Graph Data

Data as objects & connections between objects

objects ↑ instaces, nodes, vertices

connections ↑ relationships, links, edges

Eg:

- socal network
- web pages
- biological networks
- relationships between things : numerical, categorical, time series

Graph $G = (V, E)$

w/ V is vertex set

$E \subseteq V \times V$ edge set

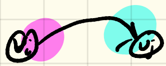
$e \in E$ as unorded pairs (e.g undirected graph)

or as ordered pairs (e.g directed graph, digraph)

directed edges (v_i, v_j)

v_i is the tail

v_j is the head



(v_i, v_i) is loop

Undirected graph w/o loop is simple

Let $e = (v_i, v_j)$ edge is incident w/ v_i and v_j

and v_i and v_j are neighbors

$|V| = n$ is the order of graph

$|E| = m$ is the size of graph

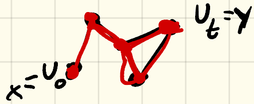
A weighted graph each edge (v_i, v_j) has weight $w_{ij} \in \mathbb{R}$

Paths & Dist

A ==walk== in a graph G between nodes x, y
ordered seq of verts

$$x = v_0, v_1, v_2, \ldots, v_{t-1}, v_t = y$$

s.t $(v_i, v_{i+1}) \in E \quad i = 0, \ldots, t-1$



==length of walk== t # of edges in walk

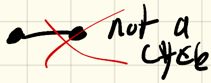
$$x = v_0 \overset{1}{,} v_1 \overset{2}{,} v_2 = y \quad \text{len } 2$$

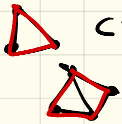
walk is ==closed== $x = y$

==trail== walk w/ distict edges

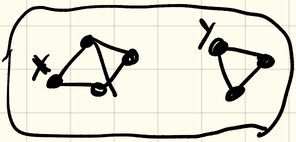
==path== walk w/ distinct vert (except x, y)

==Cycle== closed path w/ length $t \geq 3$





==Shortest path== between $x, y \in V$
is a path w/ min length
length of the shortest path is the ==distance==
between x, y written as $d(x,y)$



$d(x,y) = \infty$ if there is
no path

Two nodes x, y are connected
if $\exists$ a path between x and y

Graph is connected
if v_i is connected to v_j $\forall$ $v_i, v_j \in V$

A connected component (component for short)
is a maximal connected subgraph

1 connected component $\Rightarrow$ graph is connected
2 or more connected components $\Rightarrow$ graph is disconnected

Adj matrix

Let $G = (V, E)$ be a graph
$|V| = n$
write G as $n \times n$ matrix called adjacency matrix

$$A(i,j) = \begin{cases} 1 & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases}$$

G undirected $\Rightarrow$ A is sym
G directed $\Rightarrow$ A may not be sym

weighted adj matrix

$$A(i,j) = \begin{cases} w_{ij} & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases}$$

turn back into binary matrix using threshold w/ $\tau \in \mathbb{R}$ $A(i,j) = \begin{cases} 1 & w_{ij} \geq \tau \\ 0 & \text{otherwise} \end{cases}$

Graph from point data

data set D n points in $\mathbb{R}^d$

turn into a graph

each $\vec{x}_i \in D \longrightarrow v_i \in V$

$\vec{x}_i, \vec{x}_j \in D \longrightarrow (v_i, v_j) \in E$

w/ weight $w_{ij} = \text{Sim}(\vec{x}_i, \vec{x}_j)$

using dist or othe sim

optionally
sparsify graph w/ threshold

Degree

degree of vertex v_i (local)

# of neighbors

more generally $d_i = \sum_{j=1}^{n} A(i,j)$

average degree (global)

$$\mu_d = \frac{1}{n} \sum_i d_i$$