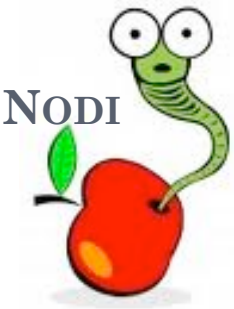
Il Problema di Infettare (e difendere!) una Rete con un Worm ovvero La Minima Copertura di Nodi



Prof. Tiziana Calamoneri

Corso di Algoritmi per le reti

A.A. 2012/13

Il Problema

I Worm (1)

- Un worm (letteralmente "verme") è una particolare categoria di *malware* (malicious software, creato con il solo scopo di causare danni più o meno gravi al computer su cui viene eseguito) in grado di autoreplicarsi.
- È simile ad un virus, ma a differenza di questo, non necessita di legarsi ad altri eseguibili per diffondersi.
- Il termine deriva da un romanzo di fantascienza degli anni '70 di Brunner
- Uno dei primi worm diffusi sulla rete fu Internet Worm, creato da R. Morris nel 1988, quando Internet era ancora agli albori. Tale virus riuscì a colpire tra le 4000 e le 6000 macchine, si stima il 4-6% dei computer collegati a quel tempo in rete.

I Worm (2)

- I worm non hanno bisogno di infettare altri file per diffondersi, perché modificano il sistema operativo della macchina ospite in modo da essere eseguiti automaticamente e tentare di replicarsi sfruttando per lo più Internet.
- Il loro scopo è rallentare il sistema con operazioni inutili o dannose.

I Worm (3)

- Il worm tenta di replicarsi sfruttando Internet in diverse maniere:
  - posta elettronica: il worm ricerca indirizzi e-mail memorizzati nel computer ospite ed invia una copia di sé stesso come file allegato a tutti o parte degli indirizzi che è riuscito a raccogliere.
  - tecniche di social engineering per indurre il destinatario ad aprire l'allegato, che spesso ha un nome che permette al worm di camuffarsi come file non eseguibile.
  - bug di client di posta molto diffusi, per eseguirsi automaticamente al momento della visualizzazione del messaggio

I Worm (4)

- Tutti i worm più recenti effettuano la falsificazione dell'indirizzo mittente, creando un fastidioso effetto collaterale di proliferazione di messaggi: gli antivirus respingono il messaggio infetto e notificano il fatto al mittente, ma dato che questo è falso tale notifica arriva ad un destinatario diverso da chi ha realmente inviato la mail e che nulla ha a che fare con l'invio del worm.

Danni Causati dai Worm (1)

- Possiamo grossolanamente dividere gli effetti nocivi cagionati da un worm in due tipologie:
  - *danni diretti*, causati dall'esecuzione del worm sulla macchina vittima, e
  - *danni indiretti*, derivanti dalle tecniche utilizzate per la diffusione.

Danni Causati dai Worm (2)

Danni diretti:

- I worm semplici, composti solo dalle istruzioni per replicarsi, di per sé non creano gravi danni diretti al di là dello spreco di risorse computazionali.
- Spesso però, per nascondersi, interferiscono con il funzionamento di software volti a scovarli e a contrastarne la diffusione (antivirus e firewall) impedendo così il funzionamento normale del computer ospite.
- La maggior parte dei worm contiene una parte detta *payload*, che ha il solo scopo di causare dei danni al sistema infettato.
- Molto di frequente un worm funge da veicolo per l'installazione automatica sul maggior numero possibile di macchine di backdoor o keylogger, che potranno poi essere sfruttati da un malintenzionato o da un altro worm.

Danni Causati dai Worm (3)

Danni indiretti:

- Sono gli effetti collaterali dell'infezione da parte di un worm di un elevato numero di computer connessi in rete sul corretto funzionamento e sull'efficacia delle comunicazioni
- I messaggi di posta elettronica inviati dai worm per replicarsi aumentano la mole di posta indesiderata, sprecando risorse preziose in termini di banda e di attenzione.
- I worm che sfruttano vulnerabilità note di alcuni software causano invece malfunzionamenti di tali programmi, con conseguenze quali l'instabilità del sistema operativo e a volte spegnimenti e riavvii forzati

Diffusione di un Worm (1)

- Per semplificare, assumiamo che il tempo di trasmissione dell'informazione in ogni data connessione nella rete sia la stessa, pari a T.
- Se un worm ha infettato con successo un certo numero di nodi tali che con un solo passo di spread tutti i nodi possono essere infettati, in tempo T tutta la rete è infetta ("first propagation step").
- Conoscere l'insieme di nodi di partenza è il primo passo per proteggere la rete dall'attacco.

Diffusione di un Worm (2)

- Il problema reale è più complesso perché tutte le reti di notevole dimensione hanno connessioni dinamiche.
- Dal punto di vista del gestore della rete, ogni filtro per proteggere la rete dagli attacchi di worm del primo ordine rallenta la comunicazione e quindi è necessario minimizzarne il numero
- I computers da proteggere sono quelli che potrebbero stare nel primo insieme.

Modello del Problema su Grafi

Il Modello su Grafi (1)

- Informalmente, definiamo copertura di nodi (vertex cover) di un grafo un insieme C di nodi con la proprietà che ogni arco nel grafo abbia almeno uno dei suoi due estremi in C.
- Rilevante la minima copertura di nodi, cioè l'insieme C di minima cardinalità.
- N.B. un grafo può avere più di una copertura di nodi (stesso numero di nodi ma diversa scelta, garantendo comunque la proprietà)

Il Modello su Grafi (2)

- Intuitivamente, ogni minima copertura di nodi rappresenta un ottimo punto di partenza per un worm.
- I computer da proteggere saranno quelli che rappresentano i nodi nella minima copertura di nodi del grafo di comunicazione.
- Se il grafo ha più di una minima copertura, saranno da proteggere i computers nell'intersezione di tutte le coperture.

La Minima Copertura di Nodi

Minima Copertura di Nodi (1)

- **Def.** Dato $G=(V,E)$, V' sottoinsieme di V è una copertura di nodi (vertex cover) per G se per ogni arco (i,j) almeno una delle seguenti condizioni è verificata:
  - i appartiene a V';
  - j appartiene a V'.
- Oss. L'insieme di tutti i nodi è banalmente una copertura di nodi
- Dato $G=(V,E)$, il Problema della minima copertura di nodi consiste nel trovare una copertura di nodi per G di cardinalità minima.

Minima Copertura di Nodi (2)

- **Def.** Il Problema Decisionale della minima copertura di nodi (VC) consiste nel rispondere alla seguente domanda:

  dati G ed un intero k, esiste una copertura di nodi per G di cardinalità minore o uguale a k?
- VC è un problema NP-completo (riduzione da 3-SAT o da Clique [Karp]).
- VC rimane NP-completo anche per grafi cubici [Garey, Johnson, Stockmeyer '74] e per grafi planari di grado al più 3 [Garey & Johnson '77].

Minima Copertura di Nodi (3)

Introducendo le seguenti n variabili di decisione:

per ogni $i=1, 2, \ldots, n$, $x_i=1$ se il nodo i è in V' e 0 altrimenti

si ottiene la seguente formulazione ILP per il problema VC:

$$\min \sum_{i=1}^{n} x_i$$

soggetto a:

$$x_i + x_j \geq 1, \forall (i,j) \in E$$

$$x_i \in \{0,1\}, i = 1,2,\ldots,n$$

Minima Copertura di Nodi (4)

- Come già sottolineato, il Problema della minima copertura di nodi di G è un problema NP-hard, per cui a tutt'oggi un qualunque algoritmo esatto ha complessità computazionale superpolinomiale.
- Tuttavia, è possibile individuare una soluzione approssimata in tempo polinomiale.
- Descriviamo due algoritmi naif intuitivamente buoni ma con rapporti di approssimazione molto alti
- Descriviamo poi un algoritmo polinomiale approssimante che, preso in input $G=(V,E)$, trova con complessità $O(n+m)$ una copertura di nodi V' per G di dimensione al più due volte la dimensione di una copertura ottima V^*, ossia $|V'| \leq 2|V^*|$.

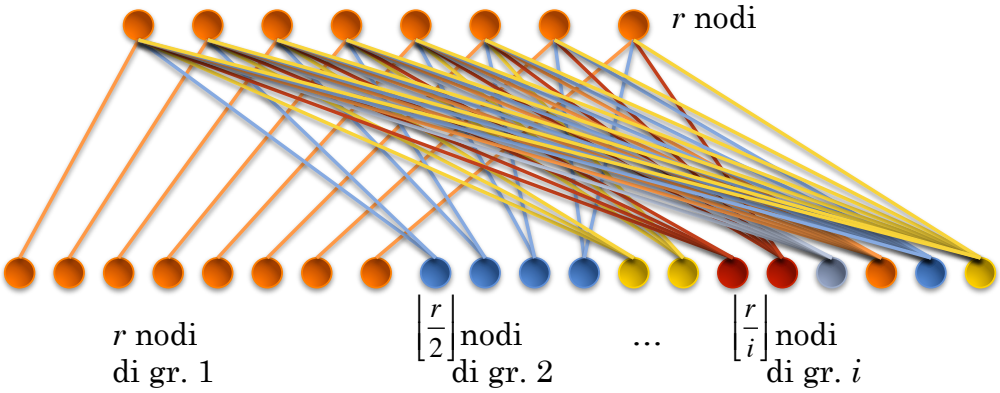
Minima Copertura di Nodi (5)

Algoritmo Greedy-VC(G)

- V'=insieme vuoto, $E'=E$
- Finché (E' non è vuoto) do
  - seleziona da E' un arco a caso (i,j) e scegli un estremo i
  - inserisci i in V'
  - elimina da E' tutti gli archi che hanno un estremo in i o in j
- Restituisci V'

Minima Copertura di Nodi (6)

L'Algoritmo Greedy-VC può trovare un vertex cover di cardinalità molto lontana dall'ottimo:

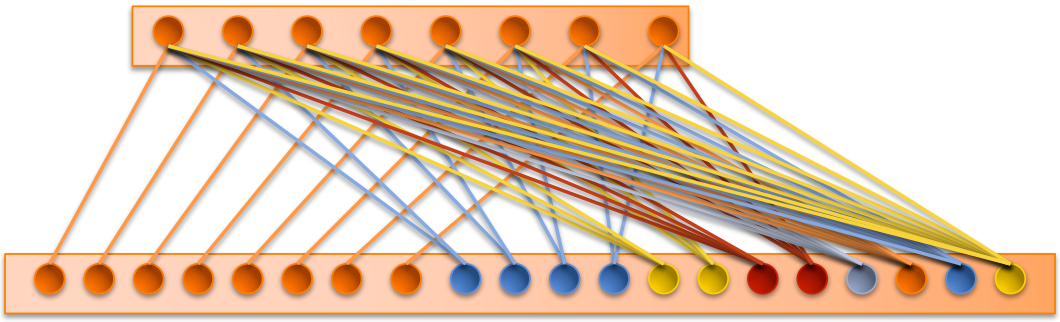


$$n = r + \sum_{i=1}^{r} \left\lfloor \frac{r}{i} \right\rfloor \le r + r \sum_{i=1}^{r} \frac{1}{i} = \Theta(r \log r)$$

Serie armonica

Minima Copertura di Nodi (7)

Vertex cover ottimo



Vertex cover che può trovare l'algoritmo

Rapporto di approssimazione $\Theta(\log r)$

Minima Copertura di Nodi (8)

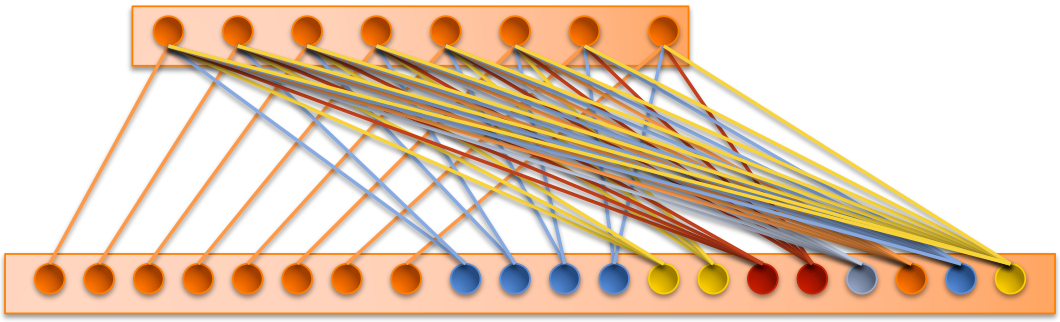
Proviamo a fare di meglio:

Algoritmo Greedy2-VC(G)

- V'=insieme vuoto, E'=E
- Finché (E' non è vuoto) do
  - Seleziona un nodo v di grado max nel grafo *corrente*
  - inserisci v in V'
  - elimina da E' tutti gli archi che hanno un estremo in v
- Restituisci V'

Minima Copertura di Nodi (9)

Il nuovo algoritmo può trovare questo vertex cover…



…ma anche questo partendo dai nodi a destra

Il rapporto di approssimazione non cambia!

Minima Copertura di Nodi (10)

Un algoritmo migliore:
Algoritmo 2-Approx-VC(G)

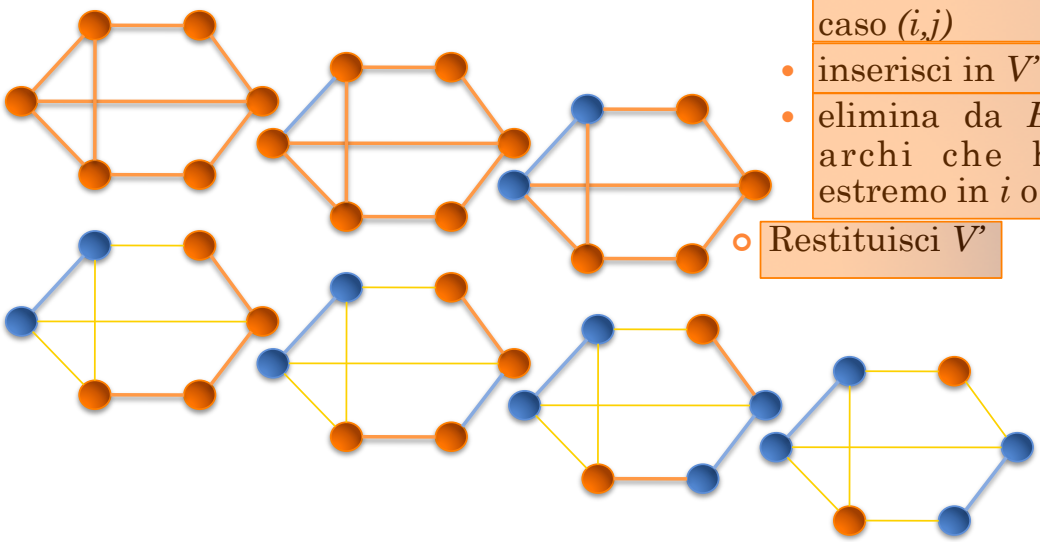
- V'=insieme vuoto
- $E'=E$
- Finché (E' non è vuoto) do
  - seleziona da E' un arco a caso (i,j)
  - inserisci in V' sia i che j
  - elimina da E' tutti gli archi che hanno un estremo in i o in j
- Restituisci V'

Complessità: $O(n+m)$

Min. Copertura di Nodi (11)

Algoritmo 2-Approx-VC(G)

- V'=insieme vuoto
- $E'=E$
- Finché (E' non è vuoto) do
  - seleziona da E' un arco a caso (i,j)
  - inserisci in V' sia i che j
  - elimina da E' tutti gli archi che hanno un estremo in i o in j
- Restituisci V'



Minima Copertura di Nodi (12)

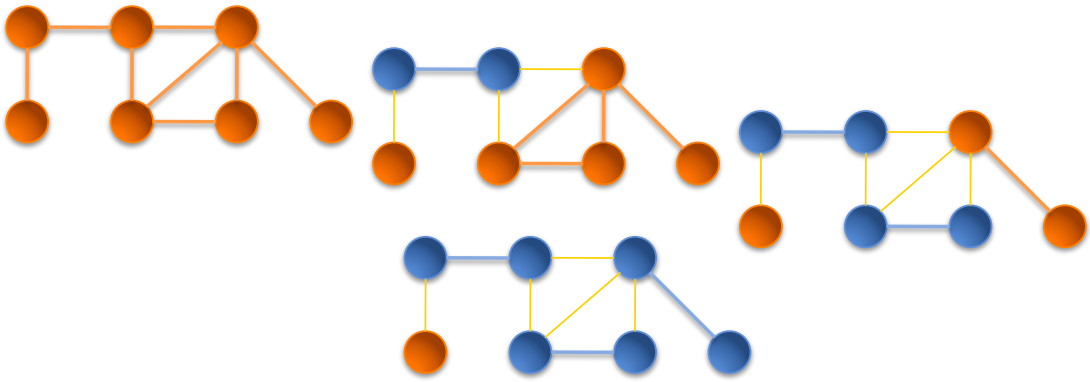
Th. L'insieme V', output di 2-Approx-VC, è una copertura di nodi per G tale che $|V'| \leq 2|V^*|$.

Dim. Sia V^* una min copertura di nodi.

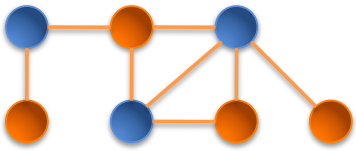
V' è una copertura di nodi per G per costruzione.
Sia A l'insieme degli archi selezionati a caso da E'.
Per ogni arco (i,j) di A, i e j sono aggiunti a V' quindi:
$|V'|=2|A|$.
Inoltre, vengono rimossi da E' tutti gli archi incidenti ad almeno uno fra i nodi i e j, per cui due archi in A non possono essere adiacenti.
Segue che $|A| \leq |V^*|$.
Unendo le due relazioni: $|V'|=2|A| \leq 2|V^*|$. *CVD*

Minima Copertura di Nodi (13)

Esempio in cui la limitazione superiore viene raggiunta:



mentre la soluzione ottima è:



Minima Copertura di Nodi (14)

- Nonostante questo algoritmo sia molto semplice, non si può fare molto meglio, infatti:
- si può dim. che VC non è approssimabile in meno di 1.1666 [Håstad '97] e poi 1.3606 [Dinur & Safra '05]
- i migliori alg. di approssimazione che si conoscano approssimano in:

$2 - \frac{\log\log|V|}{2\log|V|}$ [Monien& Speckenmeyer '85]

$2 - \frac{\log\log|V|}{\log|V|}$ [Bar-Yehuda, Even '85] e

$2 - \frac{\ln\ln|V|}{\ln|V|}(1 - o(1))$ [Halperin '00]

Proprietà della Min Copertura di Nodi (1)

Un insieme indipendente di $G=(V,E)$ è un insieme di nodi di V tale che i suoi elementi siano a coppie non adiacenti.

Th. Un insieme V' è una copertura di nodi se e solo se il suo complemento $V-V'$ è un insieme indipendente.

Dim.

V' *VC* $\Rightarrow V-V'$ *IS*

- P.A. Se esiste una coppia di nodi adiacenti in $V-V'$ quell'arco non è coperto da V'. Assurdo

$V-V'$ *IS* $\Rightarrow V'$ *VC*

- P.A. Se esiste un arco non coperto da V' i nodi ad esso incidenti sono adiacenti in $V-V'$.

CVD

Proprietà della Min Copertura di Nodi (2)

- Cor. Il numero di nodi di un grafo è pari alla cardinalità della sua min copertura di nodi + la cardinalità del suo max insieme indipendente [Gallai '59]
- Ciò nonostante, i due problemi non sono equivalenti dal punto di vista dell'approssimazione: IS non ha fattore di approssimazione costante [Håstad '99].

Proprietà della Min Copertura di Nodi (3)

Def. Un accoppiamento di $G=(V,E)$ è un sottinsieme M di E t.c. ogni coppia di archi in M non ha nodi in comune.

Th. Sia M un accoppiamento in G e sia C una copertura di nodi di G. Allora $|M| \leq |C|$.

Dim. Poiché C è una copertura, deve coprire tutti gli archi di M.

Per definizione di accoppiamento, per ogni arco di M, almeno uno dei suoi estremi deve stare in C.

Quindi $|M| \leq |C|$. CVD

Proprietà della Min Copertura di Nodi (4)

Cor. Sia M un accoppiamento in G e sia C una copertura di nodi di G. Se $|M|=|C|$ allora M è un accoppiamento massimo e C è una copertura di nodi minima.

Calcolare un accoppiamento massimo è polinomiale, quindi si potrebbe pensare di aver risolto il problema.

No, perché:

Fatto: L'inverso del precedente corollario non è vero.

Tuttavia, è stato proposto un algoritmo basato su questo risultato.

Proprietà della Min Copertura di Nodi (5)

Algoritmo Nuovo-Approx-VC(G) [Gavril '79]

- Calcola un accoppiamento massimo M
- V'= insieme vuoto
- Per ogni e in M

  Inserisci in V' entrambi gli estremi di e
- Restituisci V'

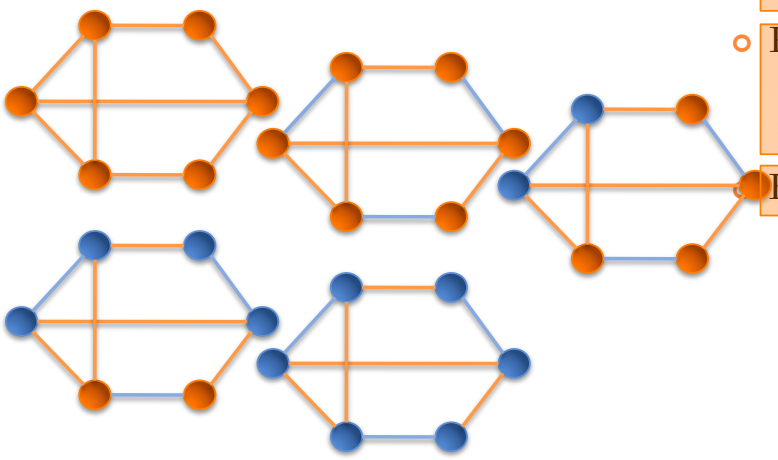
Complessità:

dipende dal calcolo dell'accoppiamento massimo :

$O(n^4)$ [Edmonds '65]

$O(m\sqrt{n})$ [Micali & Vazirani '80]

Proprietà della Min Copertura di Nodi (6)



Algoritmo Nuovo-Approx-VC(G) [Gavril '79]

- Calcola un accoppiamento massimo M
- V'= insieme vuoto
- Per ogni e in M
  Inserisci in V' entrambi gli estremi di e
- Restituisci V'

Proprietà della Min Copertura di Nodi (7)

Th. L'insieme V', output di Nuovo-Approx-VC, è una copertura di nodi per G tale che $|V'| \leq 2|V^*|$.

Dim. V' è una copertura perché un arco (u,v) di G è:

- o parte dell'accoppiamento e quindi ha entrambi gli estremi in V'
- o non fa parte dell'accoppiamento, ma ha comunque un estremo in V' altrimenti si potrebbe inserire nell'accoppiamento, che però è massimo.

Per costruz. $|V'|=2|M|$

Nota che ogni $(u,v) \in M$ deve contenere almeno un nodo della copertura ottima: $|M| \leq V^*$

Unendo le due relazioni: $|V'|=2|M| \leq 2|V^*|$. CVD

Proprietà della Min Copertura di Nodi (8)

Nel caso dei grafi bipartiti si può affermare una relazione più forte tra minima copertura e massimo accoppiamento, deducendo che:

VC è polinomiale per i grafi bipartiti.

Ne segue che l'algoritmo precedente, in questo caso, può essere modificato per trovare l'ottimo (complessità $O(m\sqrt{n})$ - dipende dal calcolo dell'accoppiamento [Hopcroft & Karp '73]).

Th. di König ['31] (Egerváry ['31]): *In un grafo bipartito, il # di archi in un accoppiamento massimo è uguale al numero di nodi in una min. copertura di nodi.*

Dim.1. (la più diffusa) Usa il th. di P.Hall per gli accoppiamenti massimi nei grafi bipartiti, che faremo più avanti.

Proprietà della Min Copertura di Nodi (8)

Th. di König ['31] (Egerváry ['31]): *In un grafo bipartito, il # di archi in un accoppiamento massimo è uguale al numero di nodi in una min. copertura di nodi*

Dim.2. $G=(X \cup Y, E)$ bipartito. M accoppiamento.

P.A. Se $|M|$ e $|C|$ sono diversi, dobbiamo dimostrare che:

- o M non è un accoppiamento massimo
- o C non è minimo ed esiste una copertura di cardinalità pari a $|M|$.

Caso 1.

Se tutti i nodi sono accoppiati da M, una parte della bipartizione forma una copertura e $|M|=|V'|$ => fine

Proprietà della Min Copertura di Nodi (9)

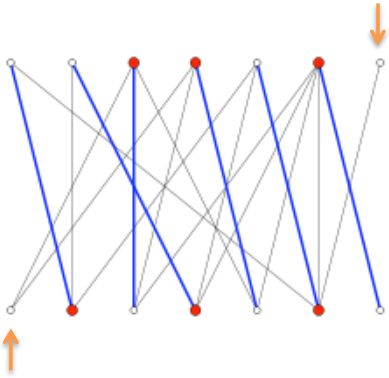
dim.2 del Th. di König (segue)

Caso 2.

Se ci sono nodi non accoppiati:

partizioniamo i nodi in insiemi S_i:

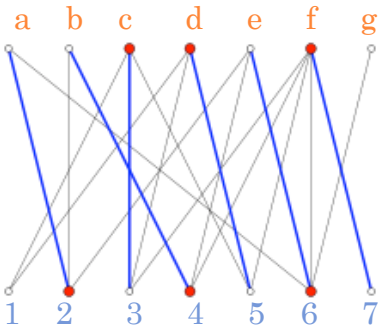
- S_0 è l'insieme di tutti i nodi non accoppiati in M (qualunque sia la partizione a cui appartengono!).
- Definito S_{2j} per qualche $j \geq 0$, S_{2j+1} è l'insieme dei nodi adiacenti ai nodi di S_{2j} tramite archi non in M e non ancora inseriti
- …

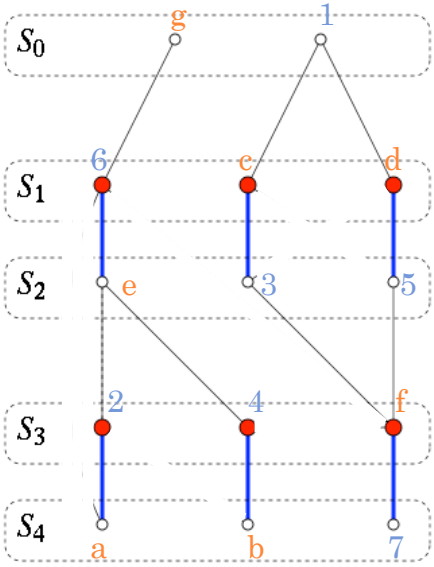


Proprietà della Min Copertura di Nodi (10)

dim.2 del Th. di König (segue)

- Ogni nodo in S_{2j+1} deve essere un estremo di un arco di M (altrimenti sarebbe in S_0); per ogni tale arco, metti l'altro estremo in S_{2j+2} se non ancora inserito.

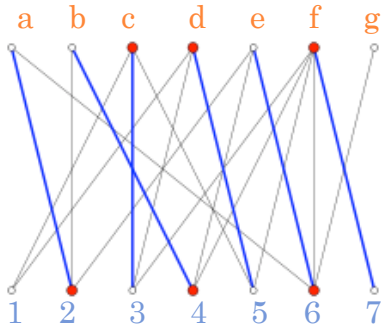


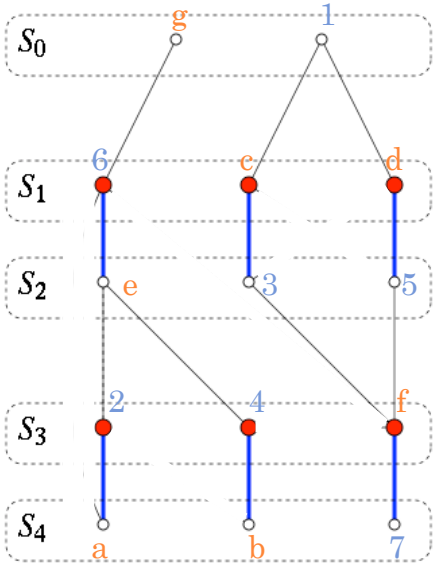


Proprietà della Min Copertura di Nodi (11)

dim.2 del Th. di König (segue)

- Si conclude la costruzione inserendo gli archi rimasti.

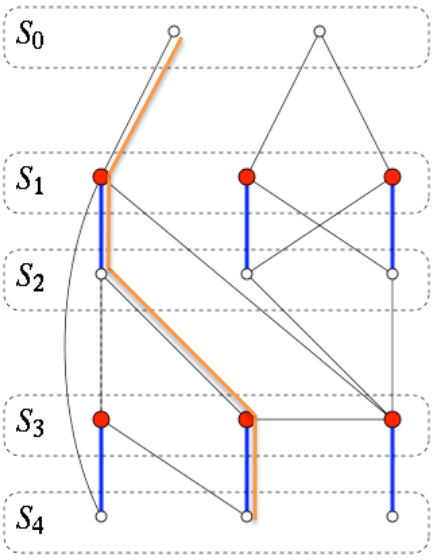




Proprietà della Min Copertura di Nodi (12)

(dim. 2 Th. di König - segue)

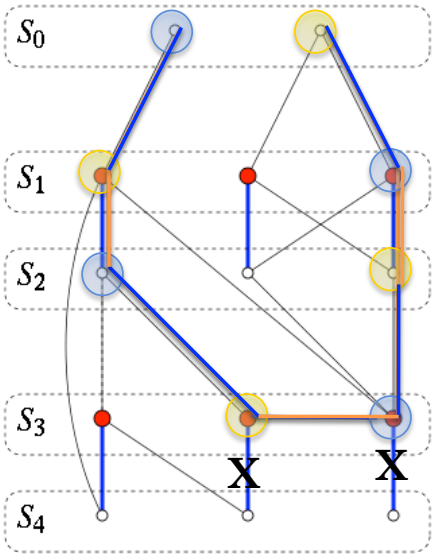
- Per ciascun nodo in un S_i esiste sempre un arco verso S_{i-1} => si può formare un cammino livello per livello fino ad un nodo di S_0.
- Il cammino ottenuto è alternante (alterna archi di M ed archi non di M)



Proprietà della Min Copertura di Nodi (13)

(dim. 2 Th. di König - segue)

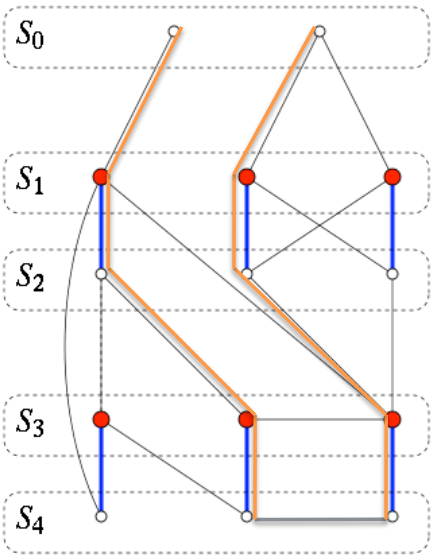
- Se esiste un arco di M (u,v) tra due nodi sullo stesso liv. dispari S_{2j+1}, i due cammini alternanti da u e da v sono connessi in un unico cammino alternante tramite (u,v).
- I due cammini non possono avere nodi in comune per via della bipartizione del grafo (u e v sono in bipartizioni diverse, quindi avranno i nodi di X uno sui liv. pari e uno sui liv. dispari).
- Ne segue che il cammino è aumentante => M non è massimo.



Proprietà della Min Copertura di Nodi (14)

(dim. 2 Th. di König- segue)

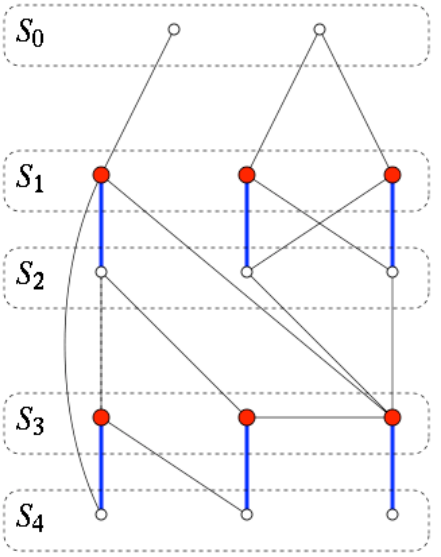
- Analogamente, se esiste un arco non in M (u,v) tra due nodi sullo stesso liv. pari S_{2j}, possiamo fare gli stessi ragionamenti ed ottenere un cammino aumentante => M non è massimo.



Proprietà della Min Copertura di Nodi (15)

(segue dim. Th. di König)

Quindi, se M è massimo:

- Ogni arco in M ha esattamente un estremo in un livello dispari S_{2j+1},
- Ogni arco non in M ha almeno un estremo in un livello dispari S_{2j+1}
- L'unione dei livelli dispari forma una copertura di nodi $|V'|$ di cardinalità $|M|$.
- Poiché V' non può avere meno nodi di $|M|$ perché deve coprire interamente M, segue che:

  $|V'|=|M|$. CVD