13/03/2024

THADOMAL SHAHANI TSEC ENGINEERING COLLEGE

Name: Ananya Garg
Class: BE
Div: C1
Batch: C12
Roll no: 2003054
Group no: 9

Assignment 6

* **Problem statement**
Develop structure based social media analytics tool for any business. (eg structure based models - community detection, influence analysis, social media graph analytics).

* Theory:

→ It refers to the use of structured data analysis techniques to extract valuable insights from social media platforms.

→ Businesses can leverage these insights to better understand their customers, improve their marketing strategies & ultimately derive more revenue.

→ Community detection algorithms:

- These are a set of techniques used to identify clusters or communities within a social network.

- Some examples include:

i) Modularity Optimization :

- This algorithm seeks to optimise the modularity of a network, which is a measure of degree to which nodes are connected within maximum cluster vs between clusters.

ii) Label Propagation :

- This algorithm assigns labels to nodes based on the labels of their neighbours and then progresses these labels through the network until each node is assigned to a community.

Its foremost focus is speed, trying to find clusters in as little time as possible.
Simply put, LPA works by iteratively updating the label of each node to a label that is most common among its neighbours.

LPA stops iterating whenever the labels of all nodes are maximal. LPA performs a number of iterations over n nodes. Each potential update of a label for a node i has a complexity $O(k_i)$ where k_i is the degree of node i, & therefore the total complexity of a single iteration is linear in the number of edges $O(m)$.

iii) Spectral Clustering :

- This algorithm uses the eigenvectors of a graph Laplacian

matrix to identify clusters within a network.

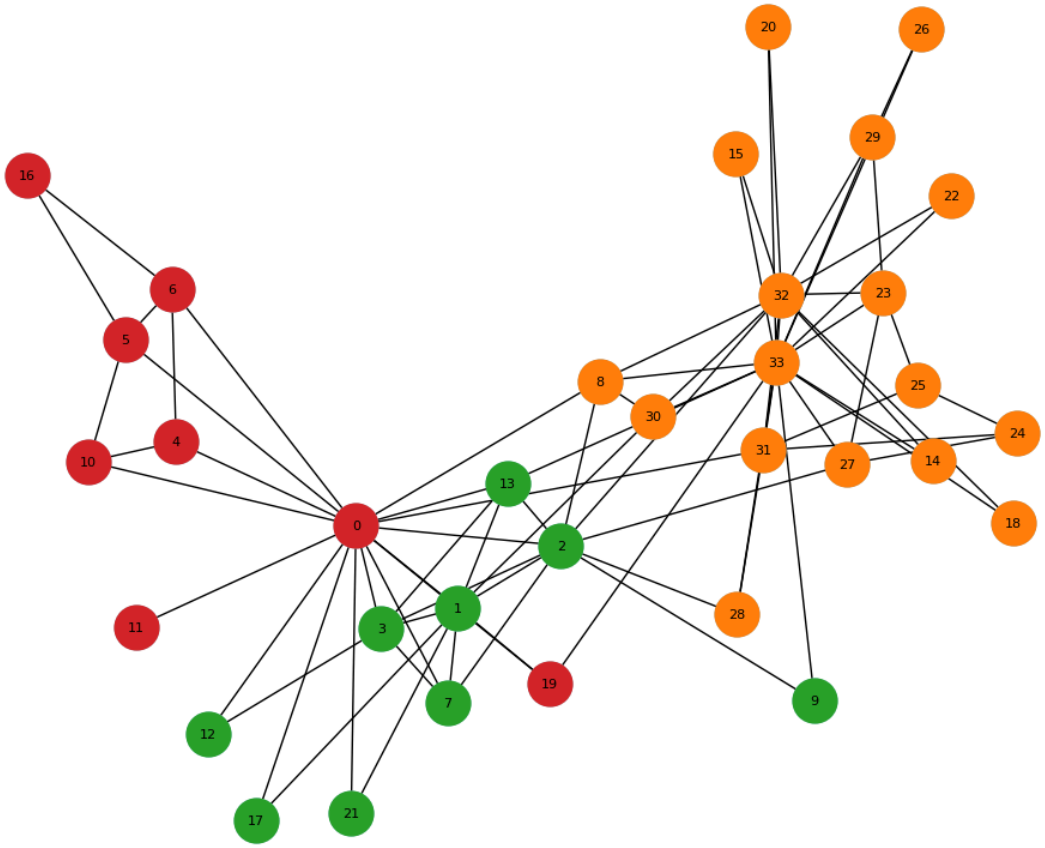
iv) Influence Analysis:

Metric to quantify the influence that certain social network users (influencers) have over other social network users because of their knowledge & reputation of a certain topic.

In [ ]:

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

# Step 1: Data Collection (Using inbuilt data from NetworkX)
G = nx.karate_club_graph()

# Step 2: Data Preprocessing (Not needed for inbuilt data)

# Step 3: Community Detection
communities = list(nx.algorithms.community.greedy_modularity_communities(G))

# Step 4: Influence Analysis
degree_centrality = nx.degree_centrality(G)

# Step 5: Social Media Graph Analytics
average_degree = sum(dict(G.degree()).values()) / len(G)
clustering_coefficient = nx.average_clustering(G)

# Step 6: Visualization
plt.figure(figsize=(10, 8))

# Plot the graph
pos = nx.spring_layout(G)
nx.draw(G, pos, with_labels=True, node_color='skyblue', node_size=700, font_size=8)

# Highlight communities
for i, community in enumerate(communities):
    nx.draw_networkx_nodes(G, pos, nodelist=list(community), node_color=f'C{i + 1}'

# Print graph properties
print(f"Average Degree of Inter-connectivity: {average_degree}")
print(f"Clustering Coefficient: {clustering_coefficient}")

plt.show()
```

```
Average Degree of Inter-connectivity: 4.588235294117647
Clustering Coefficient: 0.5706384782076823
```


20
26
29
15
22
16
6
32
23
5
33
8
25
30
24
4
31
27
14
10
13
0
18
2
1
28
11
3
19
9
7
12
17
21