Networkx Python Assignment
Sept 03, 2024

Community Detection and Anomaly Detection in Business Transactions

Scenario:
In the town of "Bizville," the local government is concerned about the financial health of the town's economy. They suspect that while most businesses operate legitimately, there may be a few that are involved in suspicious activities. These businesses might be transferring large sums of money to entities outside their regular community, which could indicate money laundering, tax evasion, or other illicit activities.

The government hires a team of data analysts to examine Bizville's business transaction records. The analysts are tasked with identifying the communities of businesses, finding the key players within these communities, and detecting any unusual transaction patterns.

Data Format

The transaction data is stored in a file named `data.txt`, where each line represents a transaction in the following format of each line:

```
A B C
```

- **A**: The sender (business initiating the transaction).
- **B**: The receiver (business receiving the transaction).
- **C**: The amount of money involved in the transaction.

For example, a line like `12 34 500` means that Business 12 sent 500 units of currency to Business 34.

Steps:

1. **Data Collection and Formatting:**
    - The transaction data is collected and stored in `data.txt` with the format `A B C`, where each line represents a transaction.
    - The data is pre-processed to ensure it is ready for analysis.
    - We need to do min max normalize on the edge values.

2. **Community Detection:**
    - The analysts use graph-based algorithms to detect communities within the network of businesses. A community is defined as a group of businesses that frequently transact with each other.
    - The goal is to understand the natural groupings within the town's economy.

3. **Identifying Key Businesses:**
    - Within each identified community, the analysts use various centrality algorithm to rank businesses based on their importance and influence. A high score indicates that a business is central to the community's transactions.

    - The top-K businesses in each community are identified as the key players.

4. **Anomaly Detection:**
    - The analysts then examine the transaction patterns of the top K businesses. They check whether these businesses are making unusually large transactions with businesses outside their community.
    - If such anomalies are detected, they are flagged for further investigation. The goal is to identify any suspicious activities that could harm the town's economy.

5. **Mitigation Strategies:**
    - Based on the detected anomalies, the analysts develop methods to mitigate their impact on the analysis. This may involve adjusting the community detection algorithm, adjusting K and so on.

Submission Guidelines:

- Submit one .ipynb file segmented into sections in-line with the assignment.
- Prepare an analysis report, like - how many communities, what the top k you took, the total amount of anomaly transactions with percentage, find anomaly nodes, whether those nodes are sending money to each other (fraud to fraud), what the top k legit businesses with rank.