Il fenomeno Small-World

Capitolo 20

Il numero di Bacon

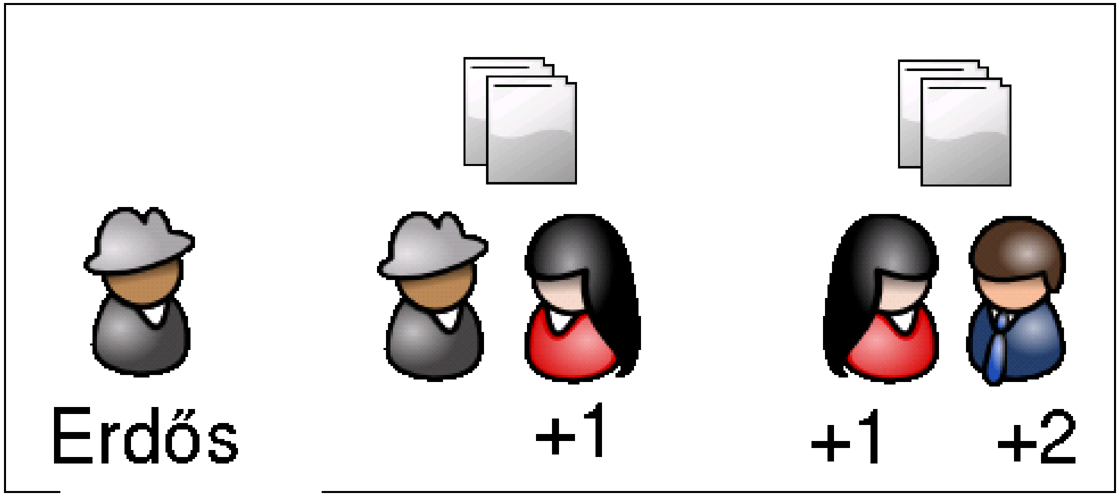
- Rete degli attori di Hollywood
- Due attori collegati se sono co-apparsi in un film
- Numero di Bacon: distanza da Kevin Bacon

Al dicembre 2007, il più alto numero (finito) di Bacon riportato era 8

Solo circa il 12% di tutti gli attori non possono essere collegati a Bacon

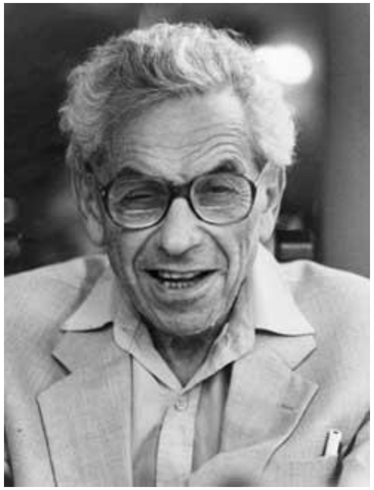
Coauthorship Network

- Nodi: tutti gli autori di almeno una pubblicazione scientifica
- Esiste un legame tra due autori se essi sono coautori in una pubblicazione.



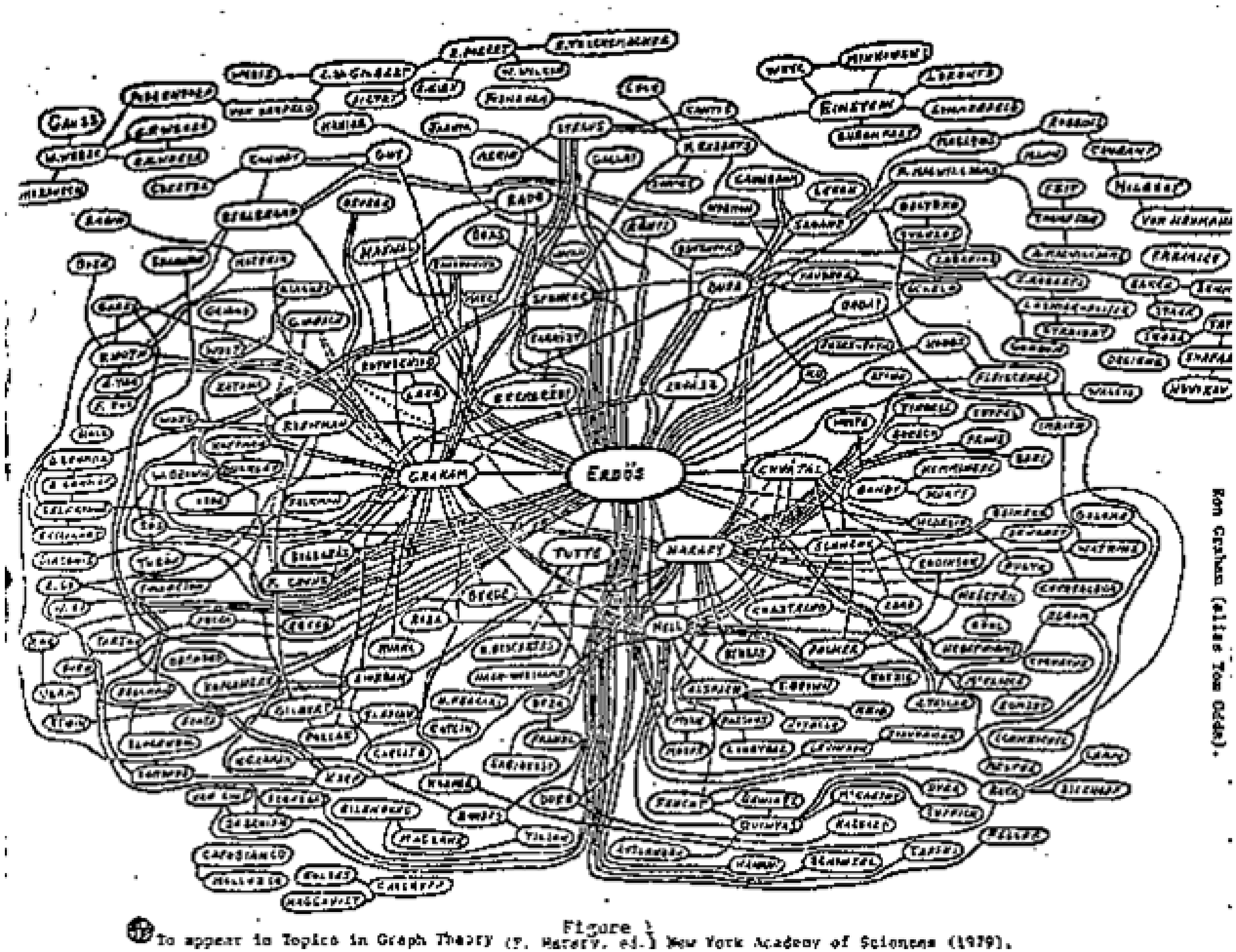
Coauthorship Nework

Il numero di Erdős di un autore è la distanza nella rete con il matematico Paul Erdős.



Importante matematico che ha trascorso gran parte della sua vita in viaggio, spesso ospitato da suoi colleghi, per scrivere i suoi lavori

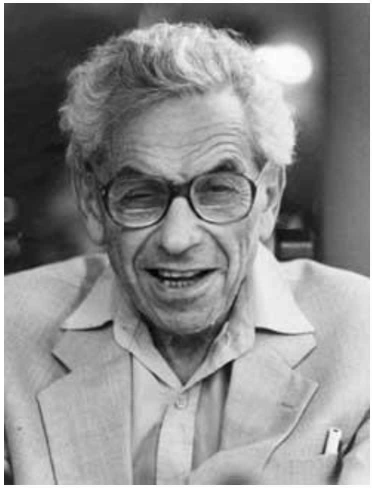
Six degrees of geekiness



- Grafo delle collaborazioni scientifiche centrato su Paul Erdös
  - Quasi tutti i matematici (ed informatici) hanno Erdös numer < 5

Coauthorship Nework

- Il numero di Erdős di un autore è la distanza nella rete con il matematico Paul Erdős.



Importante matematico che ha trascorso gran parte della sua vita in viaggio, spesso ospitato da suoi colleghi, per scrivere i suoi lavori

Erdös number	0	---	1 person
Erdös number	1	---	504 people
Erdös number	2	---	6593 people
Erdös number	3	---	33605 people
Erdös number	4	---	83642 people
Erdös number	5	---	87760 people
Erdös number	6	---	40014 people
Erdös number	7	---	11591 people
Erdös number	8	---	3146 people
Erdös number	9	---	819 people
Erdös number	10	---	244 people
Erdös number	11	---	68 people
Erdös number	12	---	23 people
Erdös number	13	---	5 people

Due persone sono collegate se sono coautori di un articolo.

Small world: Il problema della navigazione

- Sei un individuo (nodo) in una grande social network
  - Vuoi trovare una (breve) catena di amicizie verso un altro individuo
  - Non hai computer potenti e una visione globale (a volo d'uccello)
  - Tutto ciò che tu sai è: chi sono i tuoi vicini/amici ed alcune info su di loro (età, interessi, religione, indirizzo, lavoro, ...)
  - Come lo faresti?

- Questo problema è noto anche come ricerca nelle reti o "small world problem’’

- Un diametro piccolo è necessario ma non sufficiente

  ... la navigazione è un problema algoritmico

  correlato al problema del routing dei pacchetti di dati in Internet

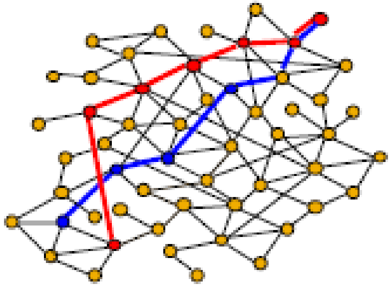
Sei gradi di separazione

L'esperimento (Milgram, 1967):

chiese a 300 persone a caso dal Nebraska e Kansas di inviare una lettera (tramite intermediari) per un agente di borsa a Boston.

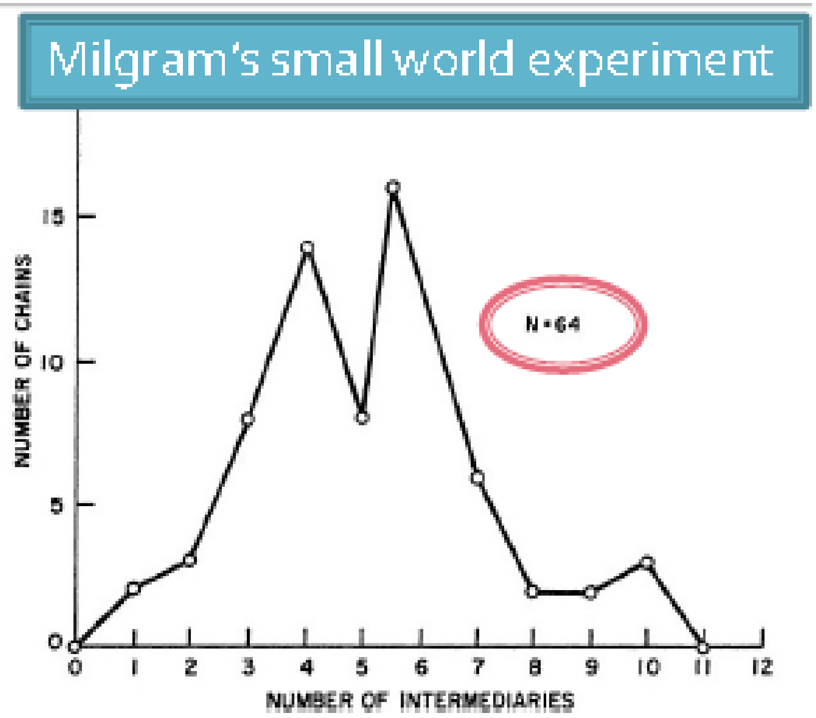
Ad ogni persona (o intermediario) era permesso di inviare la lettera esclusivamente ad un amico o conoscente.

Stanley Milgram (1933-1984)



Sei gradi di separazione

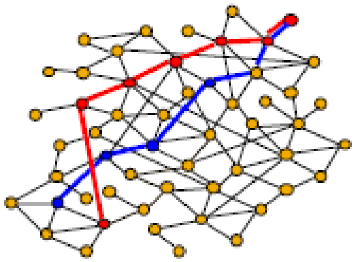
- 64 catene completate: (Vale a dire, 64 lettere hanno raggiunto l'obiettivo)
- Ci sono voluti 6,2 passi sul media, da cui "6 gradi di separazione"



Sei gradi di separazione

Note

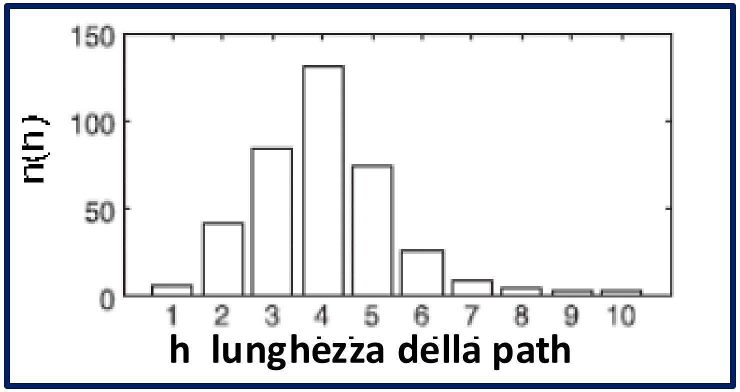
1. Punti di partenza e l'obiettivo non erano casuali
2. Non ci sono molti campioni (solo 64)
   – La gente ha rifiutato di partecipare (25% per Milgram)
   – Non tutte le ricerche finite (solo 64 su 300)
3. **Una sorta di social search**:
   – Persone non inoltrano la lettera a tutti, seguono una qualche strategia, e potrebbero avere utilizzato informazioni aggiuntive
   – **Non stanno cercando il percorso più breve!**



Small-world: esperimento via email

Nel 2003 Dodds, Muhamad e Watts eseguono l'esperimento via e-mail:

- 18 target di varia tipologia
- 24.000 primi passi (~ 1.500 per target)
- 65% di abbandono per passo
- 384 catene portate a termine (1,5%)



- Aggregando le 384 catene completate su tutti i target, la lunghezza media di una catena risulta 4.05
- Problema: la gente smette di partecipare le catene si interrompono (quindi catene brevi terminano con più probabilità)
- Per ovviare si introduce un fattore di correzione
- lunghezza del percorso **h = 7**

Navigazione small-world

Affinché le persone (o le macchine) trovino cammini brevi nelle reti:

- devono esistere cammini brevi (diametro piccolo)
- le persone devono essere in grado di trovare tali cammini tramite il solo l'inoltro a livello locale (algoritmico)

Nota: I vincoli algoritmici sono forti
- Il nodo conosce solo i propri vicini nella rete
- ha informazioni limitate sul target/destinazione (posizione fisica, qualche background)

Serve un modello che incorpori sia i vincoli strutturali che algoritmici

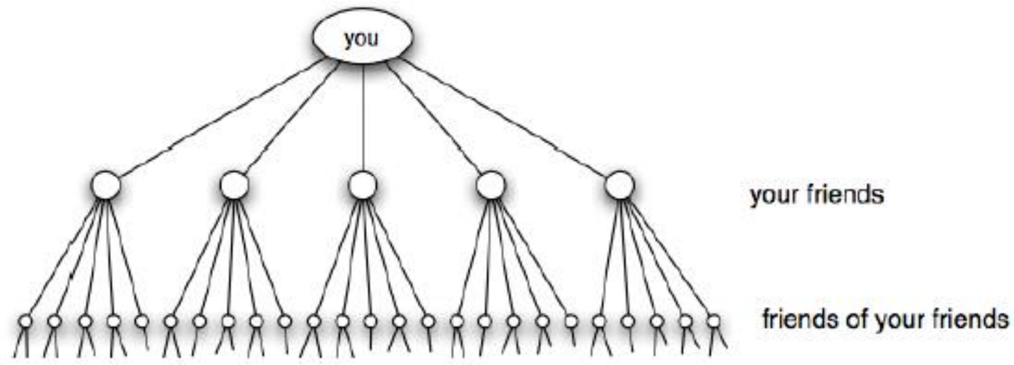
Come possiamo trovare una persona?

Potremmo

chiedere a tutti i nostri amici di

chiedere a tutti i loro amici di

chiedere a tutti i loro amici….



Assumiamo che ognuno è collegato ad altre 100 persone:

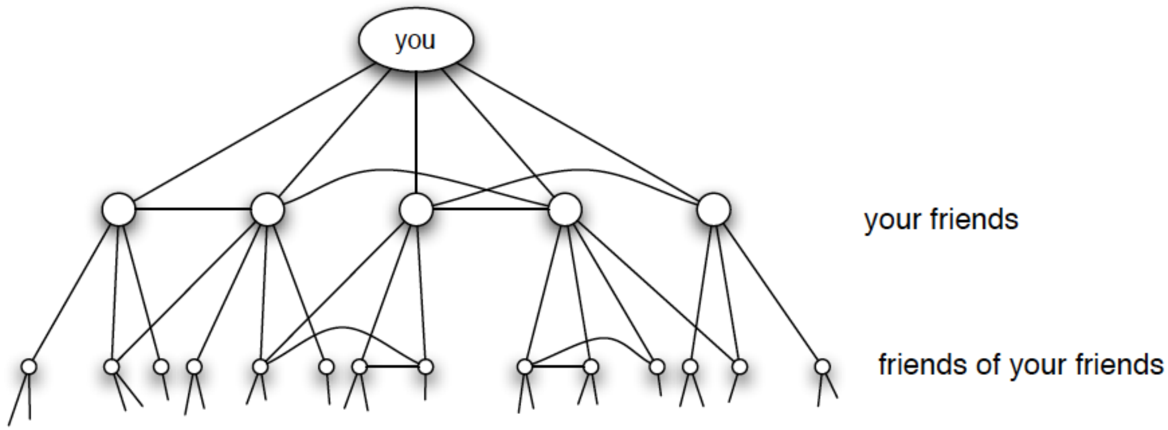
- Passo 1: possiamo raggiungere 100 persone
- Passo 2: possiamo raggiungere 100 * 100 = 10.000 persone
- Passo 3: possiamo raggiungere 100 * 100 * 100 = 1.000.000 di persone
- Passo 4: possiamo raggiungere 100 * 100 * 100 * 100 = 100 milioni di persone

In teoria in 5 passi potremmo raggiungere 10 miliardi di persone,

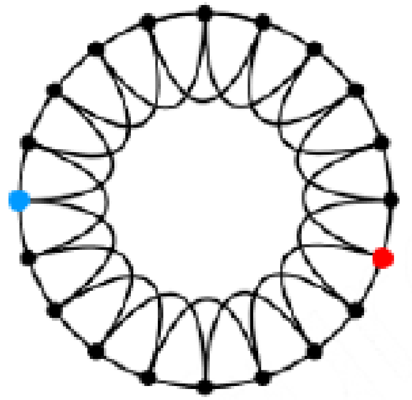
In pratica la chiusura triadica riduce molto il tasso di crescita

Come possiamo trovare una persona?

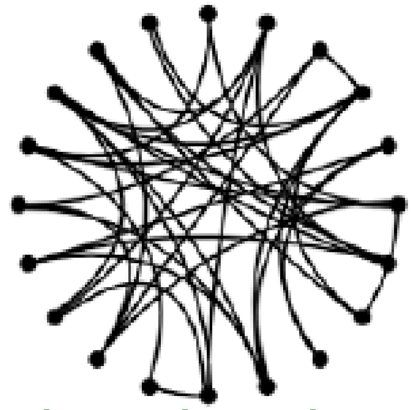
In pratica la chiusura triadica riduce molto il tasso di crescita



Small world: come?



High clustering
High diameter



Low clustering
Low diameter

- Possiamo avere un modello di rete che esibisce
  - sia **molti triangoli** (clustering elevato)
  - che **percorsi molto brevi**?

Modello di Watts Strogatz

Modello di Watts Strogatz

combinazione di

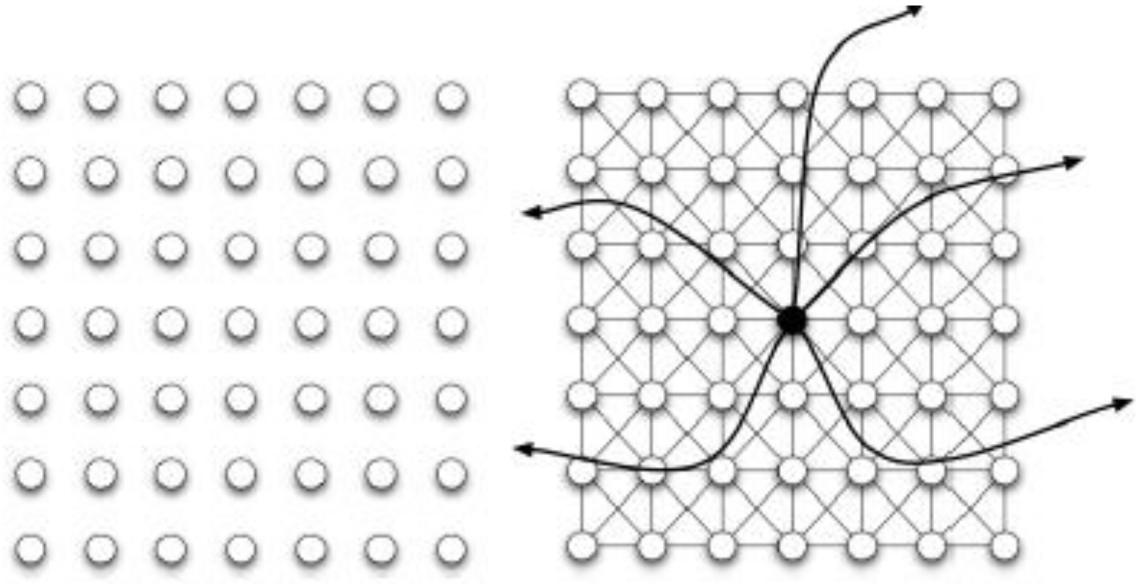
- **Legami forti (Omofilia)**: la tendenza a stabilire contatti e relazioni con i propri simili crea molti triangoli,
- **Legami deboli** producono una struttura ampiamente ramificata che raggiunge molti nodi in pochi passi.

Supponiamo che ognuno viva su una griglia bidimensionale. Creiamo una rete dando ogni nodo due tipi di link:

- quelli spiegabili per omofilia, e
- quelli che costituiscono legami deboli.

Modello di Watts Strogatz

- **Omofilia**. Questo è catturato dall' avere ogni nodo collegato a tutti gli altri nodi che si trovano entro un raggio di r passi su una griglia, per un valore costante di r.

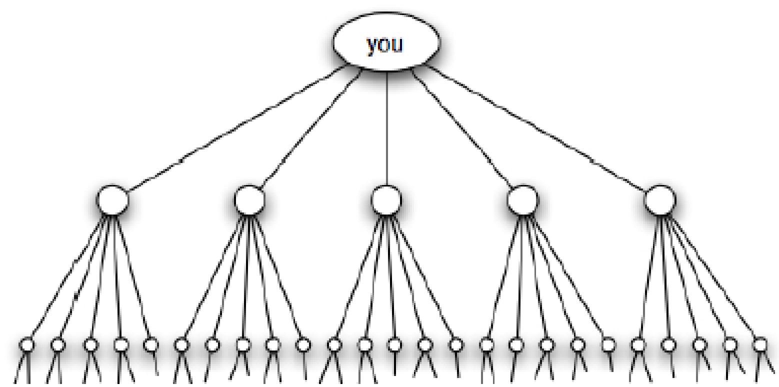


- **Legami deboli.** Per qualche costante k, ogni nodo crea anche collegamenti a k altri nodi selezionati in modo uniforme a caso dalla rete.

Idea per l’esistenza di path brevi

- Inizia tracciando cammini verso l'esterno da un nodo di partenza v, utilizzando solo i k legami deboli su ogni nodo.
- Dal momento che questi collegamenti sono scelti in modo uniforme a caso, è molto improbabile vedere un nodo due volte nei primi passi allontanandosi da v.

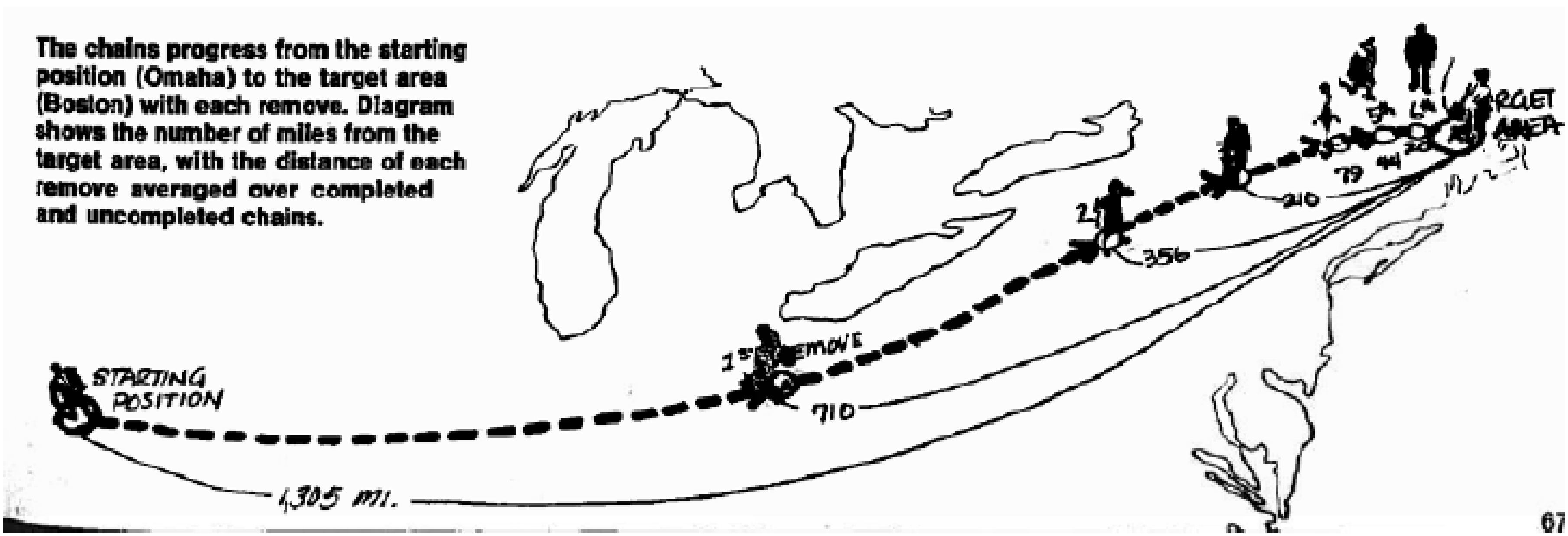
  Quasi sicuramente non vi è alcuna chiusura triadica, quindi un gran numero di nodi vengono raggiunti in pochi passi.



- Ne deduciamo che introdurre una piccola quantità di casualità - nella forma legami deboli di lungo raggio- è sufficiente a rendere il mondo "piccolo« (cioè con brevi percorsi tra ogni coppia di nodi)
- **Rimane la domanda: come trovare percorsi brevi facilmente?**

Idea per l'esistenza di path brevi

- Perché nell'esperimento di Milgram, le persone erano in grado di trovare collettivamente percorsi brevi per il target designato?
- Questa è un'immagine dall'articolo originale di Milgram, mostra un ''unione" dei percorsi completati convergenti sul target

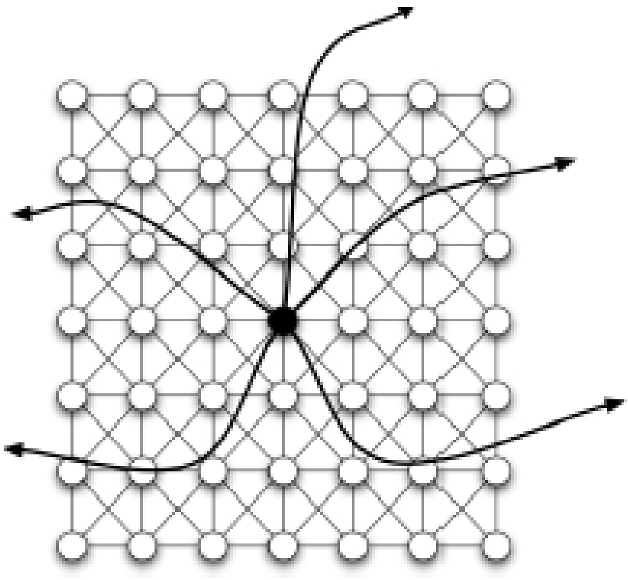


- Ogni passo intermedio è posizionato alla distanza media di tutte le catene che hanno completato quel numero di passi

Un modello per la Ricerca Decentralizzata

- Possiamo costruire una rete casuale in cui l'instradamento decentrato riesce, e se sì, quali sono le proprietà che sono cruciali per il successo?

- Prendiamo in considerazione il modello di Watts -Strogatz e supponiamo che un a un nodo *s* viene dato un messaggio che deve trasmettere ad un nodo target *t*.



Un modello per la Ricerca Decentralizzata

- Possiamo costruire una rete casuale in cui l'instradamento decentrato riesce, e se sì, quali sono le proprietà che sono cruciali per il successo?

- Prendiamo in considerazione il modello di Watts -Strogatz e supponiamo che un a un nodo *s* viene dato un messaggio che deve trasmettere ad un nodo target *t*.

  – Inizialmente *s* conosce solo la posizione di *t* sulla griglia di partenza, ma, soprattutto, *non conosce gli archi casuali di qualsiasi altro nodo*.
  – Ogni nodo intermedio lungo il percorso ha le informazioni parziali, e deve scegliere a quale dei suoi vicini inviare il messaggio successivo.

Un modello per la Ricerca Decentralizzata

- Queste scelte costituiscono una procedura collettiva per reperire percorso da *s* a *t*.

- Dato questo scenario, si può dimostrare che la ricerca decentralizzata nel modello di Watts-Strogatz richiede necessariamente *un gran numero di passi per raggiungere un obiettivo.*

- La ragione è semplice: - *Il modello non "impone" alcun tipo di progresso verso la destinazione!*

Un modello per la Ricerca Decentralizzata

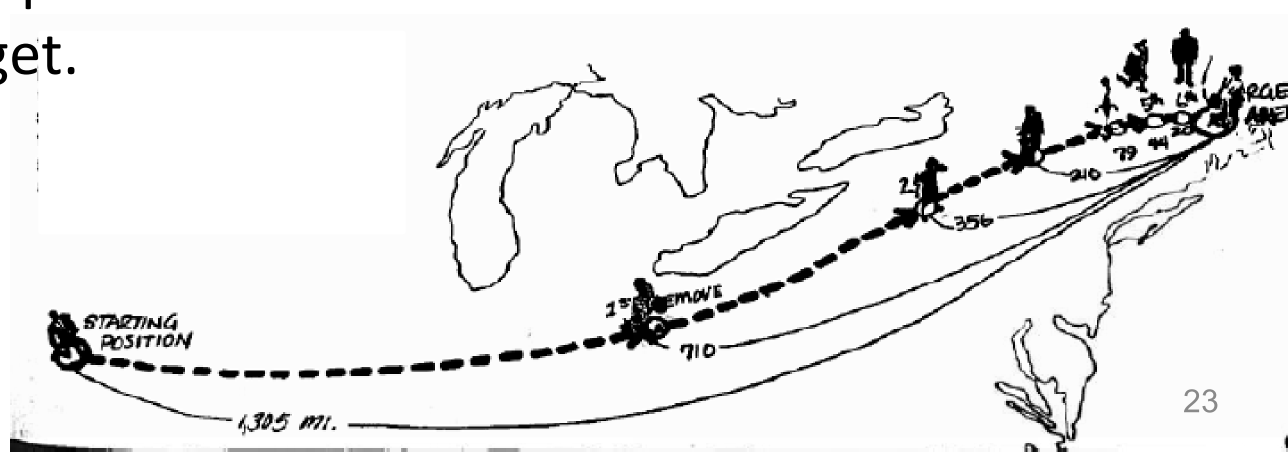
Se consideriamo la griglia bidimensionale in cui ogni nodo ha 1 arco casuale, si dimostra che

- Questo è un grafo small-world (small-world= diametro *O (log n)*)

- Ma si dimostra anche che
  - un algoritmo di ricerca decentrata nel modello di Watts-Strogatz ha bisogno $n^{2/3}$ passi per raggiungere il target *t, anche se i percorsi di O(log n) passi esistono!*

Un modello per la Ricerca Decentralizzata

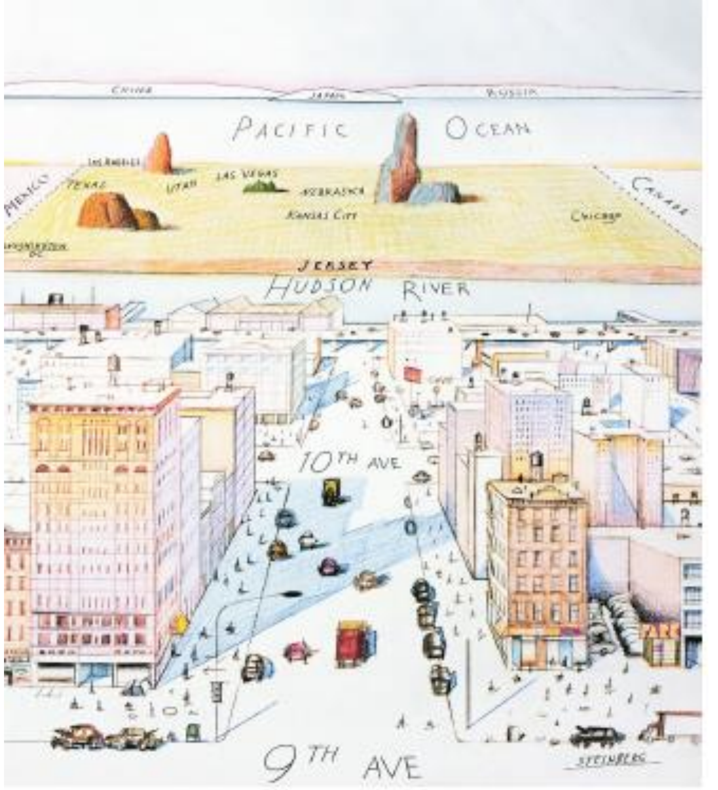
- l modello Watts-Strogatz è efficace nel catturare la densità di triangoli e l'esistenza di percorsi brevi***, ma <u>non</u> la capacità delle persone***, lavorando insieme in rete, **di trovare effettivamente i percorsi**.
- Il problema è che i legami deboli che rendono il mondo piccolo sono "**troppo casuali**" nel modello
- Una soluzione semplice di Milgram (1967):
  - Per raggiungere un obiettivo lontano, si devono usare legami deboli a lungo raggio in un modo metodico abbastanza strutturato, che permettono una costante riduzione della distanza dal target.



Un modello per la Ricerca Decentralizzata

- Come possiamo ottenere small-world navigabili?

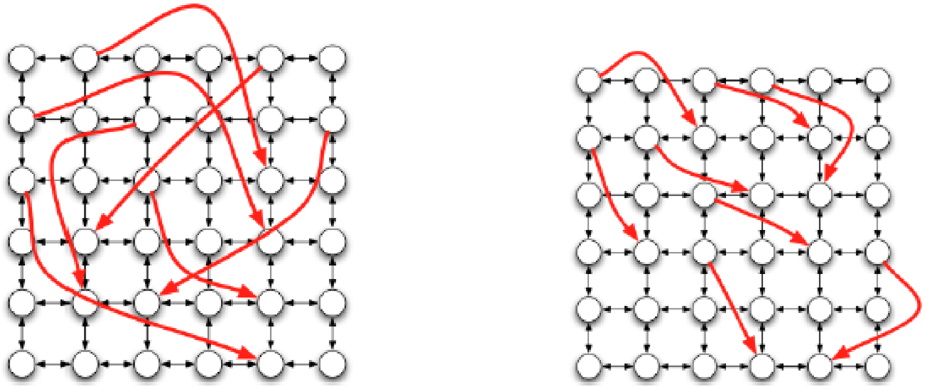
- Intuizione:
  - I nostri collegamenti a lungo raggio non devono essere casuali ma in qualche modo adeguarsi alla geografia!



Saul Steinberg, "View of the World from 9th Avenue"

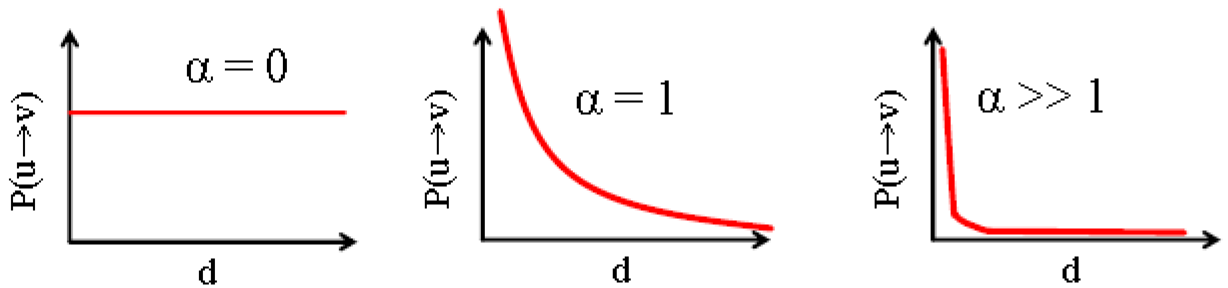
Modello di Kleinberg

- Abbiamo una griglia e ciascun nodo ha archi ad ogni altro nodo ad al più r passi di griglia.
- Ciascuna dei suoi k archi casuali viene generato in modo che la probabilità decade con la distanza, in base ad un parametro q, come segue.
  - Per due nodi *v* e *w*, sia *d(v, w)* la loro distanza sulla griglia (il numero di passi sui rami della griglia).
  - Generiamo un arco casuale da *v* a *w* con probabilità proporzionale a $d(v,w)^{-q}$.
  - Vale a dire, *selezioniamo* *w* a distanza *d* dal *v* con probabilità d^{-q}



Modello di Kleinberg

- Così abbiamo un modello diverso per ciascun valore di *q*.
- Il modello originale basato sulla griglia Watts-Strogatz corrisponde a *q = 0,* cioè i link deboli sono scelti in modo uniforme a caso

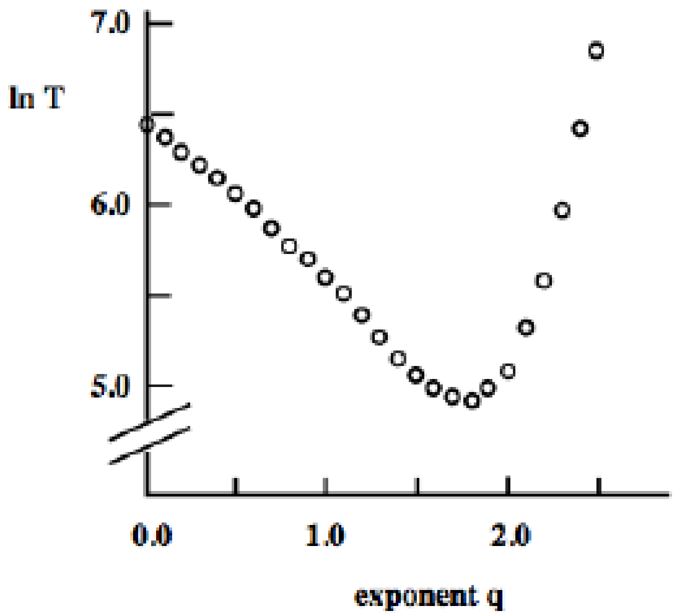


α=q

- Quando q è molto piccolo, i collegamenti a lungo raggio sono "troppo casuali" e non possono essere utilizzati efficacemente per la ricerca;
- quando q è grande, i collegamenti a lungo raggio sono "non abbastanza" per fornire abbastanza dei salti a lunga distanza che creano un grafo small-world

Ricerca decentralizzata più efficiente : q = 2

- Simulazione della ricerca decentralizzata con diversi valori di q, per un rete di diverse centinaia di milioni di nodi.
- asse x è l'esponente q, asse y è il tempo di consegna ln T.
- Il risultato principale di questo modello è che, per reti di grandi dimensioni, la ricerca decentralizzata è più efficiente quando q = 2



Theorem

(J. Kleinberg 2000)

(a) *For $0 \leq q < 2$, the (expected) delivery time T of any "decentralized algorithm" in the $n \times n$ grid-based model is $\Omega\left(n^{\frac{2-q}{3}}\right)$.*

(b) *For $q = 2$, there is a decentralized algorithm with delivery time $O(\log n)$.*

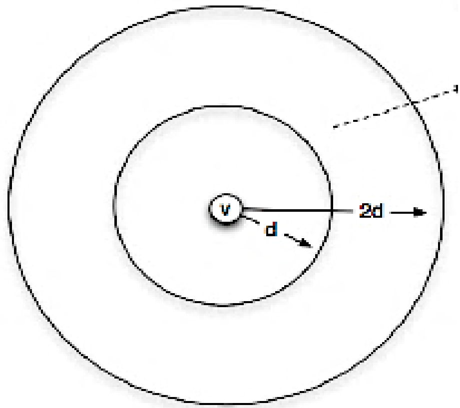
(c) *For $q > 2$, the delivery time of any decentralized algorithm in the grid-based model is $\Omega\left(n^{\frac{q-2}{q-1}}\right)$.*

(The lower bounds in (a) and (c) hold even if each node has an arbitrary constant number of long-range contacts, rather than just one.)

Ricerca decentralizzata più efficiente : q = 2

Perché *q = 2* è ottimale?

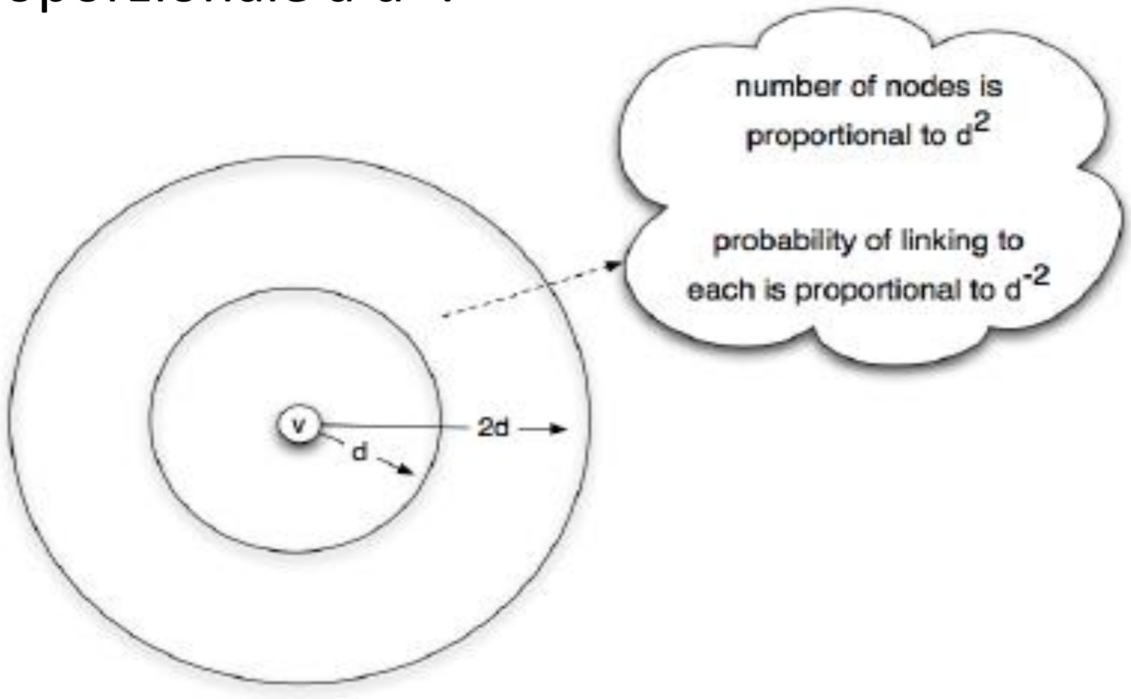
- **Possiamo organizzare distanze in diverse "scale di risoluzione":**
  – qualcosa può essere in tutto il mondo, in tutto il paese, attraverso lo Stato, dall'altra parte della città, ...
- **Un modo ragionevole per pensare a queste scale di risoluzione in un modello di rete è quello di considerare i gruppi di tutti i nodi a gamme di distanza da *v* sempre maggiori:**
  – nodi a distanza da 2 a 4, da 4 a 8, 8 a 16, e così via.
- **Come si rapporta *q = 2* con queste scale di risoluzione?**
  – Consideriamo il gruppo (anello in figura) di nodi a distanze comprese tra *d* e *2d* da *v*.
  Qual è la probabilità che v abbia un link verso alcuni nodi all'interno dell'anello?



Ricerca decentralizzata più efficiente : q = 2

- Il numero totale di nodi in questo gruppo *A* è proporzionale a d^2.

- Per $u \in A$, la probabilità *Pr [v links to u]* che v ha un link verso u, varia a seconda di quanto lontano è u, ma ogni probabilità individuale è proporzionale a d^{-2}.



Ricerca decentralizzata più efficiente : q = 2

- Ne consegue che

  il numero di nodi nel gruppo, e

  la probabilità di collegamento a uno qualsiasi di essi

  si annullano (circa)

- Quindi,

$$\begin{aligned}\Pr[\exists u \in A (v \text{ links to } u)] &= \Pr\left[\bigcup_{u \in A} (v \text{ links to } u)\right] \\ &= \sum_{u \in A} \Pr[v \text{ links to } u]\end{aligned}$$

- ... e la probabilità è approssimativamente *indipendente dal valore di d.*
- In altre parole, se si parte da v, con *probabilità costante* si sfugge alla corona circolare (rendendo così possibile il progresso)

Distribuzione armonica

- **Distribuzione r-armonica p_r, $r \geq 0$:**

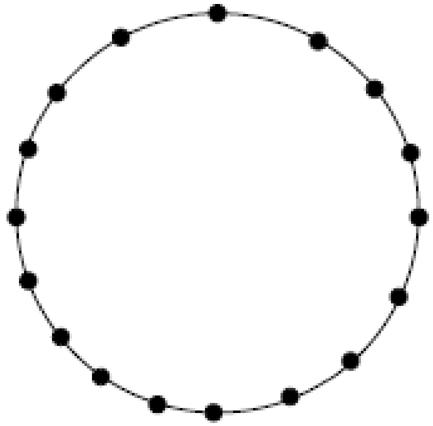
Dati due nodi *u* e *v*, la probabilità di avere *u* come il contatto a lungo raggio di *v* è data da

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

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

- Distribuzione uniforme: ottenuta per *r = 0,* *p(u, v) = 1 / (n-1),*

Small World in 1 Dimensione

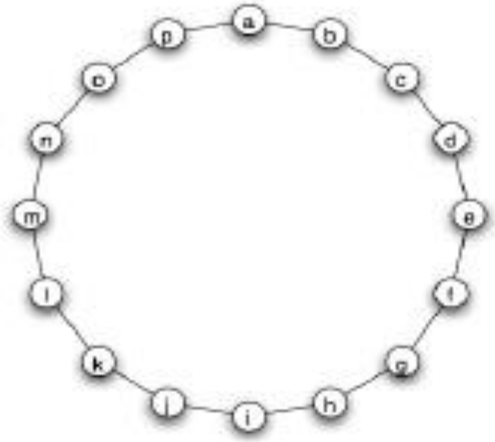


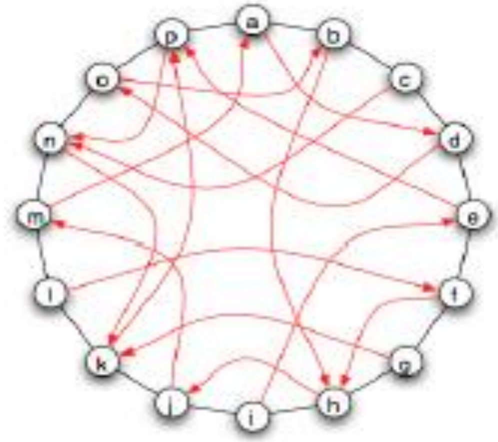
- Per fare i calcoli in maniera un po' più semplice, poniamo i nodi su una dimensione anziché due.

- In generale il migliore esponente per la ricerca in un dominio è uguale alla dimensione.
  Quindi, nell' analisi unidimensionale useremo l'esponente q = 1 (piuttosto che q = 2)

- E' possibile eseguire uno studio approfondito delle prestazioni di instradamento greedy sul ring aumentato con k contatti ''long range'' , per ogni $k \geq 0$. Noi considereremo il caso k=1.

Ring





- *n* nodi disposti su un ring
- **Contatti locali**: Ogni nodo *v* è collegato da archi orientati ai due altri immediatamente adiacente 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.

Ring

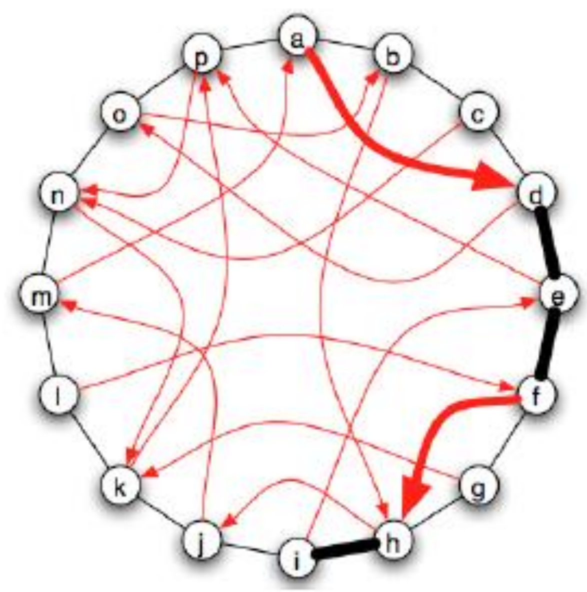
Scelti un nodo di partenza casuale *s* e un nodo target casuale *t,*

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 sull'intera a rete.

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

È una ragionevole approssimazione le strategie utilizzate dalla maggior parte d persone nell'esperimento di Milgram



Esempio: nella figura

- il nodo di origine è *a* e
- il nodo di destinazione è *i*.

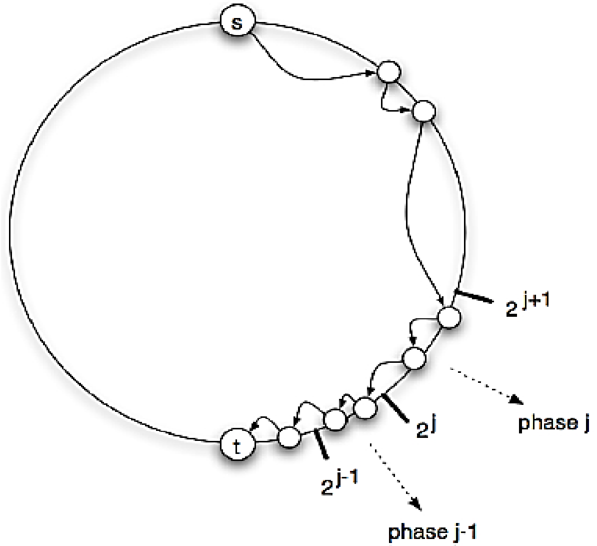
Idea

FASI

- Mentre il messaggio si sposta da *s* a *t*, diciamo che esso è nella *fase j* della ricerca se la sua distanza dal target è

  tra 2^j e 2^{j+1}

- Il numero di fasi è al massimo $\log_2 n$ (cioè il numero di duplicazioni necessarie per passare da 1 a *n*)



MEDIE

Sia X_j è la v.c. che conta il numero di passi nella fase *j*.

Il numero di pa $X = X_1 + X_2 + \cdots + X_{\log n}$

Poichè

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

risulta sufficiente a dimostrare che $E[X_j] \in O(\log n),\ \forall j \leq \log n$

Il fattore di normalizzazione

- Il nodo *v* forma il suo collegamento a lungo raggio verso w con probabilità proporzionale a $d(u,v)^{-1}$, ma qual è la costante di proporzionalità?

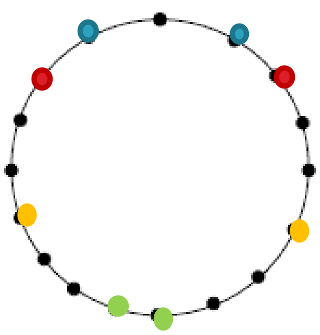
- Ricordiamo, che la distribuzione 1-armonica è definita in modo che

$$p_1(u,v) = \frac{1}{\sum_{w \neq u} d(u,w)^{-1}} d(u,v)^{-1}$$

- Ponendo

$$Z = \sum_{w \neq u} d(u,w)^{-1}$$

la frazione 1/Z dà la costante di proporzionalità



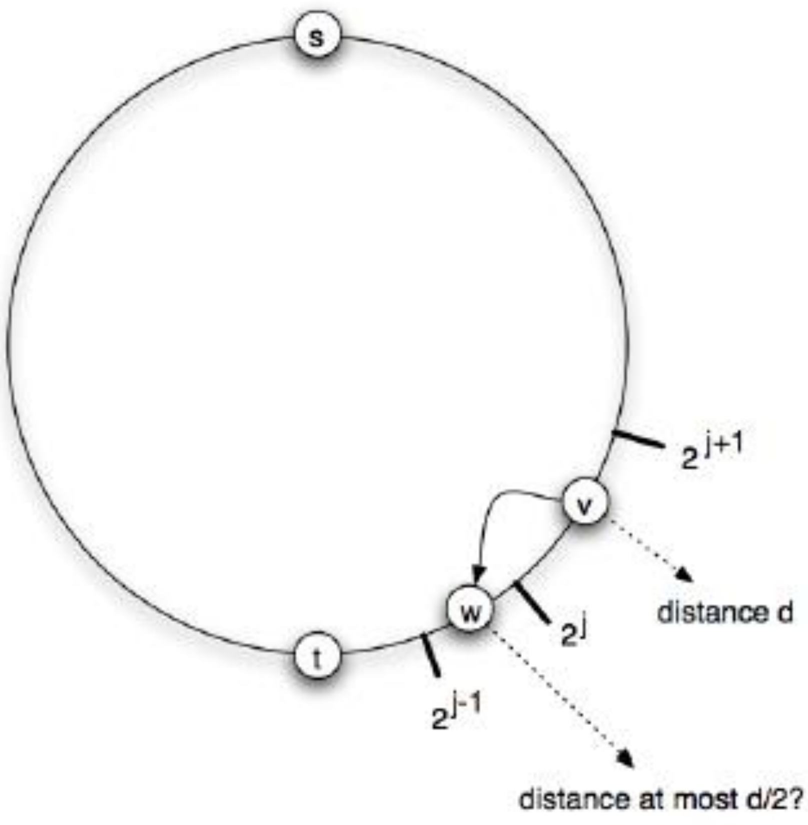
- Nell'anello ci sono **due nodi a distanza *1*** da *v*, **due a distanza *2*,** e, più in generale, **due nodi ad ogni distanza *d <n/2*.**
  Se *n* è pari, c'è anche un singolo nodo a distanza *n/2* dal *v*.
  Di conseguenza

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

Il fattore di normalizzazione

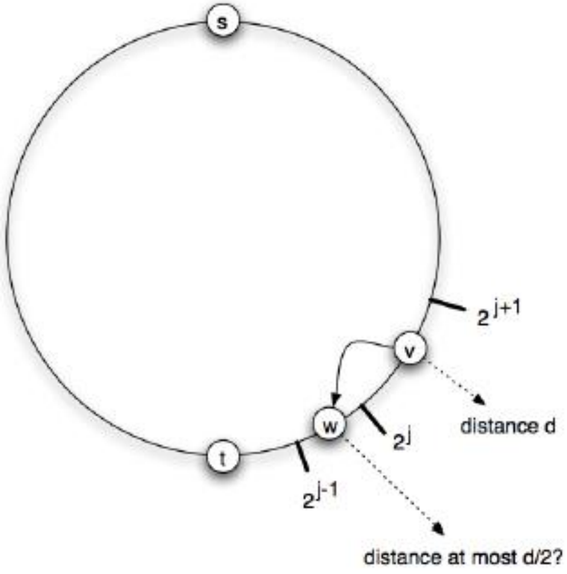
- Otteniamo quindi che la probabilità *P(v,w)* di avere il link da *v* a *w* è

$$\frac{1}{Z}d(v,w)^{-1} \geq \frac{1}{2\log_2 n}d(v,w)^{-1}$$



Tempo trascorso in una fase

Fissiamo una particolare fase *j* della ricerca, quando il messaggio è in un nodo *v* la cui distanza del target *t* è un numero *d* tra 2^j e 2^{j+1}

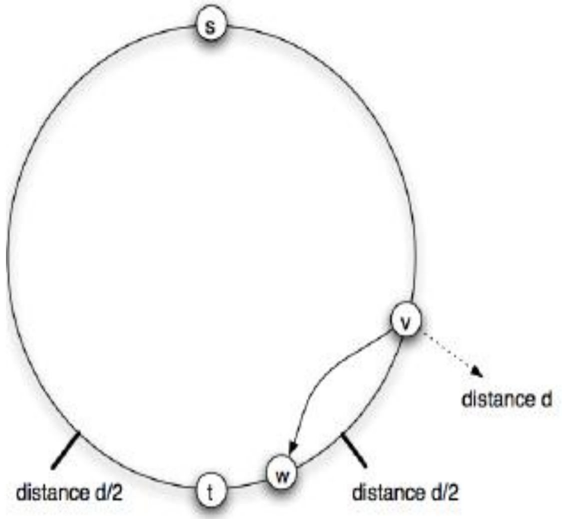


- La fase sarà giunta al termine una volta che la distanza dal target scende sotto 2^j,
- La fase finisce immediatamente se *r,* il contatto a lungo raggio di *v,* è a distanza ≤ *d / 2* da *t*.
  - In questo caso, *v* è l'ultimo nodo toccato nella fase j
- **mostriamo che un dimezzamento immediato della distanza accade con alta probabilità**

Tempo trascorso in una fase

Sia *I* l'insieme dei nodi a distanza $\leq d/2$ da t.

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



- Pertanto, per ogni nodo *w* in I la probabilità che *w* sia il contatto a lungo raggio di v è data da

$$\frac{1}{2\log n}d(v,w)^{-1} \geq \frac{1}{2\log n}\cdot\frac{1}{3d/2} = \frac{1}{3d\log n}$$

- Poiché ci sono più di *d* nodi in I, la probabilità che un qualche nodo in I sia il contatto a lungo raggio di v è almeno

$$d\cdot\frac{1}{3d\log n} \geq \frac{1}{3\log n}$$

Tempo trascorso in una fase

- Se uno di questi nodi è il contatto a lungo raggio di v, allora abbiamo che la fase j termina immediatamente

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

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

- Per eseguire almeno *i* passi , la fase *j* non deve terminare per i - 1 volte consecutive. Quindi la probabilità che la e fase j duri *almeno i* passi è al massimo

$$\left(1-\frac{1}{3\log n}\right)^{i-1}$$

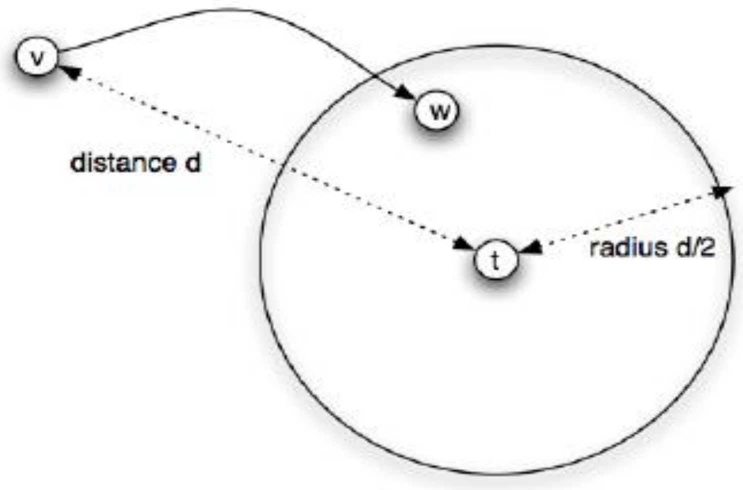
Tempo trascorso in una fase

- Otteniamo quindi che

$$
\begin{aligned}
E[X_j] &= \sum_i \Pr[X_j \geq i] \ \textbf{(check it!)} \\
&\leq \sum_i \left(1 - \frac{1}{3\log n}\right)^{i-1} \ \textbf{(from above)} \\
&\leq \frac{1}{1 - \left(1 - \frac{1}{3\log n}\right)} \ \textbf{(geometric series)} \\
&= 3\log n
\end{aligned}
$$

Small World in 2D

- In 2D con il messaggio ad una distanza   corrente d dal target t, ancora una volta guardiamo 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

Small World in 2D

1D è stato utilizzato in due posti

- Quando abbiamo determinato la costante di normalizzazione Z
- Per sostenere che ci sono stati almeno d nodi a distanza d/2 da t

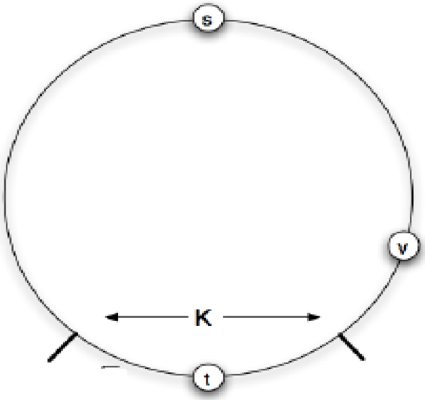
- Il fattore d annulla d -1 nella probabilità dei link, che ci permetteva di concludere che la probabilità di dimezzare il distanza del target in un determinato passo era almeno proporzionale a 1 / (logn), indipendentemente dal valore di d.

- Quando andiamo a 2D, il numero di nodi a distanza al più d / 2 da t sarà 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 quello che si usa*.

Perché la ricerca non lavora bene quando q = 0

(cioè, nel modello originale di Watts-Strogatz quando vengono scelti i collegamenti a lungo raggio uniformemente a caso)

Supponiamo di essere nel Ring
Consideriamo l'insieme di tutti i nodi entro una certa distanza del target t



- q= 1: abbiamo sostenuto che è facile entrare in insiemi sempre più piccoli centrati intorno a t,

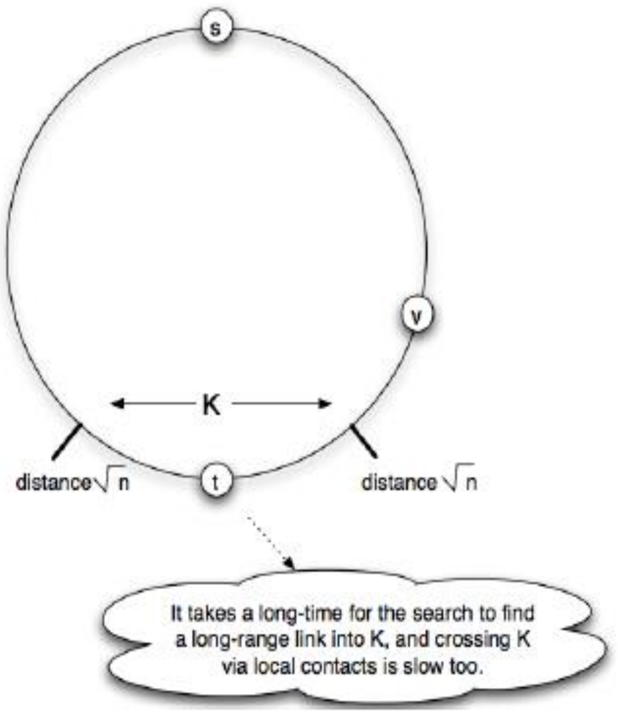
- q = 0: vogliamo identificare un insieme di nodi centrato in t che è in qualche modo "impenetrabile" (molto difficile da raggiungere per la ricerca)

Perché la ricerca non lavora bene quando q = 0

Denotiamo con

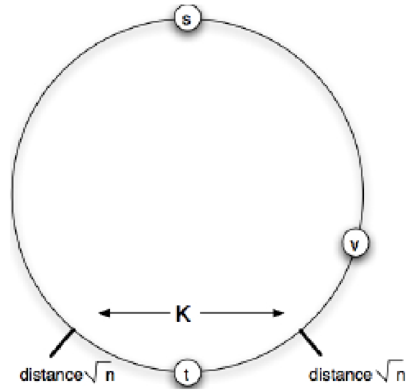
K: l'insieme dei $2\sqrt{n}$ nodi più vicini al target t

E : l'evento in cui qualsiasi dei primi k nodi visitati dall'algoritmo di ricerca ha un collegamento ad un un nodo in K



Consideriamo il tempo necessario per entrare in K

Perché la ricerca non lavora bene quando q = 0



- 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 all'interno di K è uguale a $\frac{|K|}{n} < \frac{2\sqrt{n}}{n} = \frac{2}{\sqrt{n}}$

- Sia *p* è la probabilità che un evento si verifichi in 1 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

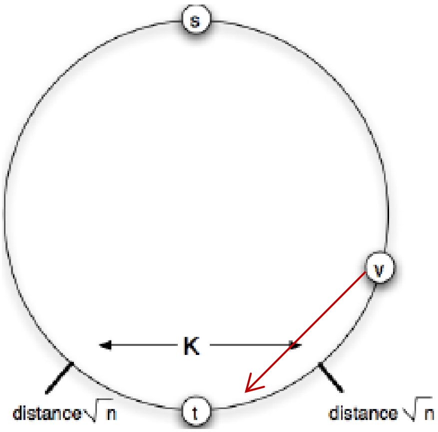
$$\sum_{i\geq 1} i(1-p)^{i-1}p = p\sum_{i\geq 1} i(1-p)^{i-1} = \frac{p}{p^2} = \frac{1}{p}$$

Quindi il numero medio di passi per trovare un nodo che ha una contatto a lungo raggio in K è $\frac{\sqrt{n}}{2}$

Perché la ricerca non lavora bene quando q = 0

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.

➔

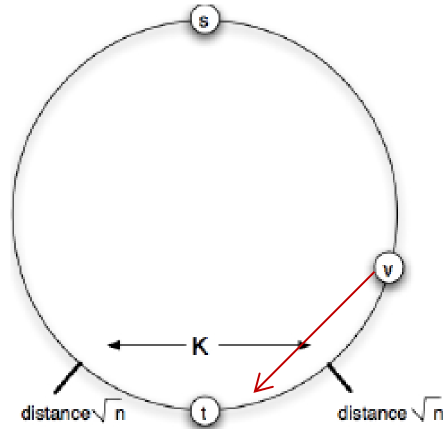


qualsiasi strategia di ricerca decentralizzata

- avrà bisogno in media di almeno $\frac{\sqrt{n}}{2}$ passi, per trovare un nodo avente un contatto a lungo raggio in K.

- Inoltre finché non trova un collegamento a lungo raggio in K, la ricerca non può che attraversare K step-by-step utilizzando solo le connessioni locali, quindi $\geq \frac{\sqrt{n}}{2}$ passi

Perché la ricerca non lavora bene quando q = 0



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.

➔

qualsiasi strategia di ricerca decentralizzata

- avrà bisogno in media di almeno $\frac{\sqrt{n}}{2}$ passi, per trovare un nodo avente un contatto a lungo raggio in K.

- Inoltre se non trova un collegamento a lungo raggio, in K la ricerca non può che procedere step-by-step utilizzando solo le connessioni locali

➔

Nessuna strategia può effettuare un numero di passi minore di $\frac{\sqrt{n}}{2}$ in media.