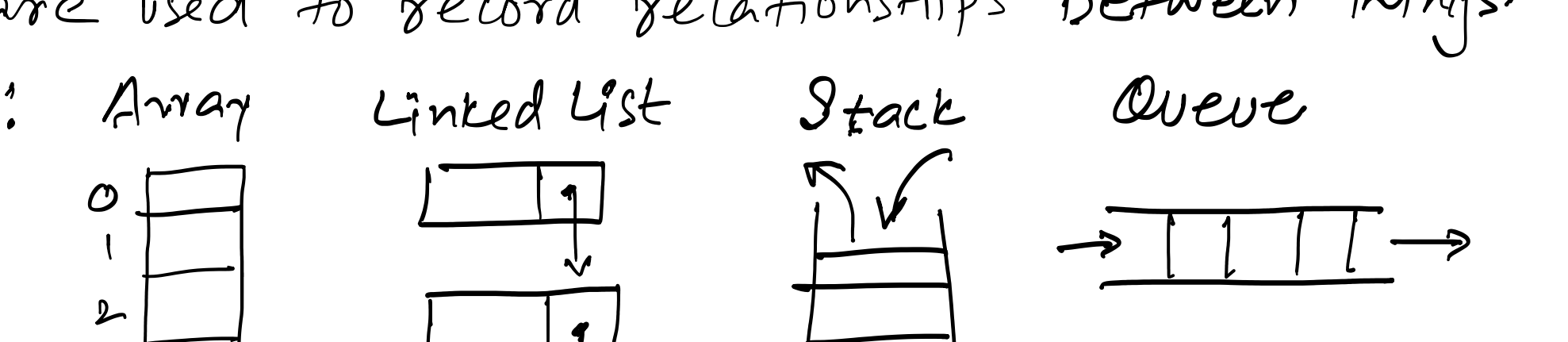
Graph DS Introduction

Wednesday, 2 March 2022   11:27 PM

- Data Structures are ways to organise and manage data. There are linear and non-linear DS for different kinds of data.
- Graphs are used to record relationships between things.

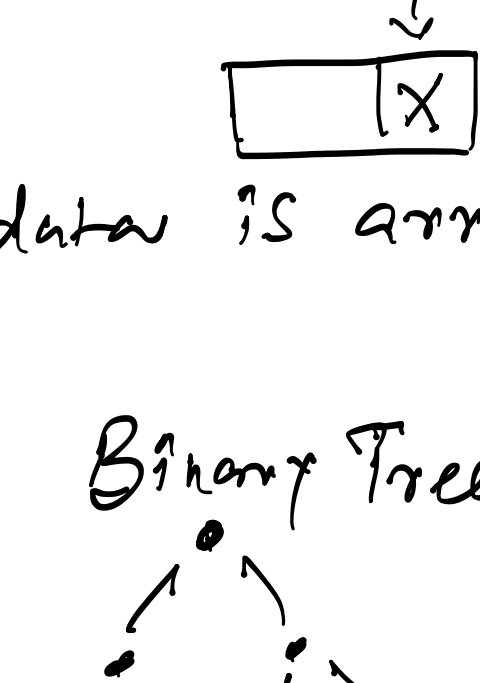
Linear DS: Array Linked List Stack Queue



In linear DS, data is arranged in linear or sequential manner.

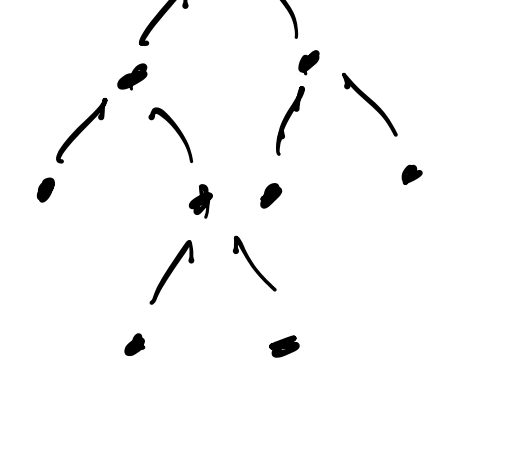
Non-linear DS:

Binary Tree



- It is a hierarchical DS.
- In BT connections are bound to be in a certain way.
- There are predefined rules defining the connections among the nodes.
- In BT with N nodes we have N-1 edges.
- All nodes are reachable from root node.
- There are exactly one possible path to a node.
- BT has directed edges.

Graph



- It is also non-linear DS that has its use in a wide number of computer science scenarios.
- It is used to model and represent a wide variety of systems.
- We study graph DS as logical or mathematical model.
- Later implementation will be discussed.
- There are no rules in graphs for defining connections among the nodes.
- A graph contains a set of nodes and a set of edges and edges can be connecting nodes in any possible way.
- Graph as a concepts is studied extensively in mathematics.
- In computer science we study & implement same concept of maths.
- The study of graph is often referred to as "graph theory".

In pure maths terms we define graph as:

"A graph G is an ordered pair of a set V of vertices and a set E of edges."

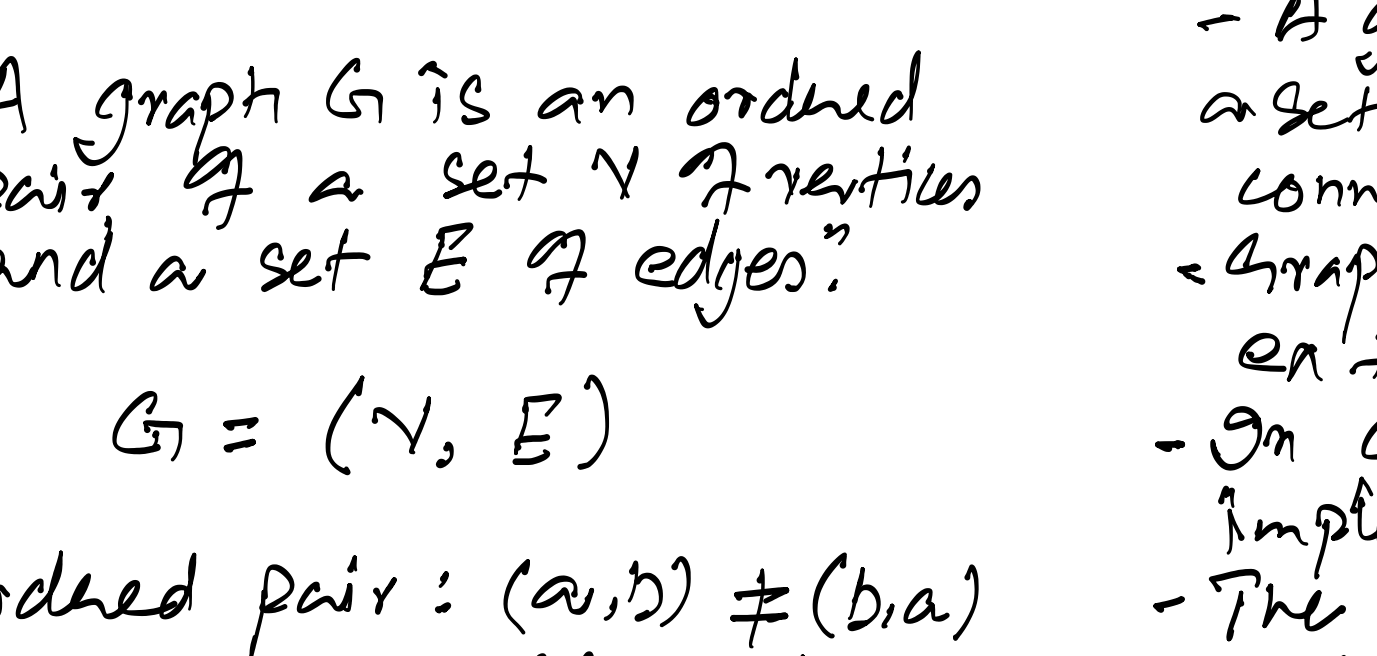
$$G = (V, E)$$

Ordered pair: $(a,b) \neq (b,a)$ if $a \neq b$.

V is set of vertices
E is set of edges.

Unordered pair: $\{a,b\}$ order is not important.

$\{a,b\} = \{b,a\}$ because order doesn't matter.



7 vertices

9 edges.

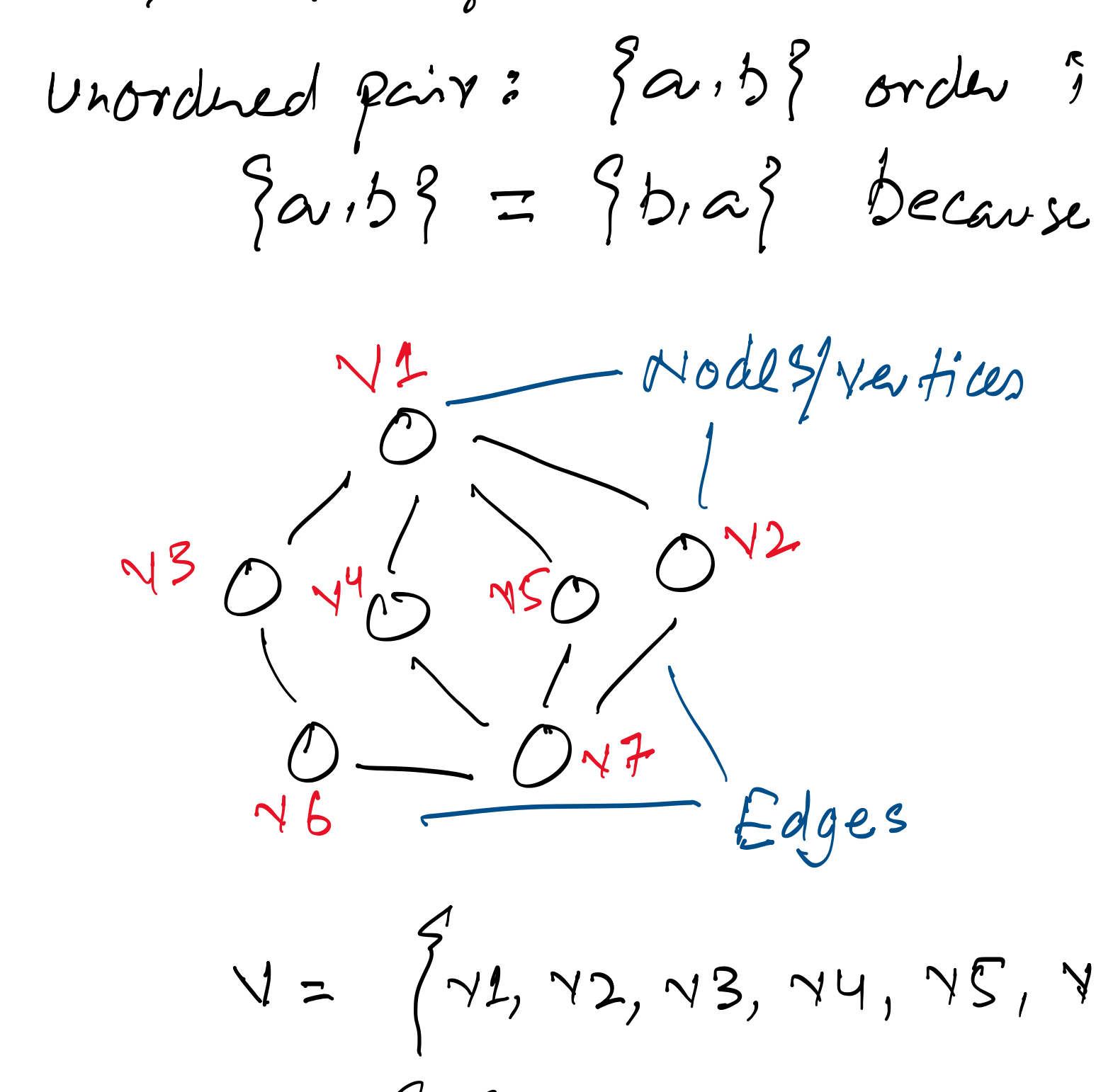
- Every vertex must have a name.

$V = \{V1, V2, V3, V4, V5, V6, V7\}$ Set of vertices for the graph above.

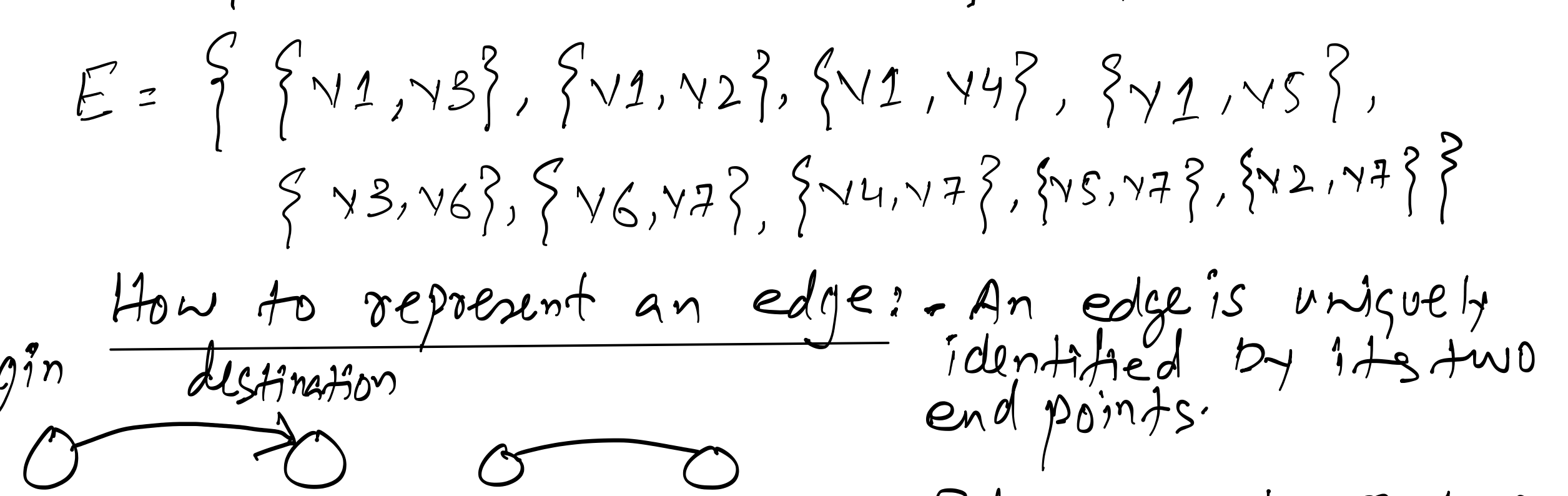
$E = \{ \{V1,V3\}, \{V1,V2\}, \{V1,V4\}, \{V1,V5\}, \{V3,V6\}, \{V6,V7\}, \{V4,V7\}, \{V5,V7\}, \{V2,V7\} \}$

How to represent an edge:

- An edge is uniquely identified by its two end points.
- Edges can be of two types; directed and undirected.
- Typically in a graph all edges can be directed or undirected.
- We study graph with all the edges either directed or undirected but not both.

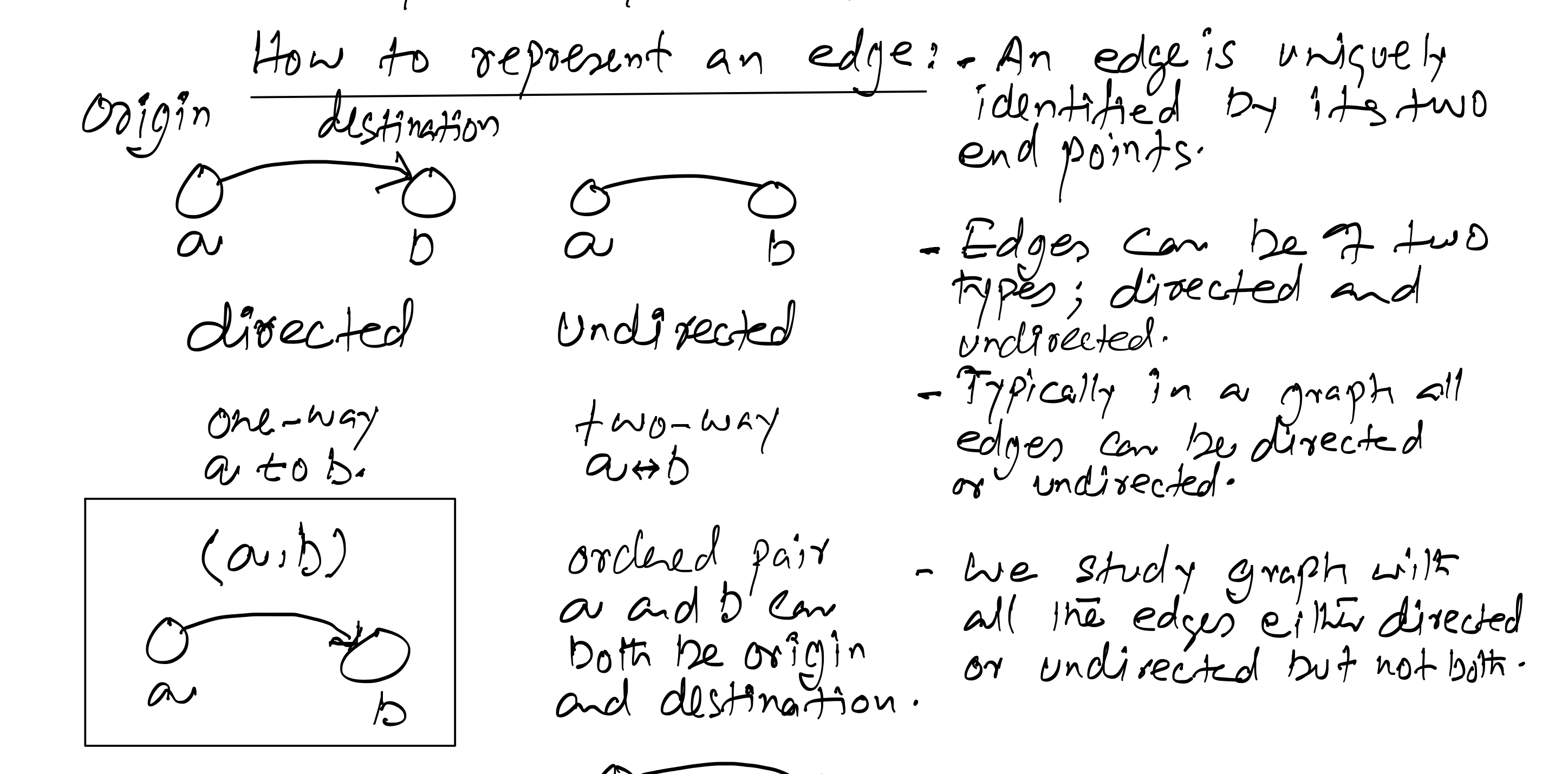


Graph Types



Directed graph. Undirected graph.

Graphs can be used to represent any real world situation/problems/systems. It can represent any collection of objects having pair-wise relationships.

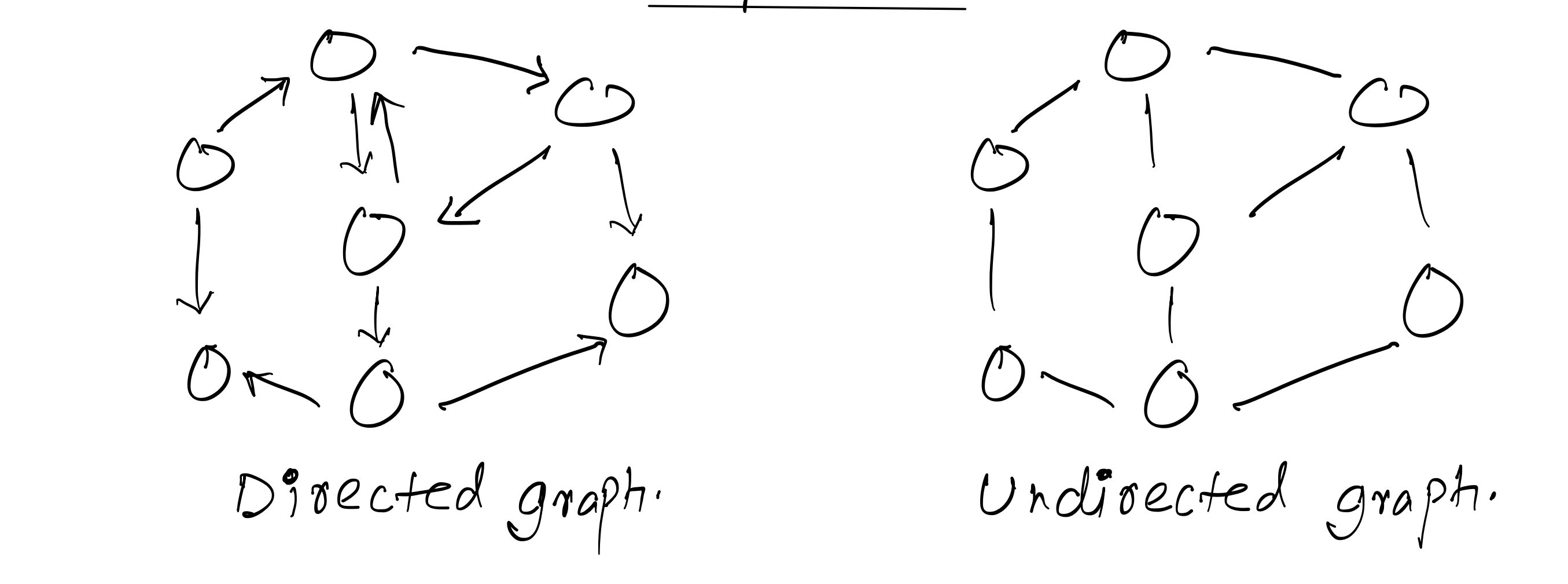


Problem: Suggest some friends to Ahmed.

Solution: Suggest friends of friends that are not already connected to Ahmed.

In graph terms: It is a standard graph problem.

Problem: Find all nodes/vertices having length of shortest path from Ahmed equal to 2.

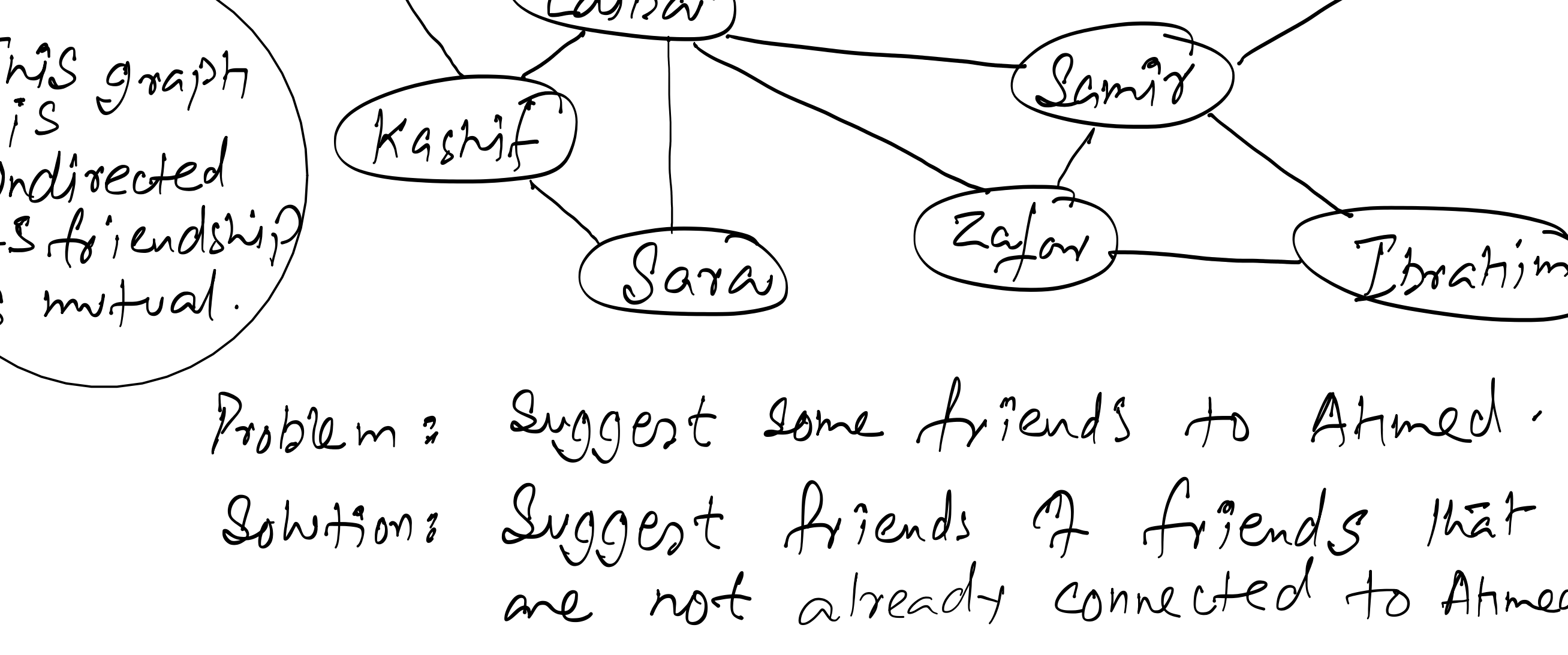


If we are able to represent world wide web as a directed graph, then we can apply standard graph theory algorithm to solve problems and perform tasks.

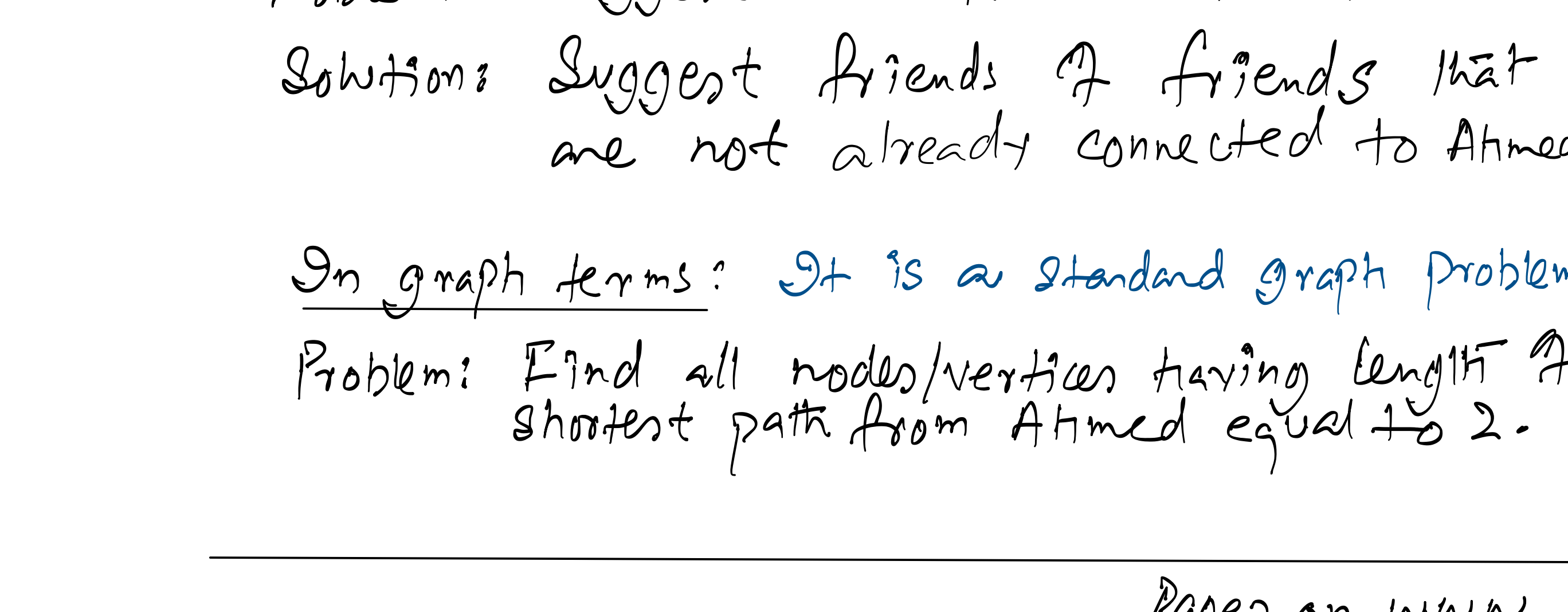
Task: web crawling to index web pages to return accurate results against search queries.

Web crawling is "Graph Traversal", that is visiting all the nodes of the graph using standard graph traversal algorithms.

Weighted and unweighted graphs:



- Intercity road network.
- All highways are bi-directional
- Graph is undirected.
- Roads/edges will be of different lengths.
- To solve any problem we need to take road length into account.
- To do so we associate some weight or cost to every edge.
- In this case weight can be length of the roads; and will become label of the edges here.



Problem: Choose a best (shortest) route from Quetta to Bahawalpur.

Solution: * If we treat the graph as unweighted then number of edges will be considered, and out of few possible routes; i.e:

1. Q – K – H – S – B
2. Q – S – B
3. Q – S – R – B
4. Q – K – S – B

etc.

Quetta – Sukkur – Bahawalpur will be considered best as it has only 2 edges.

* If we consider the graph as weighted then we will have to calculate the edge weight (road length) through all possible routes to know the best (shortest) path.

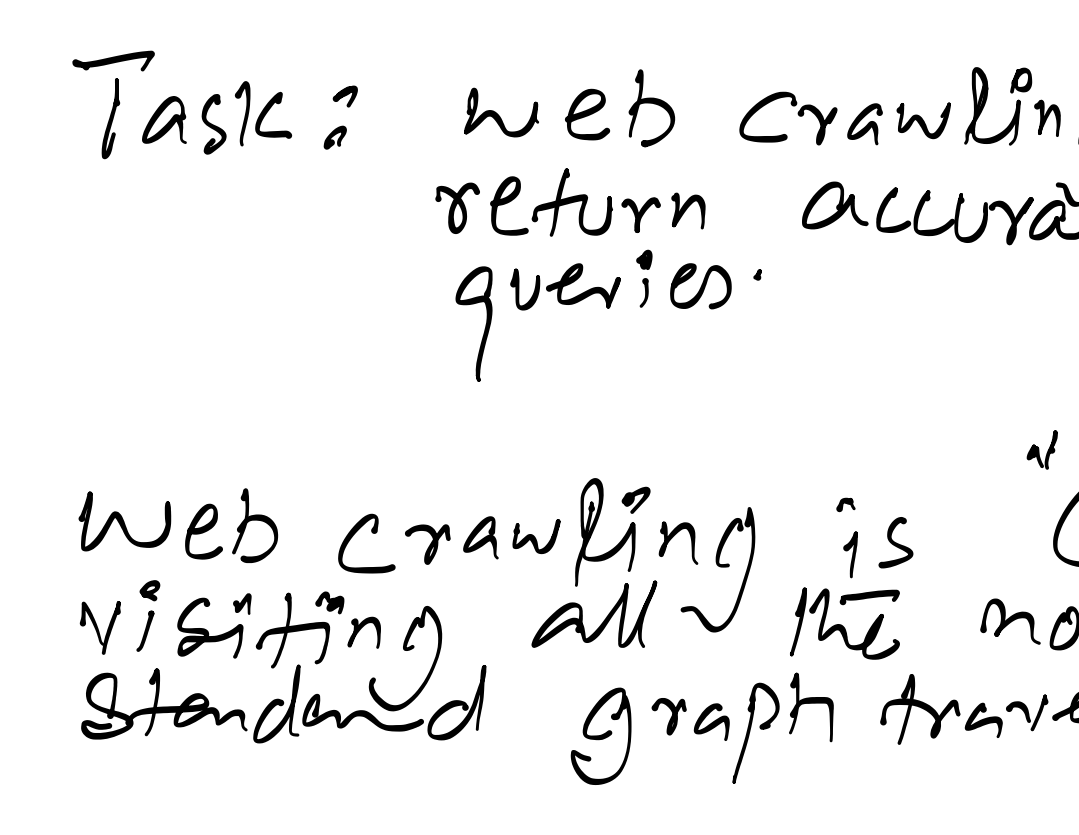
Quetta — Khuzdar — Sukkur — Bahawalpur
150 + 600 + 300 = 1050
Shortest route

- An unweighted graph is a weighted graph with all edges having weight = 1 unit.
- Actually we can look at all the unweighted graphs as weighted graphs.
- Inter city road network is a weighted undirected graph.
- Social network is an unweighted undirected graph.
- World wide web is unweighted directed graph.
- We can always redraw an undirected graph as a directed graph by having all the edges drawn in both directions.

Adjacency Matrix:

- To implement a Graph, we can use adjacency matrix or an adjacency list.
- An adjacency matrix stores the relationship between every vertex to all other vertices.
- For an unweighted graph '1' represents an edge and '0' no edge.
- For a weighted graph actual cost/weight replaces '1' and an infinity symbol '∞' replaces '0'.

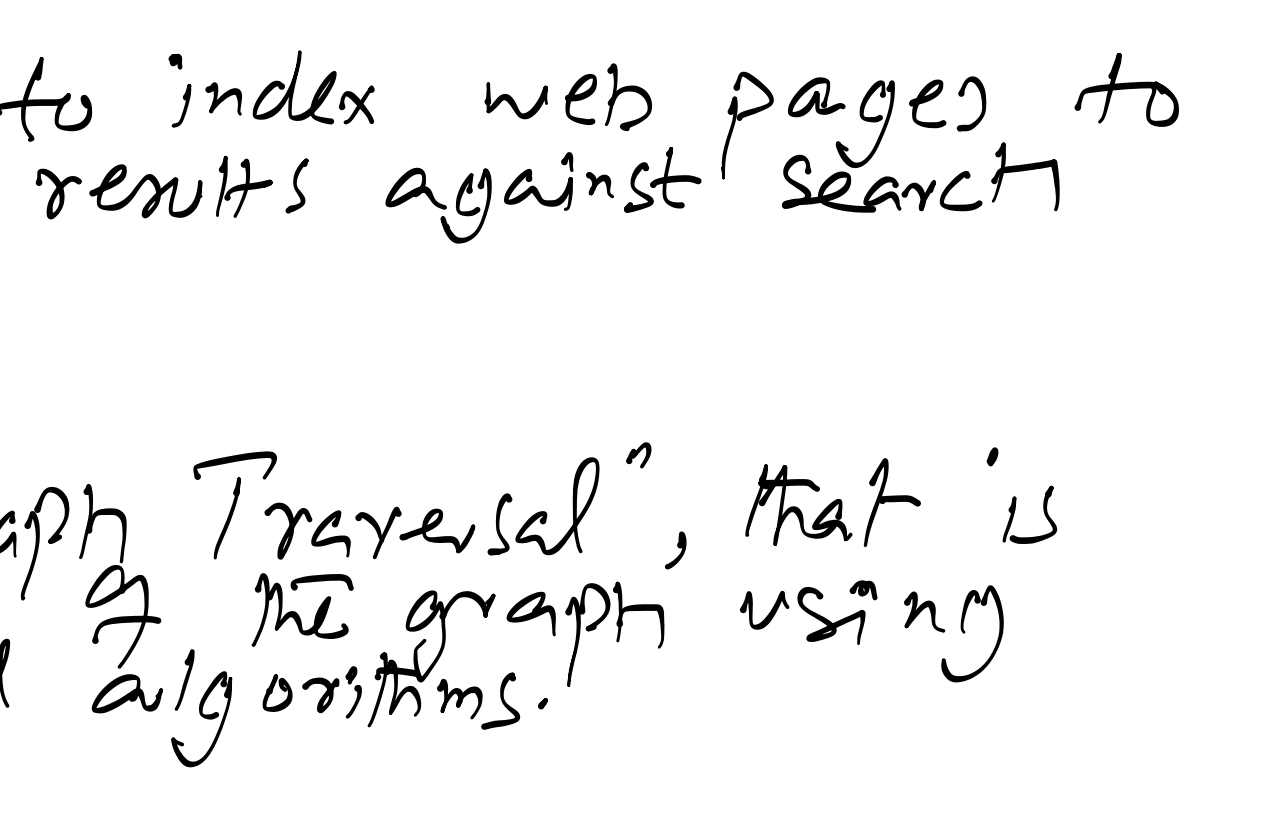
Undirected



	A	B	C	D	E	F
A	0	1	0	1	0	0
B	1	0	1	1	1	0
C	0	1	0	0	0	1
D	1	1	0	0	1	0
E	0	1	0	1	0	1
F	0	0	1	0	1	0

Matrix

directed



	A	B	C	D	E	F
A	∞	3	∞	5	∞	∞
B	3	∞	3	∞	4	∞
C	∞	3	∞	∞	∞	2
D	5	4	∞	∞	2	∞
E	∞	4	∞	4	∞	2
F	∞	∞	3	∞	3	∞

Matrix.

- An adjacency list stores the relationship between every vertex to all relevant (connected) vertices.
- An entry is made only when there is an edge between two vertices.

	Connected to
A	B, D
B	A, C, D, E
C	B, F
D	A, B, E
E	B, D, F
F	C, E

List

	Connected to
A	B,3; D,5
B	A,3; C,3; E,4
C	B,3; F,2
D	A,5; B,4; E,2
E	B,4; D,4; F,2
F	C,3; E,3

List.