

Il fenomeno Small World

Capitolo 20 del testo

1 Esperimento di Milgram

Lo psicologo sociale Stanley Milgram nel 1967 condusse il seguente esperimento:

- scelse una persona - un destinatario, una persona alla quale doveva essere recapitata una lettera della quale lo stesso Milgram era il mittente.
- per fissare le idee, l'obiettivo risiedeva molto lontano da Milgram - ad esempio, Milgram e l'obiettivo erano sulle coste opposte degli Stati Uniti.
- Milgram ha poi scelto a caso un insieme di iniziatori e ha consegnato ad ognuno di essi una copia della lettera.
- > ha anche fornito ad ogni iniziatore una serie di informazioni sul destinatario: nome, indirizzo, occupazione, e informazioni personali quali interessi, passatempi.
- Milgram ha chiesto ad ogni iniziatore di fare in modo di far giungere la copia della lettera in suo possesso al destinatario, senza, però, inviargliela direttamente a mezzo sistema postale.
- invece, ogni iniziatore doveva: scrivere il suo nome sulla lettera e, con l'obiettivo di far giungere la lettera al destinatario nel minor numero di passi possibile, consegnarla (o inviarla) a un suo diretto conoscente chiedendogli di ripetere le medesime azioni.

1.1 Risultati dell'esperimento

E cosa ha osservato Stanley Milgram al termine del suo esperimento?

- innanzi tutto, che circa un terzo delle lettere hanno raggiunto il destinatario[cite: 13].
- poi, che le lettere che hanno raggiunto il destinatario lo hanno fatto, in media, in sei passi.

Le lettere hanno viaggiato da una parte all'altra degli Stati Uniti, trasmesse da un individuo all'altro sulla base di conoscenze personali, e hanno impiegato in media 6 passi per giungere a destinazione!

> I "famosi" 6 gradi di separazione[cite: 16].

Ebbene, in base all'esito dell'esperimento di Milgram, possiamo trarre due conclusioni:

1. in una rete sociale è presente una moltitudine di percorsi molto brevi, che connettono qualunque coppia di nodi - il fenomeno Small World.
- > che i percorsi brevi non solo esistono, ma possono essere trovati con facilità [cite: 19] , ossia, da nodi che non conoscono altro della struttura della rete se non i propri immediati vicini!

2 Una moltitudine di percorsi brevi

Ci domandiamo ora: quale spiegazione intuitiva possiamo trovare al fatto che in una rete sociale esistano tanti shortest paths fra una coppia di nodi?

Intuitivamente, possiamo vederla così:

- > io ho, diciamo, 100 amici.
- ciascuno dei quali ha, diciamo, 100 amici.

- ciascuno dei quali ha 100 amici, e così via.
- questo significa che il grafo delle relazioni in questo gruppo di persone contiene 10000 percorsi di lunghezza 2 da me ad altre persone della rete - 1000000 percorsi di lunghezza 3 da me ad altre persone della rete.
- cioè, io sono collegata a mezzo di percorsi molto brevi a un sacco di gente!

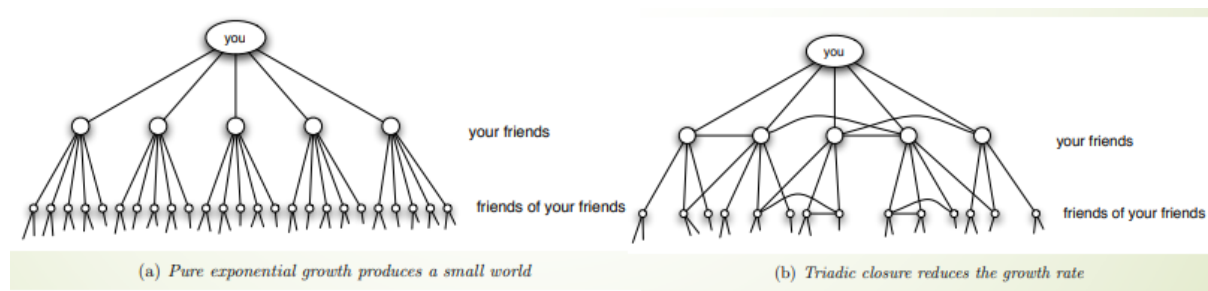
E, poiché posso ripetere lo stesso ragionamento per qualunque altro individuo che popola la rete, ecco la spiegazione dell'esistenza di tanti percorsi brevi in una rete sociale.

Bello, intuitivo, facile, ma questo ragionamento ha una pecca...

2.1 L'effetto della Chiusura Triadica

Bello, intuitivo, facile, ma il ragionamento che abbiamo illustrato ha una pecca: non tiene conto della chiusura triadica (della quale avremo modo di parlare in seguito) ossia, del fatto che in una rete sociale esistono tanti triangoli.

- ossia, se un individuo a conosce due individui b e c , allora è probabile che prima o poi anche b e c si conosceranno.
- perché, ad esempio, i miei amici, frequentandomi, avranno possibilità di incontrarsi!
- Quindi, fra i 100 amici dei miei amici, si troveranno anche alcuni dei miei amici e il grafo della relazione assomiglierà alla figura a destra, piuttosto che a quella a sinistra.



3 Il modello Watts-Strogatz

Ciò premesso, ci proponiamo di studiare un modello generativo di grafi aleatori che generi:

- Small Worlds.
- contenenti molte chiusure triadiche.

Il modello proposto da Watts e Strogatz (1998) consiste di un grafo fissato deterministicamente ed un insieme di archi casuali.

3.1 La componente deterministica: la griglia

Il grafo fissato deterministicamente è una griglia "arricchita" - che corrisponde, sostanzialmente a uno Unit Disk Graph, considerando i nodi come punti di uno spazio metrico bidimensionale.

- i nodi sono disposti sui punti a coordinate intere di un quadrato centrato nell'origine degli assi cartesiani.
- e ogni nodo è collegato a ciascuno dei nodi vicini in orizzontale, verticale e diagonale.

Formalmente:

fissato $n \in \mathbb{N}$, $V = \{(i, j) : 0 \leq i \leq n \wedge 0 \leq j \leq n\}$ e ciascun nodo (i, j) con $0 < i < n$ e $0 < j < n$ è collegato ai nodi $(i, j-1)$, $(i+1, j-1)$, $(i+1, j)$, $(i+1, j+1)$, $(i, j+1)$, $(i-1, j+1)$, $(i-1, j)$, $(i-1, j-1)$ e analogamente per i nodi (i, j) con $i \in \{0, n\}$ e/o $j \in \{0, n\}$.

3.2 La componente casuale: gli archi random

Il modello proposto da Watts e Strogatz (1998) consiste di un grafo fissato deterministicamente ed un insieme di archi casuali.

- Il grafo fissato deterministicamente è una griglia "arricchita" [cite: 60], ossia, i nodi sono disposti sui punti a coordinate intere di un quadrato centrato nell'origine degli assi cartesiani e ogni nodo è collegato a ciascuno dei nodi vicini in orizzontale, verticale e diagonale.
- Poi, fissato un valore k , ogni nodo sceglie uniformemente a caso k nodi che diventeranno suoi vicini.
- In effetti, comunque, più che a una griglia su una superficie piana, dobbiamo pensare a una griglia "appoggiata" su una superficie sferica che si "richiude" su sé stessa che chiameremo wrapped.

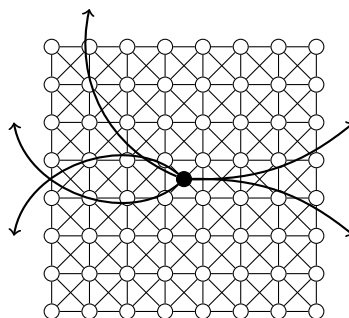


Figure 1: Rappresentazione di un grafo a griglia con collegamenti a lunga distanza.

3.3 Osservazioni sul modello

Osserviamo, ora, un grafo generato in accordo al modello di Watts-Strogatz.

- > **Dicotomia locale/distanza:** possiamo individuare una sorta di dicotomia relazioni locali / relazioni a distanza soggiacente fra gli archi deterministici e quelli random.
 - gli archi della griglia, che costituiscono "l'ossatura fissa" del grafo, rappresentano le relazioni fra nodi "fisicamente" vicini - quelli le cui coordinate differiscono di poco.
 - gli archi random esprimono relazioni fra nodi "fisicamente" lontani.
- 1. **Strong e Weak Ties:** possiamo ben immaginare che i nodi "fisicamente" vicini abbiano più probabilità di incontrarsi rispetto ai nodi "fisicamente" lontani, ossia, che i nodi vicini abbiano frequentazioni assidue, quelli lontani no. Possiamo pensare, allora, agli archi della griglia come archi che rappresentano relazioni forti (strong ties) e agli archi random come archi che rappresentano relazioni deboli (weak ties).
- 2. **Triangoli:** i triangoli sono sempre presenti a livello locale e sono poco probabili fra gli archi random.
 - > confermando l'idea di cui al punto 2): infatti, i triangoli si formano quando individui che hanno un amico comune si incontrano, e due amici di uno stesso individuo che vivono ai due capi opposti della Terra non è probabile che abbiano molte occasioni per incontrarsi e diventare amici.

Dunque, un grafo generato dal modello Watts-Strogatz contiene molti triangoli, in accordo a quel che si riscontra, generalmente, nelle reti sociali. Rimane l'altra questione: sarà vero che coppie di nodi qualunque sono collegati da numerosi percorsi brevi?

A tal proposito, Watts e Strogatz hanno osservato che: se partiamo da un nodo u e a partire da u , per un certo numero di passi ci muoviamo lungo gli archi random, poiché gli archi random sono distribuiti uniformemente nel grafo, è molto improbabile che, in questo procedimento tocchiamo due volte lo stesso nodo. Ossia, molto probabilmente, in h passi abbiamo la possibilità di raggiungere k^h nodi.

Il ragionamento di Watts-Strogatz, appena descritto, è basato su considerazioni intuitive. Successivamente, Bollobás e Chung (1988) hanno formalmente dimostrato questo punto e hanno anche individuato la lunghezza media degli shortest paths nei grafi generati in accordo al modello di Watts-Strogatz.

3.4 Riduzione della casualità

Nel modello di Watts-Strogatz da ogni nodo partono k archi random. In effetti, comunque, il ragionamento intuitivo di Watts e Strogatz può essere ripetuto su un modello in cui è presente di gran lunga meno casualità: è sufficiente che soltanto da un nodo su k partano archi random e che, inoltre da tale nodo parta un solo arco random.

Idea della motivazione: raggruppiamo quadrati di $k \times k$ nodi della griglia in "città", dove ogni città ha uno e un solo arco random uscente. Ripetiamo il ragionamento sopra a livello di città: in h passi possiamo giungere in k^h città. Infine all'interno di una città ci muoviamo attraverso gli archi della griglia. Possiamo, quindi, concludere che poca casualità è sufficiente per avere tanti shortest paths.

4 Percorsi brevi facili da trovare

Ora: siamo il nodo u in un grafo di Watts-Strogatz e vogliamo inviare un messaggio ad una certa destinazione v di cui conosciamo le coordinate. E, dalle sue coordinate, sappiamo che v è molto distante da noi. Naturalmente, cerchiamo di fare arrivare il messaggio il più velocemente possibile, ossia, cerchiamo di fare in modo che venga consegnato attraverso uno shortest path (che, probabilmente, sarà breve) che collega u (noi) a v . Ma noi (che siamo il nodo u), meschini, non conosciamo altro, della rete, che i nostri contatti, ossia, i nodi ai quali siamo collegati - i nostri vicini nel grafo.

4.1 L'approccio "ingenuo"

Potremmo fare tante copie del messaggio quanti sono i nostri vicini e mandare una copia a ciascun vicino. Tanto, abbiamo soltanto 8/9 vicini - non andiamo falliti in fotocopie. Non possiamo fare altro: conoscere l'indirizzo della destinazione non ci aiuta. Potremmo, certo, scegliere, fra i nostri vicini quello le cui coordinate sono più prossime a quelle di v , ma la casualità dei weak ties potrebbe far sì che, invece, un vicino che, al momento, mi appare peggiore ha un arco random che lo collega direttamente a v .

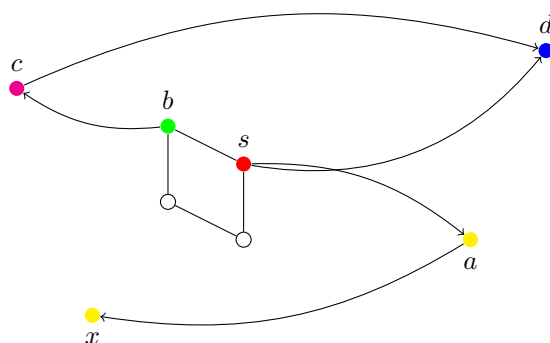


Figure 2: Esempio di ricerca decentralizzata.

In figura: s deve inviare un messaggio a d . Fra i vicini di s il più vicino a d sulla griglia è a , perché b sulla griglia è più lontano di a da d , e questo è tutto ciò che s sa. Perciò, s inoltra il messaggio ad a , che poi dovrà seguire i nodi della griglia per giungere a d . Invece, se s avesse inoltrato a b , allontanandosi momentaneamente da d , avrebbe raggiunto d in 3 passi!

4.2 Il problema del Flooding

Se mandiamo una copia a ciascun vicino che, a sua volta, conosce solo i propri vicini e non può far altro che ripetere il nostro procedimento, a parte che non è carino chiedere ai nostri amici di spendere soldi in fotocopie per spedire un nostro messaggio, dopo h passi circoleranno nel grafo $\approx 7^h$ copie del messaggio!

Tecnicamente parlando, per spedire un messaggio da un nodo u a un nodo v , è stato utilizzato un flooding: quando un nodo entra in possesso del messaggio, crea di esso tante copie quanti sono i suoi vicini e invia una copia a ciascun vicino. Il meccanismo del flooding determina un sovraccarico della rete. Un carico inaccettabile.

Di contro, nell'esperimento di Milgram, ad ogni passo, circolavano tanti messaggi quanti ne aveva creati Milgram; i messaggi non aumentavano ad ogni passo. Per consegnare il messaggio attraverso uno

shortest path, non veniva usato il flooding. Eppure, anche gli individui che parteciparono all'esperimento di Milgram avevano conoscenza solo dei propri vicini e dell'indirizzo del destinatario, esattamente le stesse informazioni che hanno i nodi di un grafo di Watts-Strogatz! Allora, evidentemente, il modello di Watts-Strogatz non riesce a descrivere qualche caratteristica di una rete sociale che rende possibile la ricerca di Milgram.

4.3 Ricerca Decentralizzata (o Miope)

Un algoritmo di ricerca decentralizzata (o ricerca miope) è tale che ciascun nodo non conosce della rete altro che i propri vicini (oltre al nodo target della ricerca) e i nodi non comunicano in alcun modo se non nell'invio del messaggio da consegnare. In sostanza, gli individui coinvolti da Milgram nel suo esperimento hanno utilizzato un algoritmo di ricerca decentralizzata, che, in quel caso, si dimostrava efficiente!

Osserviamo: probabilmente, ad ogni passo, un individuo inoltrava la copia della lettera in suo possesso a quello fra i suoi contatti che stimava essere più vicino possibile al destinatario, dove "più vicino" può riferirsi a metriche diverse: più vicino geograficamente, o rispetto alla professione, o agli interessi culturali...

Allora, la struttura della rete ha qualche caratteristica che fa sì che "inoltrare al più vicino" funziona bene. Ossia, la struttura della rete garantisce che ad ogni passo mi avvicino alla destinazione e che, probabilmente, ci sono tanti passi in cui la distanza dalla destinazione diminuisce drasticamente.

4.4 Limiti del modello di Watts-Strogatz

Riconsideriamo ora il modello di Watts-Strogatz: certamente, ciascun nodo ha un vicino sulla griglia più vicino di sé alla destinazione. D'altra parte, se seguiamo un percorso costituito di soli archi della griglia, impieghiamo un sacco di passi per giungere a destinazione - $O(\sqrt{n})$ passi. Allora, se voglio trovare un percorso molto breve, devo usare gli archi random.

Ma non c'è garanzia che, in un grafo di Watts-Strogatz, usando la regola "invia al tuo vicino che è il più vicino alla destinazione", incontrerò una serie di archi random in modo tale che in un piccolo numero di passi giungerò a destinazione. In effetti, si può dimostrare che nel modello di Watts-Strogatz la ricerca decentralizzata di un percorso da s a t individua mediamente un percorso molto più lungo di uno shortest path [Kleinberg, 2000].

Perché nel modello di Watts-Strogatz l'estremo di un arco random uscente da un nodo è scelto uniformemente a caso fra tutti gli altri nodi. Gli archi random non tengono conto in alcun modo di quanto sono "vicini" i nodi che congiungono. Detto altrimenti, gli archi random sono troppo random!

5 Un modello per la ricerca decentralizzata

Vogliamo definire un modello generativo cui corrispondano grafi che contengono molti triangoli, nei quali esistono molti shortest paths fra le coppie di nodi, e nei quali trovare gli shortest paths mediante ricerca decentralizzata sia possibile. Per soddisfare l'ultimo punto, è necessario che gli archi random siano scelti in modo da tener conto di quanto sono "vicini" i nodi che congiungono.

Il nostro nuovo modello è ancora basato su un'ossatura deterministica: la stessa griglia arricchita e wrapped del modello di Watts-Strogatz, e da ogni nodo esce un arco random. Ma ora la probabilità che l'arco random uscente dal nodo u sia (u,v) è inversamente proporzionale alla distanza sulla griglia dei nodi u e v :

$$> \text{ossia, } P((u,v) \in E) = \frac{1}{Z_u} \frac{1}{d(u,v)^q}.$$

5.1 Definizione del modello

La probabilità che l'arco random uscente dal nodo u sia (u,v) è inversamente proporzionale alla distanza sulla griglia dei nodi u e v :

$$P((u,v) \in E) = \frac{1}{Z_u} \frac{1}{d(u,v)^q}$$

dove:

- $d(u,v)$ indica la lunghezza di uno shortest path fra u e v sulla griglia (ossia, un percorso che non contiene gli archi random).

- Z_u è un fattore di normalizzazione: poiché da ogni nodo deve uscire uno e un solo arco random, deve essere $\sum_{v \in V - \{u\}} P(\{u, v\} \in E) = 1$. Ossia, $\sum_{v \in V - \{u\}} \frac{1}{Z_u} \frac{1}{d(u, v)^q} = 1$, e quindi $Z_u = \sum_{v \in V - \{u\}} \frac{1}{d(u, v)^q}$.
- Poiché la griglia wrapped è simmetrica, allora Z_u ha lo stesso valore per tutti i nodi e, pertanto, indicheremo il fattore di normalizzazione, semplicemente, come Z .
- q è un parametro che prende il nome di **esponente di clustering**. Abbiamo un modello diverso per ogni valore di q : ad esempio, per $q = 0$ abbiamo il modello di Watts-Strogatz.
- in generale, gli archi random sono "troppo random" quando q è piccolo, "non abbastanza random" quando q è grande.

Naturalmente, la ricerca decentralizzata funziona meglio con alcune scelte di q e peggio con altre. Quel che ci proponiamo è mostrare che esiste una scelta di q che rende efficiente la ricerca decentralizzata, anzi, che esiste un valore di q ottimale.

5.2 Il parametro ottimale $q = d$

Nel caso in cui la componente deterministica del grafo è una griglia (wrapped) bidimensionale (ossia, i nodi sono immersi in una superficie bidimensionale), allora l'esponente di clustering ottimale è $q = 2$, quando il numero di nodi è molto, molto grande.

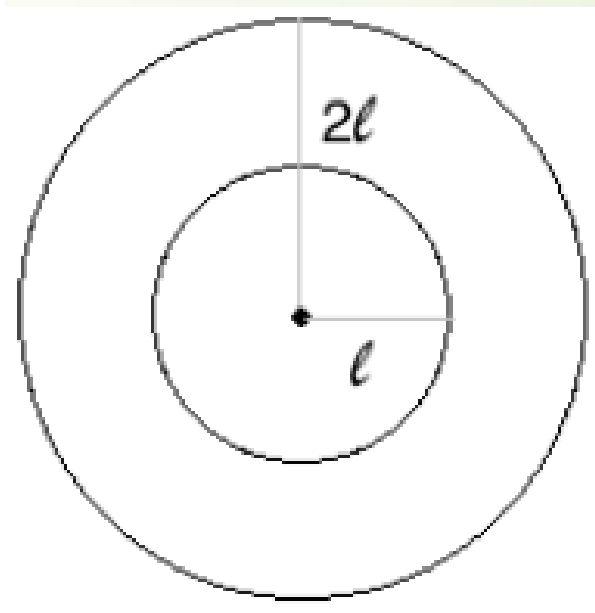
In generale, se la componente deterministica è una griglia (wrapped) d -dimensionale (ossia, i nodi sono immersi nello spazio \mathbb{R}^d), allora l'esponente di clustering ottimale è $q = d$. Vediamo qualche intuizione in supporto di questa affermazione.

Ripensiamo l'esperimento di Milgram: il tragitto della lettera segue, grosso modo, uno schema a "scale di risoluzione". Prima viaggia da un continente all'altro, poi all'interno del continente, poi all'interno della nazione, della regione, della città.

5.3 Intuizione per l'ottimalità di $q = 2$ (caso 2D)

Proviamo a formalizzare questa osservazione. Fissiamo un nodo u e partizioniamo i nodi rimanenti per blocchi definiti in base alla distanza da u : i nodi a distanza da u compresa fra 2^h e 2^{h+1} .

- Il numero di nodi nel blocco anulare tra 2^h e 2^{h+1} è $\approx \pi(2^{h+1})^2 - \pi(2^h)^2 = 3\pi 2^{2h}$, ossia è proporzionale a 2^{2h} .
- Scelto v in tale blocco, la probabilità che l'arco random uscente da u sia (u, v) è proporzionale a $1/d(u, v)^q \approx 1/(2^h)^q$. Se poniamo $q = 2$, questa probabilità è proporzionale a $1/2^{2h}$.
- Allora la probabilità totale che l'arco random uscente da u cada in questo blocco è (numero di nodi) \times (probabilità per nodo) $\propto 2^{2h} \times \frac{1}{2^{2h}} = 1$.
- Questa probabilità è indipendente da h , ossia, è indipendente da quale blocco si stia considerando. La probabilità di raggiungere un nodo a distanza 2, o 4, o 64, o 1024 è la stessa!



Ossia, i weak ties sono distribuiti uniformemente su tutte le scale di risoluzione. Questo fa sì che non occorreranno molti passi prima di arrivare ad un nodo il cui arco random diminuisce drasticamente la distanza dall'obiettivo.

6 Prestazioni della ricerca miope nel modello (Caso $d=1$)

Analizziamo formalmente le prestazioni dell'algoritmo di ricerca decentralizzata applicata al modello generativo nel solo caso $d = 1$, perché l'analisi risulta più semplice. Ossia, analizziamo il caso in cui i nodi sono disposti su un anello, al quale sono aggiunti gli archi random con coefficiente di clustering $q = 1$.

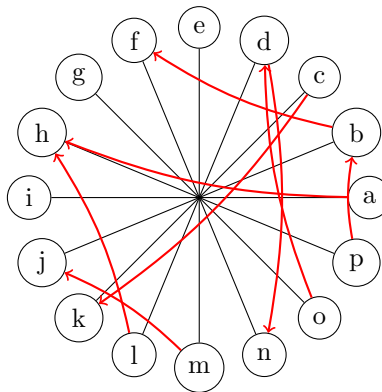


Figure 3: Un anello aumentato con collegamenti random a lunga distanza.

Consideriamo un grafo G tale che i nodi sono disposti su un anello, $V = \{1, \dots, n\}$, e per $u, v \in V$, $P((u, v) \in E) = \frac{1}{Z} \frac{1}{d(u, v)}$ (poiché $q = 1$). Osserviamo che, in questo caso, per ogni $u \in V$:

$$Z = \sum_{v \in V - \{u\}} \frac{1}{d(u, v)} = 2 \sum_{h=1}^{n/2} \frac{1}{h} \approx 2 \ln n$$

Scegliamo uniformemente a caso due nodi s e t in G ed utilizziamo l'algoritmo di ricerca decentralizzata. Indichiamo con X la variabile aleatoria che denota la lunghezza di tale percorso.

Teorema 1. *Dato un percorso calcolato con l'algoritmo di ricerca decentralizzata in un grafo ad anello con n nodi e coefficiente di clustering $q = 1$, la sua lunghezza attesa X è limitata da:*

$$E[X] \in O(\ln^2 n)$$

Proof. La dimostrazione si articola in tre passaggi principali: prima suddividiamo il processo di ricerca in fasi, poi calcoliamo la durata attesa di ogni fase e infine sommiamo i risultati per ottenere la stima totale.

1. Suddivisione in Fasi Il processo di trasmissione del messaggio dal nodo di partenza s al nodo di destinazione t viene suddiviso in fasi. Un processo si trova nella **fase j** se il nodo corrente u che possiede il messaggio si trova a una distanza da t tale che:

$$\frac{d(s, t)}{2^{j+1}} < d(u, t) \leq \frac{d(s, t)}{2^j}$$

Il processo inizia in fase $j = 0$ e termina quando la distanza diventa 1. Poiché ad ogni fase la distanza dalla destinazione viene almeno dimezzata, il numero totale di fasi è al più $\log_2(d(s, t)) \leq \log_2 n$.

Definiamo X_j come la variabile aleatoria che rappresenta la durata (cioè il numero di passi) della fase j . La lunghezza totale del percorso è $X = \sum_j X_j$, e il suo valore atteso è $E[X] = \sum_j E[X_j]$. Per dimostrare il teorema, è sufficiente mostrare che per ogni fase j , il valore atteso della sua durata è $E[X_j] \in O(\ln n)$.

2. Calcolo della Durata Attesa di una Fase ($E[X_j]$) Consideriamo un generico passo all'interno della fase j , in cui il messaggio si trova nel nodo v . La fase termina se il passo successivo porta il messaggio a un nodo z la cui distanza da t è sufficientemente piccola da iniziare la fase successiva. Una condizione sufficiente perché ciò avvenga è che il nodo v possieda un collegamento (un "salto") verso un nodo z che dimezzi la sua distanza attuale da t , cioè $d(z, t) \leq d(v, t)/2$.

Sia p la probabilità che un tale salto si verifichi in un singolo passo. Definiamo l'insieme dei nodi "bersaglio" come $\mathcal{I} = \{u \in V \mid d(u, t) \leq d(v, t)/2\}$. La probabilità p è quindi:

$$p = P(\text{la fase } j \text{ termina in un passo}) \geq P(\exists z \in \mathcal{I} : (v, z) \in E)$$

Per calcolare un limite inferiore per p , seguiamo questi passaggi:

- **Limite sulla distanza $d(v, z)$:** Per ogni $z \in \mathcal{I}$, usando la disuguaglianza triangolare, abbiamo $d(v, z) \leq d(v, t) + d(t, z) \leq d(v, t) + \frac{d(v, t)}{2} = \frac{3}{2}d(v, t)$.
- **Probabilità di un singolo arco random:** La probabilità che esista un arco random tra v e z è $P((v, z) \text{ è random}) = \frac{1}{Z} \frac{1}{d(v, z)}$. Sostituendo il limite superiore per $d(v, z)$ e il valore $Z \approx 2 \ln n$, otteniamo:

$$P((v, z) \text{ è random}) \geq \frac{1}{2 \ln n \cdot \frac{3}{2}d(v, t)} = \frac{1}{3d(v, t) \ln n}$$

- **Dimensione dell'insieme bersaglio $|\mathcal{I}|$:** L'insieme \mathcal{I} contiene tutti i nodi entro una distanza di $\lfloor d(v, t)/2 \rfloor$ da t in entrambe le direzioni lungo l'anello, più il nodo t stesso. Quindi, $|\mathcal{I}| = 2\lfloor d(v, t)/2 \rfloor + 1 \geq 2(\frac{d(v, t)-1}{2}) + 1 = d(v, t)$.

Combinando questi risultati, possiamo stimare la probabilità p :

$$\begin{aligned} p &\geq \sum_{z \in \mathcal{I}} P((v, z) \text{ è random}) \\ &\geq |\mathcal{I}| \cdot \min_{z \in \mathcal{I}} P((v, z) \text{ è random}) \\ &\geq d(v, t) \cdot \frac{1}{3d(v, t) \ln n} = \frac{1}{3 \ln n} \end{aligned}$$

La durata X_j di una fase segue una distribuzione simile a quella geometrica con probabilità di successo $p \geq \frac{1}{3 \ln n}$. Il suo valore atteso può essere calcolato come $E[X_j] = \sum_{h=1}^{\infty} P(X_j \geq h)$. Poiché $P(X_j \geq h) \leq (1-p)^{h-1}$:

$$E[X_j] \leq \sum_{h=1}^{\infty} \left(1 - \frac{1}{3 \ln n}\right)^{h-1} = \frac{1}{1 - (1 - \frac{1}{3 \ln n})} = 3 \ln n$$

Abbiamo quindi dimostrato che $E[X_j] \in O(\ln n)$.

3. Calcolo del Valore Atteso Totale ($E[X]$) Infine, sommiamo i valori attesi di tutte le fasi. Poiché il numero di fasi è al più $\log_2 n$, otteniamo:

$$E[X] = \sum_{j=1}^{\log_2 n} E[X_j] \leq \sum_{j=1}^{\log_2 n} (3 \ln n) = (\log_2 n) \cdot (3 \ln n)$$

Dato che $\log_2 n = \frac{\ln n}{\ln 2}$, il risultato finale è:

$$E[X] \leq \frac{3}{\ln 2} (\ln n)^2 \in O(\ln^2 n)$$

Questo completa la dimostrazione. □

7 Analisi dell'esponente di clustering q

7.1 Perché $q = d$ funziona bene

- Gli “ingredienti” che permettono di dimostrare che la ricerca decentralizzata si comporta “bene” nell’anello ($\mathbf{q} = \mathbf{d} = \mathbf{1}$) sono:
 1. fissato d , il numero di nodi a distanza (al più) d dalla destinazione \mathbf{t} è, all’incirca, d .
 2. il fattore di normalizzazione è $Z \leq 2 \ln n$.
- Questi due “ingredienti” permettono di dimostrare che, trovandoci in un nodo \mathbf{v} a distanza \mathbf{d} da \mathbf{t} , la probabilità che \mathbf{v} abbia un vicino \mathbf{z} a distanza $\leq \frac{d}{2}$ da \mathbf{t} è proporzionale a $\frac{1}{\ln n}$ **indipendentemente da d** .
- Infatti, detto $\mathcal{I} = \{u \in V : d(u, t) \leq \frac{d(v, t)}{2}\}$, con $|\mathcal{I}| = d(v, t)$:

$$P(\exists z \in V : (v, z) \in E \text{ e } d(z, t) \leq \frac{d(v, t)}{2}) \geq \frac{1}{3d(v, t) \ln n} |\mathcal{I}| = \frac{1}{3 \ln n}$$

- considerazioni analoghe possono essere fatte nel **caso bidimensionale con $\mathbf{q} = \mathbf{2}$** :
 1. ora, $|\mathcal{I}| = \alpha d(v, t)^2$, per qualche costante α (area di un quadrato di lato $\approx d(v, t)$).
 2. ora $Z = \sum_{v \in V - \{u\}} \frac{1}{d(u, v)^2}$ e si dimostra che $Z \leq \beta \ln n$, per qualche costante β .
- allora,

$$P(\exists z \in V : (v, z) \in E \text{ e } d(z, t) \leq \frac{d(v, t)}{2}) \geq \sum_{z \in \mathcal{I}} \frac{1}{Z d(v, z)^2} \geq \frac{4}{9 d(v, t)^2} |\mathcal{I}| \geq \frac{4\alpha}{9\beta \ln n}$$

- allora, anche nel caso $\mathbf{q} = \mathbf{d} = \mathbf{2}$ la probabilità che \mathbf{v} abbia un vicino a distanza $\leq \frac{d}{2}$ da \mathbf{t} è proporzionale a $\frac{1}{\ln n}$ **indipendentemente da d** .
- Allo stesso modo, considerazioni analoghe valgono anche per $\mathbf{d} \neq \mathbf{2}$.

7.2 Perché $q \neq d$ funziona male

- Cerchiamo, ora, di capire perché il nostro modello nel caso $\mathbf{d} = \mathbf{1}$ e $\mathbf{q} = \mathbf{0}$ mal si presta alla ricerca decentralizzata
 - ossia: quando si esegue la ricerca decentralizzata di un percorso in un anello con archi random generati in accordo al modello con $q = 0$, si trova mediamente un percorso la cui lunghezza è parecchio elevata
- La probabilità che l’arco random uscente dal nodo u sia (u, v) è, in questo caso, $P((u, v) \in E) = \frac{1}{Z} \frac{1}{d(u, v)^0} = \frac{1}{Z}$ dove
 - poiché deve essere $\sum_{v \in V - \{u\}} P((u, v) \in E) = 1$, allora

– ossia, $\sum_{v \in V - \{u\}} \frac{1}{Z} = 1$ e quindi $Z = \frac{1}{n-1}$

- da cui la probabilità che l'arco random uscente dal nodo u sia (u, v) è

$$P((u, v) \in E) = \frac{1}{n-1}$$

- La probabilità che l'arco random uscente dal nodo u sia (u, v) è ora

$$P((u, v) \in E) = \frac{1}{n-1}$$

- Come abbiamo già osservato, nel caso “anello & $q=1$ ” è “facile” entrare in regioni del grafo contenenti nodi sempre più vicini a t – gli insiemi \mathcal{I}

– più precisamente, la probabilità di dimezzare la nostra distanza da t è la stessa sia quando siamo molto lontani da t che quando siamo molto vicini a t

- Mostriamo ora (molto informalmente, tanto per farci un'idea) che nel caso “anello & $q=0$ ” è “difficile” entrare nell'insieme

$$R = \{u \in [n] : d(u, t) \leq \sqrt{n}\}$$

- Mostriamo ora (molto informalmente, tanto per farci un'idea) che nel caso “anello & $q=0$ ” è “difficile” entrare nell'insieme $R = \{u \in [n] : d(u, t) \leq \sqrt{n}\}$

- scegliamo t , e poi scegliamo $s \notin R$

- allora $\forall u \in R$ e $\forall v \in [n] - R$, $P((v, u) \in E) > \frac{1}{n}$

– “>” sia perché $\frac{1}{n-1} > \frac{1}{n}$ sia perché $\frac{1}{n-1}$ è la probabilità di esistenza di un arco random, ma u e v potrebbero anche essere vicini lungo l'anello

- allora, $\forall v \notin R$, $P(\exists u \in R : (v, u) \in E) > \frac{|R|}{n} = \frac{2\sqrt{n}}{n} = \frac{2}{\sqrt{n}}$

- allora, detta Y la variabile aleatoria che rappresenta il numero di passi per raggiungere da s un nodo in R , $P(Y \geq h) < \left(1 - \frac{2}{\sqrt{n}}\right)^h$

- e $E[Y] = \sum_{h \geq 0} P(Y \geq h) \leq \sum_{h \geq 0} \left(1 - \frac{2}{\sqrt{n}}\right)^h = \frac{1}{1 - (1 - \frac{2}{\sqrt{n}})} = \frac{\sqrt{n}}{2}$

- che è **esponenzialmente** più grande di $E[X] \in O(\ln^2 n)$ del caso $q = 1$!

- Abbiamo visto (informalmente) che nel caso “anello & $q = 0$ ” entrare nella regione $R = \{u \in [n] : d(u, t) \leq \sqrt{n}\}$ richiede, mediamente, $\frac{\sqrt{n}}{2}$ passi.

- A questo punto, vediamo in quanti passi, mediamente, entriamo nella regione $R_2 = \{u \in [n] : d(u, t) \leq \frac{\sqrt{n}}{2}\}$:

– analogamente a quanto fatto nel caso di R , $\forall v \in R - R_2$,

$$P(\exists u \in R_2 : (v, u) \in E) \geq \frac{|R_2|}{n} = \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$$

– e, quindi, $P(Y_2 \geq h) \leq \left(1 - \frac{1}{\sqrt{n}}\right)^h$ - indicando con Y_2 il numero di passi per entrare in R_2

– e $E[Y_2] = \sum_{h \geq 0} P(Y_2 \geq h) \leq \sum_{h \geq 0} \left(1 - \frac{1}{\sqrt{n}}\right)^h = \frac{1}{1 - (1 - \frac{1}{\sqrt{n}})} = \sqrt{n}$

- Questo significa che, una volta entrati nella regione R , per entrare in R_2 utilizzare gli archi random è mediamente equivalente a muoversi lungo gli archi dell'anello!

- Ossia, quando si raggiunge una distanza dall'obiettivo dell'ordine di \sqrt{n} , gli archi random **non** sembrano giocare più alcun ruolo.

- e la ragione di ciò è il fatto che gli archi random sono troppo random
 - come avevamo già osservato per il modello di Watts-Strogatz
- In effetti, quel che abbiamo visto per il caso “anello & $q = 0$ ” può essere generalizzato a tutti i casi “anello & $0 \leq q < 1$ ”
 - in tutti questi casi, gli archi random sono troppo random...
- Di contro nel caso “anello & $q > 1$ ” gli archi random sono troppo corti
 - e, quindi, è “difficile” imbattersi in archi random che coprano grandi distanze
 - e che permettano di avvicinarsi alla destinazione in pochi passi
- di conseguenza, la ricerca decentralizzata nel caso “anello & $q > 1$ ”, riesce a fare poco meglio che utilizzare soltanto archi dell’anello per raggiungere la destinazione
- In effetti, si può dimostrare il seguente

Teorema 2. *comunque si scelga $q \neq 1$, esistono due costanti positive a_q e c_q tali che, detta X la variabile aleatoria che esprime la lunghezza del percorso trovato dall’algoritmo di ricerca decentralizzata in un anello di n nodi cui sono aggiunti archi random in accordo al modello che abbiamo descritto,*

$$E[X] \geq a_q n^{c_q}$$