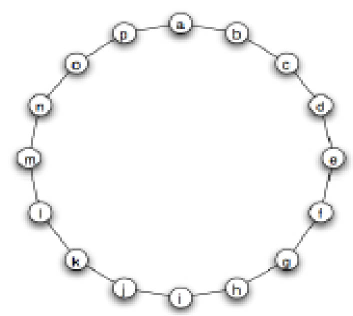
Small World in 1 Dimensione: Ring

Per fare i calcoli in maniera un pò più semplice, poniamo i nodi su una dimensione anziché due.

- Supponiamo di avere n nodi disposti su un ring. I contatti locali ed a lungo raggio di ogni nodo sono definiti come segue.

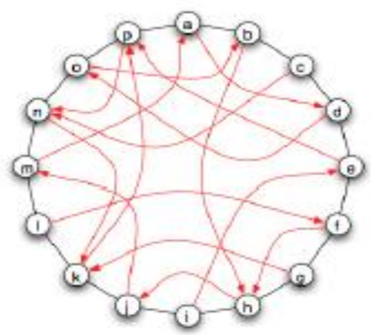
- **Contatti locali**: Ogni nodo v è collegato da archi orientati ai due altri nodi immediatamente adiacenti ad esso.



- **Contatti a lungo raggio**: Ogni nodo v dispone anche di un arco orientato verso qualche altro nodo sull'anello; la probabilità che v sia collegato ad un particolare nodo w è *proporzionale* a

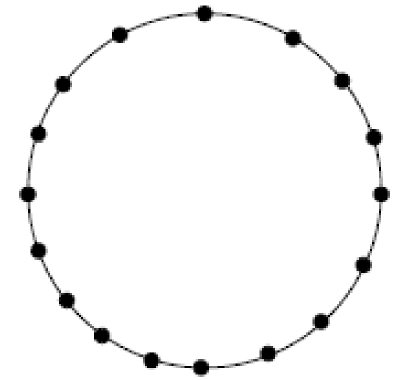
$$d(v, w)^{-1},$$

dove d(v, w) è la loro distanza sul ring.



- **Distribuzione r-armonica p_r, r ≥ 0:**

  Dati due nodi u e v, la probabilità di avere u come il contatto a lungo raggio di v è data da



$$p_r(u,v) = \frac{d(u,v)^{-r}}{\sum_{w \neq u} d(u,w)^{-r}}$$

  dove d (·, ·) è la funzione distanza nella rete sottostante

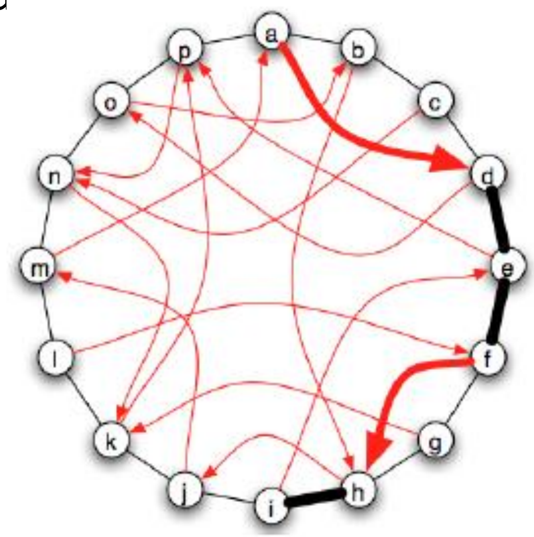
- E’ possibile eseguire uno studio approfondito delle prestazioni di instradamento greedy sul ring, aumentato con i contatti ‘’long range’’, per ogni r ≥ 0.

  In generale si mostra che il migliore esponente per la ricerca in un dominio è uguale alla dimensione del dominio stesso. Quindi, nella nostra analisi unidimensionale useremo un esponente q = 1 (piuttosto che q = 2 come nelle griglie bidimensionali). Scriveremo p(u,v)=P(v→u) per indicare la distribuzione 1-amonica.

Scelti un nodo di partenza casuale *s* e un nodo target casuale *t*, come nella esperimento di Milgram, l'obiettivo è inoltrare un messaggio da *s* a *t*; ogni nodo intermedio conosce solo le posizioni dei propri vicini e di *t*, ma nient'altro sulla rete completa.

Definizione. Ricerca Miope: ogni nodo intermedio v inoltra un messaggio al contatto che si trova *più vicino a t* sull'anello

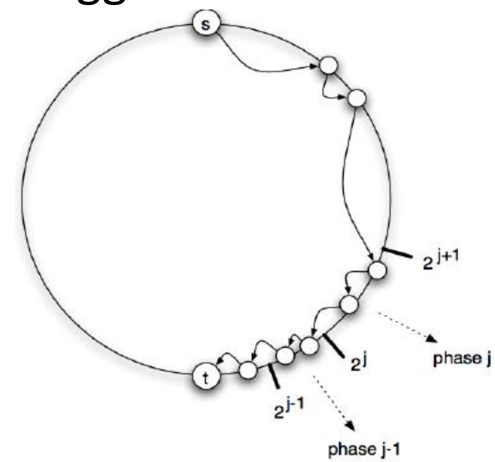
La ricerca miope è una strategia greedy che costituisce una ragionevole approssimazione delle strategie utilizzate dalla maggior parte delle persone nell'esperimento di Milgram



Esempio: nell'anello in figura

- il nodo di origine è *a* e
- il nodo di destinazione è *i*
- Il cammino risultante dalla ricerca miope è evidenziato in grassetto

Vogliamo ora valutare il numero medio di passi utilizzati dal metodo di ricerca miope su di un ring di *n* nodi.
Siano *s* e *t,* rispettivamente, l'origine e la destinazione del messaggio.



IDEA

- Mentre il messaggio si sposta da *s* a *t,* diciamo che esso è nella *fase j* della ricerca se la sua distanza dal target *t* è compresa tra 2^j e 2^{j+1}-1
- Poichè il numero di duplicazioni necessarie per passare da 1 a *n* è al massimo $\log_2 n$, segue che vi sono $\log_2 n$ fasi

Indichiamo con X la variabile casuale che conta il numero di passi effettuati per raggiungete t da s. Sia X_j è la variabile casuale che conta il numero di passi effettuati durante fase j, per ogni j.
È chiaro che il numero di passi totali è dato da $X = X_1 + X_2 + \cdots + X_{\log n}$
Ricordando che il valor medio di una somma di variabili casuali è pari alla somma dei valori medi delle singole variabili casuali, abbiamo

$$\begin{aligned} E[X] &= E[X_1 + X_2 + \cdots + X_{\log n}] \\ &= E[X_1] + E[X_2] + \cdots + E[X_{\log n}]. \end{aligned}$$

Dimostreremo che $E[X_j] \in O(\log n), \ \forall j \leq \log n$

Come prima cosa andiamo a valutare il fattore di normalizzazione nella distribuzione di probabilità dei collegamenti long-range di un nodo v. Ricordiamo, che la distribuzione 1-armonica è definita in modo che

$$P(v \to u) \ = \frac{D(u,v)^{-1}}{\sum_{w \neq v} D(w,v)^{-1}}$$

Ponendo

$$Z = \sum_{w \neq v} D(w,v)^{-1}$$

la frazione 1/Z dà la costante di proporzionalità. Calcoliamo quindi *Z.*

Nell'anello ci sono due nodi a distanza *1* da *v*, due nodi a distanza *2*, e più in generale due nodi ad ogni distanza *d* <*n/2*. Inoltre se *n* è pari, c'è anche un singolo nodo a distanza *n/2* dal *v*. Otteniamo quindi che

$$Z \leq 2\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n/2}\right) \approx 2\ln(n/2) \leq 2\log_2 n.$$

Utilizzando la limitazione ottenuta su Z, abbiamo

$$P(v \to u) \ = \frac{D(u,v)^{-1}}{\sum_{w \neq v} D(w,v)^{-1}} \geq \frac{D(u,v)^{-1}}{2 \ log_2 \ n}$$

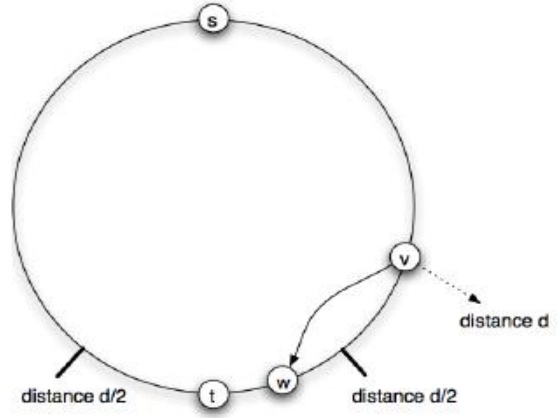
Valutiamo ora il tempo trascorso in fase j, per *j=1,…,log n.*
Ricordiamo che la fase *j* della ricerca si ha quando il messaggio è in un nodo *v* la cui distanza del target *t* ha un valore *d* compreso tra 2^j e 2^{j+1}-1.

Quindi la fase j sarà giunta al termine una volta che la distanza dal target del nodo che ha il messaggio scende sotto 2^j. Vogliamo dimostrare che ciò accade in tempi relativamente brevi.

Notiamo che la fase finisce immediatamente se *r,* il contatto a lungo raggio di *v,* è a distanza *≤ d / 2* da *t* (infatti d/2< 2^{j+1}/2= 2^j). In questo caso, *v* è l'ultimo nodo toccato nella fase j.
Mostriamo ora che un dimezzamento della distanza si verifica con alta probabilità

Sia *I* l'insieme dei nodi a distanza *≤ d/2* da t.
Se uno dei nodi in I è il contatto a lungo raggio di v, allora la fase j termina immediatamente. Valutiamo quindi la probabilità che il contatto a lungo raggio di v sia un nodo in I.

Ci sono *d + 1* nodi in *I*: il nodo t stesso e *d/2* nodi consecutivi su ogni lato di t.
Ogni nodo *w* ε I ha distanza al massimo *3d/2* da v: il più lontano è sul "lato opposto" di t rispetto a v ed ha distanza *d + d / 2*.



Pertanto, ogni nodo *w* in I ha probabilità di essere il contatto a lungo raggio di v data da

$$\frac{D(u,v)^{-1}}{2\, log_2\, n} \geq \frac{1}{(2\, log_2\, n)} \frac{1}{(\frac{3d}{2})} = \frac{1}{3\, d\, log_2\, n}$$

Poiché ci sono più di *d* nodi in I, la probabilità che uno di questi è il contatto a lungo raggio di v è

$$\frac{|I|}{3\, d\, log_2\, n} \geq \frac{1}{3\ \ log_2\, n}$$

Ne consegue che in ogni passo, la fase *j* ha probabilità almeno $\frac{1}{3\ \ log_2\, n}$ di terminare, indipendentemente da quanto è accaduto prima
(in altri termini, la probabilità che la fase j non termina ad un dato passo è limitata superiormente da (1- $\frac{1}{3\ \ log_2\, n}$))

Affinchè nella fase j vengano eseguiti almeno *i* passi , si deve verificare che la fase non deve termina per i - 1 volte consecutive. Quindi la probabilità che la e fase *j* duri almeno *i* passi è

$$\Pr[X_j \geq i] \geq \left(1 - \frac{1}{3\ log_2\ n}\right)^{i-1}$$

Otteniamo quindi che

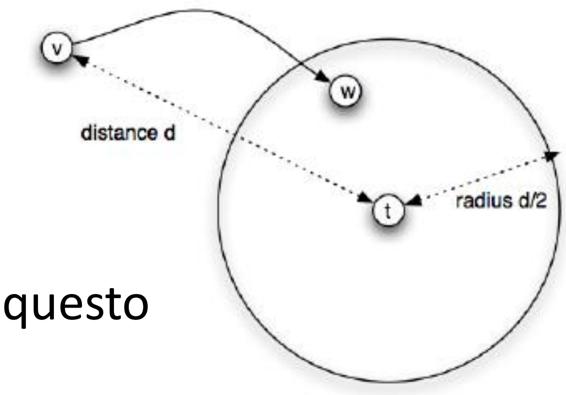
$$E[X_j] = \sum_i i\ \Pr[X_j = i] = \sum_i\ \Pr[X_j \geq i]$$

$$= \sum_i \left(1 - \frac{1}{3\ log_2\ n}\right)^{i-1}$$ (Nota: si sa che $\sum_i\ x^i$=1/(1-x))

$$= \frac{1}{1-\left(1 - \frac{1}{3\ log_2\ n}\right)} = 3\ log_2\ n$$

Small World in 2 dimensioni (2D)

Ritornando al 2D, con il messaggio ad una distanza corrente d dal target t, ancora una volta possiamo guardare l'insieme dei nodi entro distanza d/2 da t.



Noi argomenteremo che la probabilità di entrare in questo insieme in un unico passo è abbastanza grande

L'assunzione di 1 dimensione è stata utilizzata in due casi:

1. quando abbiamo determinato la costante di normalizzazione Z
2. per sostenere che ci sono almeno d nodi a distanza al più d/2 da t

Il fattore d annulla d^{-1} nella probabilità dei link, ciò ci permetteva di concludere che la probabilità di dimezzare la distanza dal target in un passo era almeno proporzionale a 1 / (logn), indipendentemente dal valore di d.

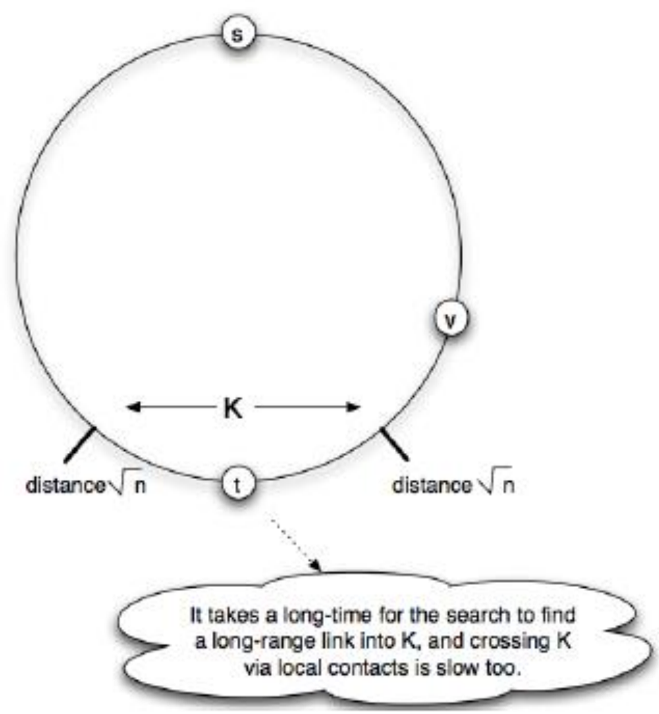
Quando passiamo a 2D, il numero di nodi a distanza al più *d/2* da *t* è proporzionale a d^2. Questo suggerisce che per ottenere la stessa proprietà di cancellazione, dovremmo avere un link da *v* per ogni nodo *w* con probabilità proporzionale al quadrato della distanza; questo è esattamente quello che si usa (cioè esponente q=2).

Perché la ricerca non lavora bene quando q = 0

Riconsideriamo il modello originale di Watts-Strogatz, quando vengono scelti i collegamenti a lungo raggio uniformemente a caso e supponiamo di essere su di un ring.

Consideriamo l'insieme di tutti i nodi entro una certa distanza del target t. Mentre nel caso q= 1 abbiamo visto che è facile entrare in insiemi sempre più piccoli centrati intorno a t, nel caso q = 0, vogliamo identificare un insieme di nodi centrato in t che è in qualche modo "impenetrabile" (molto difficile da raggiungere per la ricerca)

In particolare si può vedere che per q = 0, è difficile attraversare l'insieme dei $\sqrt{n}$ nodi più vicini al target.
A tal fine, denotiamo con *K* l'insieme dei $2\sqrt{n}$ nodi più vicini al target *t*.
Valutiamo ora il tempo necessario affinchè la ricerca riesca ad entrare nell' intervallo *K*.



Poiché i contatti a lungo raggio sono creati in modo uniforme a caso (q = 0), la probabilità che un nodo ha una contatto a lungo raggio in K è $\frac{|K|}{n} = \frac{2\sqrt{n}}{n} = \frac{2}{\sqrt{n}}$
Quindi il numero atteso di passi per trovare un nodo che ha una contatto a lungo raggio in K risulta $\sqrt{n}$ /2 [a)]

Supponiamo che il nodo *s* iniziale sia al di fuori di K. L'algoritmo di ricerca può prendere un link verso un nodo in K oppure deve attraversare K usando i link locali.
Pertanto qualsiasi strategia di ricerca decentralizzata avrà bisogno di almeno $\sqrt{n}$ /2 passi in media, per trovare un nodo avente un contatto a lungo raggio in K.
D'altra parte, finché non trova un collegamento a lungo raggio che porta in K, la ricerca non può raggiungere t in meno $\sqrt{n}$ /2 passi, in quanto questo tempo serve per attraversare K utilizzando solo le connessioni locali.

In conclusione, in nessun modo possiamo effettuare un numero di passi minore di $\sqrt{n}$ /2 in media.

----

a) Nota: Se p è la probabilità che un evento si verifichi in un passo, la probabilità che l'evento si verifichi in esattamente *i* passi è $(1-p)^{i-1}$ p.
Quindi il numero medio di passi da attendere prima che l'evento si verifichi è

$$\sum_{i\geq 1} i\,(1-p)^{i-1}\,p = p\sum_{i\geq 1} i\,(1-p)^{i-1} = \frac{p}{p^2} = \frac{1}{p}$$