Reti e modelli random

Sistemi complessi: sistemi in cui varie entità interagiscono tra loro.

Esempio: infrastrutture di comunicazione

"Complesso": riferito a numero entità e tipo di interazione.

Reti: modello matematico che rappresenta il funzionamento di un sistema complesso.

Grafo $\equiv (V, E)$ $\qquad V \equiv$ insieme vertici (entità)

$E \equiv$ insieme archi (interazioni entità)

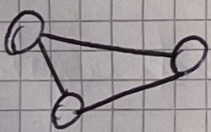
Network science: scienza che studia i grafi
E' una disciplina molto giovane che abbraccia varie materie come matematica, fisica, informatica ...

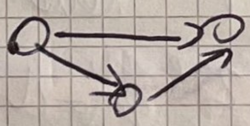
Grafi diretti e indiretti

$(a, b) \in E \rightarrow b$ è adiacente a a

Grafo diretto: se $(a,b) \in E$ vicendevolmente $(b,a) \in E$

Grafo indiretto: non è detto

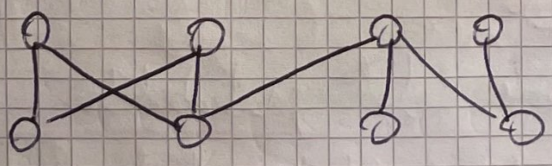




Si parla di **rete pesata** se ogni arco ha un numero associato che ne rappresenta il peso

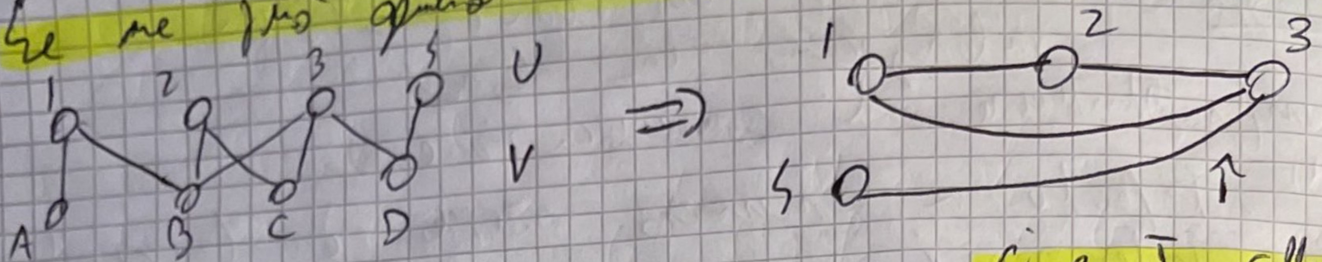
Si possono associare valori anche ai nodi, potrebbero essere delle etichette

Grafo bipartito: nodi divisi in U e V disgiunti ogni arco collega un nodo a un nodo dell'altro classe.



Proiezione del grafo bipartito

Se ne può generare una relativa ad U e ad una a V



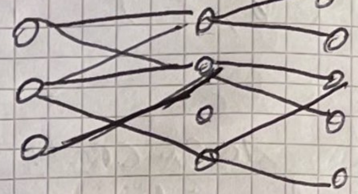
Proiezione

i e j collegati se hanno almeno un nodo di V in adiacenza

Si può costruire anche la proiezione V

Grafo multipartito

Estensione del bipartito



(1° non va a 3°)

Grafo completo

Grafo in cui tutte le coppie di punti sono collegate da un arco

Grafo regolare

Tutti i nodi hanno lo stesso grado

(grado di un nodo è il numero di archi collegati)

Se un nodo ha grado 0 è detto isolato

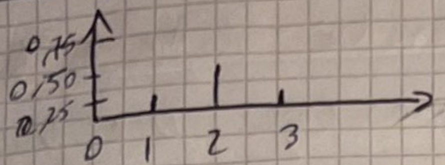
Nei grafi diretti abbiamo 2 gradi

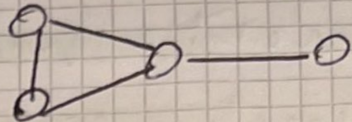
- Grado uscente di n: nodi adiacenti a n
- Grado entrante di n: nodi a cui n è adiacente
- Grado totale di n: grado uscente + grado entrante

Grado medio del grafo $\langle K \rangle = \dfrac{|E|}{|V|}$

Distribuzione dei gradi

È una distribuzione di probabilità dove p_k è il numero relativo di nodi con grado k





$P_K = N_K / N$ dove $N_K \equiv$ nodi con grado k
$N \equiv$ nodi

Cammino tra 2 nodi e distanza

Cammino tra i e U e V è una seq. ordinata di n archi $(i_0, i_1) \dots (i_{n-1}, i_n)$ dove $i_0 = U$ e $i_n = V$

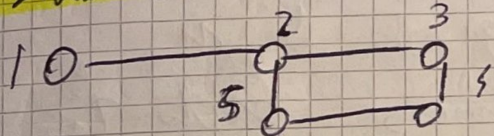
$n \equiv$ lunghezza cammino

Il cammino minimo è quello con lung. minore

Distanza D tra 2 nodi: lung del cammino minimo

Diametro di un grafo

Distanza massima tra 2 nodi in un grafo



$d = D(1,4) = 3$

Ciclo

Particolare cammino che torna al nodo di partenza

Un particolare ciclo è quello di lung 1 qualora esistesse un cappio

Un grafo senza cicli è detto aciclico o albero.

Connettività

i e j sono connessi se c'è un cammino tra i e j se no sono disconnessi

Un grafo è connesso se ogni coppia è connessa se no è disconnesso.

Un grafo disconnesso è l'unione di tanti sottografi connessi detti componenti.

Nei grafi orientati vi è connettività forte e debole

G è fortemente connesso se per ogni u e v esiste il cammino tra u a v

G è debolmente connesso se per ogni u e v esiste un cammino tra u a v o se esiste un cammino in G' (grafo come G in cui non c'è direzione)

Componente ⇒ può essere debolmente o fortemente connessa

Coefficiente di clustering

$C_n \Leftarrow n$ Misura di quanto gli adiacenti di n sono connessi tra loro.

$C_n = \frac{2L_n}{k_n \times (k_n - 1)}$

k_n : grado di n

L_n : numero archi tra i k_n adiacenti di n

C_n va da 0 a 1

Il coefficiente di clustering misura la densità locale della rete in un nodo n

Più densamente interconnesso è il vicinato di n più alto è il coefficiente di clustering

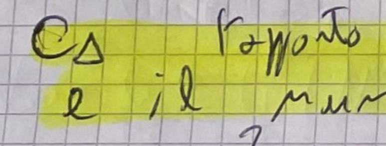
Coefficiente di clustering medio

Media dei coefficienti di clustering di ogni nodo

$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$

Coefficiente di clustering globale

C_Δ rapporto tra numero triangoli nella rete e il numero di triple di nodi connessi tra loro



1-2-3 è una tripla (1 e 3 hanno il cammino)

Centralità di un nodo

Misura l'importanza di un nodo nella rete
Esistono diverse misure di centralità

Degree centrality: è la più semplice, è data semplicemente dal grado del nodo
Più è alto il grado più è importante
È una misura troppo semplice.

Betweenness centrality

Dato v e i, j si calcola σ_{ij} ovvero la frazione di cammini minimi tra i e j che passano per v
La Betweenness di v è ottenuta sommando σ_{ij} per ogni i, j

Dunque un nodo è centrale/importante se sta in mezzo a molti path

Closeness centrality

È basata sulla vicinanza media di un nodo rispetto agli altri nodi

Dato v si calcola L_v lung media dei cammini minimi da v agli altri nodi.

La closeness centrality di v è $1/L_v$

Questa misura la velocità media con cui un'informazione partendo da v arriva agli altri nodi

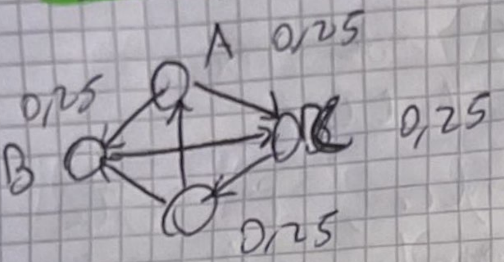
Page Rank centrality Un nodo è più importante se collegato a un nodo importante

Assumiamo che nodi con elevato grado hanno connessioni con peso maggiore rispetto alle connessioni a nodi di grado minore

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{K_v}$$

K_v: grado uscente di v
B_u: nodi con u attaccato a v

Simulazione Page Rank (Tesoretto)



Ogni nodo inizialmente ha peso uguale.

Ogni nodo deve cedere il suo tesoretto in parti uguali agli adiacenti

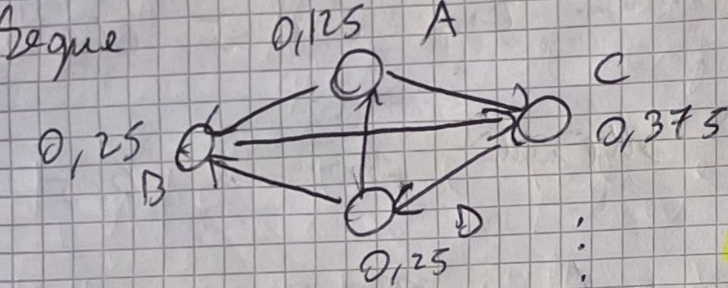
A: cede 0,125 a B e C

C: cede 0,25 a D

D: 0,25 a A e B

B: 0,25 a C

Segue



⋮

Iterando poi varie volte si arriva a una situazione di equilibrio in cui i pesi non variano più e quindi ci fermiamo.

Modelli Random

Permettono di generare dei grafi/reti con determinate proprietà.

Erdos-Renyi

Fissata p e numero di Nodi N

1) Si creano N nodi isolati
2) Per ogni coppia di nodi genera un arco con probabilità p

Si ottiene alla fine un grafo random o rete di Erdos-Renyi

Variante G(N, L): qui non c'è p ma L che è il numero di archi che dobbiamo ottenere

Si generano dunque N nodi isolati e si prendono 2 nodi casuali e vi si mette un arco.
Questo finché non ho L archi

Proprietà del grafo random

Con N nodi, P_k : prob che un nodo abbia k grado

P_k è il prodotto di 3 probabilità

- prob i sia connesso a k nodi distinti $\equiv p^k$
- prob i non sia connesso ai restanti $N-1-k$ nodi ovvero $(1-p)^{N-1-k}$
- numero modi in cui si possono selezionare k nodi da collegare a i, da scegliere tra $N-1$
  $\binom{N-1}{k}$

La distribuzione dei gradi in un grafo random è una binomiale $P_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$

Il grado medio della rete è $\langle k \rangle = p(N-1)$

Per $N \gg \langle k \rangle$ si può approssimare a una distribuzione di Poisson. $P_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$

Dato che la distribuzione è binomiale tutti i nodi hanno circa lo stesso grado

In reti grandi molti nodi hanno grado vicino a $\langle k \rangle$ mentre pochi avranno un grado diverso.

Distanza media tra nodi

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$$

$\ln N \ll N$ dunque le distanze tra nodi nelle reti random sono piccole in media.

$1/\ln\langle k \rangle$ indica che più è densa la rete più sono piccole le distanze.

Se k_n è il grado di n il numero atteso di archi è
$$\langle L_n \rangle = p \frac{k_n(k_n-1)}{2}$$
Il coefficiente di clustering di n è
$$C_n = \frac{\langle L_n \rangle}{k_n(k_n-1)/2} = p = \frac{\langle k \rangle}{N-1}$$
Più è grande la rete più è basso il coefficiente di clustering. La stessa considerazione è fatta per il coefficiente di clustering della rete.

NB: il coefficiente di clustering è indipendente dal grado del singolo nodo

I grafici di reti reali e random sono significativamente diversi

Nelle reti reali: molti nodi con grado basso e pochi nodi con grado alto

I grafici presentano forti differenze sia per la distribuzione dei gradi sia per il coefficiente di clustering (nodo e rete)

- In particolare il coeff. di clustering di una rete reale è più alto di quello di una rete random

SMALL WORLD

Unica proprietà che i grafi random riescono a cogliere abbastanza bene dei grafi reali.

Il fenomeno ci indica che in una rete grande 2 nodi sono collegabili tramite pochi nodi

Da qui nasce la teoria dei 6 gradi di separazione

Da qui viene l'esperimento di Milgram che conferma che effettivamente tra 2 individui ci sono praticamente sempre massimo 6 nodi di separazione

Definizione di Small World

Una rete soddisfa questa proprietà se $\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$

Le reti reali soddisfano le la proprietà
ad esempio internet ha una $\langle d \rangle = 6,58$

Modello di Watts - Strogatz

Erdos-Renyi approssima la proprietà "small world" ma non il coefficiente di clustering.

Il modello di Watts - Strogatz è un'estensione del 1°
il Modello permette di creare un'interpolazione tra 2 grafi

- Un grafo regolare = alto coeff. di clustering ma non ha proprietà small world
- Un grafo random

La distribuzione dei gradi è simile una distr. di Poisson.

Il tipo di grafo dipende da un parametro p detto parametro di rewiring. (p è una probabilità)

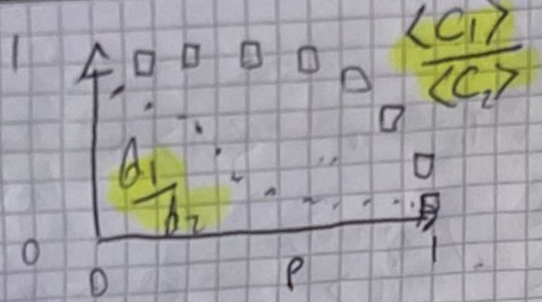
Algoritmo

Dati N (num nodi), il parametro p e un $d \geq 1$ intero

1) Siano $1, 2 \dots N$ gli id dei nodi. Collega ogni nodo i a $i+1, i+2 \dots i+d$. Otteniamo un grafo regolare
2) Per ogni arco, con probabilità p, scambia il nodo destinazione dell'arco con un altro nodo random

Il grafo ottenuto ha caratteristiche intermedie tra il regolare e il random

p = basso $\Rightarrow$ rete regolare $\quad$ p = alto $\Rightarrow$ rete random



$\langle C_1 \rangle$: coeff di clustering con parametro p che cresce →

$\langle C_2 \rangle$: coeff di clustering in una rete regolare.

d_1: d media con parametro p che cresce →

d_2: d media in una rete regolare.

NB: Se una rete è regolare vuol dire che ha parametro $p = 0$

Analizzando il grafico si può capire per quale p posso ottenere un coefficiente di clustering alto e una d media tra i nodi bassa contemporaneamente

Distribuzione power-law

Nel modello di Watts - Strogatz la distribuzione dei gradi è simile a una Poisson.

Nelle reti reali la distribuzione è diversa ed è detta power-law $P_k \sim k^{-\gamma}$ dove γ è detto esponente di grado

Nelle reti reali in genere $2 < \gamma < 3$

Dunque nemmeno questo modello approssima bene della distribuzione dei gradi

Rete Scale - free

Una rete è scale-free se la distribuzione dei gradi segue una una power-law

Nelle scale-free vi è una porzione piccola di nodi con grado elevato e una grande porzione con grado basso.

Nodi con grado alto: Hub

Una rete scale-free è caratterizzata dalla presenza di certi Hub, che sono assenti nei modelli di reti random e small-world.

Il comportamento ~~delle~~ scale-free delle reti reali è riassunto nella regola di Pareto (80/20)

Nelle reti scale-free dunque vi sono nodi più "centrali"

"Scale-free" si riferisce al fatto non c'è una scala (un valore di riferimento) che permette di stabilire il grado di un nodo.

In una rete random la media si trova facilmente ma in una rete scale-free non vi è alcuna informazione per stimare un valore, infatti la varianza è troppo alta.
In una rete scale-free preso un nodo è difficile/impossibile stimarne il grado.

Robustezza

Gli hub rendono la rete più robusta per gli attacchi esterni e nel caso un nodo vada "giù"

In una rete random la "caduta" di un nodo può compromettere tutta la rete.

In una rete scale-free il problema è grave solo se colpisce gli hub.

Per questo in una rete si proteggono meglio gli hub.

Proprietà Ultra small-world

La presenza ~~media~~ degli hub riduce la d media tra i nodi.
Per $2 < \gamma < 3$ (valori tipici nella rete reale) si ottiene la proprietà ultra small world ovvero la d media è ancora più bassa rispetto alla d small-world

Per $\gamma > 3$ la rete è small-world ed è veramente molto simile ad una rete random.

Per $\gamma < 2$ si avrebbero cose strane. La media e la varianza divergono. Perciò non possono esistere reti grandi con $\gamma < 2$.

Perché le reti reali sono scale-free?

Molte reali reti assumono il comportamento scale-free.

Caratteristiche reti reali

- Crescita: nelle reali il numero di nodi cresce continuamente nel tempo
- Preferential attachment: nelle reali un nodo nuovo tende a legarsi agli hub piuttosto che ai nodi periferici (nodi con più connessioni)

Modello di Barabasi-Albert

Il preferential attachment è ciò che favorisce la formazione degli hub.

Un nodo con alto grado ha più probabilità di stabilire nuove connessioni e diventare quindi sempre più "importante"

- "Rich gets richer"

Col modello di Barabasi-Albert si possono creare reti scale-free.

Crescita e preferential attachment sono ciò che ci permette di creare un modello per creare/generare reti scale-free

Algoritmo

1) Crea rete random iniziale con m_0 nodi dove ogni nodo ha grado almeno 1
2) Aggiungi nuovo nodo e collegalo a $m \le m_0$ nodi della rete.

Prob. che si colleghi a un nodo i con grado k_i:

$$P(k_i) = \frac{k_i}{\sum_j k_j}$$

3) Ripeti 2 finché non si arriva al numero M di archi desiderato.

Il passo 2 garantisce la proprietà del preferential attachment