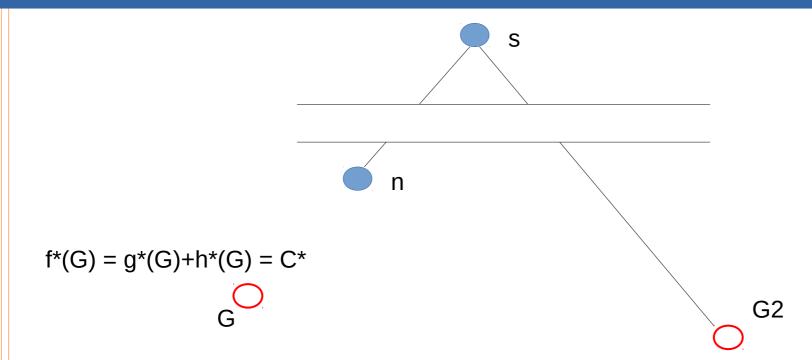
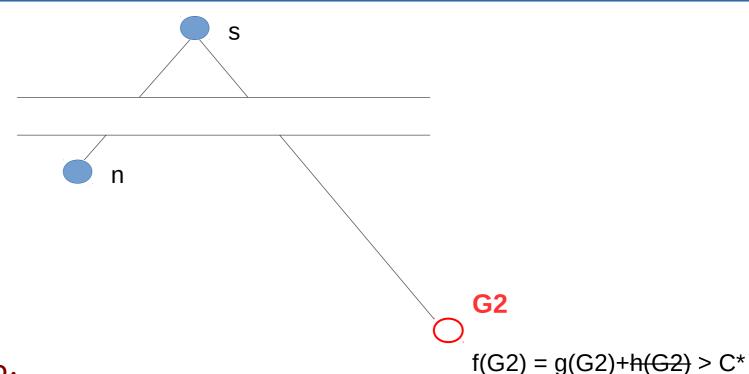
Ottimalità dell'algoritmo A*

- Si dimostra che quando:
 - 1) l'euristica h è ammissibile
 - 2) e tutti i passi hanno un costo maggiore di una costante positiva piccola a piacere
- Allora:
 - A* termina e trova una soluzione ottima (di costo minimo)
 - In altri termini in questo caso A* è completo e ottimale

- Nota: in un albero di ricerca ogni nodo ha un solo cammino assoluto
- Perché manchi l'ottimalità deve accadere che durante la ricerca l'algoritmo:
 - scelga un nodo obiettivo sub-ottimo (G2)
 - al posto di un nodo (n)
 - che si trova su un cammino ottimo (che porta al goal G)
- Dimostriamo che in un albero di ricerca ciò non può capitare laddove l'euristica è ammissibile e i passi hanno costo non nullo



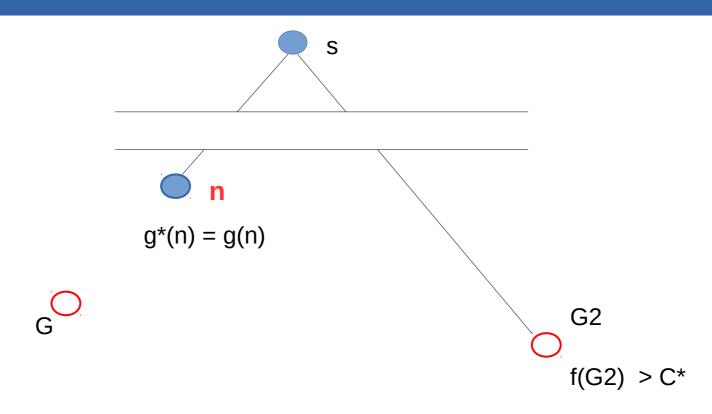
- Chiamiamo:
 - G2 = obiettivo subottimo
 - C* il costo reale della soluzione ottima



Consideriamo G2:

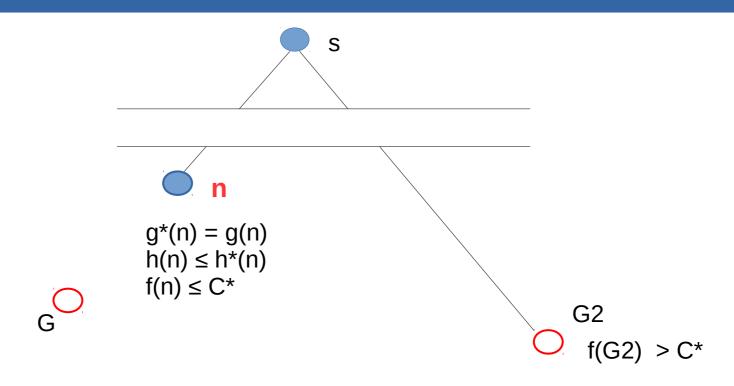
- 1) h(G2) = 0 perché G2 è un nodo obiettivo
- 2) f(G2) = g(G2) + h(G2) = g(G2) + 0 = g(G2)
- 3) Poiché per assunto G2 è <u>subottimo</u>, si ha che:

$$f(G2) = g(G2) > C^*$$

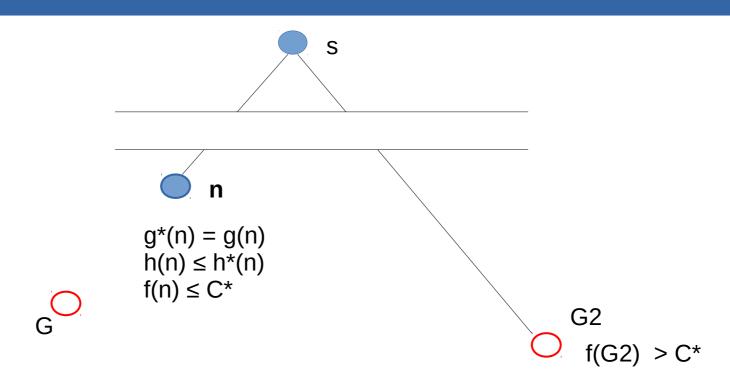


Sia n un generico nodo appartenente a un cammino ottimo:

 g*(n) = g(n) perché lavoriamo su di un albero e ogni nodo è raggiungibile dalla radice lungo un solo percorso



- hèammissibile per ipotesi, quindi h(n) ≤ h*(n)
- Quindi: $f(n) = g(n) + h(n) \le g^*(n) + h^*(n) = C^*$
- Quindi $f(n) \le C^* < f(G2)$

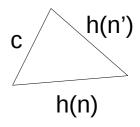


• Quindi:

1)
$$f(n) < f(G2)$$

- 2) A* sceglie il nodo aperto con f(.) minima,
- Di conseguenza fra n e G2 verrà scelto n (q.e.d.)

- Nei grafi vi è molteplicità di cammino:
 il primo cammino trovato durante la ricerca che porta a un certo stato
 non è necessariamente quello ottimo (nota: nell' esempio della
 Romania A* non si ferma quando incontra Bucarest la prima volta)
- Questa proprietà invalida la precedente dimostrazione!
- La dimostrazione generale è piuttosto complessa, occorre aggiungere
 l'ipotesi che h sia monotòna (o consistente), cioè che dati un qualsiasi
 nodo n e un qualsiasi suo successore n'prodotto eseguendo l'azione a
 in n vale che h(n) ≤ c(n, a, n') + h(n')
- Tale disuguaglianza è una <u>disuguaglianza triangolare</u>



- Si dimostra che quando l'euristica è monotona, per la disuguaglianza triangolare i costi f(n) lungo un cammino sono non decrescenti
- A* espande i nodi in ordine non decrescente di f:
 - se un nodo CHIUSO viene incontrato più volte lungo un percorso, i nuovi valori di f saranno superiori a quello calcolato la prima volta (cfr. Arad nell' esempio della Romania)
 - Di conseguenza sono i primi incontri con i nodi obiettivo che permetteno di individuare una soluzione ottima
- Su questo assunto si dimostra l'ottimalità

• Dimostriamo che se l'euristica è monotona, i costi di f(n) lungo un cammino sono <u>non decrescenti</u>:

- 1) Sia **n'** un nodo successore di **n**, per definizione: g(n') = g(n) + c(n, a, n')
- 2) Sempre per definizione: f(n') = g(n') + h(n')
- 3) E sostituendo g(n'): f(n') = g(n) + c(n,a,n') + h(n')

4) Applichiamo ora la disuguaglianza triangolare:

$$f(n') = g(n) + c(n, a, n') + h(n') \ge g(n) + h(n)$$

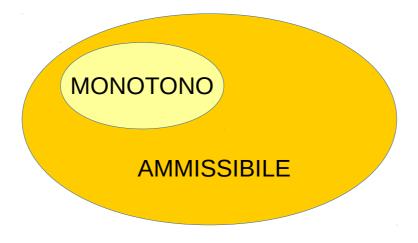
- 5) Sappiamo però che, per definizione: g(n) + h(n) = f(n)
- 6) Di conseguenza: $f(n') \ge f(n)$ (q.e.d.)

Euristiche e impatto sulla ricerca



Euristiche monotone e ammissibili

- La monotonicità è una proprietà più stringente dell'ammissibilità
 - Si dimostra che un' euristica monotona è anche ammissibile
 - Spesso ma non sempre le euristiche ammissibili sono anche monotone



Esempio di euristica monotona

- La distanza in linea d'aria è un'euristica:
 - ammissibile e monotona per il problema della Romania, infatti data una città (genericamente indicata da luogo) e un suo possibile successore (indicato da luogo1) avremo sempre che:
 - h(luogo) < c(luogo, vai, luogo1) + h(luogo1)
- Cioè la distanza in linea d'aria da "luogo" a Bucharest è minore della distanza via terra fra "luogo" e il confinante "luogo1" più la distanza in linea d'aria da "luogo1" a Bucharest

Ammissibile non vuol dire informativo!!

- h(n) = 0 è un' euristica sempre ammissibile ma non è informativa della desiderabilità degli stati
 - Permette di valutare solo il costo del percorso fatto per raggiungere un nodo
 - La ricerca diventa cieca e richiede l'espansione di un maggiore numero di nodi rispetto a usare un'euristica ammissibile e informativa
- In particolare se abbiamo inoltre che tutte le operazioni che permettono di passare da un nodo a un successore hanno costo uniforme pari a 1, A* diventa una ricerca in ampiezza

Valutazione

- A* è ottimamente efficiente per qualsiasi euristica: non esiste alcun altro algoritmo ottimo che garantisca di espandere meno nodi di quelli espansi da A*
- Purtroppo il <u>numero di nodi espansi aumenta</u>
 <u>esponenzialmente con la profondità della soluzione</u>
 <u>ottima</u>
- A* mantiene in memoria tutti i nodi generati

Ridurre i requisiti di memoria di A*

- IDA*: unisce A* e iterative deepening (non lo studiamo)
- RBFS: recursive best-first search:
 - Simile alla *ricerca ricorsiva in profondità* con una differenza importante: usa un "upper bound" dinamico che consente di focalizzare la ricerca sul percorso più promettente invece di continuare indefinitamente lungo lo stesso percorso.
 - Questo upper bound (limite superiore) ricorda la migliore alternativa fra I percorsi attualmente aperti

Recursive Best-First Strategy (RBFS)

Difetto:

lo stesso nodo può essere <u>visitato più volte</u>, se l'algoritmo, dopo aver cambiato percorso, ritorna al percorso precedentemente abbandonato

Pregio:

poche esigenze di spazio, come la *ricerca in profondità senza* backtracking: mantiene solo i nodi del percorso di ricerca corrente e i loro fratelli. È <u>un vantaggio rispetto ad A*</u> che mantiene informazione su tutti i nodi aperti e chiusi

Recursive best first strategy (RBFS)

- RBFS utilizza:
 - Un nodo
 - Un limite superiore (upper bound) locale del nodo
- Esplora:
 - il sottoalbero del nodo finché nella sua frontiera ci sono nodi i cui costi non eccedono il limite superiore
- <u>Upper bound</u> di un nodo:
 - min(upper_bound_parent, valore_fratello_di_costo_minimo)
 - memorizza la <u>stima del costo del percorso alternativo</u> <u>migliore</u>,

Dall'articolo di Korf 1993

- L'algoritmo ha 3 argomenti:
 - Un nodo N,
 - Un valore F(N) ad esso associato,
 - un upper bound B
- f(n): la funzione di valutazione, è statica cioè il valore restituito non cambia nel tempo
- F(n): valore associato al nodo n, è <u>dinamico</u> cioè cambia nel tempo e dipende dai discendenti di N:
 - F(N) = f(N) se N è esplorato per la prima volta
 - $min(F(S_N))$ con S_N sottoalbero di N altrimenti

Linear-space best-first search, Richard E. Korf, Artificial Intelligence 62 (1993) 41-78 Elsevier

Dall'articolo di Korf 1993

- L'algoritmo ha 3 argomenti:
 - Un nodo N,
 - Un valore F(N) ad esso associato,
 - un upper bound B
- f(n): la funzione di valutazione, è statica cioè il valore restituito non cambia nel tempo
- F(n): valore associato al nodo n, è dinamico
- B: è calcolato basandosi sui valori F(.) dei nodi fratelli, ricorda il secondo migliore

Linear-space best-first search, Richard E. Korf, Artificial Intelligence 62 (1993) 41-78 Elsevier

Dall'articolo di Korf 1993

- L'algoritmo ha 3 argomenti:
 - Un nodo N,
 - Un valore F(N) ad esso associato,
 - un upper bound B
- La chiamata iniziale a RBFS è:

RBFS
$$(r, f(r), \infty)$$

dove r è il nodo radice r, il valore F(r) è pari a quello statico f(r) e l'upper bound è pari a ∞

• Esempio: RBFS(Arad, 366, ∞)

Linear-space best-first search, Richard E. Korf, Artificial Intelligence 62 (1993) 41-78 Elsevier

Intuizione

- RBFS lavora come A* fintantoché la soluzione che costruisce rispetta l'upper bound (è la migliore)
- Sospende la ricerca lungo un cammino quando questo non appare più il migliore
- Il cammino viene <u>dimenticato</u> (nodi cancellati)
- Viene solo conservata traccia nella sua radice del costo che si era stimato esplorando quella via

Algoritmo RBFS (Korf 1993)

```
RBFS (node: N, value: F(N), bound: B)

IF f(N)>B, return f(N) // la fz. val. supera il limite superiore, cambia percorso

IF N is a goal, EXIT algorithm // successo!

IF N has no children, RETURN infinity // vicolo cieco, cambia percorso

FOR each child Ni of N, // inizializza F(.) per i successori di N

IF f(N)<F(N), F[i] := MAX(F(N), f(Ni)) // N è il nodo padre, su cui siamo focalizzati ELSE F[i] := f(Ni)

sort Ni and F[i] in increasing order of F[i] // ordinamento per F crescenti, dopo

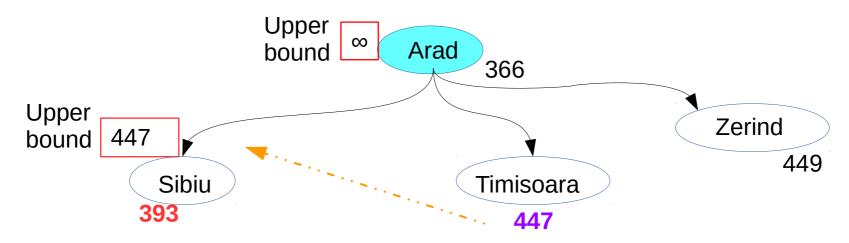
// F[1] sarà l'F del figlio più promettente

IF only one child, F[2] := infinity

WHILE (F[1] <= B and F[1] < infinity) // discesa ricorsiva solo se upper bound rispettato F[1] := RBFS(N1, F[1], MIN(B, F[2])) // applico ricorsivamente RBFS insert N1 and F[1] in sorted order
```

NOTA: F[i] è il valore F del nodo in posizione i dopo l'ordinamento. La posizione 1 è quella del nodo più promettente, la 2 quella dell'alternativa migliore e così via.

Esempio 1/5



Viene scelta Sibiu perché è il nodo sul percorso ritenuto più conveniente Viene associato a tale nodo il valore 447, che è la stima di costo della sua alternativa più conveniente

```
F[Arad] = f(Arad) = 366
Upper bound[Arad] = \infty
```

For each child Ni of N:

f(N) < F(N)?

366 < 366? No quindi F[i] := f(Ni)

esempio F[Sibiu] := f(Sibiu) cioè 393

Sort basato su F:

Sibiu (393), Timisoara (447), Zerind (449),

quindi F[1] è F[Sibiu]

CICLO WHILE:

F[Sibiu] < Upper bound[Arad] ? Sì!

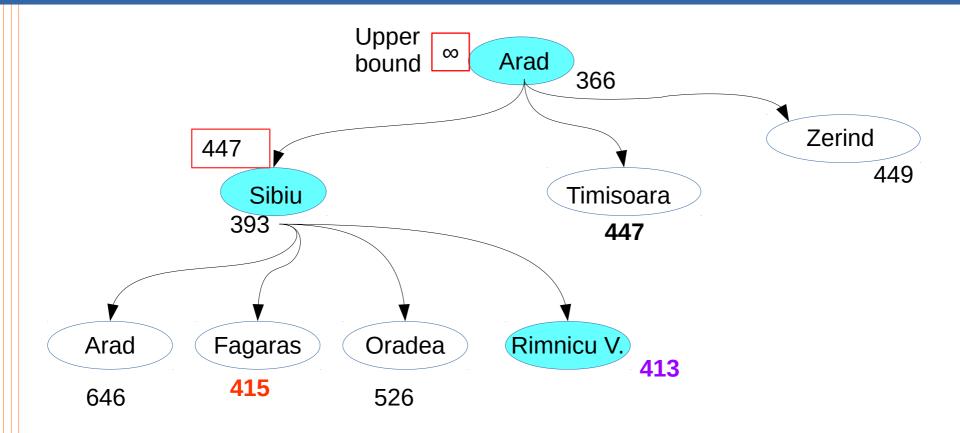
F[Sibiu] < infinity? Sì!

F[1] := RBFS(N1, F[1], MIN(B, F[2]))

Dove B=∞, 2 corrisponde a Timisoara e F[2]=447:

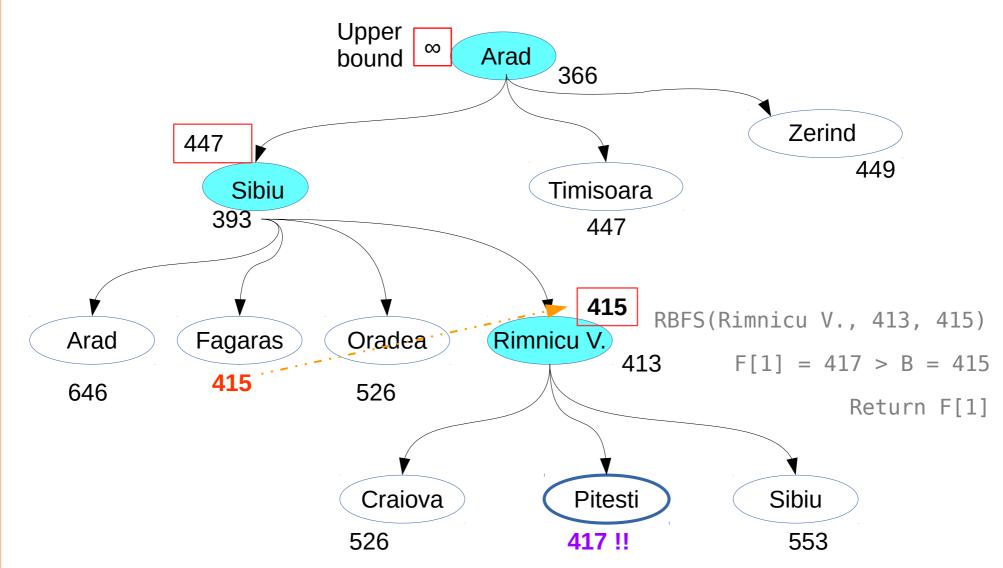
F[1] := RBFS(Sibiu, 393, 447)

Esempio 2/5



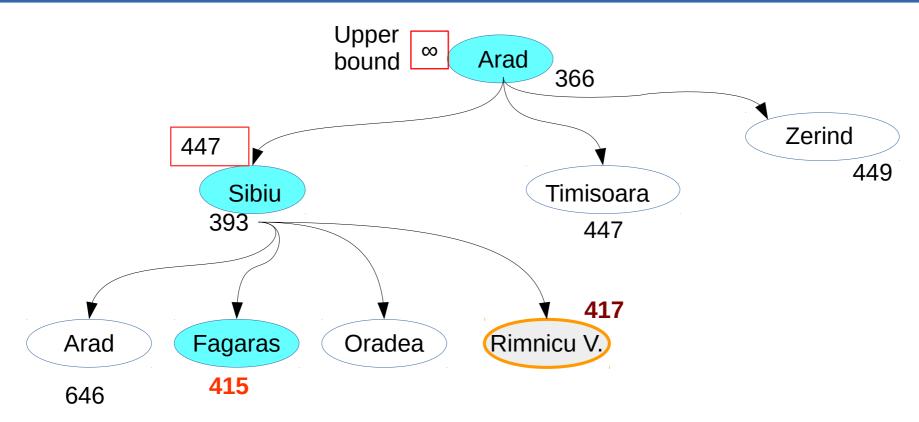
Rimnicu Vilcea è il nodo più promettente. Fagaras è il secondo migliore, quindi 415 sarà l'upper bound per il richiamo ricorsivo su Rimnicu V.

Esempio 2/5



Pitesti ha una stima di costo più elevata dell'alternativa più conveniente a Rimnicu Vilcea, la condizione del while fallisce: si cambia percorso

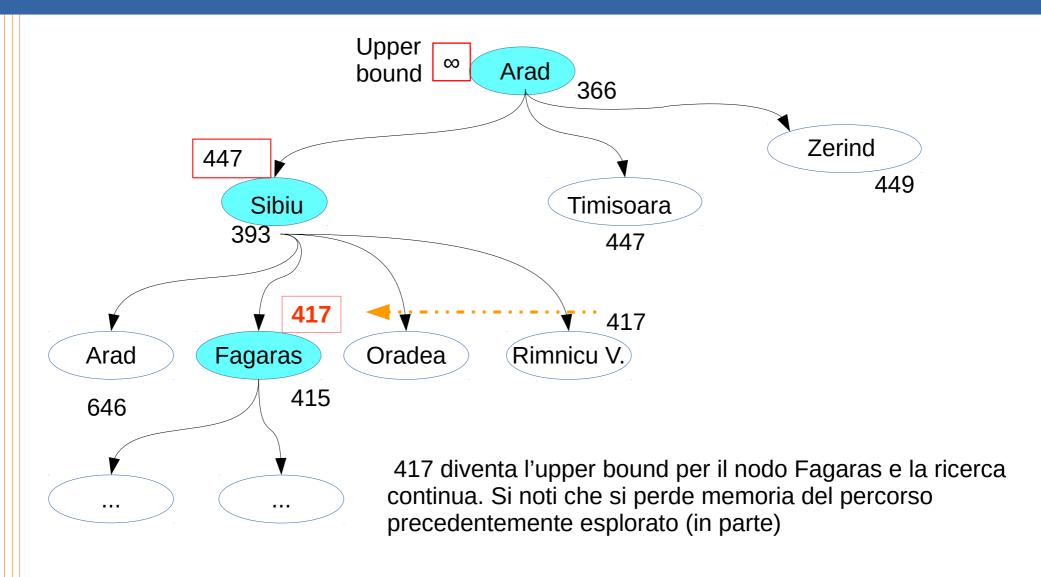
Esempio 3/5



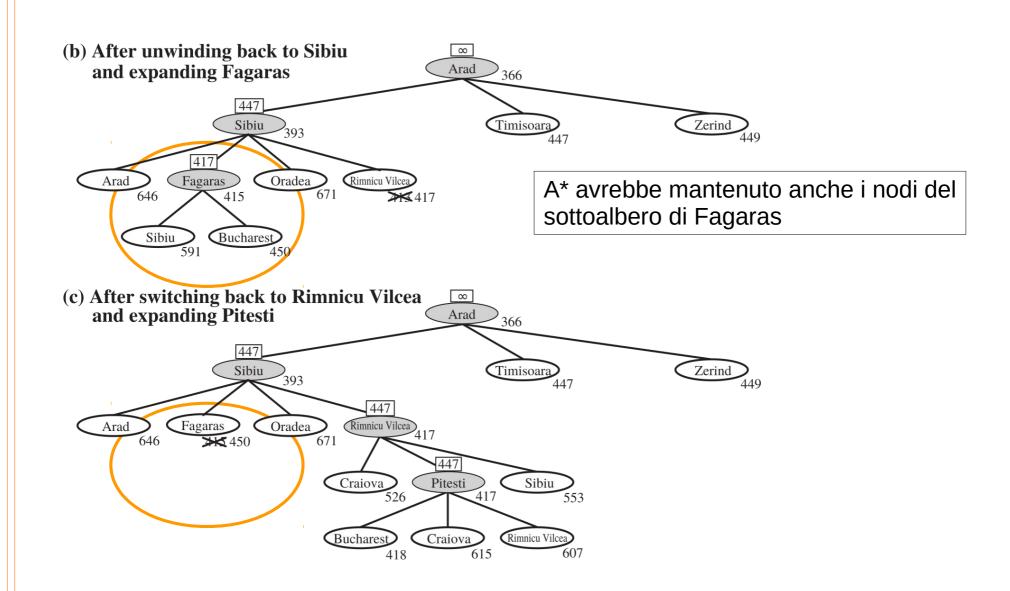
La valutazione di Rimnicu Vilcea viene aggiornato alla valutazione del suo discendente più promettente (Pitesti). I nodi figli di Sibiu sono riordinati di conseguenza. Ora il nodo più promettente è Fagaras

Il sottoalbero di Rimnicu Vilcea è stato rimosso

Esempio 4/5



Esempio 5/5



Valutazione

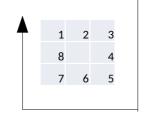
- RBFS è ottimo se l'euristica è ammissibile
- Complessità spaziale lineare: O(bd)
- Complessità temporale: difficile da dafinire perché dipende dall'accuratezza dell'euristica
- Non si accorge di cammini ripetuti
- <u>Sfrutta poco la memoria</u>: Non riesce a sfruttare aumenti nella disponibilità della memoria per incrementare l'efficienza

Altro esempio: gioco dell'8 (vedere da soli)

• Stato iniziale

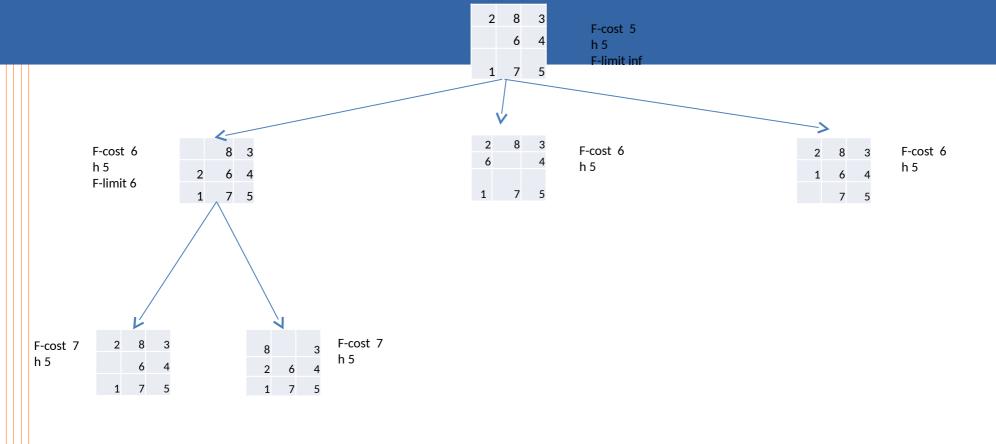


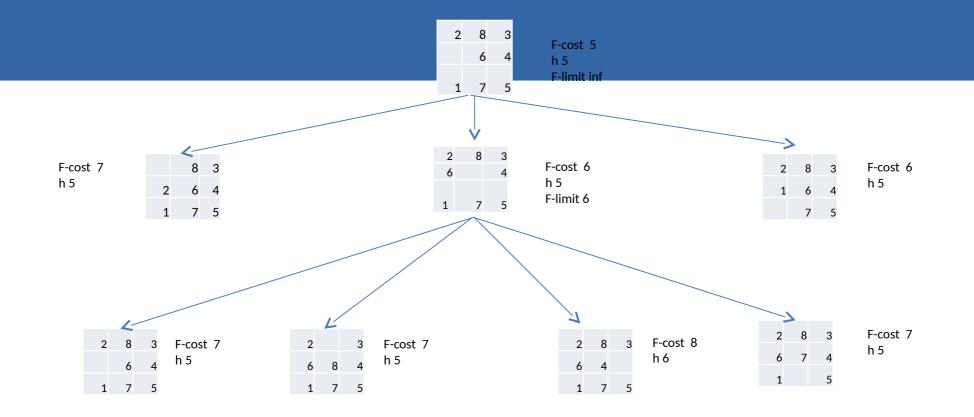
Stato finale

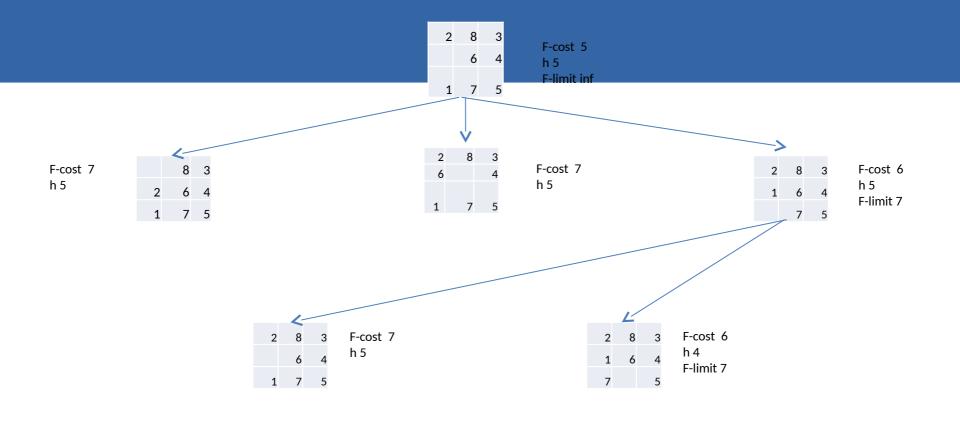


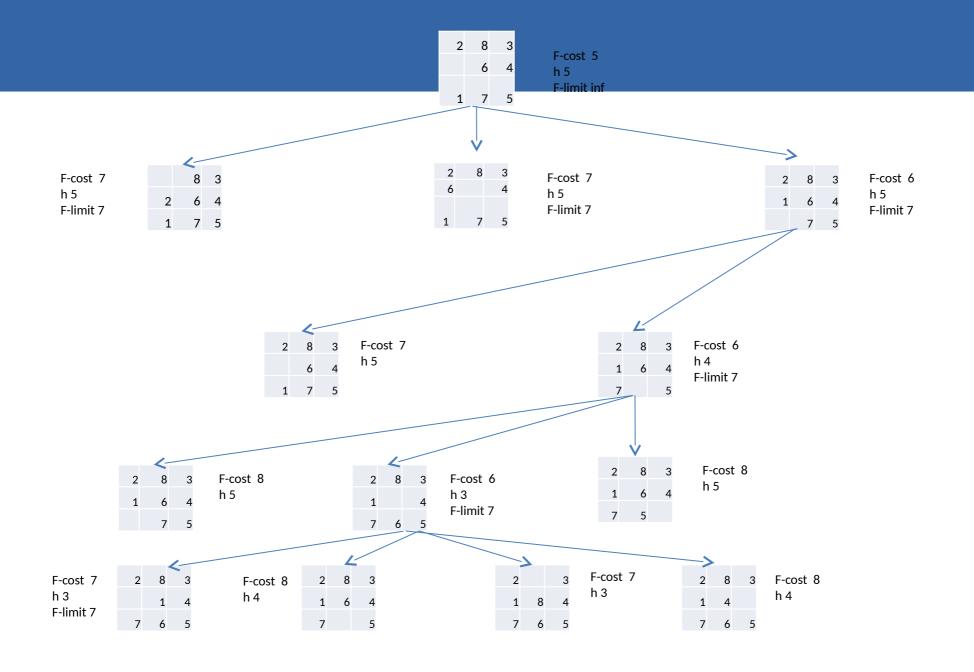
Costo operatori: unitario

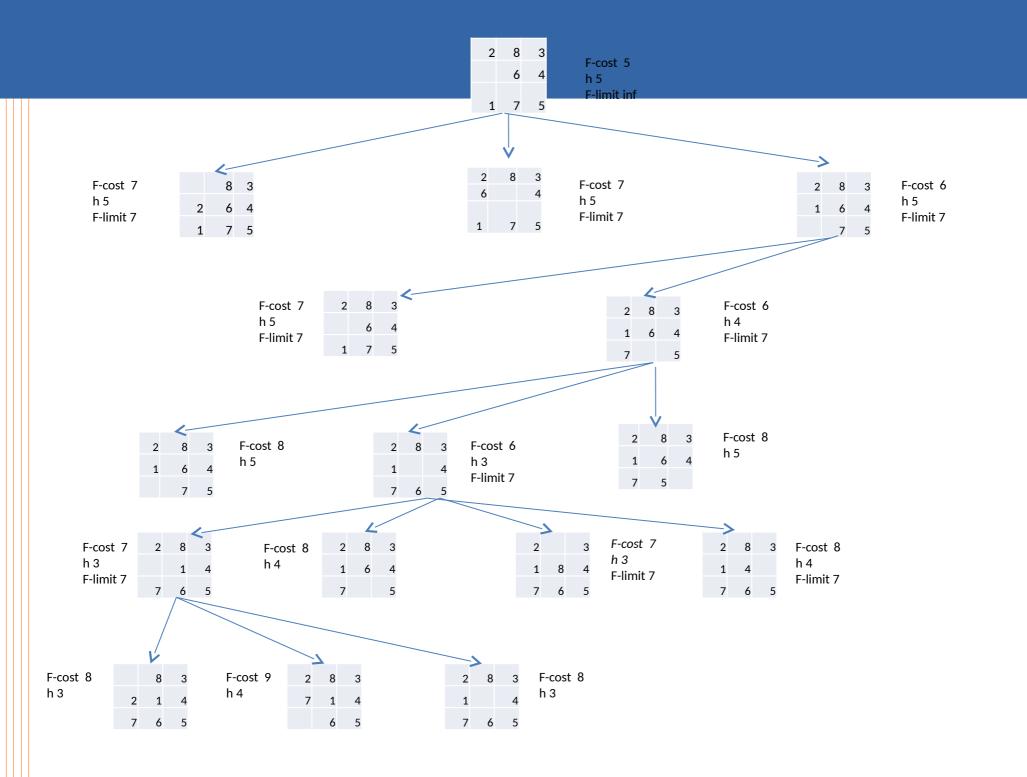
• h = numero tessere fuori posto, esempio: stato iniziale h=5 (sono a posto solo 3, 4 e 5)

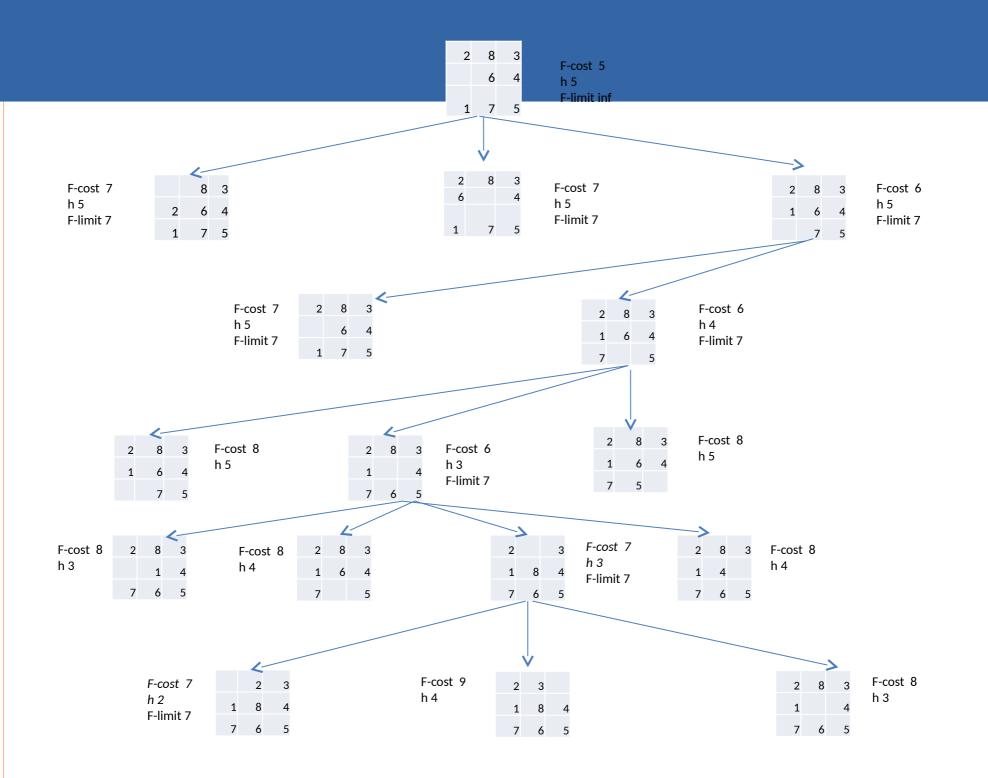












Funzioni euristiche

- Studiamo la natura delle euristiche usando il gioco dell'8, uno dei primi problemi sui quali si è sperimentata la ricerca informata
- In media, generando in modo casuale lo stato iniziale:
 - occorrono 22 mosse per arrivare alla soluzione
 - Il branching factor è pari a 3:



Ho 4 mosse per la tessera centrale



Ho 2 mosse per gli spigoli

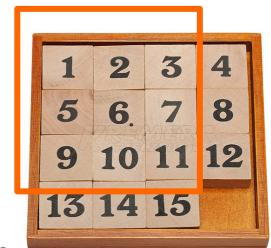


Ho 3 mosse per le tessere sui lati

Albero e grafo esaustivo di ricerca

- Albero esaustivo di ricerca:
 contiene 3²² nodi (oltre 30.000.000.000)
- Grafo esaustivo di ricerca:
 contiene "solo" ~ 180.000 nodi, perché si evitano i duplicati
- E se passiamo dal problema dell'8 al problema del 15? Non sembra molto più complesso, invece:
 - Il grafo esplode: avrebbe circa 10²³ nodi

Problema dell'8



Slocum and Sonnenveld, The 15 Puzzle Book: How it drove the world crazy, 2006

Problema del 15

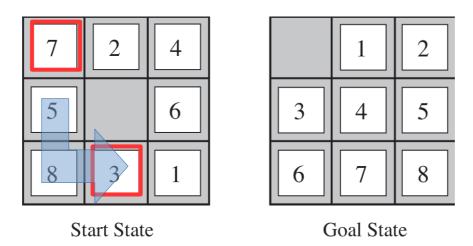
A*: euristiche per il problema del 15 (dell'8)

- A* necessita di euristiche <u>ammissibili</u>, cioè tali da non sovrastimare mai il numero dei passi che portano al goal
- Due possibili euristiche:
 - h1 = numero di tessere fuori posto.
 È ammissibile perché ogni tessera fuori posto deve essere spostata almeno una volta.
 - h2 = distanza di Manhattan (o block distance). È la somma della distanza di una tessera dalla posizione desiderata, contata in numero di tessere attraversate (originariamente di isolati attraversati) sulle ascisse più numero di tessere attraversate sulle ordinate.
 - È ammissibile perché ogni mossa può spostare una tessera al più di una posizione più vicina al goal.



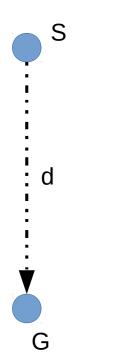
Esempio

- h1(s) = 8
 tutte le tessere sono fuori posto
- h2(s) = 3 + 1 + 2 + ... = 18
 si sommano le distanze di Manhattan calcolate per ogni tessera

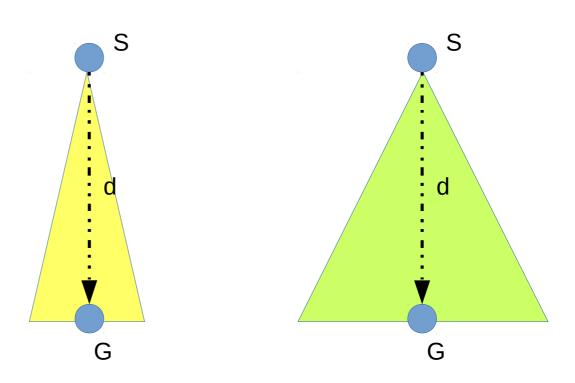


Confronto sperimentale

Scopo: vogliamo decidere quale sia l'euristica migliore basandoci sui risultati derivanti dal loro utilizzo, in particolare i numeri di nodi che uno stesso algoritmo di ricerca informato (nel nostro caso A*) produce

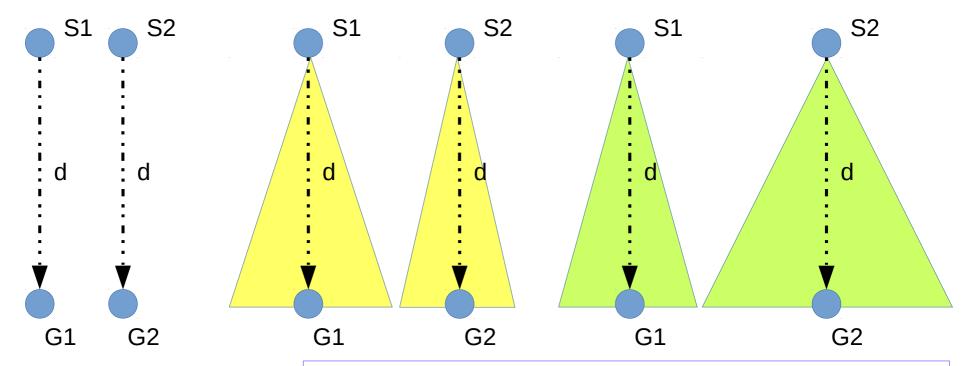


Sia d la profondità della soluzione



Euristiche diverse causeranno in generale l'esplorazione di numeri di nodi differenti.

Ogni istanza produce un risultato: come combinare questi dati?



Se considero due diverse istanze del problema con soluzioni a pari profondità Il numero di nodi espansi su ciascuna istanza del problema cambia perché anche stato iniziale e goal hanno un'influenza. Come confrontare le euristiche in presenza di tanta variabilità?

Valutazione sperimentale

- La valutazione sperimentale delle euristiche h comprende i seguenti passi:
 - Generare un numero significativo di casi
 - Applicare lo stesso algoritmo di ricerca a ogni caso, tante volte quante sono le euristiche da valutare (una per ogni euristica)
 - Raccogliere i dati risultanti (numero di nodi generati, profondità della soluzione ...)
 - Calcolare i valori medi dei risultati ottenuti in casi affini (esempio quelli in cui la profondità della soluzione è la stessa)
 - Valutare e confrontare le prestazioni

Qualità delle euristiche

- La qualità di un' euristica può essere calcolata computando il branching factor effettivo b*
- Supponiamo di avere eseguito A* su un certo problema, siano:
 - N = numero di nodi generati a partire da un nodo iniziale
 - d = profondità della soluzione trovata
- b* = branching factor di un albero uniforme di profondità d che contiene N+1 nodi
- $N+1=1+(b^*)+(b^*)^2+...+(b^*)^d$

Branching factor effettivo (una stima)

$$N + 1 = 1 + b^* + (b^*)^2 + \dots + (b^*)^d$$

$$N + 1 = ((b^*)^{d+1} - 1) / (b^* - 1)$$

$$N \approx (b^*)^d \Rightarrow b^* \approx \sqrt[d]{N}$$

Valutazione sperimentale: qualità delle euristiche

- Qualità di un' euristica calcolata <u>a posteriori</u>, a partire da alcuni casi
 d' uso, cioè problemi in cui viene applicato A*
- Ogni istanza può produrre b* differenti ma tali valori saranno tendenzialmente consistenti
- Quindi bastano alcune misure su un campione (piccolo insieme di problemi) per calcolare la bontà di un' euristica
- Le euristiche migliori hanno <u>b* bassi</u>, vicini a 1
- Esse permettono di risolvere problemi complessi in tempi ragionevoli

Esempio: meglio h1 o h2?

- Sono stati generati in modo casuale 1200 problemi del 15 con profondità di soluzione compresa fra 2 e 24
- Sono stati risolti con Iterative Deepening e A*, usando h1 e poi h2
- I numeri di nodi generati e l'effective branching factor sono stati calcolati caso per caso
- Sono state prodotte le medie per ogni profondità di soluzione

Esempio

	Search Cost (nodes generated)			Effective Branching Factor		
d	IDS	$A^*(h_1)$	$A^*(h_2)$	IDS	A*(h ₁)	$A^*(h_2)$
2	10	6	6	2.45	1.79	1.79
4	112	13	12	2.87	1.48	1.45
6	680	20	18	2.73	1.34	1.30
8	6384	39	25	2.80	1.33	1.24
0	47127	93	39	2.79	1.38	1.22
2	3644035	227	73	2.78	1.42	1.24
1	-	539	113		1.44	1.23
1	-	1301	211		1.45	1.25
1	10 11-12	3056	363		1.46	1.26
	209-13	7276	676		1.47	1.27
	-	18094	1219		1.48	1.28
	-	39135	1641	1991 12	1.48	1.26

Figure 3.29 Comparison of the search costs and effective branching factors for the ITERATIVE-DEEPENING-SEARCH and A* algorithms with h_1 , h_2 . Data are averaged over 100 instances of the 8-puzzle for each of various solution lengths d.

Esempio

	Search Cost (nodes generated)			Effective Branching Factor		
d	IDS	$A^*(h_1)$	$A^*(h_2)$	IDS	$A^*(h_1)$	$A^*(h_2)$
2	10	6	6	2.45	1.79	1.79
1	112	13	12	2.87	1.48	1.45
	680	20	18	2.73	1.34	1.30
1	6384	39	25	2.80	1.33	1.24
1	47127	93	39	2.79	1.38	1.22
1	3644035	227	73	2.78	1.42	1.24
1	The second second	539	113		1.44	1.23
	344 24-44	1301	211		1.45	1.25
	10 11-12	3056	363	200 2000	1.46	1.26
	- 1	7276	676		1.47	
	-	18094	1210		1.47	1.27
	100 100 100 100 100 100 100 100 100 100	39135	1641	1000	1.48	1.26

Compa son of the sea h costs and effective branc ng factors for ne Figure 3.29 ITERATIVE-DEEPENIN -SEARCH and algorithms with h_1 , h_2 . Da are averaged of er 100 instances of the 8-r zzle for each of arious solution lengths d.

con A(h1)

Nodi prodotti → Nodi prodotti con A(h2)

 b^* per A(h1) \gg b^* per A(h2)