Il fenomeno Small-World

Capitolo 20

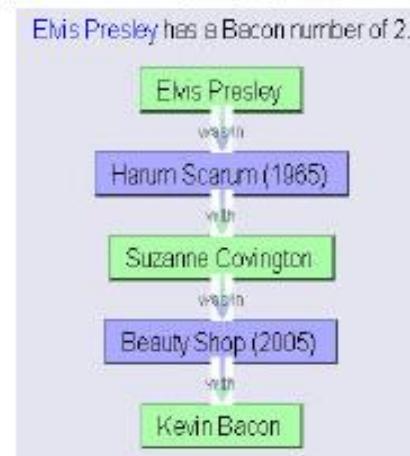
Il fenomeno Small-World

Le origini dell'idea small-world: Il numero di Bacon:

- Creare una rete degli attori di Hollywood
- Collegare due attori se sono co-apparsi in un film
- Numero di Bacon: serie di passi per Kevin Bacon

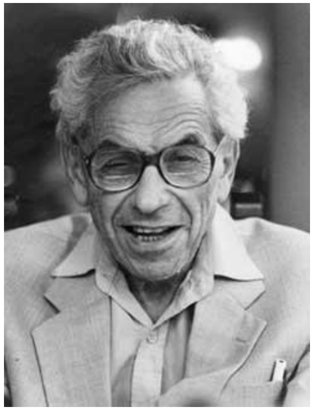
Al dicembre 2007, il più alto numero (finito) di Bacon riportato è 8

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



Esempio: Coauthorship Network

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

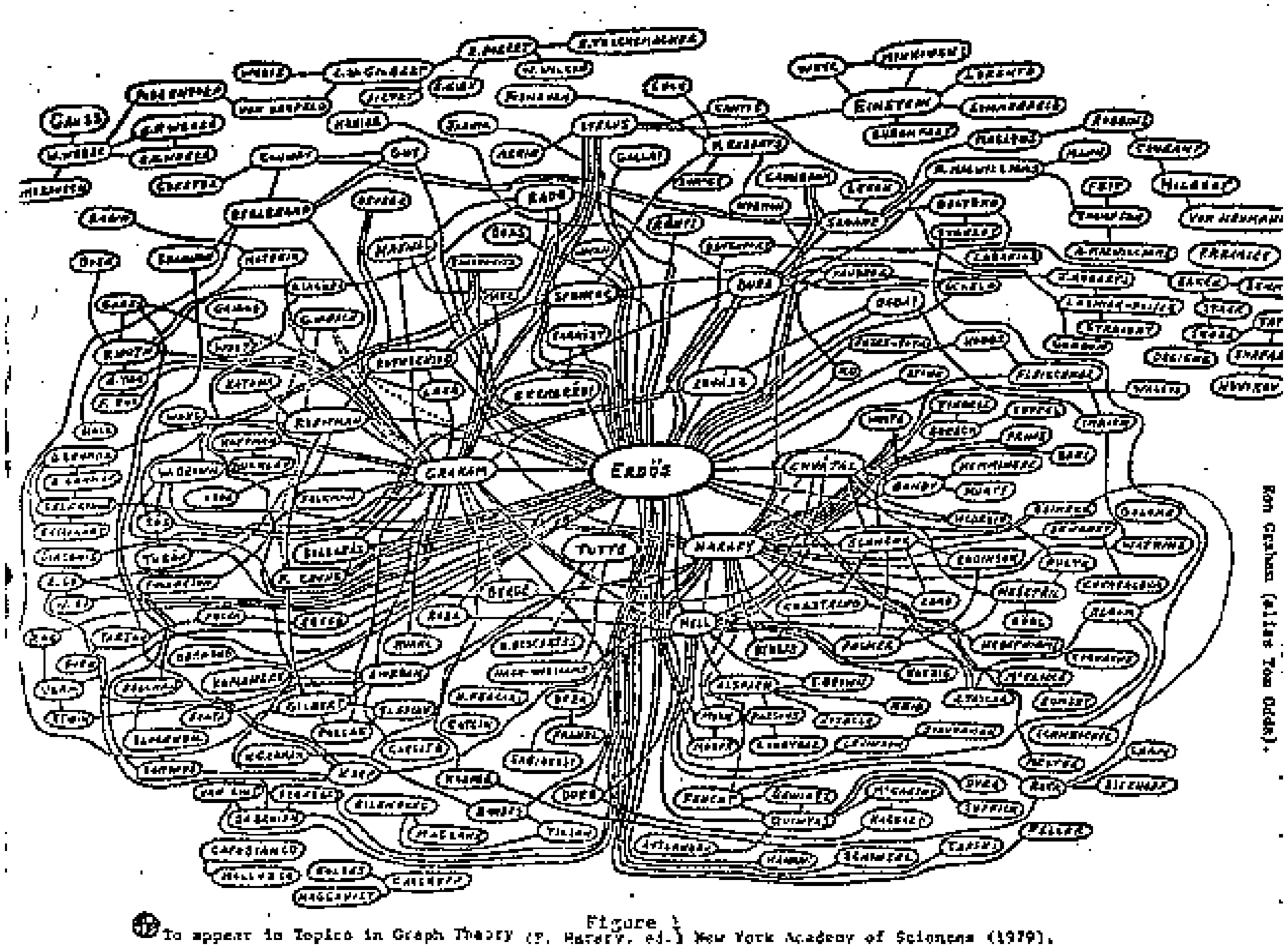


Paul Erdős

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

N is a Number: A Portrait of Paul Erdős (trailer)

Erdős Number



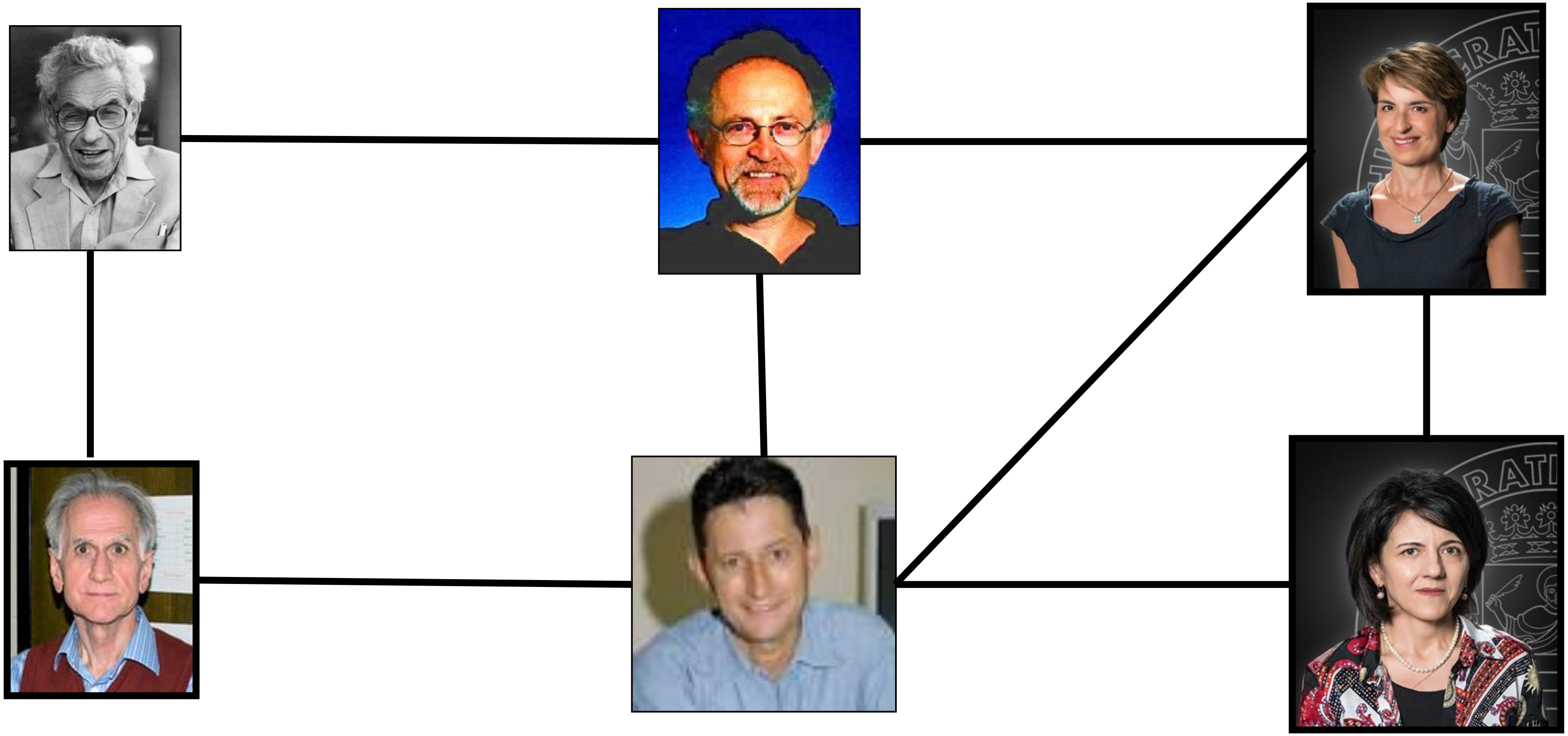
Grafo delle collaborazioni scientifiche centrato su Paul Erdös

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

Erdős Number

Erdős number 2.

Erdős Number



Coauthorship Nework

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

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.

Sei gradi di separazione

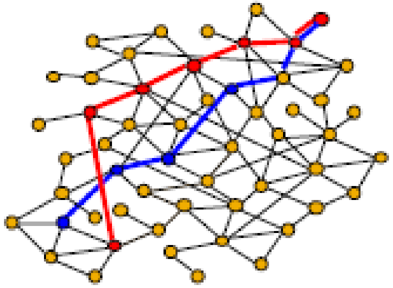
L'esperimento (Milgram, 1967):

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.

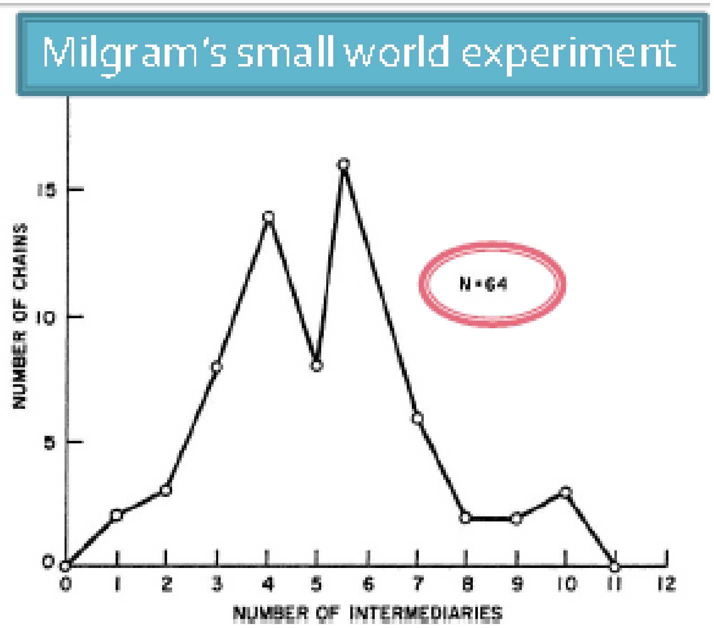
Sono stati necessari quanti invii?

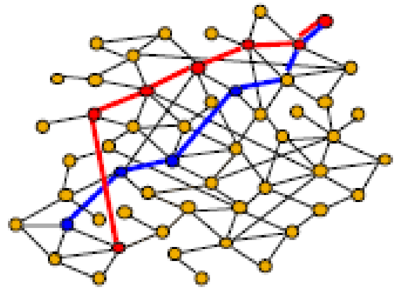
Stanley Milgram (1933-1984)



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"
- Ulteriori osservazioni:
  - Le persone che possedevano azioni avevano percorsi più brevi per l'agente di cambio di persone a caso: 5.4 rispetto a 6.7
  - Le persone della zona di Boston avevano i percorsi più corti: 4.4

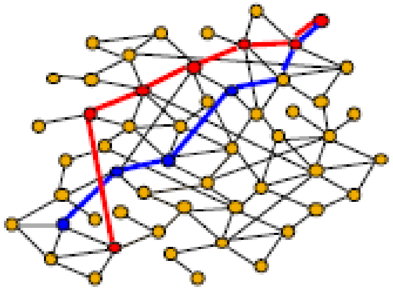




Sei gradi di separazione

Note

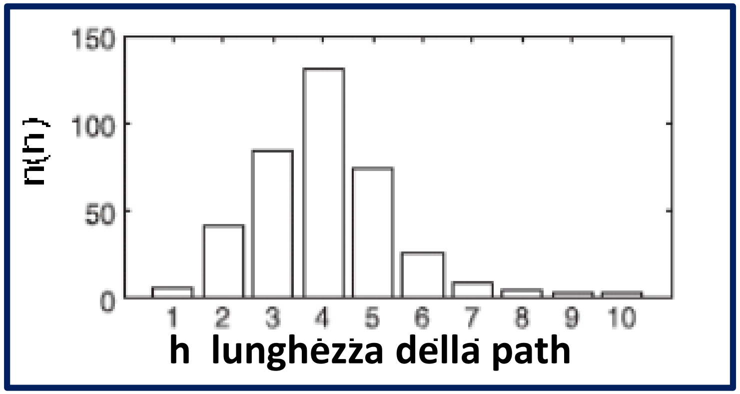
- 31 su 64 catene passate attraverso 1 di 3 persone nel loro ultimo passo
  - *Non tutti i link / nodi sono uguali*
- Punti di partenza e l'obiettivo non erano casuali
- Non ci sono molti campioni (solo 64)
  - La gente ha rifiutato di partecipare (25% per Milgram)
  - Non tutte le ricerche finite (solo 64 su 300)
- **Una sorta di social search**: Persone seguono una qualche strategia, non inoltrano la lettera a tutti.
  - **Non stanno cercando il percorso più breve!**
  - Potrebbero avere utilizzato informazioni aggiuntive



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%)



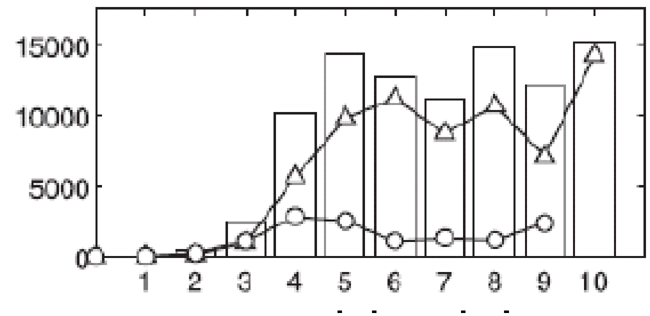
- Aggregando le 384 catene completate su tutti i target, la lunghezza media di una catena risulta **4.05**
- Problema: la gente smette di partecipare le catene si interrompono (quindi catene brevi terminano con più probabilità)
- Per ovviare si introduce un fattore di correzione:

$$n^*(h) = \frac{n(h)}{\prod_{i=0}^{h-1}(1-r_i)}$$

dove n (h) è il numero di catene completate lunghe h e r_i è il tasso di abbandono al passo i

Small-world: esperimento via email

Dopo la correzione: $$n^*(h) = \frac{n(h)}{\prod_{i=0}^{h-1}(1 - r_i)}$$



lunghezza del percorso h = 7

Alcuni fenomeni non ben compresi nelle reti sociali:

Effetto di focalizzazione: Alcuni amici del target hanno maggiori probabilità di costituire il passo finale

Congettura: alta reputazione / autorità

Effetti delle caratteristiche del target: perché target di alto status costituiscono un obiettivo più facile da trovare

Congettura: struttura di rete centro-periferia

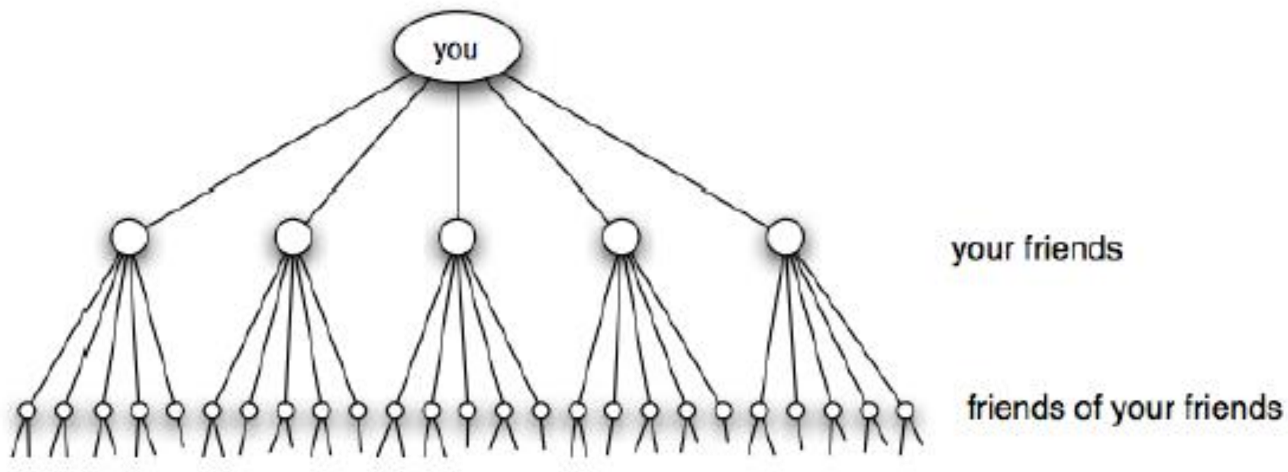
Struttura di una rete sociale

Small world: come?

Assumiamo che ogni umano 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 5 passi possiamo raggiungere 10 miliardi di persone

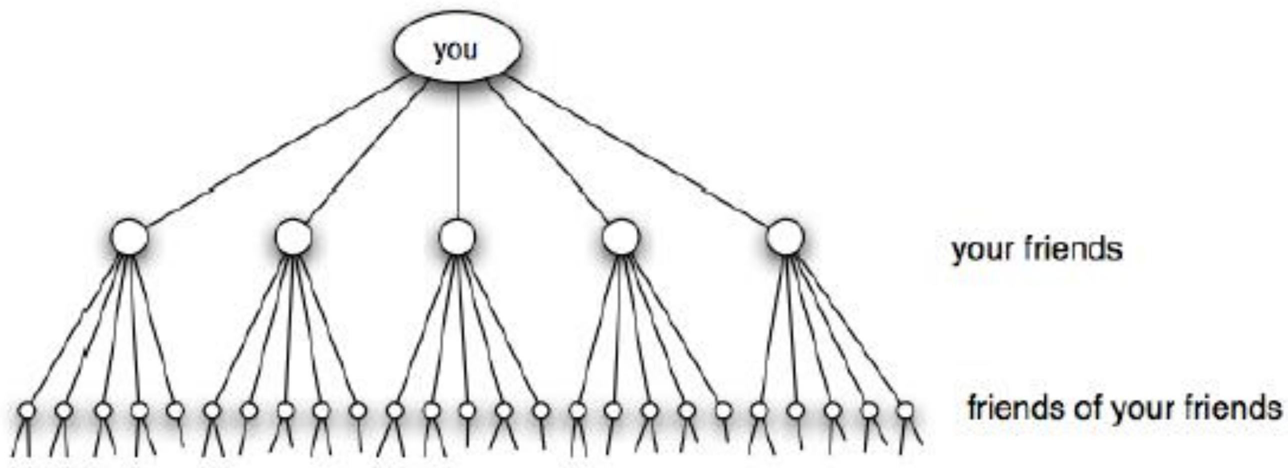


Small world: come?

Assumiamo che ogni umano è collegato ad altre 100 persone:

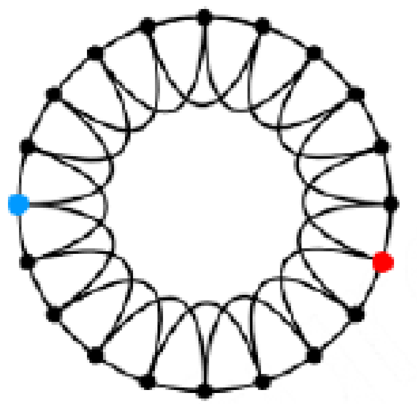
- Passo 1: possiamo raggiungere 100 persone
- Passo 2: possiamo raggiungere 100 * 100 = 10.000 persone
- Passo 3: possiamo raggiungere 100 * 100 * 100 = 1.000.000 di persone
- Passo 4: possiamo raggiungere 100 * 100 * 100 * 100 = 100 milioni di persone

In 5 passi possiamo raggiungere 10 miliardi di persone

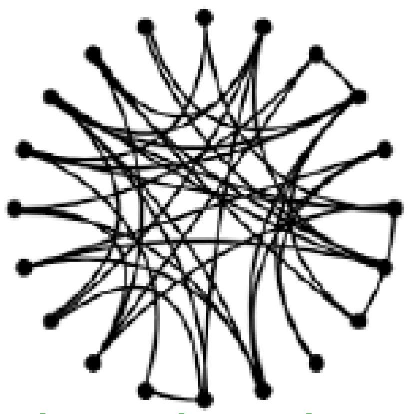


Qualche problema nel ragionamento?

Small world: come?



High clustering
High diameter



Low clustering
Low diameter

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

Small world: come?

Possiamo avere un modello di rete che esibisce sia molti triangoli, sia percorsi molto brevi?

- **Si! Modello di Watts Strogatz**

Tale modello segue naturalmente da una combinazione di due idee fondamentali : **legami forti e legami deboli.**

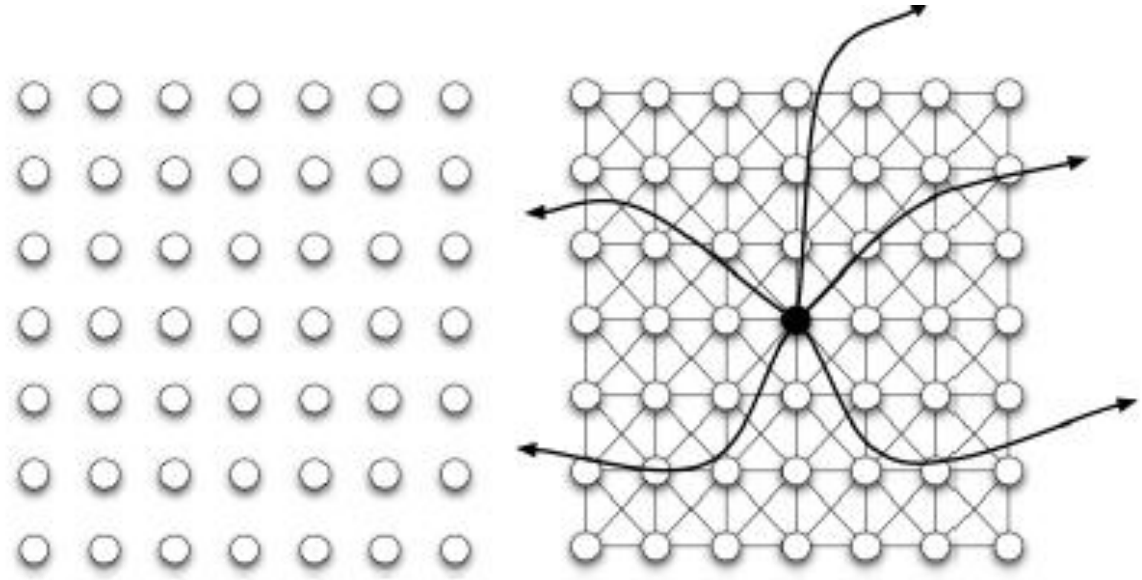
- **Omofilia** (tendenza a stabilire contatti e relazioni con i propri simili) crea molti triangoli,
- **Legami deboli** producono una struttura ampiamente ramificata che raggiunge molti nodi in pochi passi.

Supponiamo che ognuno vive 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 di una griglia, per un valore costante di r.



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

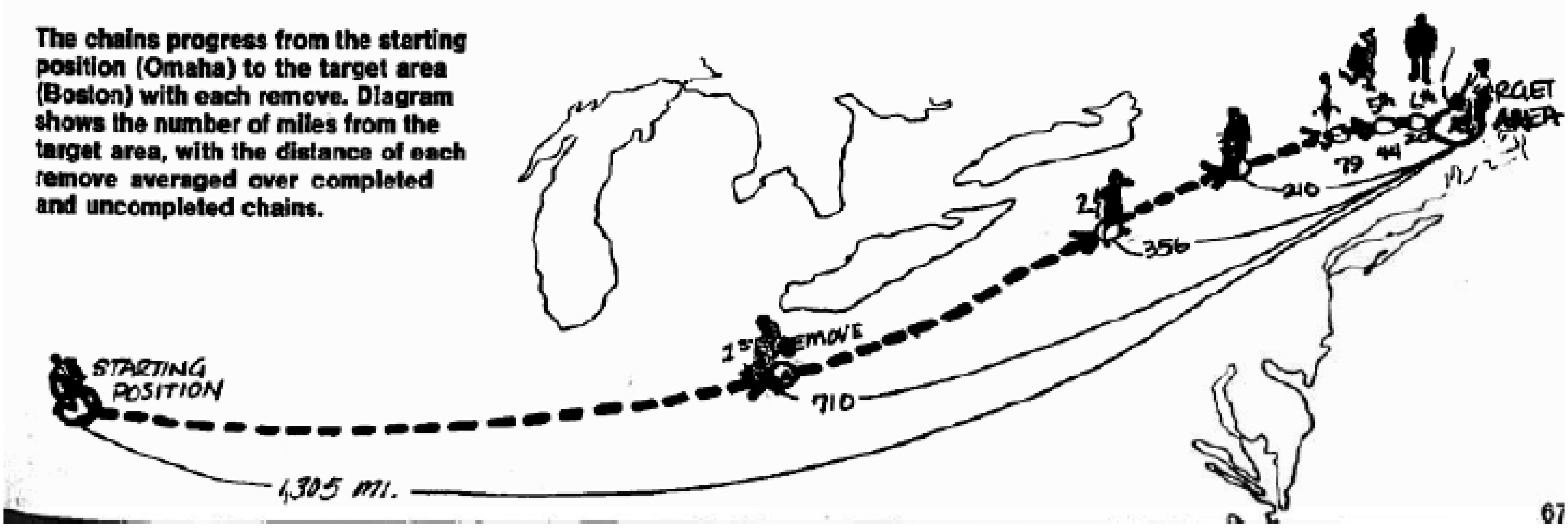
Idea per l’esistenza di path brevi

- Iniziate tracciando percorsi 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 uniformemente a caso, è molto improbabile vedere un nodo due volte nei primi passi di allontanamento da *v*.
- In questi primi passi quasi sicuramente non vi è alcuna chiusura triadica, quindi un gran numero di nodi vengono raggiunti in un piccolo numero di passi.

- Questo ci fa supporre (ed in effetti si può dimostrare) che introdurre una piccola quantità di casualità - nella forma legami deboli di lungo raggio- è sufficiente a rendere il mondo *piccolo*, cioè con percorsi brevi tra ogni coppia di nodi
- **Rimane la domanda: come trovare percorsi brevi facilmente?**

Idea per l’esistenza di path brevi

- Perché 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 di successo 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*.

- 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

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

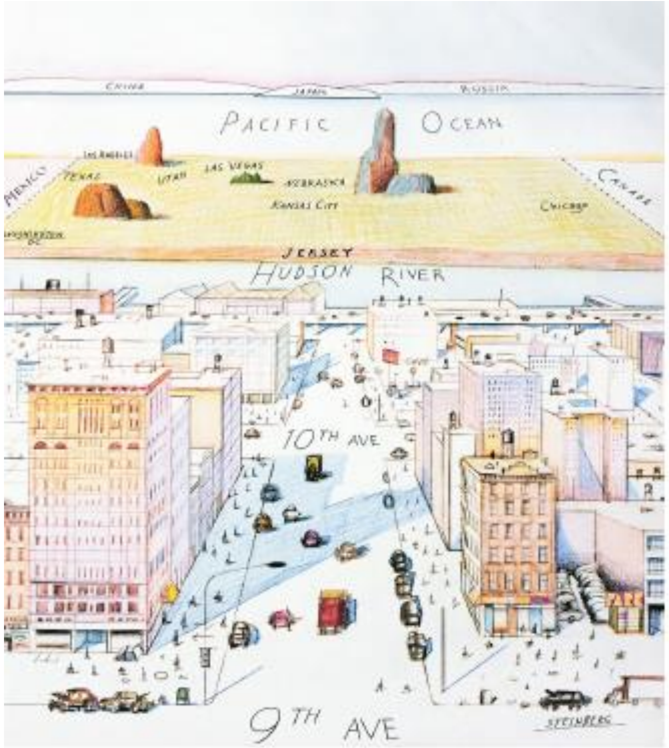
- Ma si dimostra che
  - un algoritmo di ricerca decentrata nel modello di Watts-Strogatz ha bisogno $n^{2/3}$ passi per raggiungere il target t
  - Nota: anche se i percorsi di *O(log n)* passi esistono!

Un modello per la Ricerca Decentralizzata

- Watts-Strogatzg graphs non sono navigabili

- Come possiamo ottenere small-world navigabili?

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

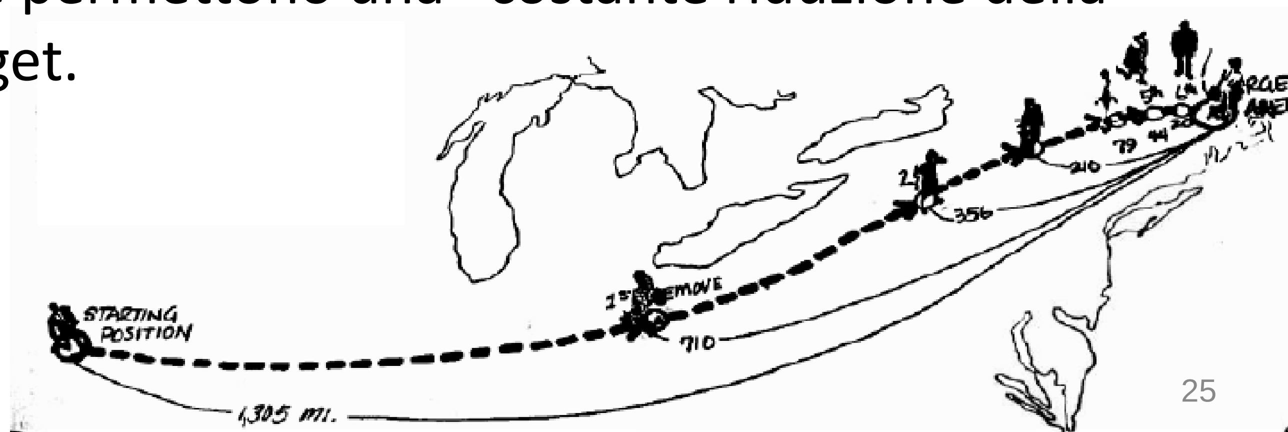


Saul Steinberg, "View of the World from 9th Avenue"

Un modello per la Ricerca Decentralizzata

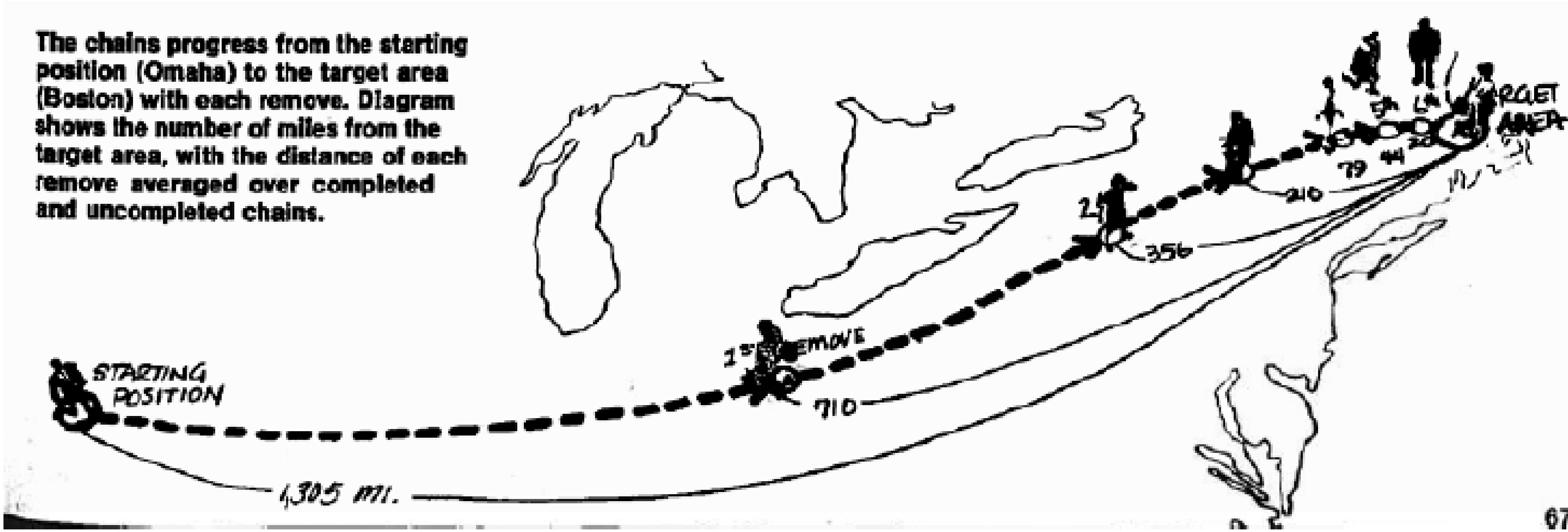
- l modello è efficace nel catturare la densità di triangoli e l'esistenza di percorsi brevi, *ma non la capacità delle persone*, lavorando insieme in rete, di trovare effettivamente i percorsi.

- Il problema è che i legami deboli che rendono il mondo piccolo sono "troppo casuali" in questo 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.



Significato dell'esperimento di Milgram

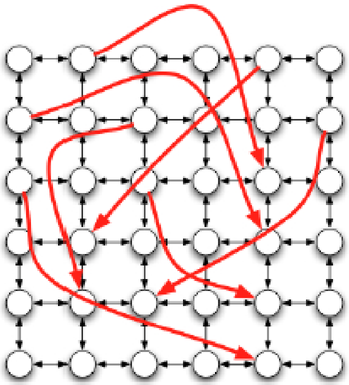
- Milgram ha osservato che "Il movimento geografico del messaggio dal Nebraska al Massachusetts è impressionante. C'è una chiusura progressiva sulla l'area di destinazione ogni volta che una nuova persona viene aggiunta alla catena "

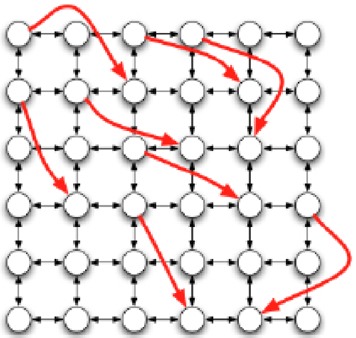


- Non è sufficiente avere un modello di rete in cui i legami deboli sono molto lunghi; è necessario abbracciare anche tutte le *lunghezze intermedie!*

Modello di Kleinberg

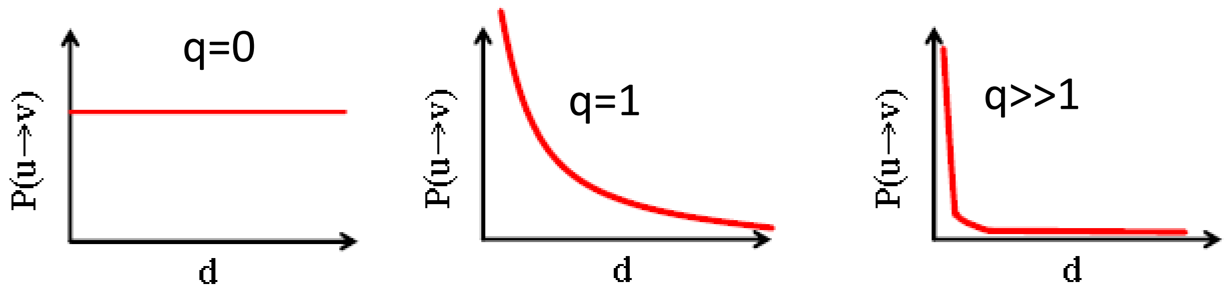
- Abbiamo una griglia e ciascun nodo ha archi ad ogni altro nodo ad al più r passi di griglia.
- Ciascuna dei suoi k archi casuali viene generato in modo che la probabilità decade con la distanza, in base ad un parametro q, come segue.
  - Per due nodi *v* e *w*, sia *D (v, w)* la loro distanza sulla griglia (il numero di passaggi su rami della griglia tra loro).
  - 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

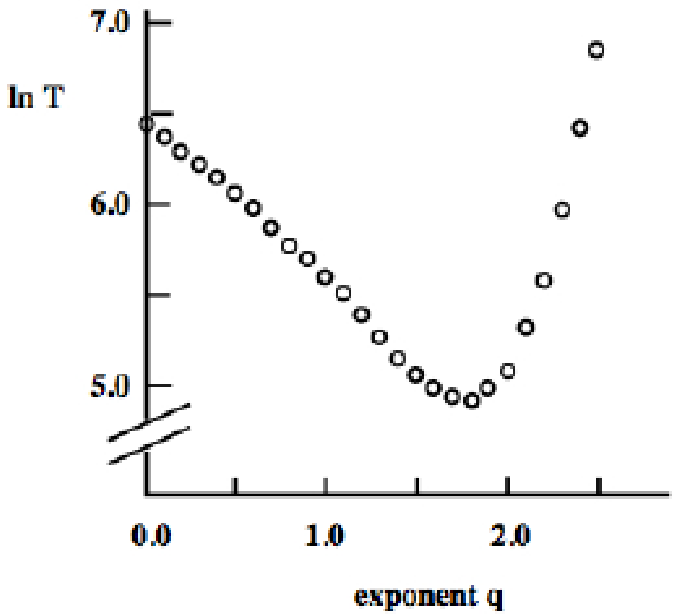
- Così abbiamo un modello diverso per ciascun valore di *q*.
- Il modello originale basato sulla griglia Watts-Strogatz corrisponde a *q = 0*, cioè i link deboli sono scelti in modo uniforme a caso



- 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 casuali" per fornire abbastanza dei salti a lunga distanza che creano un grafo small-world

Ricerca decentralizzata più efficiente : q = 2

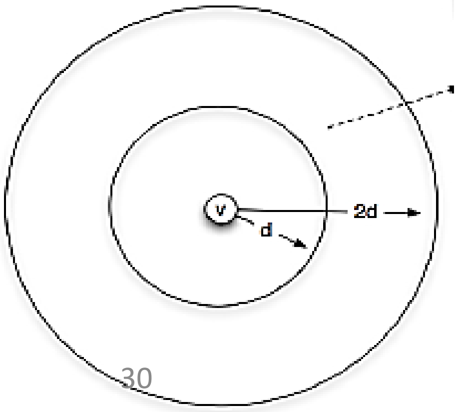
- metodo di ricerca decentralizzata tra diversi valori di q, per un rete di diverse centinaia di milioni di nodi.
- asse x è l'esponente q, asse y è il tempo di consegna ln T.
- Il risultato principale di questo modello è che, nel limite di grande dimensioni della rete, ricerca decentralizzata è più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



Ricerca decentralizzata più efficiente : q = 2

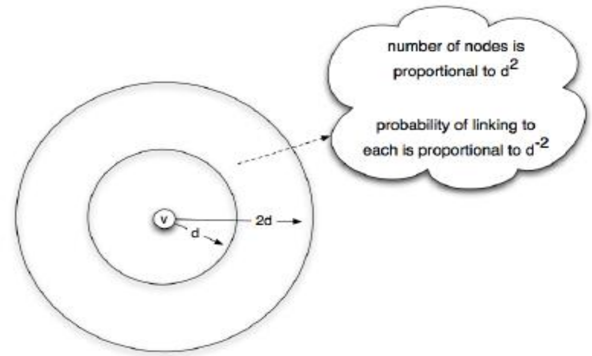
Perché *q = 2* è ottimale?

- **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, che chiamiamo *A,* è proporzionale a d^2.
- Per u ∈ A, la probabilità *Pr[v links to u]* che v ha un link verso u, varia a seconda di D(v,u), ma ogni probabilità individuale è proporzionale a d^{-2}.



- Ne consegue che

  - il numero di nodi nel gruppo A di nodi della corona circolare, e

  - la probabilità di collegamento a uno qualsiasi di essi

  si annullano (approssimativamente)

  In altre parole, questo ragionamento *ci suggerisce* che se si parte da v, con probabilità costante si sfugge alla corona circolare e questo rende possibile il progresso verso il target.