

Lezione n.15

Random Graphs, Small-Worlds, Scale-Free Networks

Materiale didattico:
Peer-to-Peer Systems
and Applications
Capitolo 6

SMALL WORLD NETWORKS

- Esperimento di Milgram: evidenza in modo empirico il fenomeno "small world" in una rete sociale
- small world = esistono delle "catene di conoscenze" di lunghezza limitata che connettono una qualsiasi coppia di persone sconosciute
- Six Degree of Separation
- Caratterizzazione del fenomeno:
 - random networks
 - modello di Watts-Strogatz
 - modello di Kleinberg

Random Networks

- posseggono diametro limitato
- la lunghezza media dei cammini è $O(\log(n))$, dove n è il numero di nodi della rete
- adatte a modellare le "small world network"
- caratterizzate da un basso coefficiente di clusterizzazione: la probabilità che i vicini di un nodo siano essi stessi vicini è bassa
- situazione reale: i vicini dei miei amici sono i miei amici.....

MODELLO DI WATTS STROGATZ

Modello di Watts Strogatz

- definisce di una rete regolare
griglia
anello con collegamenti con i nodi a distanza minore o uguale a k ...
- "sovrappone" a questa struttura regolare un insieme limitato di collegamenti generati in modo casuale
- la struttura regolare definisce un buon grado di clusterizzazione
- i collegamenti casuali garantiscono la proprietà di "small world"

SMALL WORLD: ROUTING

Caratteristiche rilevate empiricamente dall'esperimento di Milgram:

- esistono cammini di lunghezza limitata che connettono una qualsiasi coppia di individui
- gli individui sono in grado di **scoprire** questi cammini mediante una **conoscenza parziale (locale)** della rete
- perchè una qualsiasi coppia di individui è in grado di individuare in modo decentralizzato la catena limitata di conoscenza che li connettono?
- Quali caratteristiche della rete garantiscono l'esistenza di tale algoritmo di routing decentralizzato?

SMALL WORLD: ROUTING

Algoritmo di routing eseguito da ogni individuo:

- supponiamo che l'individuo sia posizionato nel vertice v di posizione $Pos(v)$ all'interno di una griglia in uno spazio d -dimensionale $Pos(v) = (x_1, x_2, \dots, x_d)$ dove tutti gli x_i sono interi. $x_i(v)$ è la posizione di v nella dimensione i .
- ogni individuo conosce la propria posizione, quella delle conoscenze dirette e quella del destinatario t (routing distribuito)
- il messaggio (la lettera) viene consegnata al conoscente w che è più vicino al destinatario. La misura della distanza $d_M(w, t)$ è la somma delle differenze in valore assoluto

$$|x_i(w) - x_i(t)| \text{ (Manhattan Distance).}$$

SMALL WORLD: ROUTING

- L'esperimento di Milgram suggerisce che la rete "incorpori" una conoscenza che consente di "guidare" il messaggio dalla sorgente alla destinazione, utilizzando ad ogni passo un insieme di conoscenze limitate.
- Milgram "Il messaggio si avvicina dal Nebraska al Massachusetts in modo progressivo. Ogni volta che una persona si aggiunge alla catena, il messaggio si avvicina (geograficamente) al target.

Il modello di Watts Strogatz

- garantisce la presenza di cammini brevi tra coppie di nodi, ma non garantisce l'esistenza di un algoritmo di routing decentralizzato che individui tali cammini
- la rete generata non "incorpora" la conoscenza necessaria per definire il routing decentralizzato



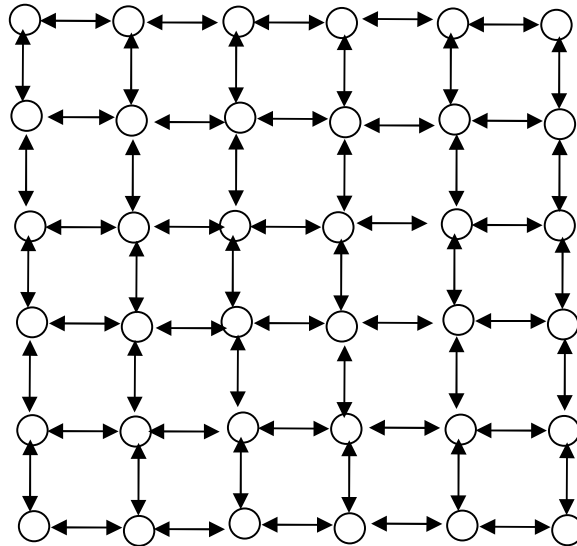
IL MODELLO DI KLEINBERG

- Reti con basso diametro
- Alto livello di clusterizzazione
- I contatti remoti vengono stabiliti tenendo conto della geometria della rete
- E' possibile definire un algoritmo di routing decentralizzato
- Utilizzato per la overlay networks P2P

IL MODELLO DI KLEINBERG

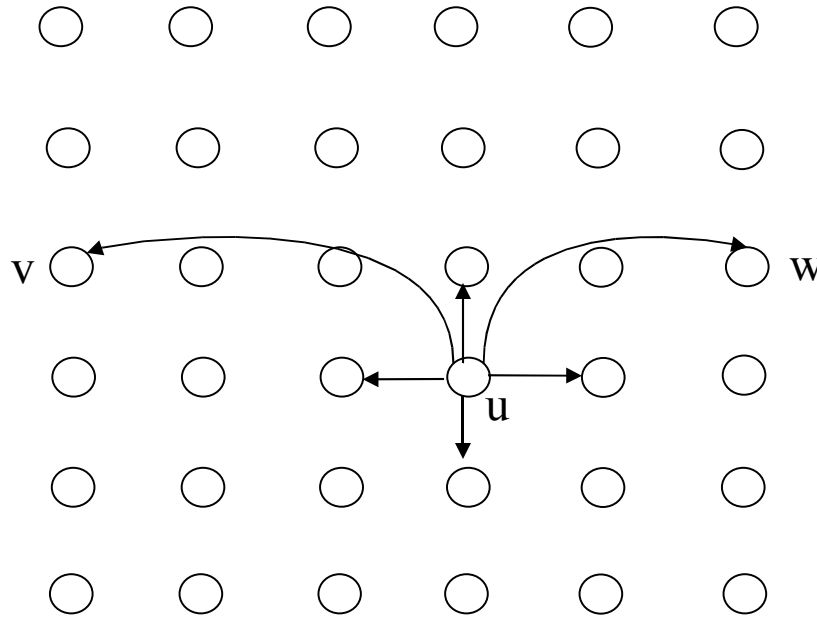
- considera un insieme di nodi corrispondenti ai punti di una griglia k dimensionale
- esempio: $k=2$ considera una griglia $n \times n$, che contiene i punti $\{(i,j) : i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,n\}\}$
- $d((i,j) \text{ e } (k,l)) = |k-i| + |l-j|$ distanza tra i nodi (i,j) e (k,l) .
- **Contatti locali:** data una costante $p \geq 1$, per ogni nodo u si definiscono archi diretti da u verso qualsiasi nodo che si trova entro una distanza p .
- **Contatti remoti:** data due costanti $q \geq 0, r \geq 0$, per ogni nodo u si definiscono q archi diretti da u verso q altri nodi.
l' i -esimo arco collega u al nodo v con probabilità $[d(u,v)^{-r}]$

IL MODELLO DI KLEINBERG



Una griglia bidimensionale, $n=6$, $p=1$, $q=0$

IL MODELLO DI KLEINBERG



Contatti del nodo u , $p=1$, $q=2$

IL MODELLO DI KLEINBERG

Teorema (griglie bidimensionali): Esiste un algoritmo di routing decentralizzato A ed una costante α_2 , indipendente da n , tale che il numero di passi impiegati da A per trasportare un messaggio dal mittente al destinatario è al più $\alpha_2 (\log n)^2$.

- **Modello di Watts Strogatz:**

i contatti remoti sono stabiliti in modo uniforme sulla griglia
non tengono conto della "geometria della griglia"

- **Modello di Kleinberg:**

La distribuzione dei contatti remoti è uniforme su diverse fasce di distanza

La probabilità che il contatto remoto di un nodo u si trovi ad una distanza compresa tra 2^{j-1} e 2^j è approssimativamente la stessa per tutti i valori di j compresi tra 1 ed $\frac{1}{2} \log n$

IL MODELLO GENERALE DI KLEINBERG

Modello generale:

si parte con una griglia a d dimensioni

sia $d_M(v,w)$ la distanza tra due nodi della rete. Tra due vertici v e w viene aggiunto un arco con probabilità

$$P(v, w) \sim d_M(v, w)^{-\alpha}$$

Teorema: L'algoritmo di routing riesce ad individuare un „cammino corto“, cioè un cammino di lunghezza $O(\log n)$ tra una coppia di vertici se e solo se $\alpha = d$.

IL MODELLO DI KLEINBERG: APPLICAZIONI P2P

- Il modello è stato utilizzato per la definizione di reti P2P
 - Symphony
 - Small World Overlay Protocol
 - Viceroy
- Freenet: alcuni risultati sperimentali hanno dimostrato che, per un carico medio/basso della rete (in termini di numeri di files condivisi), il sistema evolve autonomamente in una rete small world

SYMPHONY

- struttura di base analoga a Chord: identificatori associati, mediante funzione hash, a nodi e a files. Gli identificatori sono assegnati nell'intervallo $[0,1]$
- i nodi vengono inseriti logicamente su un anello, in base al loro identificatore
- *segment of responsibility*: ogni nodo è responsabile di tutti quei files con identificatore *maggiore o uguale* al proprio identificatore e minore dell'identificatore del nodo successivo
- Ogni nodo stabilisce $O(1)$ collegamenti con gli altri nodi
- Utilizza il modello di Kleinberg per:
 - ottimizzare numero di links
 - definire routing decentralizzato

SYMPHONY

Inserimento di un nuovo nodo nell'anello:

- scelta (funzione hash) dell'identificatore del nodo nell'intervallo $[0,1]$
- connessione ai vicini sull'anello (**contatti locali**)
- connessione a k nodi scelti in modo casuale (**contatti remoti**)

si sceglie un valore $x \in [0,1]$ con probabilità

$$P(X = x) = 1/(x * \log n)$$

dove n è il numero di nodi della rete

si stabilisce un collegamento remoto con il nodo che gestisce x

il sistema garantisce che i contatti remoti di ogni nodo siano limitati da una costante k

SYMPHONY

- **Algoritmo di routing greedy** (alla Kleinberg): ogni richiesta viene instradata verso il nodo che gestisce il segmento "più vicino" alla chiave contenuta nella richiesta
- **Teorema:** il numero medio di passi effettuati dall'algoritmo di routing di Symphony con $k = O(1)$ connessioni remote stabilite da ogni nodo è inversamente proporzionale a k e proporzionale a $(\log n)^2$

POWER LAW/SCALE FREE NETWORKS

- $K(v)$ grado di un nodo : numero di archi incidenti in quel nodo (entranti + uscenti in un grafo diretto)
- Analisi di alcune reti: www (Barabasi, 99) Internet (Faloutsos,99)
- Dai risultati sperimentali, in queste reti:

la maggior parte dei nodi sono connessi ad un insieme ristretto di nodi

un numero limitato di nodi della rete sono caratterizzati da un alto grado, mentre la maggior parte dei nodi è caratterizzata da un basso grado

POWER LAW/SCALE FREE NETWORKS

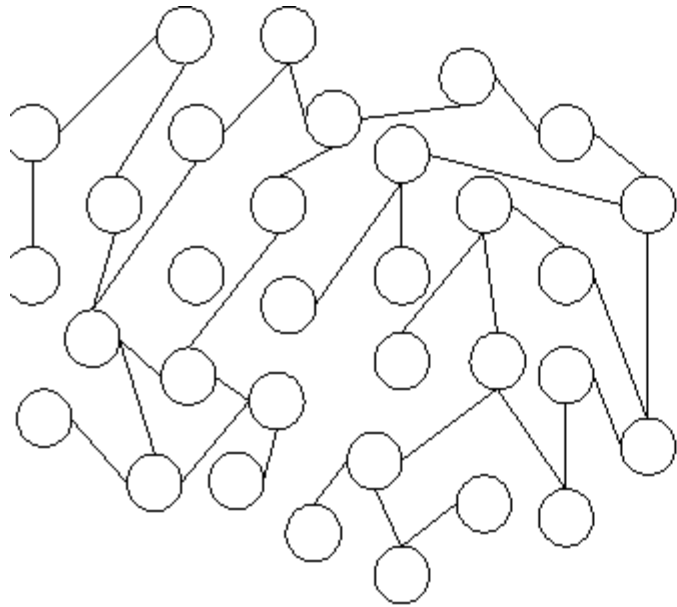
Esperimento di Barabasi:

- viene esaminato parte del www
- si costruisce la rete corrispondente, in cui i nodi corrispondono alle pagine, gli archi ai links tra pagine.
- si calcola $P(K)$ = probabilità che un nodo della rete abbia grado k
- risultato:

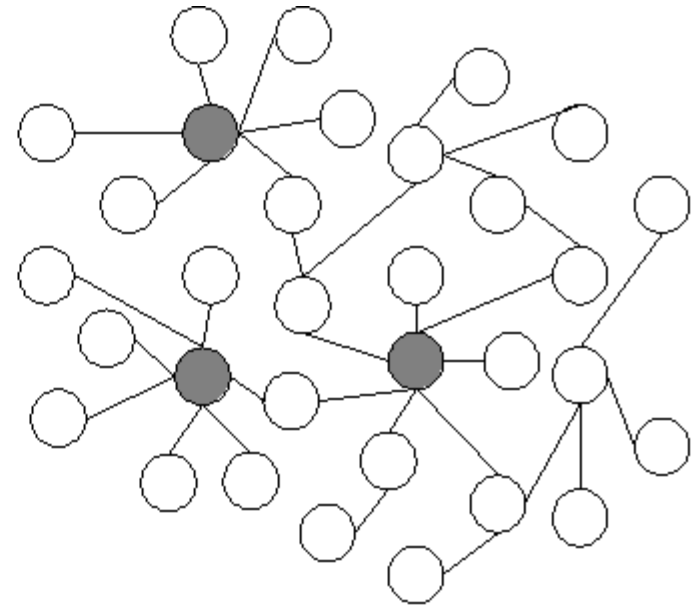
$$P(k) \cong k^{-\gamma}$$

valori tipici di γ compresi tra 2 e 3.

POWER LAW/SCALE FREE NETWORK



(a) Random network



(b) Scale-free network

IL MODELLO DI BARABASI-ALBERT

Generazione di Scale Free Networks:

- Il tempo è discretizzato: la rete cresce nel tempo, ad ogni passo un nuovo nodo entra nella rete
- la generazione di un arco tra un nuovo nodo w ed uno vecchio v non avviene in modo casuale, ma segue il seguente principio
più alto è il grado di v , più alta è la probabilità che w stabilisca un collegamento con v
- Con una metafora.....**the rich get richer**

Preferential Attachment: La probabilità $\Pi(v)$ che si stabilisca un collegamento tra il vertice v ed il nuovo vertice w è definita come

$$\Pi(v) = \frac{\deg(v)}{\sum_{w \in V} \deg(w)}$$

IL MODELLO DI BARABASI-ALBERT

- si inizia con una rete di piccole dimensioni (ad esempio 10 vertici, 20 archi, generati in modo casuale)
- ad ogni passo, si aggiunge un nuovo vertice x .
- si aggiungono m archi che connettono x ai vertici v che si trovano già nel grafo. La probabilità di connessione ad un vertice esistente segue il **preferential attachment**

SCALE FREE NETWORKS: CARATTERISTICHE

- robustezza rispetto a guasti.

se si rimuove in modo casuale (distribuzione uniforme) un nodo v da una rete scale-free, con alta probabilità, tale vertice è un vertice di grado basso

la rimozione di un vertice di grado basso non costituisce un evento „disastroso“

Le scale-free networks presentano buone caratteristiche di tolleranza ai guasti

- sensibilità verso gli attacchi

un attacco può colpire i vertici di grado più alto

evento disastroso: provoca il frazionamento della rete in tante componenti di piccola dimensione

P2P E SCALE FREE NETWORKS

- Diversi protocolli P2P generano overlay network con caratteristiche di tipo scale free
- **Gnutella**: ogni nodo si connette inizialmente ad alcuni servers, memorizza nella cache i nodi contattati in sessioni precedenti

alcune analisi mostrano che questo induce una struttura scale free nell'overlay network creata

positivo per la tolleranza ai guasti, negativo per la sensibilità agli attacchi

definiti protocolli "ad hoc" per recuperare la rete in caso di attacchi ai suoi nodi vitali.

RETI P2P, SMALL WORLDS, SCALE FREE

Quali dovrebbero essere le caratteristiche di una rete peer to peer?

- se si vuole rendere la rete „navigabile“ in modo decentralizzato si può definire una rete „small world“ alla Kleinberg. In questo modo si può definire un opportuno algoritmo di routing
- se la rete P2P è definita in un contesto piccolo e sicuro (es una intranet di un'organizzazione) si può renderla una rete scale free, utilizzando un numero limitato di hubs, di alta banda
- se la rete P2P può essere soggetta ad attacchi si può definire una rete „small world“, in cui la maggior parte dei vertici sono caratterizzati da un basso grado

CONCLUSIONI

- La struttura di una rete peer-to-peer system influenza
 - la lunghezza media dei cammini,
 - la possibilità di definire algoritmi di routing greedy, decentralizzati
 - la stabilità della rete verso guasti
 - la sensibilità della rete verso gli attacchi
 -
- Caratteristiche importanti per la valutazione di una rete:
 - la lunghezza media dei cammini,
 - il coefficiente di clusterizzazione,
 - la distribuzione del grado dei nodi
- E' importante stabilire le regole per la generazione degli archi della overlay network in modo tale che la struttura della rete definita garantisca le proprietà desiderate