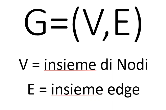
Slide 4

Grafo → metodo di rappresentazione di insiemi di dati, in cui esistono delle relazioni tra i nodi.

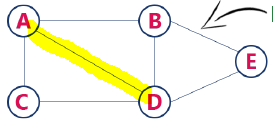
I **nodi** sono i punti del dataset, possono essere dei concetti, dei pazienti, degli utenti ad esempio di Twitter che si inviano un messaggio…

Gli **archi** (anche detti ***edge*** o ***link***) indicano una relazione tra i due nodi.

Tendenzialmente un grafo lo si vede indicato così:



UNDIRECTED GRAPH

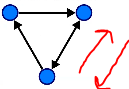
I nodi del grafo sono uniti da archi che non hanno una direzione.

Questo indica una relazione biunivoca: il nodo A è in relazione con il nodo D ed il nodo D è in relazione con il nodo A.

Slide 5

DIRECTED GRAPH

I grafi che ad esempio rappresentano i messaggi di Twitter sono grafi diretti, ovvero grafi in cui il nodo A manda un messaggio al nodo B, ma il nodo B non rimanda un messaggio al nodo A.

 I nodi diretti possono avere archi che rappresentano entrambe le direzioni, in questo caso si dicono **archi paralleli**.

(Solitamente si trovano come in figura, ma dovrebbero essere indicati come in rosso.)

WEIGHTED GRAPH

Alcuni grafi possono essere **pesati**: gli edge trasportano un peso che tendenzialmente indica la forza della relazione.

Esempio: se si ha una matrice di similarità, quindi una matrice quadrata in cui ogni nodo, ogni elemento della matrice rappresenta la similarità tra due campioni; i campioni sono i nodi del grafo , l’elemento sulla matrice di similarità può essere visto come l’elemento di una matrice di adiacenza che dice quanto quei nodi sono simili ovvero questa similarità può essere il peso sull’arco che connette nodi adiacenti.

• Quindi per elementi che hanno similarità 0, non ci saranno archi.

• Per elementi che hanno una similarità >0, ci sarà un arco che li unisce e questo arco è pesato dalla similarità.

Esempio

In campo medico: i nodi di un grafo rappresentano i pazienti che sono collegati sulla base della loro similarità, si fa una network di pazienti.

I grafi pesati vengono utilizzati per la predizione dei rischi sui pazienti.

Altro esempio

I nodi sono i potenziali terroristi, sono collegati se tra loro c’è una relazione e il peso indica la forza della relazione che indica quanto i due personaggi interagisono o la loro similarita in termini di nazionalità, potenza dell’attacco..

COMPLETE GRAPH che possono essere:

* FULLY CONNECTED GRAPH: ogni nodo è collegato ad ogni altro nodo.
* DENSELY CONNECTED GRAPH: ogni nodo ha un alto *degree*.

REGULAR GRAPH

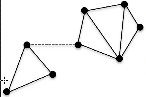
Ogni nodo è connesso allo stesso numero di nodi, ovvero (in figura) ogni nodo è connesso a due vicini.

Quindi i vicini sono i nodi direttamente connessi ad ogni nodo.

Questo è un caso particolare di grafo formato da due triangoli: la loro presenza indica che ogni nodo ha due vicini, che sono anche vicini tra di loro; ossia ogni nodo è connesso a due vicini che sono anche connessi tra di loro.

Questo grafo è composto da due componenti connesse, ovvero due sottografi in cui ogni nodo è raggiungibile tramite un cammino nel nodo.

Cammino: insieme di edge che portano da un nodo ad un altro.

Ad esempio qui si ha un grafo connesso perché ogni coppia di nodi è connessa da un grafo.

Se si parte da un qualsiasi nodo, si può raggiungere un qualsiasi altro nodo, attraverso un cammino.

Il link tratteggiato indica l’edge, che se rimosso, crea due componenti connesse.

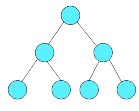
Il numero di componenti connesse all’interno di un grafo è fondamentale.

Il diametro di un grafo che ha una sola componente connessa è il cammino minimo maggiore (più lungo) all’interno del grafo.

Se si prende ogni nodo e si calcolano i cammini che connettono quel nodo a tutti gli altri nodi del grafo, si calcola per ogni nodo il cammino minimo che lo porta a quel nodo, poi si prende un altro nodo si calcola il cammino minimo che lo porta a quel nodo e così via; quindi per ogni nodo si ha il nodo a lui più distante, ossia il nodo per cui il cammino minimo è più lungo.

Se si fa questo per tutti i nodi del grafo si ha il cammino minimo più lungo, ossia il diametro del grafo.

Slide 6

TREE

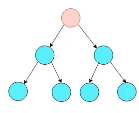
Ci sono dei grafi particolari ossia gli alberi che sono generalmente formati da un nodo che è quello più importante ossia il nodo padre, da questo nodo padre (nel caso di grafi diretti) sono raggiungibili le foglie, ma delle foglie non si può riraggiungere il padre perciò non esistono cicli all’interno del grafo.

I trangoli formano un ciclo, ossia si può tornare al nodo di partenza.

I cicli sono meno frequenti nei grafi direzionati, cioè grafi con edge orientati (con una direzione), però esistono anche in grafi direzionati.

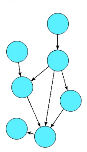
Gli alberi sono particolari perché non ci sono cicli, cioè si parte da un nodo e si raggiunge la coda.

Nel caso di non direzione, dalla coda si può raggiungere il padre.

ROOTED TREE

Nel caso di alberi decisionali si parte da un nodo padre, che è il più importante perché da lui solamente escono gli archi orientati e portano ai figli che portano poi alle foglie.

Le foglie sono raggiungibili tramite uno ed un solo cammino.

DAG

In questo caso non è un albero (di decisione) perché ci sono due nodi padre.

Slide 7

Perché visualizzare un grafo?

Quando si hanno dataset rappresentabili come grafi, la visualizzazione consente di rappresentare le comunità.

Consente di identificare quegli edge che sono importanti perché connettono due componenti connesse, ossia quegli edge che se tolti eliminano qualsiasi relazione tra due sottografi.

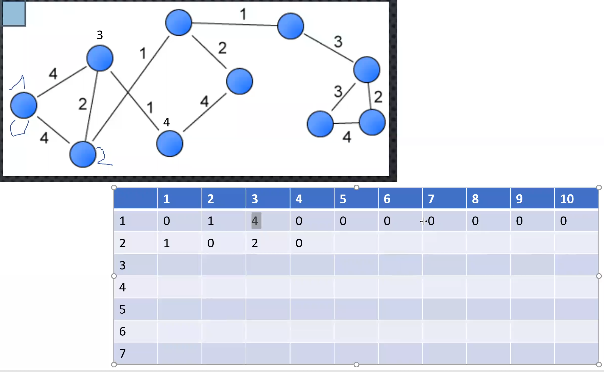
Una visualizzazione ci permette di capire quali sono quelle componenti connesse per cui con pochi link, o la rimozione di un link, può generare nuove componenti sconnesse tra di loro oppure per cui basterebbe l’inferenza di un link per generare un’unica componente connessa.

Si possono anche generare delle feature.

Quindi se si vede il grafo come una matrice di adiacenza allora si possonio usare le coordinate come feature.

Slide 8

Molte volte interessa fare la proiezione del grafo in uno spazio multidimensionale (matrice di adiacenza) per cui poi lo si riesce a visualizzare.

(Gli edge possono anche essere delle similarità)

Si numerano i nodi, gli si dà un indice, e poi si può creare una tabella con 10x10 elementi (la matrice di adiacenza) in cui si mostra quali nodi sono connessi tra loro.

Si mette 0 quando il nodo non è in relazione con nessun altro nodo o è in relazione con sé stesso (si ha un ciclo – vedi nodo 1).

La matrice di adiacenza ha tante colonne quanti sono i nodi (cardinalità insieme V → G=(V,E) ).

Ha tante righe quanti sono i nodi.

La matrice di adiacenza è un pò come la matrice di similarità: per ogni elemento della matrice si vede se il nodo j-esimo (colonna) è linkato da un edge al nodo i-esimo (riga) e in caso di archi pesati si mettte il peso dell’arco; nel caso di archi non pesati si mette 1 cioè un valore booleano che indica che quei due nodi sono legati.

Se non si ha un numero grande di nodi e di archi, solitamente di un grafo si ha la sua matrice di adiacenza.

Tante volte si vuole anche visualizzare la matrice di adiacenza ed il modo più semplice di farlo è vedere tale matrice come un’immagine e visualizzare l’immagine: se un grafo non diretto si visualizzerà un’immagine in bianco e nero dove si vedono i punti bianchi dove c’è l’1.

Se si tratta di un grafo diretto si vedrà un’immagine a diversi livelli di grigio.

Slide 12

I grafi nascono dal problema dei 7 ponti.

Eulero formulò i 7 ponti come nodi e dimostrò che non esiste un cammino che passa dai 7 ponti una sola volta.

Slide 13

Un altro problema è il problema dei 4 colori, per cui vengono utilizzati i grafi per dimostrarlo.

La separazione del piano deve essere fatta in modo che ciascuna ragione tocchi gli altri colori.

Queste sono le teorie da cui sono nate le visualizzazioni dei grafi.

Slide 15

***Graph visualization***, ***graph prediction***, ***graph analysis*** vengono anche dette ***network visualization***, ***network analsysis***, ***network prediciton***.

Le predizioni fatte sui grafi sono tendenzialmente fatte sui nodi, quindi ***node prediction***, oppure ***link prediction***, si vuole quindi predire se tra due nodi esiste un edge oppure no.

NODE PREDICTION

Sui nodi tante volte è indicata una label, ad esempio in campo medico il nodo rappresentante il paziente ha una label con indicato “deceduto” o meno.

Dato un nuovo nodo che viene inserito nel grafo, bisogna predire se quel nodo sarà 1 o 0; cioè se per la diagnosi, il paziente ha una malattia o non ce l’ha.

Slide 17

Ci sono diversi benefici nella *graph visualization*:

si possono visualizzare dei trend, fare correlazione tra i dati, condividere ciò che si ha trovato, fare delle analisi dei grafi.