L23: Graph Embeddings

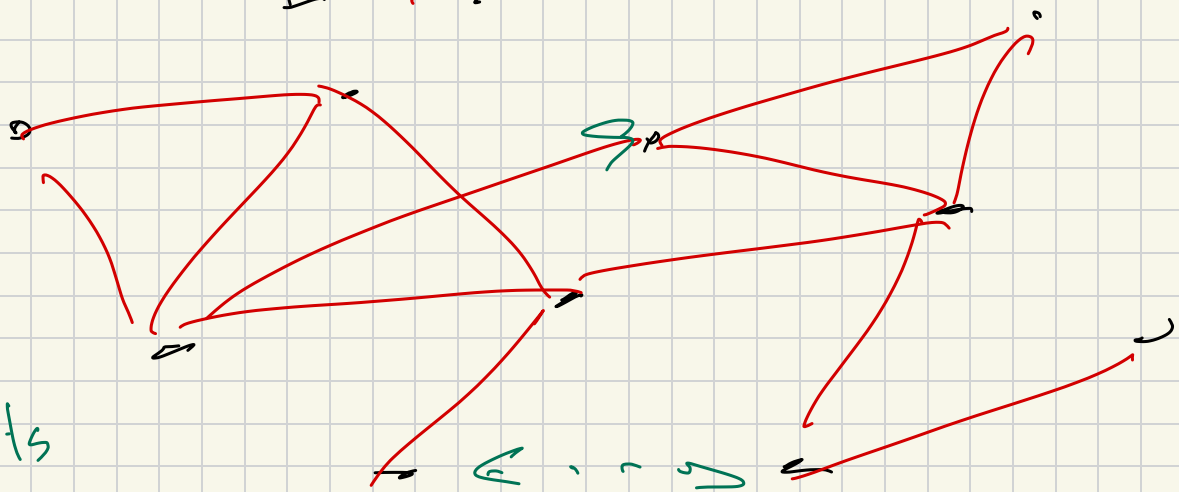
+ Ethics Assessment

Nov 19, 2025
Data Mining
Jeff M. Phillips

Data graph $G = (V, E)$

V = vertices = webpages, companies, people, products

E = edges = connections



Similarity

$S(v_1, v_2) = e_{12}$

$\in [0, 1]$

Graph Embeddings

$$G = (V, E) \Longrightarrow V \in \mathbb{R}^d$$

$$dist(v_1, v_2) \approx f(E)$$

Input

list vertices $V = \{v_1, v_2 \dots\}$

list edges $E = \{e_1 = (v_1, v_7),\ e_2 = (v_2, v_3) \dots\}$

$d = 2 \Longrightarrow$ draw picture

$d = k \Longrightarrow$ clustering, classification

MDS input V, distance function

$$D(v_1, v_2) = \begin{cases} 0 & \text{if edge } e=(v_1,v_2) \in E \\ 1 & \text{o.w.} \end{cases}$$

not comparing w.r.t

Klent

$$D^{(2)} \in \mathbb{R}^{|V| \times |V|} \qquad D^{(2)}_{ij} = D(v_i, v_j)^2$$

$\text{eigs}(D^{(2)}) \Longrightarrow$ eigenvectors $v_1, v_2 \ldots$

Project $\phi(v_i) \in \mathbb{R}^k$

$(v_1(i), v_2(i) \ldots v_k(i))$

Laplacian Matrices

$|V| = n \geq 10{,}000$

D = diagonal matrix degrees

$$\begin{bmatrix} deg(v_1) & 0 & 0 \\ 0 & deg(v_2) & \\ 0 & & \ddots \\ & & deg(v_n) \end{bmatrix}$$

$L_0 = D - A$

$L = I - D^{-1/2} A D^{-1/2}$

A = adjacency matrix

$A_{ij} = 1$ if edge $e_{ij} \in E$

$= 0$ o.w.

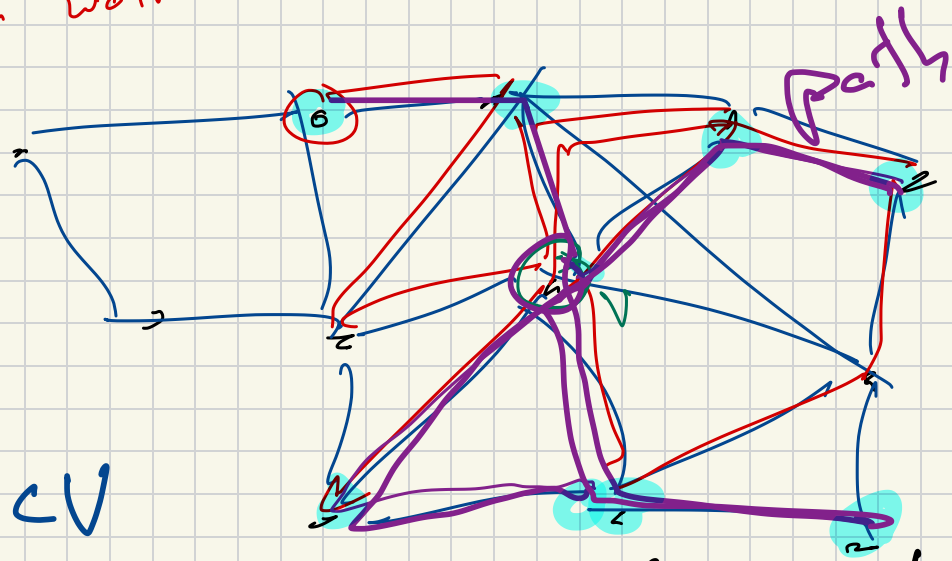
Graph Neural Network

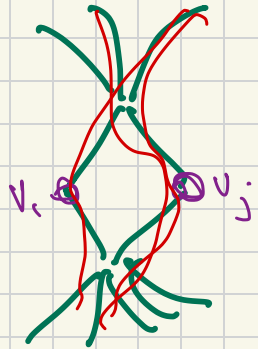
DeepWalk , node2vec

Learn embedding $G=(V,E) \Rightarrow \phi(V) \in \mathbb{R}^k$

Random Walk

Neighborhood Length $L=2$





neighbor hood $N(v_i, P) \subset V$

$$Sim(v_i, v_j) = Similarity\left(\bigcup_{\text{Paths}} N(v_i, P),\ \bigcup_{\text{Paths}} N(v_j, P)\right)$$

In Practice Scalable

↳ optimize $\phi(v) \in \mathbb{R}^k$

with SGD

goal $\langle \phi(v_i), \phi(v_j) \rangle \approx Sim(N(v_i), N(v_j))$