Batch: D-2 **Roll No.: 16010123325**

Experiment No. 6

TITLE : Performing Graph Analytics

AIM: To analyze the structural properties of a real-world social network by constructing a graph representation, identifying key players and influential individuals through centrality measures, and detecting communities within the network using appropriate algorithms.

Expected OUTCOME of Experiment:
CO3: Perform the social data analytics

Books/ Journals/ Websites referred:

Students have to list.

Pre Lab/ Prior Concepts:

Students should have a basic understanding of:

Graph theory: Nodes, edges, directed and undirected graphs, weighted graphs.
Data structures: Lists, dictionaries.
Python programming: Basic syntax, data manipulation, libraries like NetworkX.
Statistical concepts: Mean, standard deviation, correlation.
Visualization techniques: Basic plotting using libraries like Matplotlib.

Procedure:

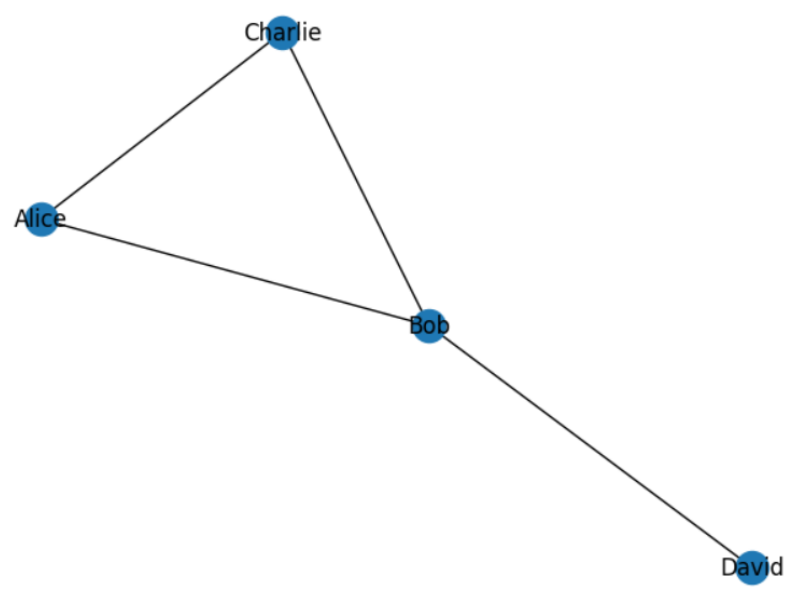
Building a Social Network Graph with NetworkX

```
import networkx as nx

# Create an empty graph
G = nx.Graph()

# Add nodes (individuals)
G.add_nodes_from(['Alice', 'Bob', 'Charlie', 'David'])

# Add edges (relationships)
G.add_edge('Alice', 'Bob')
G.add_edge('Alice', 'Charlie')
G.add_edge('Bob', 'Charlie')
G.add_edge('Bob', 'David')

# Print the graph
print(G.nodes())
print(G.edges())
```

```
['Alice', 'Bob', 'Charlie', 'David']
[('Alice', 'Bob'), ('Alice', 'Charlie'), ('Bob', 'Charlie'), ('Bob', 'David')]
```

Visualizing the Graph

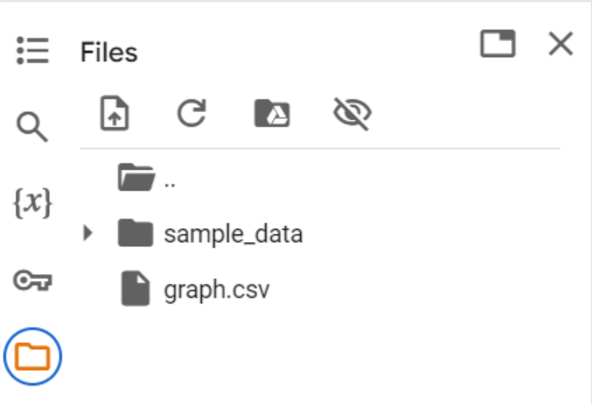
```
import matplotlib.pyplot as plt

# Draw the graph
nx.draw(G, with_labels=True)
plt.show()
```



Exporting a NetworkX Graph to CSV

```python
import networkx as nx
import pandas as pd

# Create a sample graph
G = nx.Graph()
G.add_edges_from([('A', 'B'), ('A', 'C'), ('B', 'C'), ('B', 'D')])

# Convert to edge list
edgelist = nx.to_edgelist(G)

# Create a pandas DataFrame, including a column for edge attributes
df = pd.DataFrame(edgelist, columns=['source', 'target', 'attributes'])

# Export to CSV
df.to_csv('graph.csv', index=False)
```

The csv file gets created



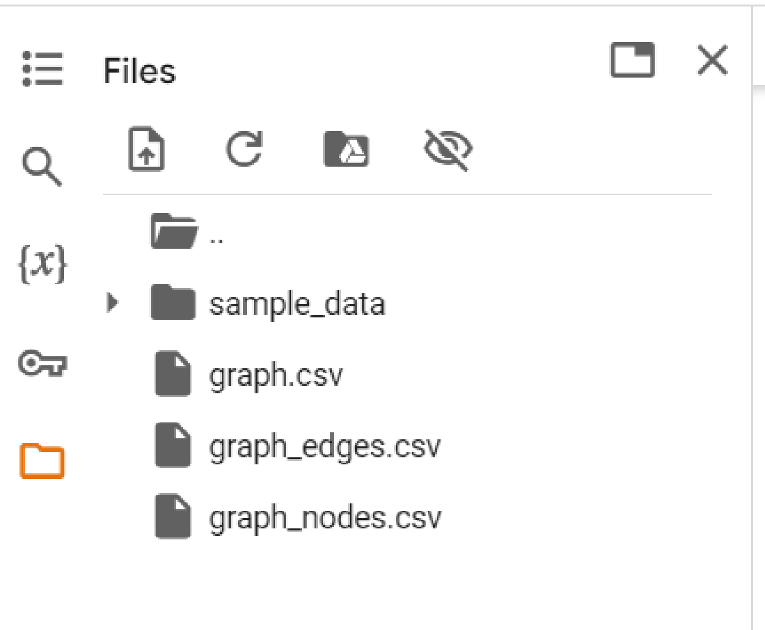
Contents of the csv file

	A	B	C	D
1	source	target	attributes	
2	A	B	{}	
3	A	C	{}	
4	B	C	{}	
5	B	D	{}	
6				
7				

Creating and exporting a NetworkX Graph with edge attributes and node attributes to a csv file

```python
import networkx as nx
import pandas as pd

# Create a graph with node and edge attributes
G = nx.Graph()
G.add_edge('A', 'B', weight=2.5)
G.add_edge('A', 'C', weight=1.0)
G.nodes['A']['color'] = 'red'

# Convert to edge list with attributes
edgelist = [(u, v, d) for u, v, d in G.edges(data=True)]

# Create a pandas DataFrame
df = pd.DataFrame(edgelist, columns=['source', 'target', 'weight'])

# Add node attributes as a separate DataFrame if needed
node_attributes = pd.DataFrame.from_dict(dict(G.nodes(data=True)), orient='index')
node_attributes.columns = ['color']

# Export to CSV
df.to_csv('graph_edges.csv', index=False)
node_attributes.to_csv('graph_nodes.csv')
```

csv files get created



Contents of graph_edges.csv

	A	B	C	D
1	source	target	weight	
2	A	B	{'weight': 2.5}	
3	A	C	{'weight': 1.0}	
4				
5				

Contents of graph_nodes.csv

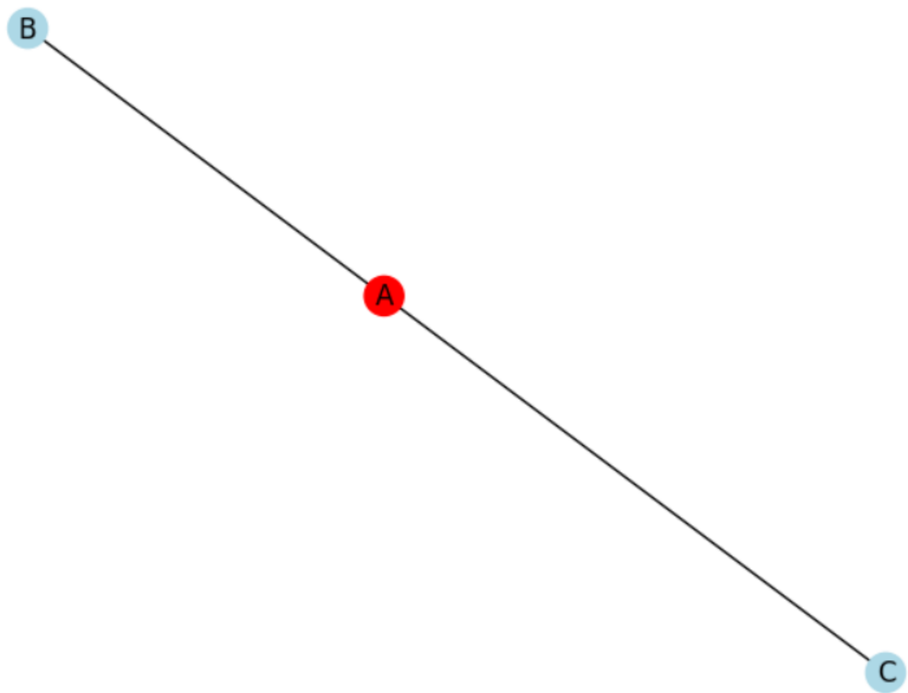
	A	B
1		color
2	A	red
3		

Importing a graph from a csv file

```python
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt

# Read edge list from CSV
df_edges = pd.read_csv('graph_edges.csv')

df_nodes = pd.read_csv('graph_nodes.csv', index_col=0)

# Extract numeric weights from the 'weight' column (assuming they are stored as dictionaries)
# If your 'weight' column is just a number, remove this line
df_edges['weight'] = df_edges['weight'].apply(lambda x: float(x.strip("{}").split(": ")[1]) if isinstance(x, str) else x)

# Create a graph from the edge list
G = nx.from_pandas_edgelist(df_edges, source='source', target='target', edge_attr='weight')

# Add node attributes if available
if df_nodes is not None:
    nx.set_node_attributes(G, df_nodes.to_dict('index'))

# Print the graph
print(G.nodes(data=True))
print(G.edges(data=True))

# Draw the graph
nx.draw(G, with_labels=True, node_color=[n[1]['color'] if 'color' in n[1] else 'lightblue' for n in G.nodes(data=True)])
plt.show()
```

Output (List of nodes and edges, and visualizing the imported graph)

```
[('A', {'color': 'red'}), ('B', {}), ('C', {})]
[('A', 'B', {'weight': 2.5}), ('A', 'C', {'weight': 1.0})]
```



Graph Analytics

1. Degree centrality : The degree centrality for a node v is the fraction of nodes it is connected to. The degree centrality values are normalized by dividing by the maximum possible degree in a simple graph n-1 where n is the number of nodes in G.
2. Betweenness centrality : Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v. The betweenness centrality is normalized by dividing by the total number of shortest paths.
3. Edge betweenness centrality : Betweenness centrality of a node e is the sum of the fraction of all-pairs shortest paths that pass through e. The betweenness centrality is normalized by dividing by the maximum possible number of edges in a graph G.
4. Communities can be identified using the Girvan Newman algorithm, by successively deleting the edges with the highest betweenness centrality values.

Importing a graph from csv file and performing graph analytics

The graph in csv file:

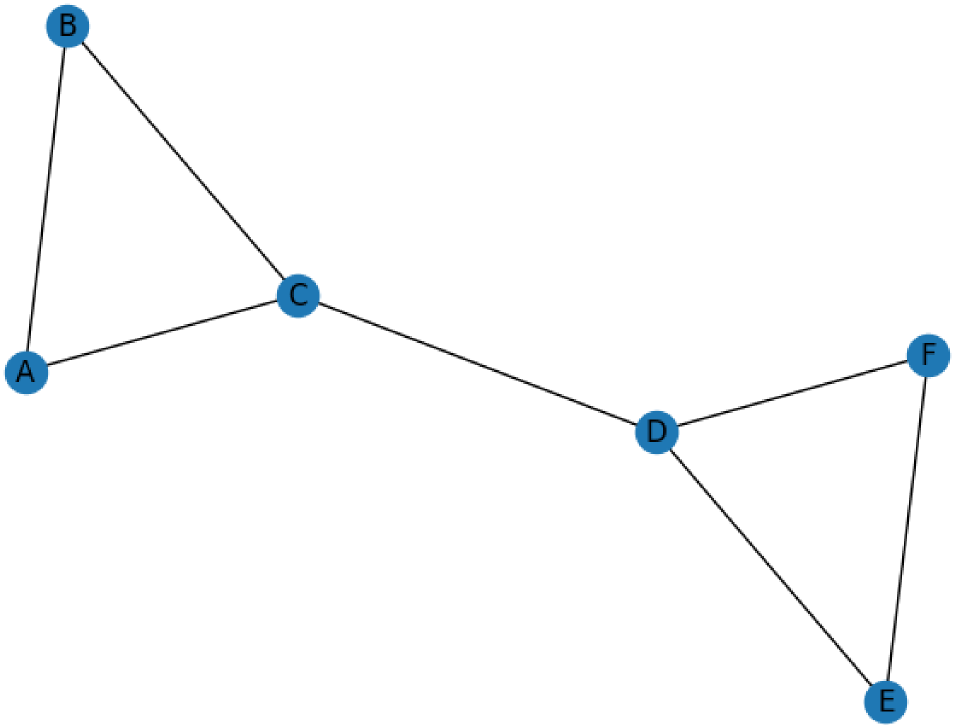
	A	B	C
1	node1	node2	attribute
2	A	B	{}
3	A	C	{}
4	B	C	{}
5	C	D	{}
6	D	E	{}
7	D	F	{}
8	E	F	{}

Importing the graph, printing its edge list and visualizing it:

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt

# Read edge list from CSV
df_edges = pd.read_csv('new_graph_edges.csv')

# Create a graph from the edge list
G = nx.from_pandas_edgelist(df_edges,source='node1', target='node2')

# Print the graph
print(G.nodes(data=True))
print(G.edges(data=True))

# Draw the graph
nx.draw(G, with_labels=True)
plt.show()
```

Output (graph details and visualization):

```
[('A', {}), ('B', {}), ('C', {}), ('D', {}), ('E', {}), ('F', {})]
[('A', 'B', {}), ('A', 'C', {}), ('B', 'C', {}), ('C', 'D', {}), ('D', 'E', {}), ('D', 'F', {}), ('E', 'F', {})]
```



Performing analytics on this graph:

```
# Basic graph properties
print("Number of nodes:", G.number_of_nodes())
print("Number of edges:", G.number_of_edges())

# Degree centrality
degrees = dict(G.degree())
print("\nDegree Centrality:", degrees)

# Betweenness centrality
betweenness = nx.betweenness_centrality(G, normalized=False)
print("\nBetweenness Centrality:", betweenness)
betweenness = nx.betweenness_centrality(G)
print("Normalized Betweenness Centrality:", betweenness)

# Closeness centrality
e_betwenness = nx.edge_betweenness_centrality(G,normalized=False)
print("\nEdge Betweenness Centrality:", e_betwenness)
e_betwenness = nx.edge_betweenness_centrality(G)
print("Normalized Edge Betweenness Centrality:", e_betwenness)
```

```python
# Community detection (Girvan-Newman)
communities = nx.algorithms.community.girvan_newman(G)
try:
    top_level_communities = next(communities)
    print("\nCommunities after 1 step:", top_level_communities)

    top_level_communities = next(communities)
    print("\nCommunities after 2 steps:", top_level_communities)

    top_level_communities = next(communities)
    print("\nCommunities after 3 steps:", top_level_communities)

    top_level_communities = next(communities)
    print("\nCommunities after 4 steps:", top_level_communities)

    top_level_communities = next(communities)
    print("\nCommunities after 5 steps:", top_level_communities)

except StopIteration:
    print("\nNo more splits are possible.")
```

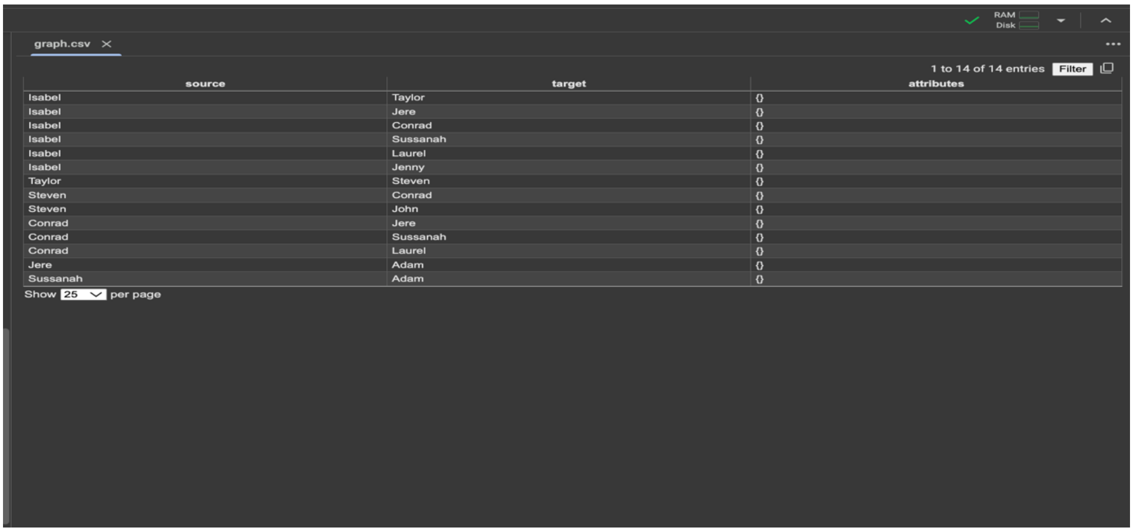
Output:

```
Number of nodes: 6
Number of edges: 7

Degree Centrality: {'A': 2, 'B': 2, 'C': 3, 'D': 3, 'E': 2, 'F': 2}

Betweenness Centrality: {'A': 0.0, 'B': 0.0, 'C': 6.0, 'D': 6.0, 'E': 0.0, 'F': 0.0}
Normalized Betweenness Centrality: {'A': 0.0, 'B': 0.0, 'C': 0.6000000000000001, 'D': 0.6000000000000001, 'E': 0.0, 'F': 0.0}

Edge Betweenness Centrality: {('A', 'B'): 1.0, ('A', 'C'): 4.0, ('B', 'C'): 4.0, ('C', 'D'): 9.0, ('D', 'E'): 4.0, ('D', 'F'): 4.0, ('E', 'F'): 1.0}
Normalized Edge Betweenness Centrality: {('A', 'B'): 0.06666666666666667, ('A', 'C'): 0.26666666666666666, ('B', 'C'): 0.26666666666666666, ('C', 'D'): 0.6, ('D', 'E'): 0.26666666666666666, ('D', 'F'): 0.26666666666666666, ('E', 'F'): 0.06666666666666667}

Communities after 1 step: ({'A', 'C', 'B'}, {'E', 'F', 'D'})

Communities after 2 steps: ({'A'}, {'C', 'B'}, {'E', 'F', 'D'})

Communities after 3 steps: ({'A'}, {'B'}, {'C'}, {'E', 'F', 'D'})

Communities after 4 steps: ({'A'}, {'B'}, {'C'}, {'D'}, {'E', 'F'})

Communities after 5 steps: ({'A'}, {'B'}, {'C'}, {'D'}, {'E'}, {'F'})
```

Students have to perform all the tasks illustrated above by creating a social network graph with nodes labelled with their own names and their friends' names. The graph should have at least 10 nodes.
Students have to paste their code and screenshots of output and csv file below.

Implementation details:
https://colab.research.google.com/drive/1acLhYzu-KszUY4QCzBSFh2oPhxCQ_v6M?usp=sharing

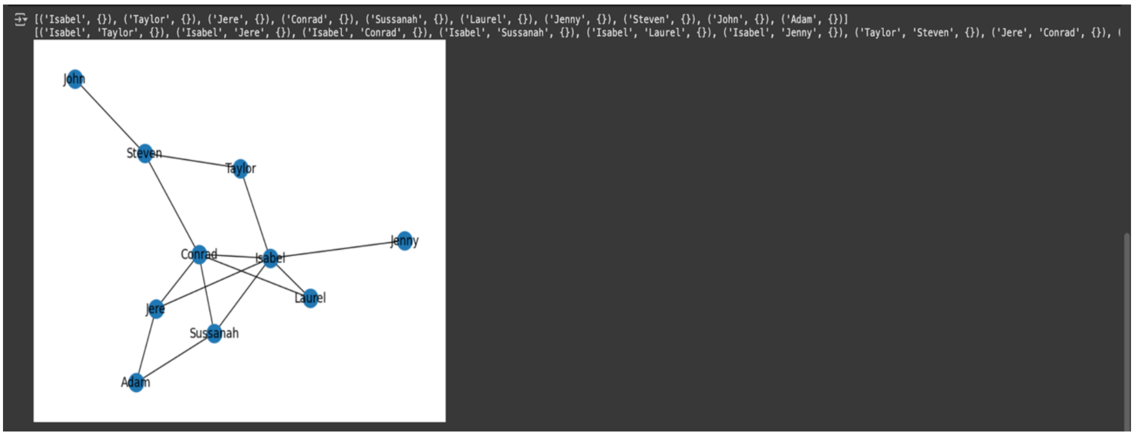
Output:


graph.csv

1 to 14 of 14 entries

source	target	attributes
Isabel	Taylor	{}
Isabel	Jere	{}
Isabel	Conrad	{}
Isabel	Sussanah	{}
Isabel	Laurel	{}
Isabel	Jenny	{}
Taylor	Steven	{}
Steven	Conrad	{}
Steven	John	{}
Conrad	Jere	{}
Conrad	Sussanah	{}
Conrad	Laurel	{}
Jere	Adam	{}
Sussanah	Adam	{}

Show 25 per page



Date: ______________ **Signature of faculty in-charge**

Post Lab Descriptive Questions:

1. Analyze the centrality measures you calculated. Which nodes were identified as the most influential? What does this mean in the context of the social network?

- Nodes with the **highest betweenness centrality** are the most influential they act as bridges connecting groups.
- Nodes with high **degree centrality** are well-connected hubs (popular or socially active).
- High **closeness centrality** means the node can reach others quickly.

For my graph,

- **Isabel** and **Conrad** stand out with the highest degree and betweenness centrality.
Steven also plays an important connector role.

- They are the most influential because they connect many others and act as bridges across the network.

2. Describe the communities identified using the Girvan-Newman algorithm. What are the characteristics of these communities? How do they relate to the social network's structure?

- The algorithm split the graph into **clusters of tightly connected nodes** (friends who interact more within their group).

- Communities show **natural divisions** in the network—like circles of friends, departments in a company, or interest groups.

For my graph,

- The network first split into two groups: one large cluster around Isabel/Conrad and a smaller group (Steven, Taylor, John).
- As edges were removed, smaller isolated communities appeared (Jenny, Jere, Adam).
- This shows that Isabel/Conrad’s group is more **tightly knit**, while Steven’s side is more **loosely connected**.

3. Discuss the implications of identifying influential nodes in the network. How can this information be used?

- Identifying them helps in **targeted communication** (e.g., marketing, spreading awareness).
- Useful for **network security** (monitoring key bridges to prevent disconnection).
- Helps understand **group dynamics**, leadership roles, and potential **information bottlenecks**.