NATIONAL INSTITUTE OF TECHNOLOGY SILCHAR

Cachar, Assam

B.Tech. VIth Sem

Subject Code: CS-321

Subject Name: Social Network Analysis Lab

Submitted By:

Name : Subhojit Ghimire

Sch. Id. : 1912160

Branch : CSE – B

AIM: TO SHOW THE NUMBER OF PREDICTED LINKS WITH FOLLOWING LINK PREDICTION METHODS:

PREDICTION METHODS: COSINE SIMILARITY OR SALTON INDEX, ADAMIC-ADAR INDEX, JACCARD COEFFICIENT, PREFERENTIAL ATTACHMENT, RESOURCE ALLOCATION INDEX

THEORY:

1. **Salton Index (Cosine Similarity):** It measures the cosine of the angle between columns of the adjacency matrix, corresponding to the given vertices. The measure is commonly used in information retrieval. This measure is given by:

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k_x \times k_y}}$$

2. **Adamic-Adar Index:** This measure was introduced in 2003 to predict missing links in a Network, according to the number of shared links between two nodes.

$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{log\ |\Gamma(z)|}$$

3. **Jaccard Index:** It is calculated by number of common neighbors normalized by total number of neighbors. It is used to measure the similarity between two finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets.

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

4. **Preferential Attachment:** Preferential attachment means that the more connected a node is, the more likely it is to receive new links Nodes with higher degree gets more neighbors.

$$S_{xy} = k_x \cdot k_y$$

5. **Resource Allocation Index:** Among a number of similarity-based methods to predict missing links in a complex network, Research Allocation Index performs well with lower time complexity. It is defined as a fraction of a resource that a node can send to another through their common neighbors.

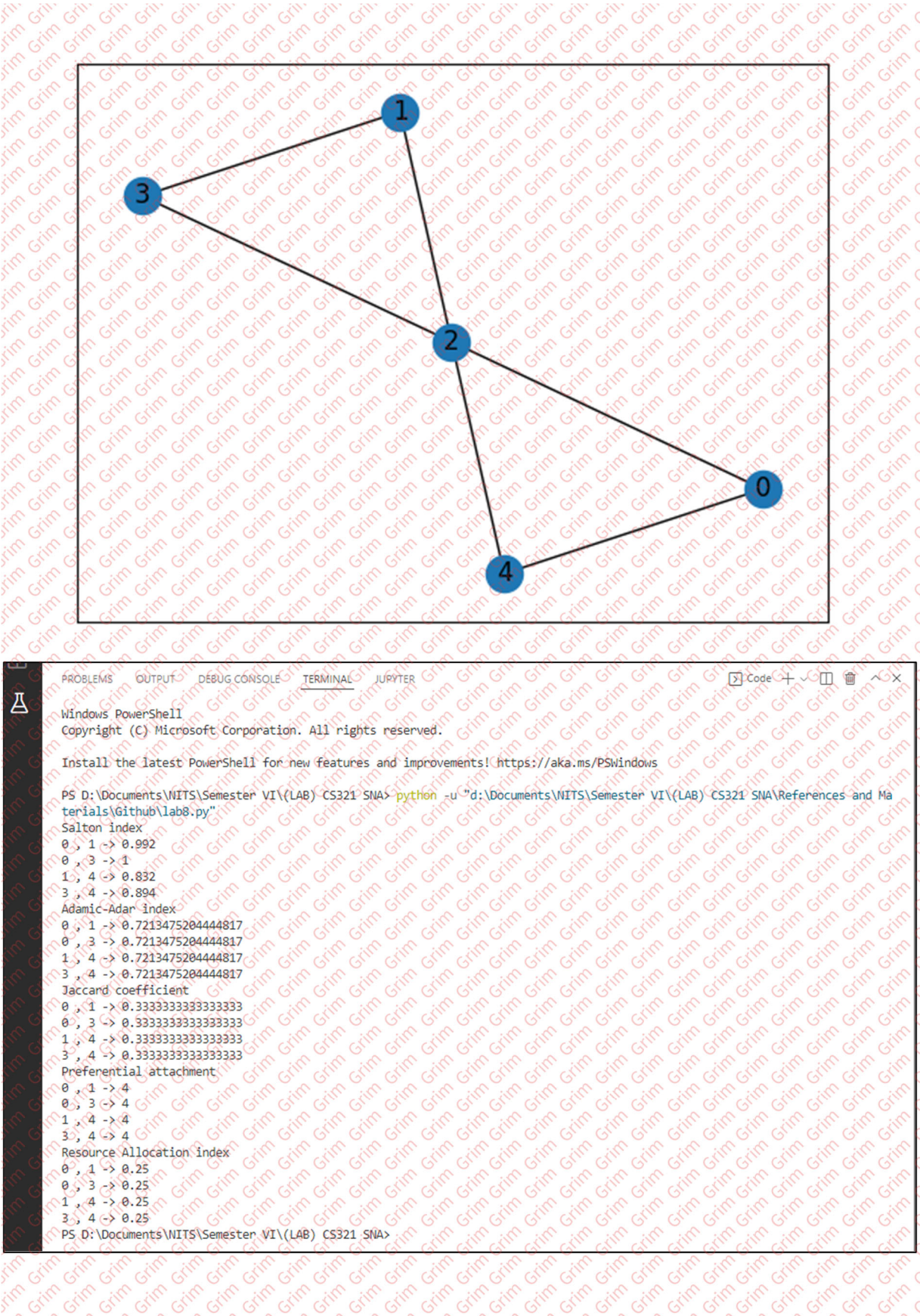
$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|}$$

CODE:

```
# FOR A GRAPH WITH HIGH NUMBER OF NODES, (LIKE KARATE CLUB GRAPH), Salton
Index gives error.
# However, even for graphs with high number of nodes, the other meausres work.
# The problem is with Salton Index only.
# Instead of calling lab8Graph.gml, you can make a random graph by your own,
but sometimes random graphs give error:
# ERROR IS LIKE: ValueError: operands could not be broadcast together with
shapes (x, ) (y, )
# In case of such error, keep re-running the porgram. For different graphs,
this error is resolved or doens't exist.
# Again, this error is associated with Salton Index only. Delete the Salton
Index part, and there are no errors for other index measures.

import networkx as nx
import matplotlib.pyplot as plt
import numpy as np
from scipy import spatial

def saltonList (G):
    edges = nx.edges(G)
    n = G.number_of_nodes()
    list = [[]]
    for i in range (n-1):
        list += [[]]
    for i in range (n):
        for j in edges:
            if (j[0] == i):
                list [i].append (j[1])
            if (j[1] == i):
                list [i].append (j[0])
    _list = [[]]
    var = 0
    for i in range (n):
        for j in range (i+1, n):
            if j not in list[i]:
                _list[var].append (i)
                _list[var].append (j)
                var += 1
                _list += [[]]
    _list.remove ([])
    var = 0
    for i in _list:
        cosineSimilarity = 1 - spatial.distance.cosine (list [i[0]], list
[i[1]])
        _list[var].append(cosineSimilarity)
        var += 1
    return _list
```

```
G = nx.read_gml ('lab8Graph.gml')

print ("Salton index")
preds = saltonList(G)
for p in preds:
    print (p[0], ",", p[1], "->", p[2])

print ("Adamic-Adar index")
preds = nx.adamic_adar_index(G)
for p in preds:
    print (p[0], ",", p[1], "->", p[2])

print ("Jaccard coefficient")
preds = nx.jaccard_coefficient (G)
for p in preds:
    print (p[0], ",", p[1], "->", p[2])

print ("Preferential attachment")
preds = nx.preferential_attachment (G)
for p in preds:
    print (p[0], ",", p[1], "->", p[2])

print ("Resource Allocation index")
preds = nx.resource_allocation_index (G)
for p in preds:
    print (p[0], ",", p[1], "->", p[2])

nx.draw_networkx (G, with_labels = 'True')
plt.show ()
```

OUTPUT AND OBSERVATIONS:




```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  JUPYTER                    Code

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS D:\Documents\NITS\Semester VI\(LAB) CS321 SNA> python -u "d:\Documents\NITS\Semester VI\(LAB) CS321 SNA\References and Ma
terials\Github\lab8.py"
Salton index
0 , 1 -> 0.992
0 , 3 -> 1
1 , 4 -> 0.832
3 , 4 -> 0.894
Adamic-Adar index
0 , 1 -> 0.7213475204444817
0 , 3 -> 0.7213475204444817
1 , 4 -> 0.7213475204444817
3 , 4 -> 0.7213475204444817
Jaccard coefficient
0 , 1 -> 0.3333333333333333
0 , 3 -> 0.3333333333333333
1 , 4 -> 0.3333333333333333
3 , 4 -> 0.3333333333333333
Preferential attachment
0 , 1 -> 4
0 , 3 -> 4
1 , 4 -> 4
3 , 4 -> 4
Resource Allocation index
0 , 1 -> 0.25
0 , 3 -> 0.25
1 , 4 -> 0.25
3 , 4 -> 0.25
PS D:\Documents\NITS\Semester VI\(LAB) CS321 SNA>
```