PYRAMID SCHEME ANALYSIS.

Insights into Structure, Influence, and Community Dynamics

Done by:

22PD10 - CHINNAHYATA POORVIKA

22PD31 - SARNIKA SANJIV KUMAR

INTRODUCTION

Pyramid schemes are unsustainable business models that promise returns based on the recruitment of new members rather than the sale of products or services. These schemes have caused significant financial harm and social instability. Recently, with the growth of the global economy, the spread of pyramid schemes has accelerated, becoming increasingly destructive and threatening the economies and social order of many countries. Some of these schemes have even exhibited violent tendencies. The rise of internet technology has further amplified the reach and impact of network marketing, making it more harmful.

Social Network Analysis (SNA) is a method for analyzing relational data, focusing on the relationships between actors and their roles in shaping individual and group behavior. It uncovers hidden relationships and structural characteristics within complex social systems. Two classical SNA methods are **motif analysis** and **exponential random graph model analysis**. Motifs analysis identifies frequent subgraphs in a network to reveal key microscopic features, while the exponential random graph model examines the network's construction process to explore interactions within the network. We apply motif analysis to identify micro-flow characteristics in a financial flow network, particularly those linked to indicted individuals, and the exponential random graph model to study the formation mechanisms of financial flows, especially between communities and indicted persons.

This report investigates the hierarchical structure and interconnections within a network dataset using graph theory and statistical analysis. The data include individual nodes representing people within a network, characterized by "depth_of_tree," which indicates their position in a hierarchical structure, and "profit_markup," which reflects an economic value associated with each individual. The network's relationships are modeled as a directed graph, where edges are formed based on hierarchical depth, and weighted by the economic influence or profit markup.

1. Data Loading and Initial Inspection

We begin by loading the dataset and conducting an initial inspection to verify data quality and structure:

- Data Import: The dataset is loaded using Pandas, enabling the use of DataFrame operations to handle and inspect the data efficiently.
- Missing Value Analysis: A check on missing data is done which counts the number of null values per column. The presence of missing values could impact the analysis by reducing the completeness of the network model, so this step is essential.

Observations:

The inspection shows a manageable dataset without significant issues, confirming the feasibility of proceeding to network construction and visualization.

2. Network Graph Constraints and Setup

Given that there are 500 nodes in the dataset, we calculate the maximum possible edges for a fully connected undirected graph using:
max_edges = n×(n−1), where n is the number of nodes. This constraint provides a baseline for understanding the network's density—the proportion of actual edges relative to the maximum potential edges. Calculating this maximum helps determine the sparsity or connectivity of the network, as a network close to this limit would be fully connected, while a network with far fewer connections would indicate sparsity.

3. Graph Construction: Defining Hierarchical Connections

A directed graph G, is constructed to model relationships between individuals based on the “depth_of_tree” attribute:

- **Node Representation**: Each individual in the dataset becomes a node in the graph, labeled with their hierarchical depth.
- **Edge Formation**: Directed edges connect nodes based on their depth values. Specifically, if an individual has a depth greater than 1, they connect to someone from the previous depth level, simulating a hierarchical chain of command or influence.
- **Weighted Edges**: Each connection carries a weight equal to the "profit_markup" of the connected node, thus emphasizing connections where profit influence is higher.

This hierarchical setup reflects an organizational or social pyramid structure where individuals influence those at lower levels.

4. Visualizing the Network Graph

To visualize the graph, the notebook utilizes two libraries: **PyVis** for an interactive visualization and **Matplotlib** for a static representation.

Interactive Visualization with PyVis

- **Custom Layout**: The PyVis library allows users to interact with the network, exploring individual nodes and their connections dynamically.
- **Graph Customizations**: Node repulsion, continuous edge smoothing, and a force-directed layout enhance clarity and readability. These options

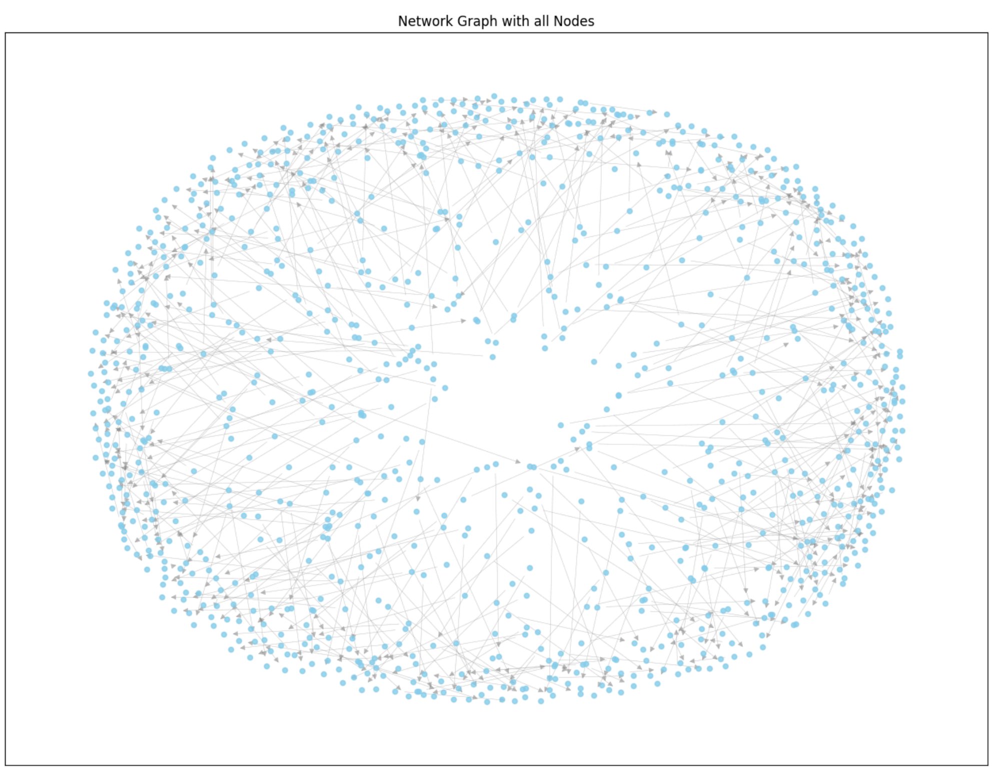
help prevent node overlap and improve spacing, allowing users to focus on significant nodes and clusters.

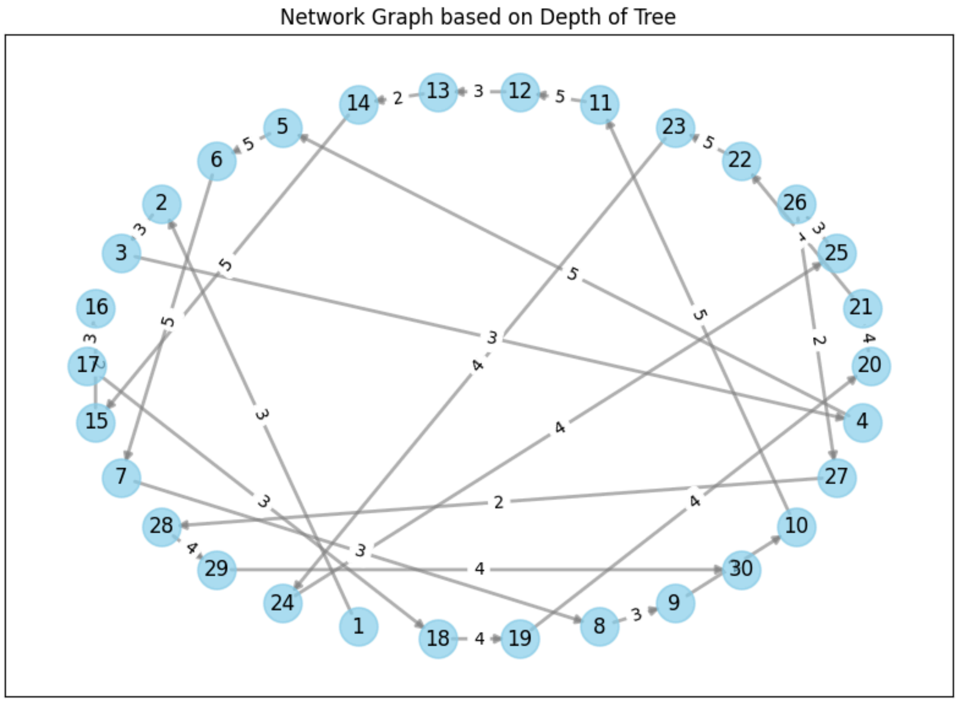
- **Edge Weight Representation**: Weighted edges show the impact of profit markup visually, helping identify influential individuals in the network.

Static Visualization with Matplotlib

- **Spring Layout**: Using NetworkX and Matplotlib, the graph is also visualized in a static format. A spring layout positions nodes based on connectivity, spreading them evenly for improved visibility.
- **Node and Edge Design**: Node size, color, and edge transparency are adjusted to highlight the network's structure. Higher-depth nodes appear centrally connected, while lower-depth nodes are spread out, illustrating the hierarchical tree.

Visualization of the network:





5. Network Metrics and Statistical Analysis

A range of network metrics from NetworkX provides insight into the network's structural properties:

Density

Network density, calculated as the ratio of actual edges to possible edges, is a measure of how connected the network is. A high density implies close connections across all levels, whereas low density indicates sparse connections, typical of hierarchical structures.

Average Distance

The average shortest path length, or **average distance**, is calculated to understand how many steps it typically takes to move between two nodes. A lower average distance signifies a more compact network, while higher values reflect dispersed connections.

Connectivity and Transitivity

Connectivity describes how easily nodes can reach each other, with particular interest in whether the network is connected in one or multiple components. **Transitivity** (or clustering coefficient) examines how interconnected a node's neighbors are, indicating local clusters or tightly-knit groups.

Degree Centralization

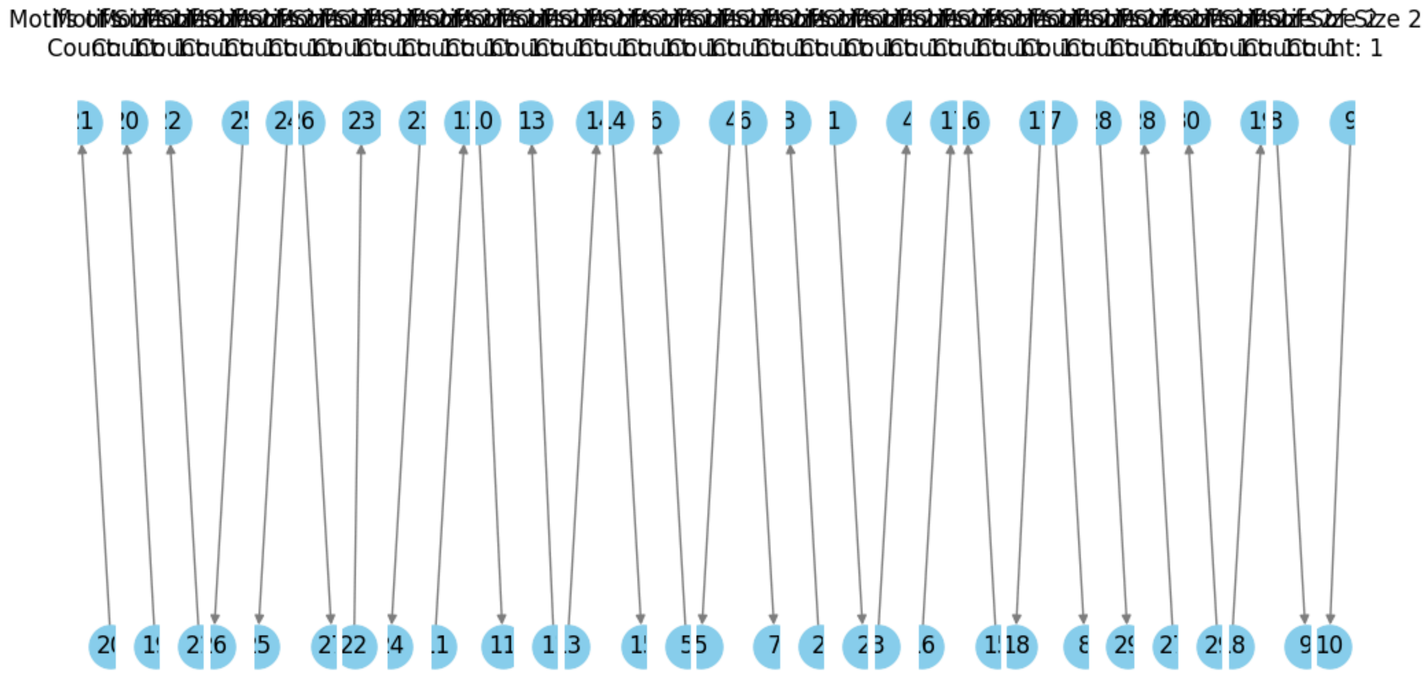
Degree centralization highlights influential nodes by examining the distribution of connections across nodes. High centralization suggests a few nodes have significantly more connections, indicating influential individuals within the hierarchy.

```
Density: 0.03333333333333333
Average Distance: inf
The graph is not strongly connected; average distance is undefined.
Number of Strongly Connected Components: 30
Transitivity: 0
Degree Centralization: 8.493290300662476e-05
```

6. Motif Analysis

A motif is a small, recurring, and significant subgraph pattern within a larger graph. For example, in directed graphs, a motif could be a simple pair of nodes connected by an edge, or a triplet of nodes forming a directed cycle. Motifs can analyze the network from the micro-level. As a typical, local, and functionalized frequent subgraph, motifs can represent local information of the network and have strong statistical significance.

Motifs of size 2 Visualization :



Inference :

Unique Connections:

The 24 distinct size-2 motifs suggest that every connection between two financial entities is unique and direct. This implies a one-to-one relationship in transactions, which could indicate:

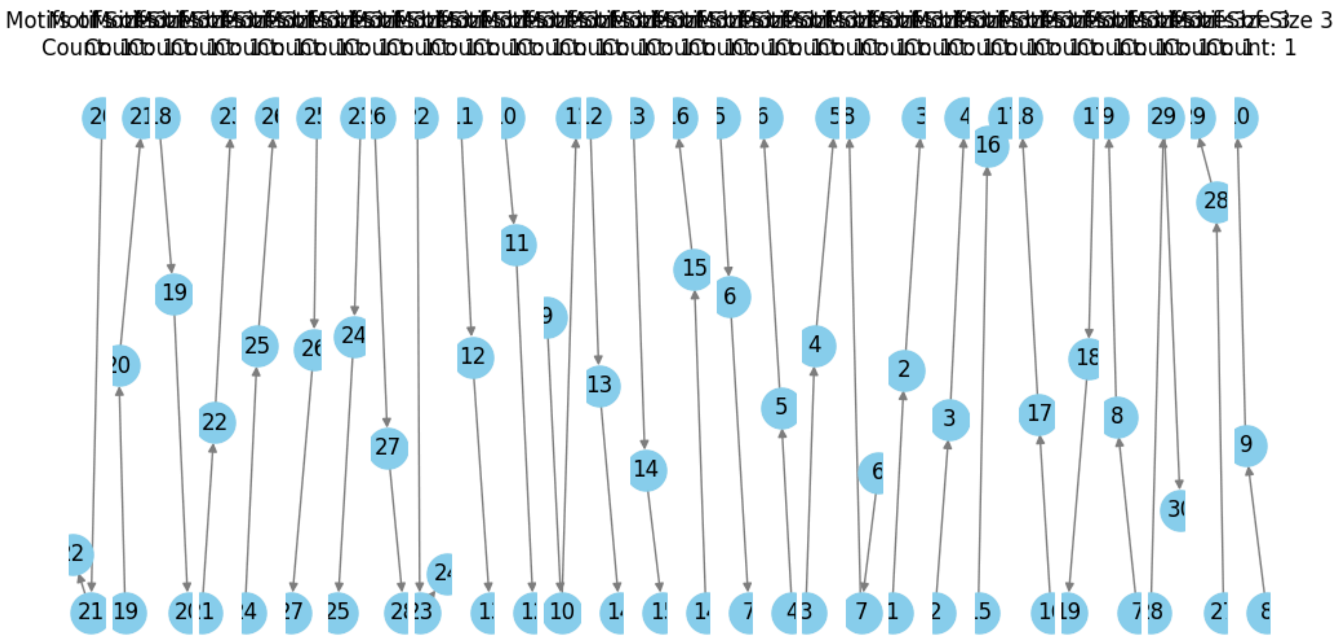
- Direct Transfers:

  Each pair of nodes may represent a direct transfer of funds. For example, if node 20 is a supplier and node 21 is a retailer, the edge could represent payments made from the retailer to the supplier.

- Limited Redundancy:

  The lack of parallel edges suggests that funds are not being rerouted through multiple channels unnecessarily, indicating a streamlined transaction process.

Motifs of size 3 Visualization :



Inference :

Sequential Transactions:

The 27 distinct size-3 motifs, each representing a sequence of transactions, highlight a structured flow of funds through multiple entities. This can be interpreted as:

Pathways of Cash Flow :

The sequences imply that transactions are likely occurring in a set order, potentially following a sales process (e.g., from manufacturer to distributor to retailer) or a financial hierarchy (e.g., department budget allocations).

Cascading Effects :

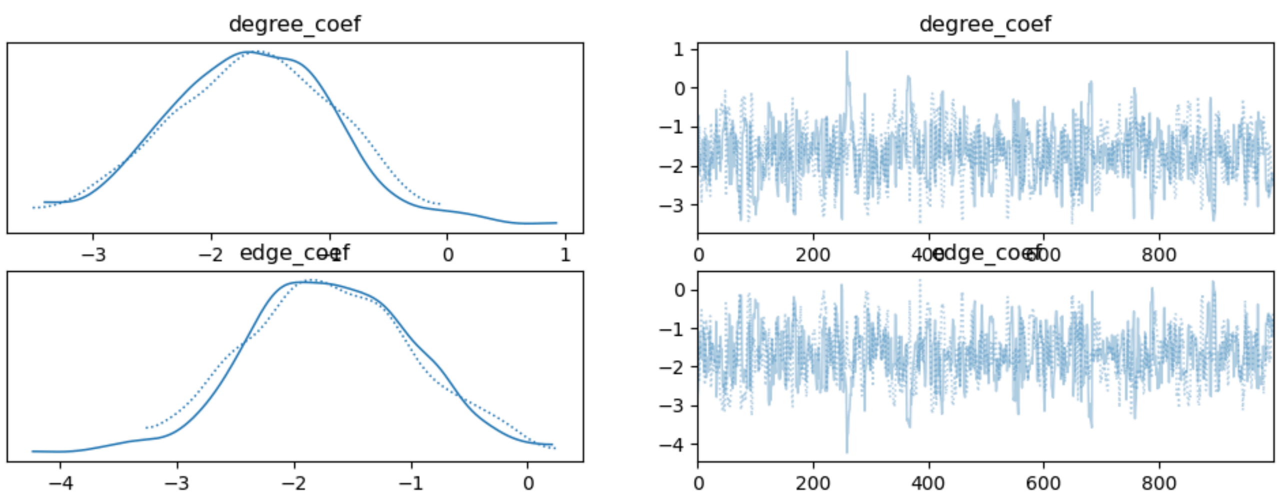
This flow structure may indicate cascading effects, where a change in one transaction could impact subsequent transactions in the chain. For instance, a delay in payment from node 21 to node 22 could affect the financial health of node 22 and subsequently influence its transaction with node 23.

7. Exponential Random Graph Model (ERGM) :

ERGM helps to explore and understand the underlying interactions that shape the network, particularly focusing on the relationships between different communities and between individuals who have been indicted.

In this case study, ERGM analysis is used to examine the formation mechanism of the financial flow network within a pyramid scheme.

Visualization :



	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
degree_coef	-1.674	0.698	-3.064	-0.480	0.031	0.022	521.0	490.0	1.01
edge_coef	-1.685	0.689	-2.821	-0.251	0.030	0.023	534.0	448.0	1.00

Inference :

1. Edge Coefficient (edge_coef):

A positive value here would suggest that connections across hierarchical depths (e.g., nodes in higher depth levels connecting to lower depth levels) are prevalent.

This could imply that hierarchical relationships are dense, meaning that there’s a strong tendency for each level to interact with others, perhaps to propagate profits or engage in commissions. A negative value would indicate a sparse connection structure, suggesting that the hierarchy is more isolated by level, with limited connections between nodes at different depth levels.

2. Degree Coefficient (degree_coef):

A positive value here would indicate that nodes tend to connect to others based on existing connections. For example, if certain nodes at specific depth levels are highly connected, they attract even more connections, reinforcing a pyramidal structure where influential nodes (higher degrees) gain more connections as they advance down the hierarchy. A negative value would suggest that nodes with many connections are not necessarily attracting more, implying a flatter network structure rather than a typical pyramid.

3.General Network Patterns:

If both coefficients are positive and significant, it would imply a strong hierarchical structure with a high density of connections and a strong tendency for nodes to form connections with nodes of similar influence or depth. If either coefficient is insignificant, it might suggest that the network does not strictly follow hierarchical patterns, with connections forming more randomly or based on other factors not captured by depth or degree alone.

Research Paper :

https://www.mdpi.com/2071-1050/11/16/4370

Colab Link :

SNDA.ipynb

THANK YOU.