Lezione 6

Status	Done
Attach file	6_Grafi.pdf

1- Introduzione

Nella realtà esistono molti esempi di sistemi complessi, caratterizzati da entità che interagiscono (fisicamente, chimicamente, ecc...) tra loro.

Il termine "complesso" si riferisce non solo alla dimensione di tali sistemi, ma anche
alla complessità delle loro interazioni e alla difficoltà nel comprenderli.

Le reti (o grafi) sono il modello matematico che permette di "codificare" le interazioni tra le componenti di un sistema complesso.
Formalmente, un grafo è una coppia $G(V, E)$ dove:

- V è l’insieme dei vertici, ossia le componenti del sistema;
- E è l’insieme degli archi ossia le interazione tra una coppia di vertici (u,v);

Network Science è la scienza che analizza le reti, per comprendere le leggi fondamentali che regolano il funzionamento dei sistemi complessi. La maggior parte
delle reti reali si formano seguendo principi organizzativi comuni.

2- Concetti di base sui grafi

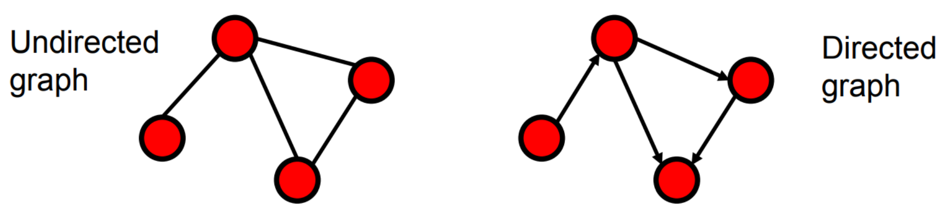
Tipi di grafi

Grafi diretti e indiretti

Un grafo è indiretto (o non orientato) se $\forall(a,b) \in E,\ (b,a) \in E$, altrimenti è diretto (o orientato).

In un grafo diretto si fa distinzione tra:

- Grado uscente del nodo n: numero di nodi adiacenti a n nel grafo;
- Grado entrante del nodo : numero di nodi a cui il nodo n è adiacente;
- Grado totale del nodo: grado uscente di n + grado entrante di n.



Grafi pesati e etichettati

In un grafo pesato, a nodi e/o archi sono associati valori numerici.
In un grafo etichettato, a nodi e/o archi sono associate etichette o stringhe.

Grado dei nodi e distribuzione dei gradi

Se $(a,b) \in E$, allora b è adiacente ad a.

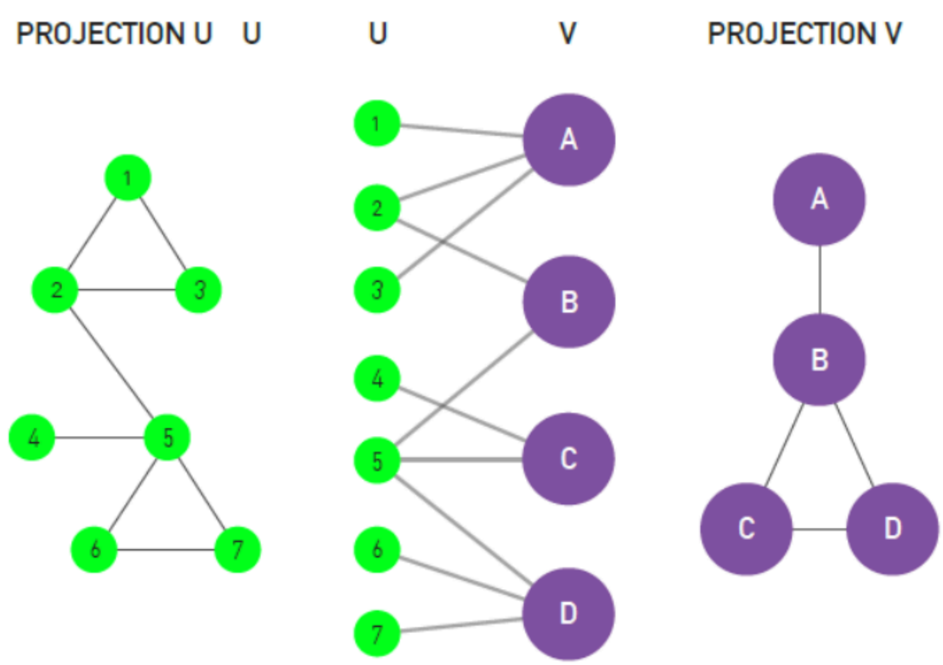
Il grado di un nodo è il numero di nodi adiacenti al nodo considerato. (se =0 allora il nodo è isolato).

La distribuzione dei gradi P è una distribuzione di probabilità e rappresenta la probabilità che un nodo random della rete abbia grado k. Si ottiene, banalmente, in maniera frequentista.

Grafi bipartiti

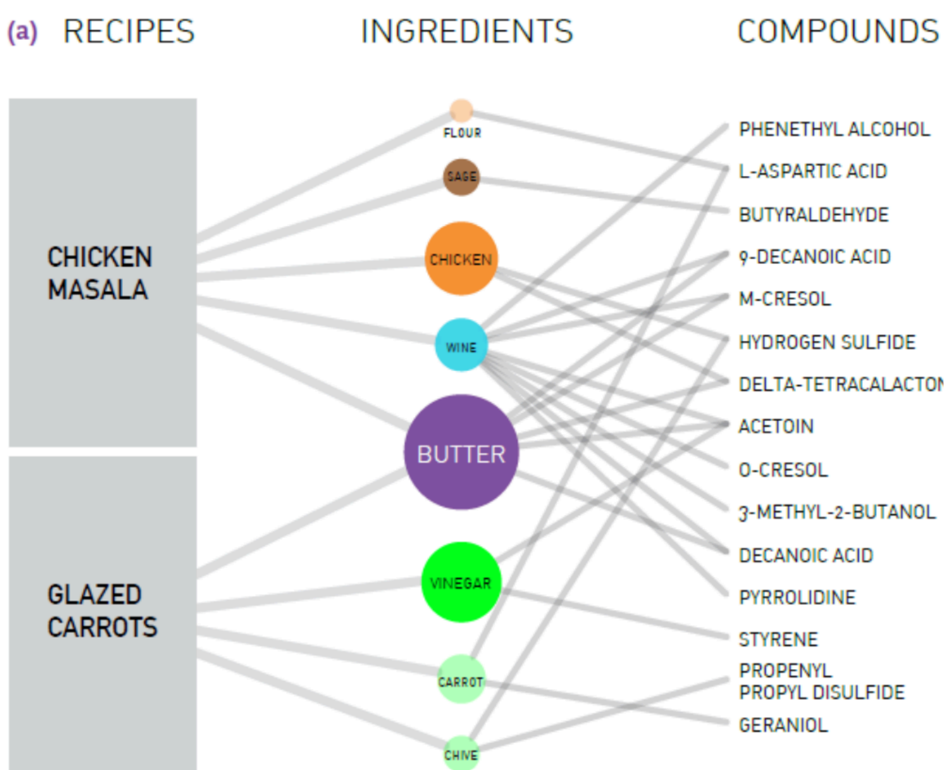
Un grafo bipartito è un grafo in cui sono presenti due insiemi disgiunti di nodi U e V e ciascun arco del grafo collega un nodo di U ad un nodo di V.
Non esistono archi che collegano tra loro nodi dello stesso insieme.

Da un grafo bipartito è possibile generare due reti chiamate proiezione rispetto a U e proiezione rispetto a V.
Nella proiezione rispetto a U due nodi di U sono connessi se hanno almeno un adiacente in comune nel grafo bipartito. Definizione analoga per la proiezione rispetto a V.



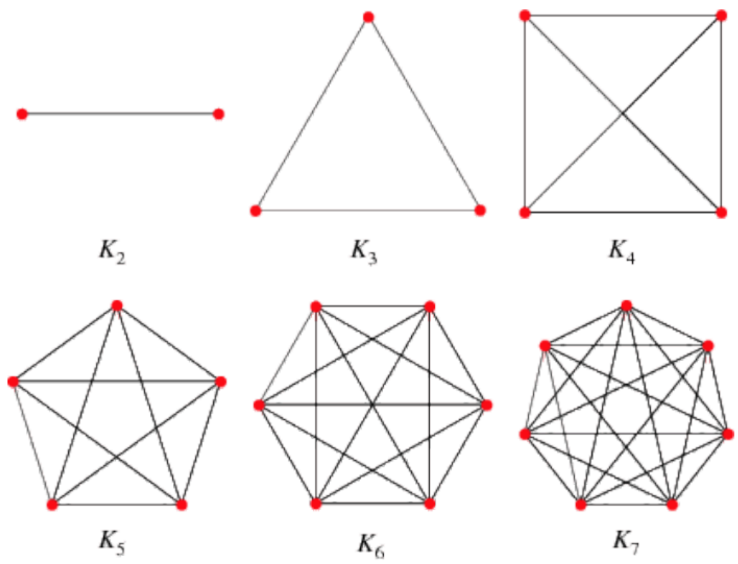
Grafi multipartiti

La definizione di grafo bipartito si può estendere facilmente al caso di tre (grafi tripartiti) o più (grafi multipartiti) insiemi disgiunti di nodi.

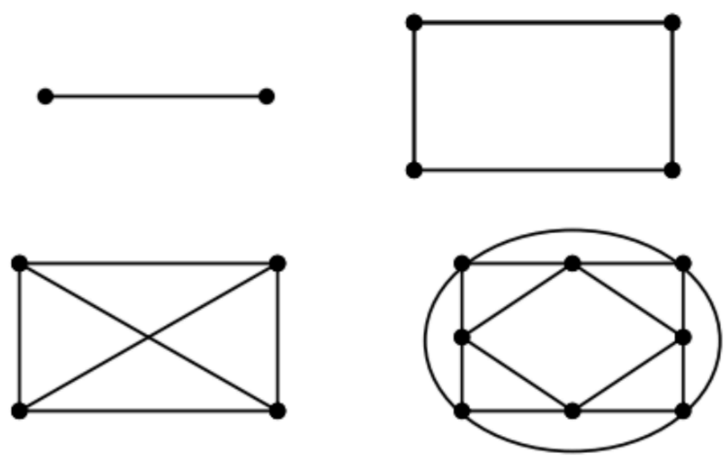


Grafi completi e regolari

Un grafo completo (o clique) è un grafo in cui tutte le coppie distinte tra loro sono collegate da un arco.



Un grafo regolare è un grafo in cui ogni nodo ha lo stesso grado.



Connettività del grafo

Un cammino tra due nodi è una sequenza ordinta di archi che portano dal primo al secondo. La lunghezza del cammino è il numero di archi che vengono attraversati.

I cammino minimo (possono essere anche più di uno) tra due nodi è il cammino di lunghezza minima. La distanza tra due nodi è la lunghezza del cammino minimo.

Il diametro di un grafo è la massima distanza tra due nodi del grafo.

Un ciclo è un cammino in cui il nodo di partenza e il nodo finale coincidono.
Un ciclo di lunghezza 1 (ovvero un arco da un nodo a se stesso) è detto cappio o

self-loop.
Un grafo che non contiene cicli è detto grafo aciclico, o anche albero.

🌴 Due nodi i e j si dicono connessi se esiste un cammino tra essi, disconnessi se tale cammino non esiste.

Un grafo è connesso se tutte le coppie di nodi sono connesse, disconnesso altrimenti.
Un grafo disconnesso G risulta formato dall'unione di due o più sottografi (cioè grafi formati da un sottoinsieme di nodi di G e dagli archi che li collegano) connessi, chiamati componenti connesse.

Un grafo orientato G è fortemente connesso se, per ogni coppia di nodi u e v, esiste un cammino
tra u e v in G.

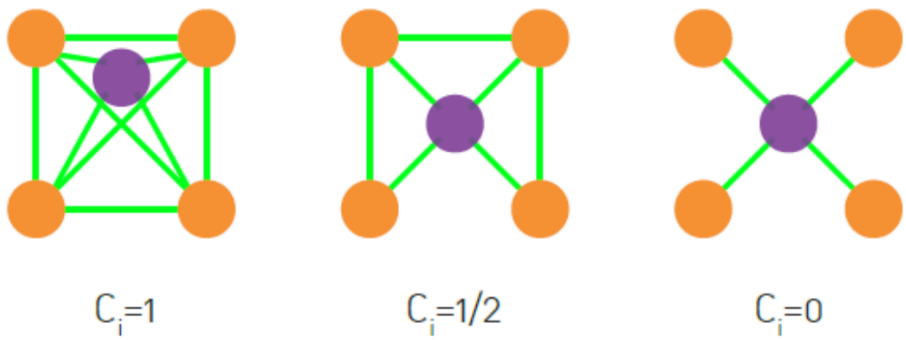
Un grafo orientato G è debolmente connesso se, per ogni coppia di nodi u e v, esiste un cammino
tra u e v in G, oppure esiste un cammino nel grafo G′ ottenuto da G sostituendo gli archi direzionati con archi non direzionati.

Coefficente di clustering

Il coefficiente di clustering di un nodo n è una misura di quanto gli adiacenti di n siano connessi tra loro. Misura la densità locale della rete in un nodo n. Più intensamente è connesso il vicinato di n, più alto è il coefficiente. Formalmente:

$$C_n = \frac{2 \cdot L_n}{k_n x(k_n - 1)} \in [0, 1]$$

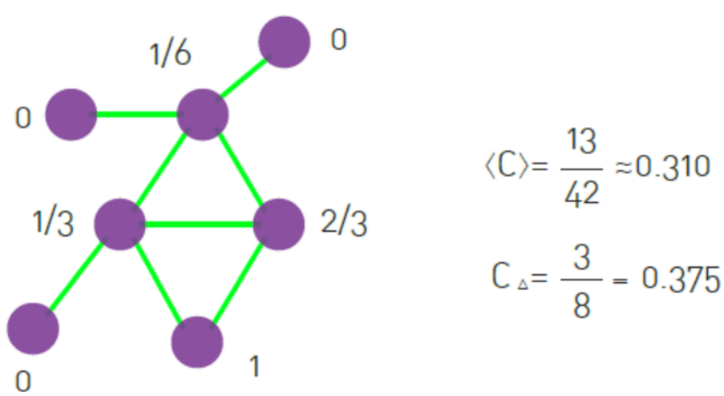
dove k_n è il grado di n e L_n è il numero di archi esistenti tra i k_n adiacenti di n.



Il coefficiente di clustering medio è la media dei coefficienti di clustering su tutti i nodi della rete:

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

Il coefficiente di clustering globale è il rapporto tra il numero di triangoli della rete (cliques con 3 nodi) e il numero di triple di nodi connessi tra loro (che possono essere cammini di 3 nodi oppure triangoli).



3- Misure di centralità

La centralità è una misura dell'importanza di un nodo nella rete.

Centralità basata sul grado

La degree centrality coincide il grado del nodo. Più alto è il grado di un nodo e più è importante.

Closeness centrality

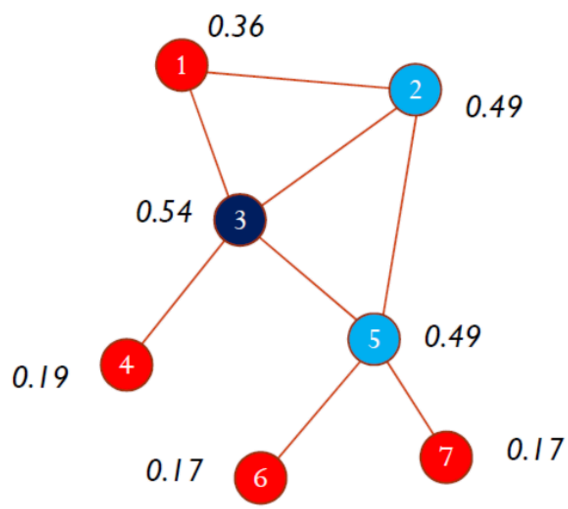
La closeness centrality si basa sulla vicinanza media di un nodo rispetto a tutti gli altri nodi del grafo.
La closeness centrality di un nodo v è definita come il reciproco di Lv, lunghezza media dei cammini minimi da v agli altri nodi del grafo.
Misura la velocità media con cui un'informazione, partendo da v, può raggiungere tutti gli altri nodi del grafo.

Betweenness centrality

La betweenness centrality si basa sul concetto di betweenness di un nodo.
Consideriamo un nodo v e due nodi qualsiasi del grafo i e j, si calcola la frazione di cammini minimi (σij) tra i e j passanti per v.
La betweenness di v è ottenuta sommando σij per tutte le coppie di nodi i e j.
Un nodo è centrale se è in mezzo a molti path di comunicazione tra nodi del grafo.

Page rank

La PageRank centrality si basa sull'osservazione che le connessioni di un nodo con gli altri nodi non hanno tutte lo stesso valore.
Connessioni a nodi con elevato grado hanno un peso maggiore rispetto alle connessioni a nodi di grado minore.
Definizione ricorsiva: un nodo è tanto più importante quanto più è connesso ad altri nodi importanti nella rete.



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

dove B_u è l'insieme dei nodi che hanno u come adiacente e kv è il grado uscente di v.