

Introduction to Data Science

Graphs



Arijit Mondal

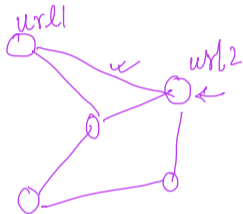
Dept. of Computer Science & Engineering

Indian Institute of Technology Patna

`arijit@iitp.ac.in`

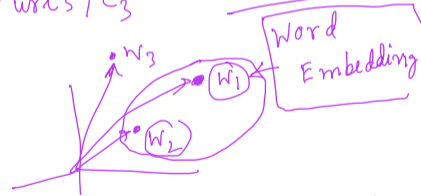
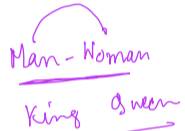
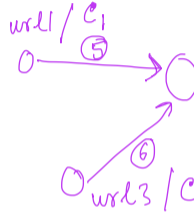
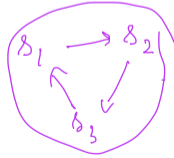
Introduction

- Graph provides an abstract representation. It will have a set of nodes and edges
 - World wide web, product/customer network, genetic, social network, etc.



Introduction

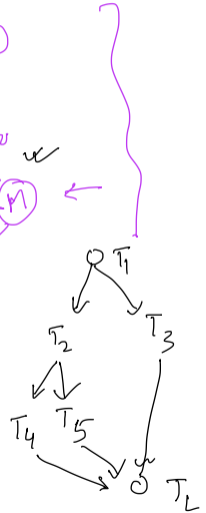
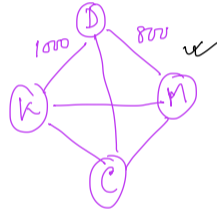
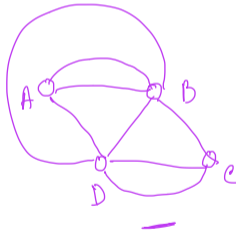
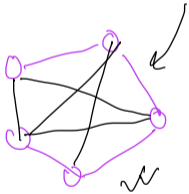
- Graph provides an abstract representation. It will have a set of nodes and edges
 - World wide web, product/customer network, genetic, social network, etc.
- Weighted graph ←
- Induced graph
- Embedding



$Man - Woman + King = Queen$

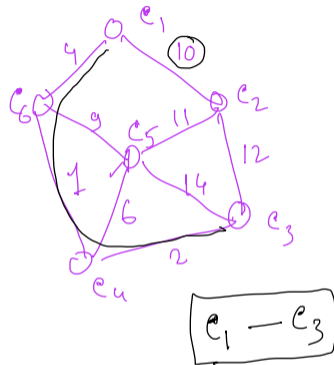
Type of graphs

- Directed vs undirected
- Weighted vs unweighted ✓
- Simple vs non-simple
- Sparse vs dense
- Embedded vs topological
- Labeled vs unlabeled



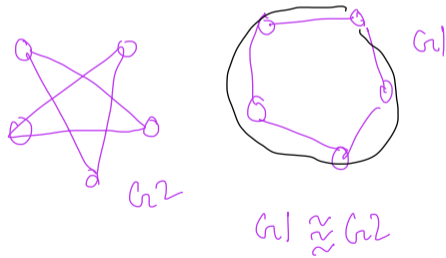
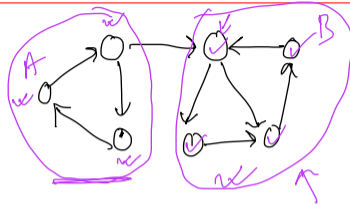
Typical graph problems

- Shortest path



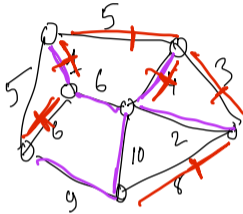
Typical graph problems

- Shortest path
- Connected component



Typical graph problems

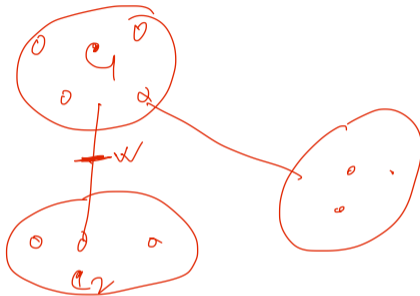
- Shortest path
- Connected component
- Minimum spanning tree



$$\begin{aligned} & \rightarrow 1+6+4+2+10+9 \\ & \quad = ? \\ & \left. \begin{aligned} & 1+6+5+3+4+8 \\ & = ? \end{aligned} \right\} \end{aligned}$$

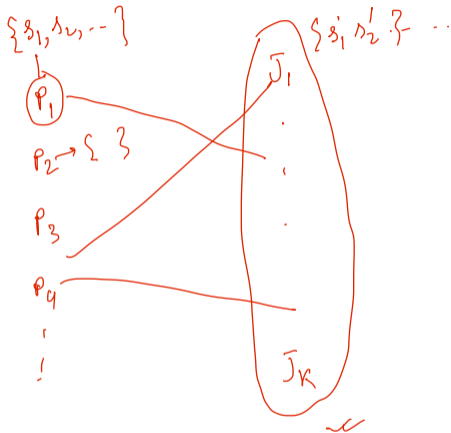
Typical graph problems

- Shortest path
- Connected component
- Minimum spanning tree
- Edge cuts



Typical graph problems

- Shortest path
- Connected component
- Minimum spanning tree
- Edge cuts
- Matching

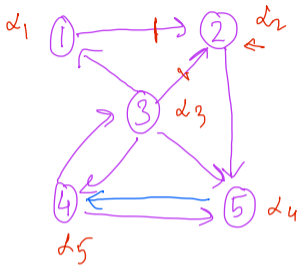


Typical graph problems

- Shortest path
- Connected component
- Minimum spanning tree
- Edge cuts
- Matching
- Topological sorting



PageRank - I

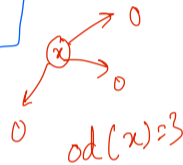


$$\alpha_i = PR(i)$$



$$\alpha_i = 1/5 = \alpha_i^1$$

$$PR(v) = \sum_{(u,v) \in E} \left(\frac{PR(u)}{\text{out-degree}(u)} \right)$$



$$\sum_v PR(v) = \underline{1}$$

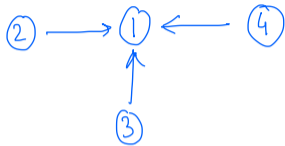
$$\alpha_1^3, \alpha_2^3, \dots$$

$$\alpha_2^2 = \frac{\alpha_3^1}{4} = 1/20 \quad | \quad v=1 \quad u=3$$

$$\alpha_2^2 = \frac{\alpha_1^1}{od(1)} + \frac{\alpha_3^1}{od(3)} = \frac{1}{5} + \frac{1}{20} = 5/20$$

$$\alpha_3^2 = 1/10, \quad \alpha_4^2 = ? \quad \left(\frac{1}{20} \right) ? \quad \alpha_5^2 = \underline{\underline{7/20}}$$

PageRank - II



$$PR(1) = \frac{1}{4}$$

$$PR^2(1) = \sum_{j=2,3,4} \frac{PR(j)}{od(j)} = \frac{3}{4}$$

$$PR^2(2) = 0$$

$$\sum PR(i) = \underline{0.75}$$

$d = \text{damping factor} = (0, 1)$

$od = \text{outdegree}$

$$PR(i) = \frac{1-d}{N} + d \times \sum_{j, (j,i) \in E} \frac{PR(j)}{od(j)}$$

$N = \text{total no. of nodes}$

$PR(u)$

$$\frac{PR(v)}{od(v)} \xrightarrow{v \rightarrow u}$$