"AntiBenford Subgraphs: Unsupervised Anomaly Detection in Financial Networks"

Authors

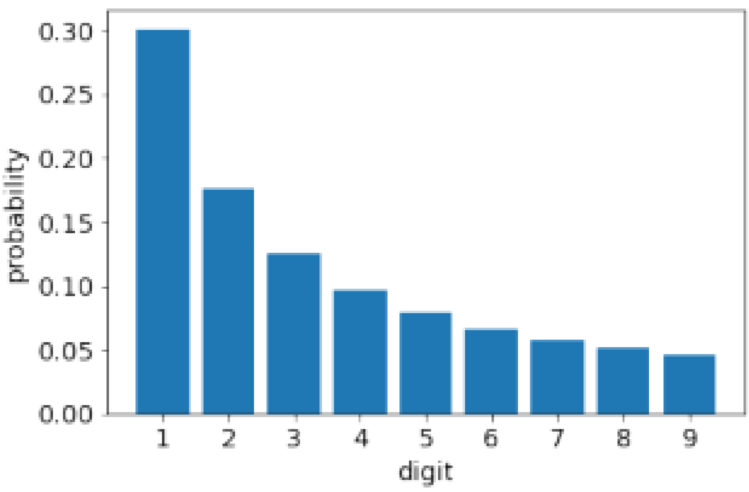
- **Tianyi Chen** (Boston University)
- **Charalampos E. Tsourakakis** (Boston University)

Abstract

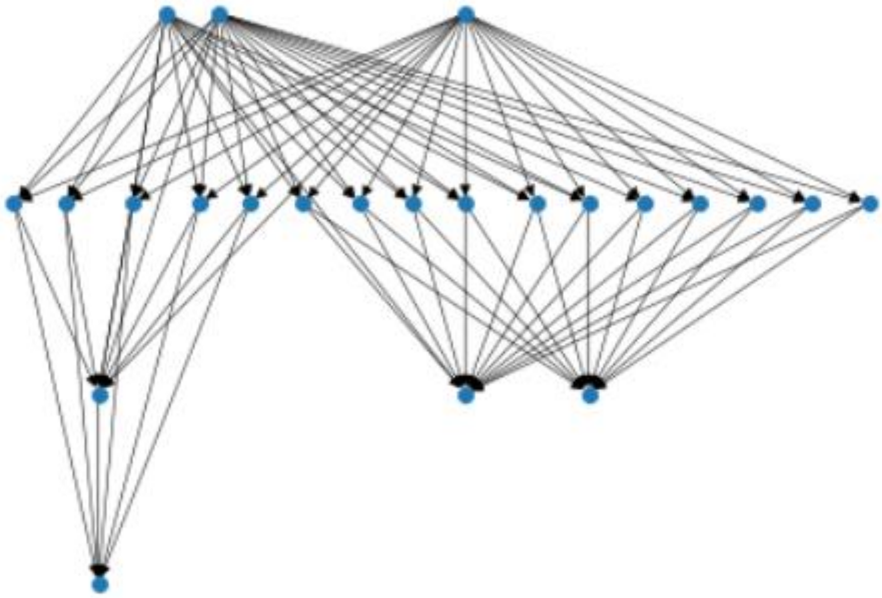
The paper introduces a novel method to detect anomalies in financial transaction networks using deviations from Benford's law. This law predicts the frequency distribution of the first digit in many naturally occurring datasets. The proposed method identifies subgraphs where transaction amounts significantly deviate from this expected distribution, suggesting possible illicit activity.

Key Concepts

1. **Benford's Law**:
    - Describes the expected frequency of the first digit in many datasets.
    - For example, the number '1' appears as the first digit about 30% of the time, while '9' appears less than 5% of the time.
    - Often used to detect anomalies in financial data, suggesting fraud or manipulation.



2. **AntiBenford Subgraphs**:
    - Defined as dense subgraphs in transaction networks where the distribution of the first digit of transaction amounts deviates significantly from Benford's law.
    - Dense subgraphs imply high transaction activity among a specific set of nodes, which may indicate coordinated illicit behavior.



3. **Methodology**:
    - The method combines two main components: adherence to Benford's law and dense subgraph discovery.
    - Dense subgraph discovery is based on the degree density of subgraphs, an approach that efficiently finds clusters of nodes with many interconnections.
    - Statistical tests are used to measure the deviation of first-digit distributions from Benford's law.

Contributions

1. **Novel Framework**:
    - Introduces an unsupervised anomaly detection framework specifically tailored for financial networks.
    - Utilizes deviations from Benford's law to pinpoint potential anomalies.
2. **Algorithmic Approach**:

    - Proposes an efficient algorithm for identifying AntiBenford subgraphs.
    - The algorithm operates in near-linear time, making it scalable for large datasets.

   Algorithm:

   The proposed algorithm combines dense subgraph discovery with statistical testing:

   - Identify dense subgraphs in the network using graph mining techniques.
   - Within these subgraphs, apply statistical tests to compare the first digit distribution of transaction amounts against Benford's law.
   - Output subgraphs where the deviation is statistically significant.

3. **Empirical Validation**:
    - The framework is tested on both real and synthetic datasets.
    - Demonstrates superior performance in detecting anomalous subgraphs in cryptocurrency transaction networks compared to existing methods.
    - Empirical tests show that Ethereum transactions generally follow Benford's law, supporting the validity of the approach.

Example Findings

- The method identified an AntiBenford subgraph in the Ethereum network from January 2018, which resembled a tripartite-like clique, a structure known to appear in money laundering schemes.

Applications

- **Fraud Detection**: Detecting tax evasion, financial fraud, and money laundering.
- **Cryptocurrency Networks**: Monitoring and analyzing blockchain transactions for suspicious activities.

Availability

- Code and datasets for the proposed method are available at GitHub.

Conclusion

The "AntiBenford Subgraphs" framework offers a novel, effective tool for detecting anomalies in financial networks by leveraging deviations from Benford's law and dense subgraph discovery. This method shows promise for uncovering illicit activities that may otherwise go undetected using traditional graph-based anomaly detection techniques.

Prepared by
-Siddhi Behere