo è sparso se m=O(n), denso se m=O(n^2)

Grafi: terminologia



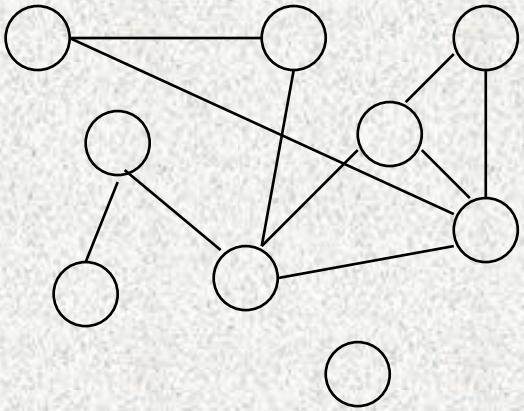
- Dato un arco (u,v), u e v sono detti adiacenti, e (u,v) è incidente a ciascuno di essi
  - Il numero di archi incidenti a un nodo è detto grado del nodo
  - Se grado=0 --> il nodo è detto isolato
- Proprietà spesso utilizzata: ∑(gradi dei nodi)=2m
  - Domanda: perché?
  - Ogni arco è incidente a due nodi
  - Quindi, ogni arco pesa 2 nella somma

Grafi: terminologia



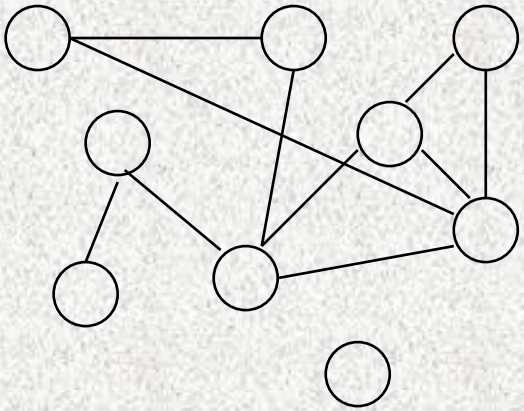
- Cammino da u a v: percorso da u a v, se esiste
  - $u=x_0=x_1=\ldots=x_k=v$
  - Lunghezza del cammino: numero di archi incontrati per andare da u a v
  - Un ciclo è un cammino per cui $x_0 = x_k$
  - Un cammino è semplice se non attraversa alcun nodo più di una volta
    - Non esiste alcun ciclo annidato al suo interno
- Un cammino minimo da u a v è il cammino di lunghezza minima tra u e v
- La distanza fra u e v è la lunghezza di un cammino minimo che li congiunge; $+\infty$ se non esiste

Grafi: terminologia



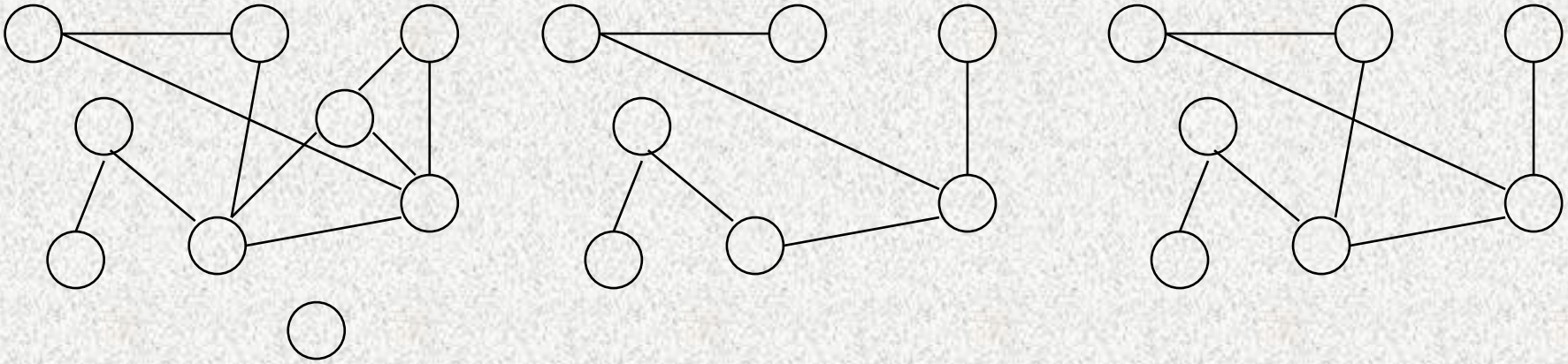
- È possibile assegnare un peso w a un arco a
  - Attraversare a costa w
- Un grafo si dice pesato se ogni suo arco ha un peso
- Il peso si utilizza nel calcolo della lunghezza dei cammini fra due nodi
  - Se il grafo non è pesato, l'attraversamento di un arco ha semplicemente peso 1

Grafi: terminologia



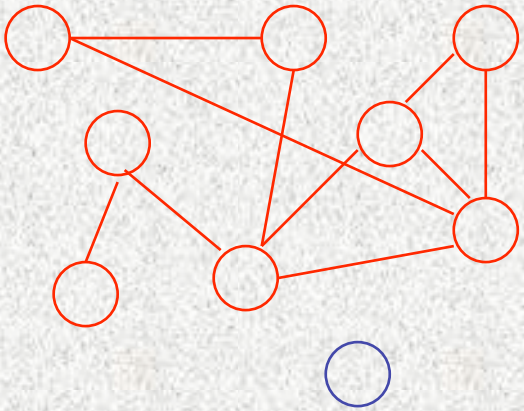
- I cammini pemettono di stabilire se i nodi sono raggiungibili
- u e v sono detti connessi se esiste un cammino fra di essi
- Un grafo in cui ogni coppia di nodi è connessa si dice connesso

Grafi: terminologia



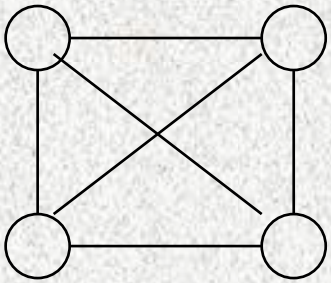
- Un sottografo di G è un grafo G’ = (V’, E’) composto da un sottinsieme dei nodi e degli archi di G
- Se vale la condizione aggiuntiva che in E’ appaiono *tutti* gli archi di E che connettono nodi di V’, allora G’ si dice grafo indotto da V’

Grafi: terminologia



- Componente connessa di G: sottografo G' connesso e massimale di G
  - Sottografo di G avente tutti nodi connessi tra loro e che non può essere esteso, in quanto non esistono ulteriori nodi in G che siano connessi ai nodi di G'
- All'interno di una componente connessa possiamo raggiungere qualunque nodo della componente stessa
  - Non possiamo passare da una componente all'altra percorrendo gli archi del grafo

Grafi: terminologia



- Un grafo completo (o cricca) è caratterizzato dall'avere tutti i suoi nodi *a due a due* adiacenti

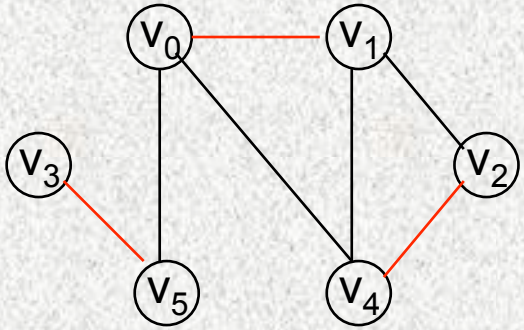
Grafi: terminologia

- I grafi visti finora sono non orientati
  - (u,v) non è distinguibile da (v,u)
- Se si vuole evitare tale simmetria, si hanno i grafi orientati
  - Es. rappresentare la viabilità stradale
- (u,v)
  - L'arco esce da u e entra in v, e si dice diretto da u a v
  - Grado in uscita (entrata): numero di archi uscenti (entranti)
  - Grado: somma del grado in uscita e in ingresso
- Tutte le definizioni viste si riadattano al caso di grafo orientato
  - Cammino
  - Ciclo

Alcuni problemi su grafi

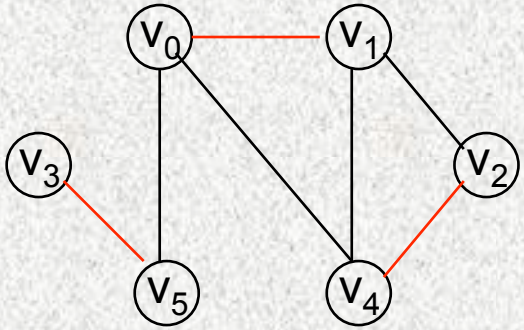
- I grafi si prestano molto bene a modellare svariati problemi
- Facciamo alcuni esempi
- Si vuole organizzare una gita in montagna per n persone
  - Supponiamo n pari
- Per il viaggio, le persone vengono accomodate in autobus a coppie
- Vincolo: due persone sono assegnate a una coppia di poltrone solo se si conoscono
- Modelliamo questa situazione con il grafo delle conoscenze

Gita in montagna



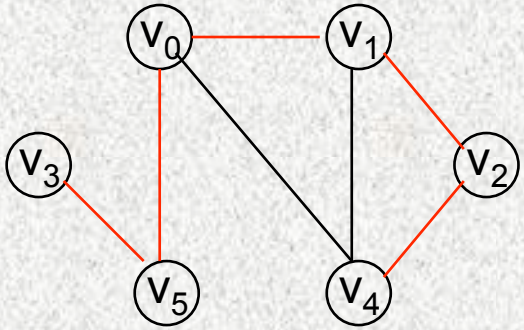
- Grafo delle conoscenze G=(V,E)
  - Nodi: turisti
  - Arco fra due nodi u e v, solo se u e v si conoscono
- Assegnamento dei posti che soddisfa il requisito della conoscenza
  - Sottoinsieme E'⊆E tale che….????
    - Tutti i nodi in V siano incidenti agli archi di E' (tutti abbiano almeno un compagno di viaggio)
    - Ogni nodo in V compaia soltanto in un arco di E' (ognuno abbia esattamente un compagno)
  - Tale sottoinsieme viene detto abbinamento (accoppiamento) perfetto….è unico????

Gita in montagna



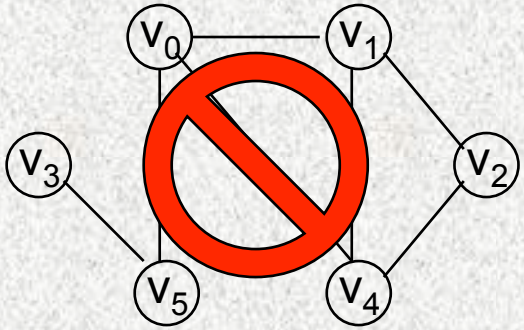
- Tale sottoinsieme viene detto abbinamento (accoppiamento) perfetto
- Problema simile che abbiamo già visto: matrimoni stabili
  - Accoppiamento su un grafo bipartito G=(V,V',E)
    - V: i clienti
    - V': le clienti

Gita in montagna



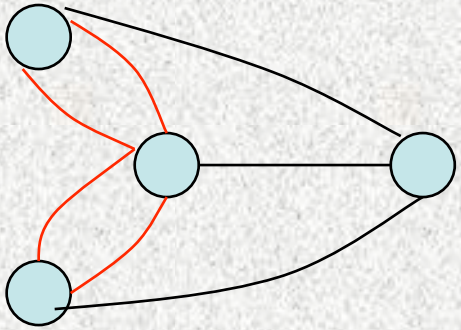
- Supponiamo ora che sia prevista un'escursione in quota
- I partecipanti devono procedere in fila indiana lungo vari tratti del percorso
- I partecipanti preferiscono che ognuno conosca sia chi lo precede che chi lo segue
- Si cerca un cammino hamiltoniano
  - Da William Rowan Hamilton, XIX secolo
  - Cammino che passi attraverso tutti i nodi una e una sola volta
  - Si tratta di trovare una permutazione dei nodi che sia un cammino
  - Non è unico….

Gita in montagna



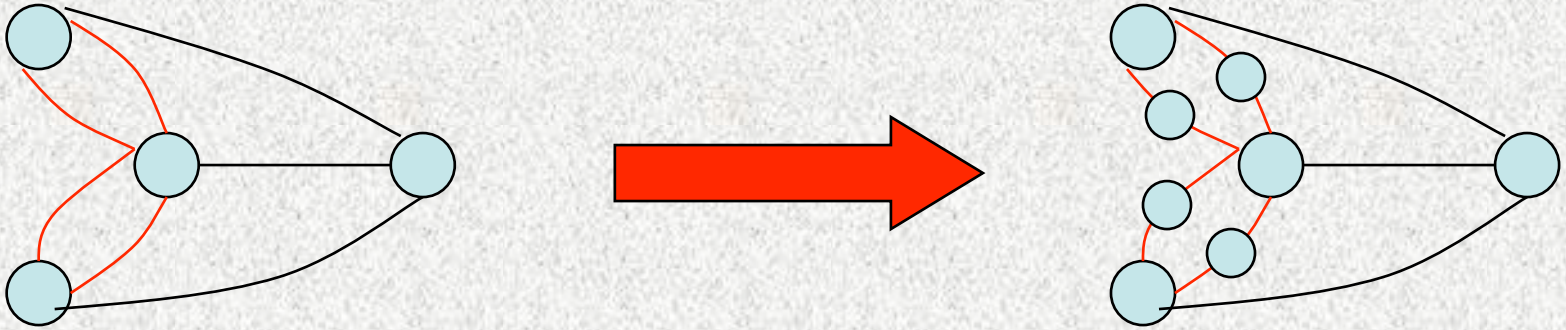
- Giunti al ristorante del rifugio montano, vogliamo disporre i partecipanti intorno a un tavolo in modo che ognuno conosca i suoi vicini
  - Vogliamo un cammino hamiltoniano dove l'ultimo nodo coincida con il primo
    - Ciclo hamiltoniano
  - Nell'esempio, tale permutazione dei nodi non esiste
  - Quindi, non esiste sempre un ciclo hamiltoniano

Gita in montagna



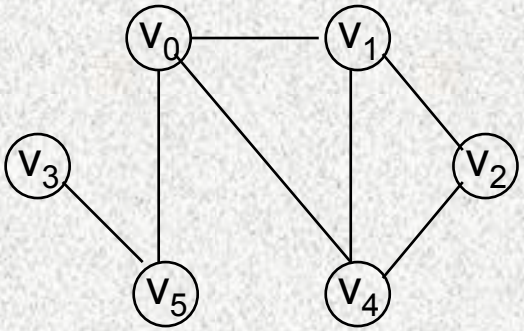
- Infine, tornati a valle, i partecipanti visitano un parco naturale ricco di torrenti che formano una serie di isole collegate da ponti di legno
- I partecipanti vogliono sapere se è possibile effettuare un giro del parco attraversando tutti i ponti una e una sola volta, tornando poi al punto di partenza della gita
- Problema studiato da Leonhard Euler, XVIII secolo, relativamente ai ponti della città di Königsberg
  - Le zone delimitate dai fiumi sono i vertici di un grafo
  - Gli archi sono i ponti da attraversare
  - Nel caso in cui più ponti colleghino due stesse zone, ne risulta un multigrafo
    - Grafo in cui la stessa coppia di vertici è collegata da più archi (in Figura, gli archi in rosso)

Gita in montagna



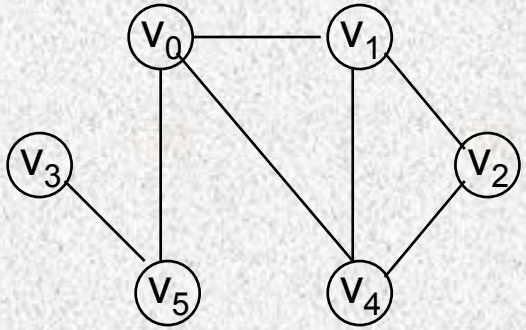
- Problema studiato da Leonhard Euler, XVIII secolo, relativamente ai ponti della città di Königsberg
  - Le zone delimitate dai fiumi sono i vertici di un grafo
  - Gli archi sono i ponti da attraversare
  - Nel caso in cui più ponti colleghino due stesse zone, ne risulta un multigrafo
    - Grafo in cui la stessa coppia di vertici è collegata da più archi (in Figura, gli archi in rosso)
  - Si trasforma in un grafo dove su ogni arco multiplo viene aggiunto un nodo
  - Ne risulta un grafo di cui vogliamo trovare il ciclo euleriano
    - Un ciclo che attraversa tutti gli archi una e una sola volta
    - Euler dimostrò che condizione necessaria e sufficiente affinché questo avvenga è che
      - G sia connesso
      - I suoi nodi abbiano grado pari

Rappresentazione dei grafi



- È un aspetto molto importante, e viene realizzato secondo due modalità
  - Matrice di adiacenza
  - Liste di adiacenza

Matrice di adiacenza

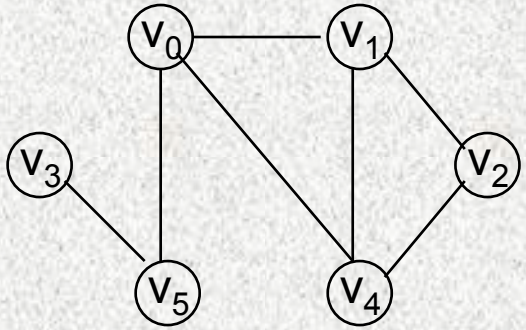


- La matrice di adiacenza A è un array bi-dimensionale di nxn elementi (con n numero di nodi), tale che
  - A[i][j]=1 $\Leftrightarrow$ (i,j) $\in$ E
    - Cioè, se esiste un arco tra i e j
- Per i grafi non orientati si ha
  - A[i][j]= A[j][i]
    - Cioè A è simmetrica

0	1	0	0	1	1
1	0	1	0	1	0
0	1	0	0	1	0
0	0	0	0	0	1
1	1	1	0	0	0
1	0	0	1	0	0

- Se il grafo è pesato, si associa ad A una matrice W dei pesi, che rappresenta in forma tabellare i pesi associati agli archi
  - Talvolta A e W vengono combinate in un'unica matrice
- La verifica dell'esistenza di un arco costa O(1)
- O(n) per scandire l'insieme dei nodi adiacenti ad un dato nodo i….cosa si scandisce????

Matrice di adiacenza

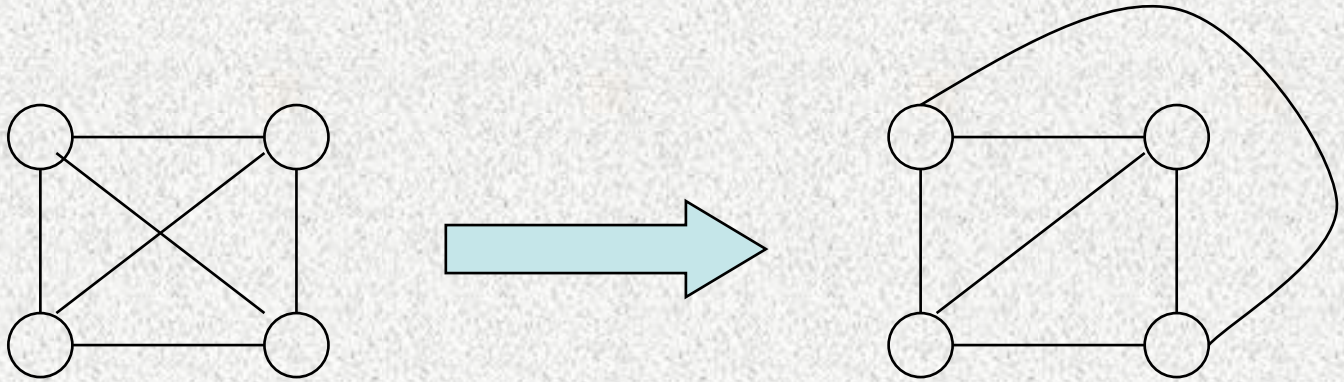


- La lista di adiacenza permette di scandire efficacemente i vertici adiacenti
- Si utilizza un array contenente n liste di adiacenza
  - Se il nodo ha grado 0 la lista è vuota
  - Altrimenti contiene tutti i nodi adiacenti
  - Lista doppia con riferimento sia all'elemento iniziale che a quello finale
- Il tempo per la scansione della lista associata a un nodo i è pari a….????
  - Grado di i
- La verifica della presenza di un arco (i,j) richiede la scansione della lista di i o j
- Ogni rappresentazione ha vantaggi e svantaggi
- Quale utilizzare dipende dall'applicazione

Spazio delle rappresentazioni

- Matrice di adiacenza
  - $\Theta(n^2)$, indipendentemente dal numero di archi presenti
  - Va bene per grafi densi
- Liste di adiacenza
  - O(n+m)
    - Infatti, ci sono n nodi
    - Inoltre, la lista per ciascun nodo è di lunghezza pari a….????
      - Al grado del nodo
    - La somma dei gradi di tutti i nodi è O(m)
  - La rappresentazione con liste può essere anche vista come una rappresentazione compatta della matrice di adiacenza

Spazio delle rappresentazioni

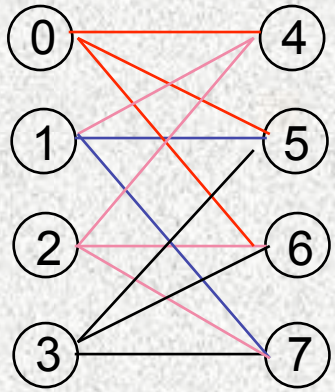


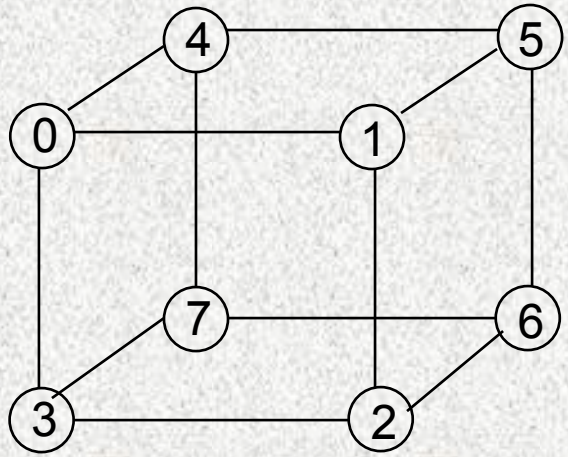
- Per grafi particolari è possibile utilizzare rappresentazioni compatte
  - Rappresentazione succinta nel caso di alberi statici
  - Rappresentazione succinta per grafi planari
    - Grafi che possono essere disegnati sul piano senza intersezioni degli archi
    - Euler dimostrò che un grafo planare di n vertici contiene O(n) archi --> è sparso
    - K_4 è planare (in Figura)

Rappresentazione dei grafi orientati

- Non presenta particolari differenze rispetto al caso di grafi non orientati
  - La matrice di adiacenza non è simmetrica
  - Vengono solitamente rappresentati gli archi uscenti nelle liste di adiacenza
    - L'arco orientato (i,j) viene memorizzato solo nella lista di adiacenza di i
    - (j,i), se esiste, viene memorizzato nella lista di adiacenza di j
    - Nei grafi non orientati, (i,j) viene memorizzato in entrambe le liste

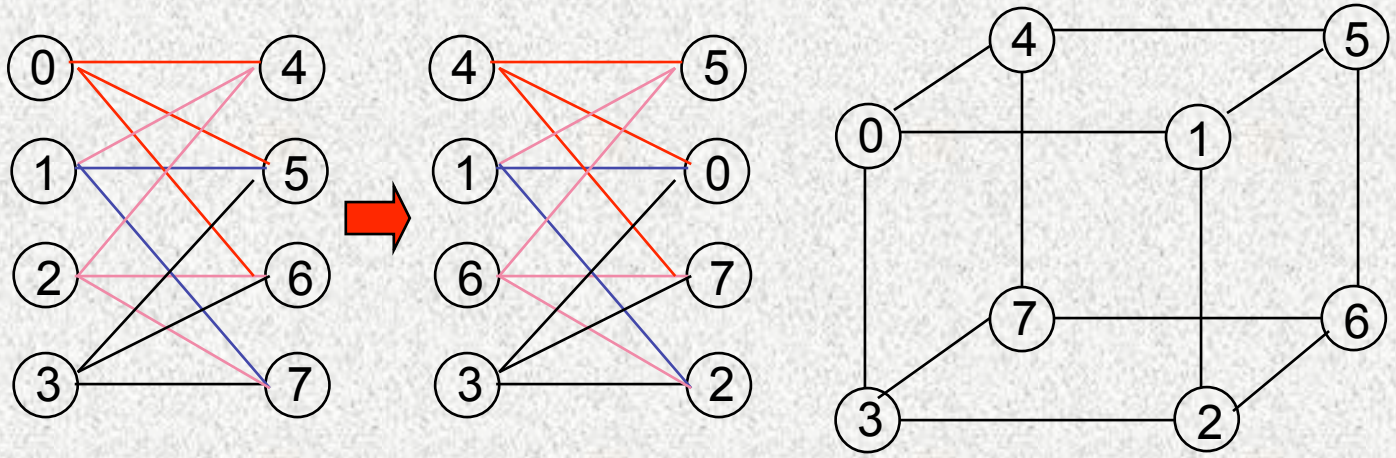
Rappresentazione dei grafi





- Mentre la rappresentazione di un grafo lo identifica in modo univoco, non è vero il viceversa
  - Cioè, data una rappresentazione, il grafo ad esso associato è univocamente determinato
  - Dato un grafo, esistono n! modi per rappresentarlo
  - Esistono, infatti, n! modi per enumerare i vertici con valori distinti in V
  - La distinzione nasce dall'artificio di enumerare arbitrariamente gli stessi vertici
    - La relazione tra i vertici resta comunque la stessa, se ignoriamo la numerazione
  - Tali grafi sono detti isomorfi

Rappresentazione dei grafi



- Come bisogna rinumerare i vertici a sx per ottenere il grafo a dx??
- Il problema di decidere in tempo polinomiale se due grafi arbitrari di n vertici sono isomorfi equivale a trovare tale numerazione, se esiste, in tempo polinomiale in n
- È uno dei problemi algoritmici fondamentali tuttora irrisolti

Chiusura transitiva

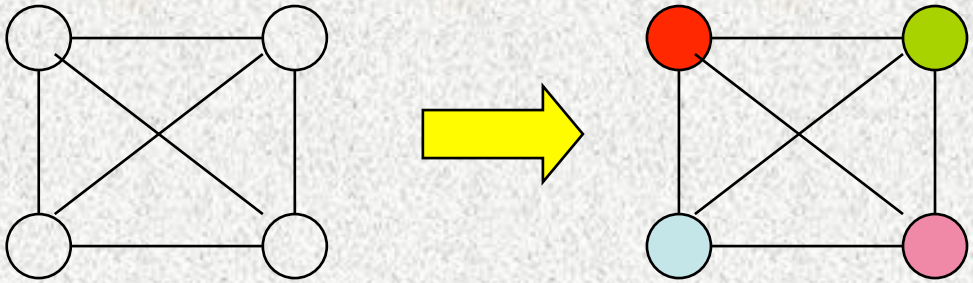
- Un grafo G*=(V,E*) è la chiusura transitiva di G=(V,E) se
  - Per ogni coppia di vertici i e j in V, vale $(i,j) \in E^* \Leftrightarrow$ esiste un cammino in G da i a j
- La matrice di adiacenza ci permette di calcolare G* facilmente
  - Sia A la matrice di adiacenza di G, dove poniamo a 1 gli elementi sulla diagonale
  - Calcoliamo $A^2=AxA$, dove la somma di elementi è l'OR, e il prodotto l'AND
  - $A^2[i][j]=1 \Leftrightarrow$ esiste un indice t tale che A[i][t] = A[t][j]=1
- Analizziamo il significato di queste matrici
  - Domanda: quando $A^2[i][j]=1$????
    - Quando esiste un t che è adiacente sia a i che a j
    - Cioè, se esiste un cammino lungo al più 2 da i a j
    - Il cammino è lungo meno di due se i=t o j=t, cioè i e j sono adiacenti. Ecco perché abbiamo posto a 1 gli elementi sulla diagonale di A

Chiusura transitiva

- Quindi la complessità per calcolare A* è $O((n^2+C_M)*\log n)$
  - C_M è la complessità per il calcolo del prodotto di due matrici

```
A* = A;
DO {
   B = A*;
   A* = BxB;
} WHILE (A* != B);
RETURN A*
```

Colorazione di grafi



- Problema: dato un grafo G, vogliamo colorare i vertici con il minimo *numero di colori* in modo tale che i vertici adiacenti siano colorati diversamente
  - Numero cromatico
- Per alcuni grafi la risposta è immediata
  - K_4?
    - 4 colori!!!!

Colorazione dei grafi

- In generale non è così semplice
- Infatti, come per il calcolo del ciclo hamiltoniano, non è noto alcun algoritmo che in tempo polinomiale determini se un grafo arbitrario ha un numero cromatico k
- Nel caso dei grafi planari è possibile trovare un 4-colorazione in tempo lineare
  - Al più 4 colori
  - Stabilire se un grafo planare ammette una 3-colorazione è nuovamente NP-completo

Modelli di reti complesse

- In molti contesti, una struttura modellabile mediante un grafo deriva come conseguenza di una serie di attività svolte in modo indipendente, senza alcun coordinamento comune
- I risultanti grafi non sono ottenuti da processi guidati da un controllo centrale
- Tali strutture sono denominate reti complesse
  - Un esempio noto di tale struttura è rappresentato dal WWW
    - Grafo orientato i cui nodi sono le pagine web, e i cui archi sono il link fra le pagine stesse
    - Tale grafo si è formato in modo *casuale*

Modelli di reti complesse

- Altro esempio: le reti sociali, usate per rappresentare persone, e un qualche tipo di relazione fra loro
  - Amicizia, conoscenza
  - Es. nel grafo *6DKB* i nodi sono attori cinematografici; esiste un arco se e solo se i nodi corrispondenti hanno recitato in uno stesso film
    - Il gioco consiste nel trovare un cammino fino all'attore Kevin Bacon
    - In alternativa, dati due attori, bisogna trovare un cammino che li collega
- Le reti biologiche
  - Modellano relazioni che sorgono in campo biologico, sia a livello di biologia molecolare e genetica che di etologia e medicina
    - Relazioni predatore-preda
    - Reti neurali, reti vascolari
- L'obiettivo dello studio di questi grafi è quello di ottenere una caratterizzazione di parametri ritenuti significativi
    - *Diametro*, *distribuzione del grado dei nodi*
  - In altre parole, si cerca di ottenere un modello matematico della struttura generale di un qualunque grafo di grandi dimensioni che rappresenta una rete complessa

Modelli di reti complesse

- Ecco alcune caratteristiche comuni a questi grafi
  - Numero di archi limitato
    - Le reti complesse tendono a essere sparse
  - Presentano raggruppamenti di nodi o aggregazioni (*cluster*)
    - Un insieme di nodi adiacenti a uno stesso nodo tende a formare una cricca
    - Le aggregazioni vengono misurate a partire dal coefficiente di aggregazione di un nodo
      - Da un nodo v, C_v è il rapporto tra il numero di archi presenti nel sottografo indotto dai nodi adiacenti a v, e il massimo numero possibile di archi tra tali nodi (cioè, quando essi formano una cricca)
      - Il coefficiente di aggregazione C del grafo è definito come la media dei C_v
    - Emerge che nelle reti complesse C assume valori elevati

Modelli di reti complesse

- Le reti complesse presentano diametro relativamente piccolo
  - Tipicamente, grafi sparsi e con alto coefficiente di aggregazione presentano un diametro elevato
  - La tendenza nelle reti complesse è opposta
  - Es.: in 6DKB la distanza media fra due nodi è 3,65 (dati del 1997)
  - Es.: Il grafo *piccolo mondo* ha diametro medio pari a 6 (chiamato anche grado di separazione)
    - Esperimento svolto negli anni ‘60
    - Per quante persone deve passare una lettera che parte dal Nebraska per arrivare nel Massachusetts, passando solo da persone che si conoscono??
      - Un mittente che riceveva la lettera poteva spedirla a un conoscente di sua scelta

Modelli di reti complesse

- Presentano una grande varietà nella distribuzione dei gradi dei nodi
  - Contengono un numero significativo di nodi per ogni possibile valore del grado, all'interno di un intervallo ampio di tali valori
- Nel testo, vengono descritti tre modelli classici di grafi casuali
  - Sono utilizzati per descrivere e generare grafi aventi caratteristiche quanto più possibile in accordo, da un punto di vista statistico, con le proprietà fondamentali delle reti complesse appena descritte
  - Semplici algoritmi per generare tali grafi in modo efficiente
  - I modelli vengono analizzati dal punto di vista della distribuzione statistica del coefficiente di aggregazione, del diametro e dei gradi dei nodi

Motori di ricerca e classificazione

- Il Web ha permesso la disponibilità di numerosi documenti
- Sono inutili se le informazioni non possono essere accedute in maniera intelligente e mirata
  - Troppa informazione non va bene se non è gestita opportunamente
- Il lavoro dei motori di ricerca è quello di gestire le richieste da parte degli utenti di documenti che contengono determinati termini (query)
- È chiaro che, data la dimensione del Web, la semplice restituzione di un elenco di tutti i documenti disponibili che contengono il termine di ricerca non è proponibile
  - Sarebbe lunghissimo!!!!

Motori di ricerca e classificazione

- Bisogna in qualche modo filtrare l'elenco, per ottenere un sottoinsieme S tale che
  - S contiene relativamente pochi documenti
  - I documenti in S sono rilevanti per l'utente
  - Molti dei documenti in S provengono da fonti autorevoli
- Per ottenere S, è necessario dunque classificare i documenti in base a un valore di significatività
  - Rango (rank) assegnato a ciascun documento
  - I documenti vengono poi restituiti in ordine decrescente di rango

Motori di ricerca e classificazione

- Il rango non viene calcolato solo in base al contenuto del documento
- Con la *tecnologia ipertestuale*, i documenti contengono infatti riferimenti ad altri documenti (link)
  - Regola empirica: quanto più un documento è riferito da altri documenti, tanti più esso è significativo
  - Essi sono ritenuti infatti *l'espressione latente di una forma di giudizio umano*
- Questa misura è difficilmente alterabile
  - Mentre è facile poter alterare il contenuto di un documento

Motori di ricerca e classificazione

- Attualmente, i più importanti motori di ricerca sul Web utilizzano una propria funzione di rango basata sull'analisi dei collegamenti
- Operano, in linea di principio, seguendo il seguente schema
  - Impiegano programmi (crawler) che visitano e raccolgono le pagine del Web seguendo i link che le collegano
  - Analizzano il testo in ogni pagina raccolta e costruiscono un indice di ricerca dei termini che compaiono nelle pagine stesse
  - Tengono traccia dei link tra le pagine, costruendo quindi la matrice di adiacenza del grafo del Web
    - Della sua porzione visitata
  - Calcolano, a partire dalla matrice, il rango delle pagine raccolte, secondo modalità che variano da motore a motore

Motori di ricerca e classificazione

- Due sono gli approcci che hanno ispirato gi attuali metodi di realizzazione della funzione rango
  - PageRank (utilizzato da Google)
    - Calcola (periodicamente) il rango di tutte la pagine raccolte
    - Le pagine restituite in seguito a un'interrogazione vengono restituite in ordine decrescente di rango
  - HITS (Hypertext Induced Topic Selection)
    - Identifica prima un opportuno insieme D di pagine che, in qualche modo, sono collegate all'interrogazione
    - Solo di queste calcola il rango, che viene utilizzato per decidere l'ordine di visualizzazione dei risultati
- I due approcci possono essere utilizzati anche insieme
  - HITS come raffinamento del PageRank

PageRank

- Il rango di una pagina è
  - Direttamente proporzionale al rango delle pagine che la riferiscono, e
  - Inversamente proporzionale al grado di uscita di tali pagine
- Se **E(i)** indica l'insieme delle pagine che riferiscono i e con **uscita(i)** il grado di uscita di i
- Sia A la matrice di adiacenza del grafo del Web
  - A[i][j] = 1 $\Leftrightarrow$ esiste un link dalla pagina i alla pagina j
  - Quindi i $\in$ E(j) $\Leftrightarrow$ A[i][j] = 1

$$\text{rango}(j)=(1-\alpha)+\alpha\sum_{i\in E(j)} \text{rango}(i)/\text{uscita}(i)$$

dove $1-\alpha$ rappresenta la probabilità che il *navigatore* arrivi a quella pagina non da un link, ma direttamente dalla barra del browser

PageRank

- Aprire una pagina all'esterno (aumentare i link in uscita)
  - Da un lato comporta la diminuzione del rank
  - Dall'altro può comportare (indirettamente) un aumento di riferimenti ad essa
- Effetto *fabbriche di link*
  - Tendono a diminuire l'affidabilità del PageRank
  - Google penalizza esplicitamente queste fabbriche
  - Non sono però noti i meccanismi secondo cui esse vengono riconosciute
- Altro svantaggio: tende a favorire le pagine vecchie
  - Le pagine nuove, anche se molto interessanti, non hanno all'inizio molti riferimenti in entrata
    - A meno che non sia parte di un sito già esistente e con un insieme di pagine fortemente connesse fra loro
  - Comunque Google non usa solo il PageRank per cacolare il rank
    - Oltre 100 fattori moltiplicativi
    - Suggerisce che il miglior modo per acquisire alto rango è di scrivere pagine di qualità

HITS

- La funzione di rango realizzata da HITS è motivata dall'osservazione che le pagine significative per certi termini di ricerca non sempre contengono quei termini stessi
  - Es.: http://www.ferrari.it e http://www.fiat.it non contengono la parola "automobile"
- Inoltre tali pagine non si riferiscono vicendevolmente in modo diretto
  - Lo fanno attraverso qualche pagina secondaria o tramite siti specializzati
- Nella terminologia HITS
  - http://www.ferrari.it e http://www.fiat.it sono delle autorità nel campo automobilistico
  - I siti specializzati vengono denominati concentratori di collegamenti (hub)

HITS

- Il fenomeno osservato è quello del mutuo rafforzamento in termini di significatività
  - Un concentratore è tanto più significativo quanto lo sono le autorità a cui si riferisce
  - Un'autorità è tanto più significativa quanto lo sono i concentratori che la riferiscono
- HITS classifica le pagine in base a questi principi
  - Ogni pagina è simultaneamente un'autorità e un concentratore
  - Sono dunque utilizzate le seguenti funzioni rango
    - rangoA(j): misura il peso della pagina j intesa come autorità
    - rangoC(j): misura il peso della pagina j intesa come concentratore

HITS

- Sono dunque utilizzate le seguenti funzioni rango

$$\text{rangoA}(j) = \sum_{i \in E(j)} \text{rangoC}(i)$$
$$\text{rangoC}(j) = \sum_{k \in U(j)} \text{rangoA}(k)$$

dove **U(j)** è l'insieme della pagine riferite da j e

E(j) l'insieme delle pagine che riferiscono j

- Questo calcolo viene effettuato a ogni interrogazione, sugli insiemi di documenti selezionati
  - Dipendono dall'interrogazione effettuata
  - In PageRank il rango viene calcolato indipendentemente dall'interrogazione, su tutte le pagine raggiunte dai crawler

Documenti selezionati

- I documenti D su cui applicare il rango di HITS vengono selezionati secondo il seguente schema
  - Si esegue un'interrogazione utilizzando un motore di ricerca con la propria funzione di rango
    - Ad es. PageRank
  - Si prendono i primi t risultati, ottenendo così un insieme S di partenza
    - t=200 nella formulazione originaria di HITS
  - S si estende con le pagine appartenenti al vicinato di quelle in S
    - HITS aggiunge a S tutte le pagine in U(j), $j \in S$, e un opportuno sottoinsieme delle pagine in E(j), $j \in S$
      - Alternativamente, è possibile aggiungere anche le pagine con i riferimenti in uscita di secondo livello
        - » U(k), $k \in U(j)$ e $j \in S$

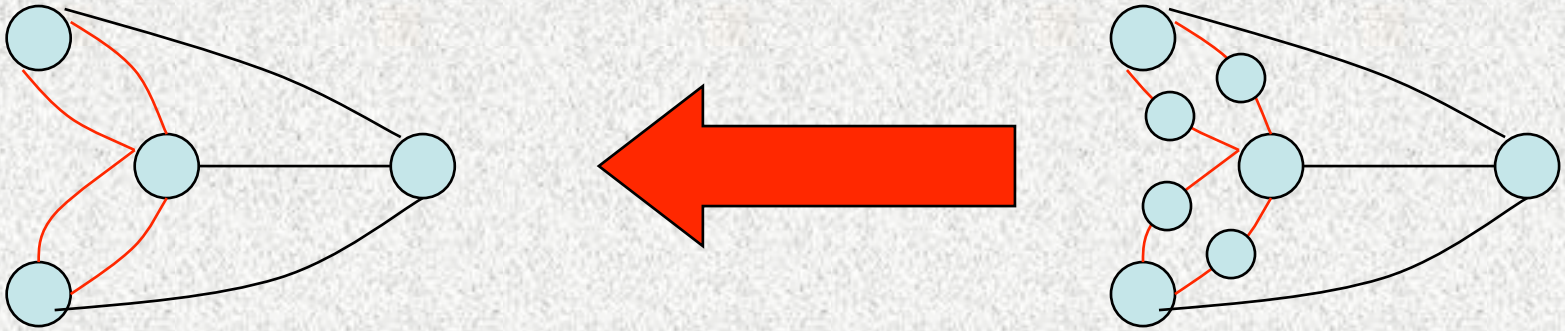
HITS

- Svantaggi
  - Come detto, il calcolo del rango viene effettuato a ogni interrogazione sui documenti selezionati
  - Presenta un problema di "deriva del soggetto" (*topic drift*)
    - È possibile che costruendo l'insieme D dei documenti venga inclusa una pagina non precisamente focalizzata sull'argomento dell'interrogazione, ma comunque con un alto punteggio di autorità
    - Il rischio è che questa pagina e quelle a essa vicina possano dominare la lista ordinata che viene restituita all'utente
      - Verrebbe così spostata l'attenzione su documenti non proprio inerenti all'interrogazione

FINE

Lucidi tratti da
Crescenzi • Gambosi • Grossi,
Strutture di dati e algoritmi
Progettazione, analisi e visualizzazione
Addison-Wesley, 2006
http://algoritmica.org

Spazio delle rappresentazioni



- K_5 e il grafo bipartito completo $K_{3,3}$ non sono planari
- Sorprendentemente questi due grafi sono sufficienti per caratterizzare tutti i grafi planari
- Infatti, il teorema di Kuratowski-Pontryagin-Wagner afferma che
  - Un grafo G non è planare ⇔esiste un suo sottografo G' la cui contrazione (si applica ai nodi di grado 2, vedi Figura) fornisce K_5 oppure $K_{3,3}$