Detecting Subtle Fraudulent Communities in Dynamic Social Network using Spectral Clustering

Candidate Numbers : 50714, 49775, 45375, 50098

Abstract

Hidden fraudulent communities, such as spam rings and bots, threaten social media platforms like Reddit by blending into reply networks. Our work investigates fraud detection through four contributions: (1) applying spectral clustering to Reddit reply graphs to uncover small-scale fraudulent communities, (2) enriching graphs with node features, like reply frequency and account age, to improve capture of behavioral fraud patterns; (3) experimenting with the impact of eigenvalue truncation in spectral clustering; and (4) employing spectral clustering on multi-relational Reddit graphs (integrating replies and mentions). We evaluate these methods on synthetic graphs with planted fraud clusters and a labeled Reddit dataset of approximately 50,000 users, using precision, recall, F1-score, accuracy, and training time. The results demonstrate superior detection of hard-to-find fraud compared to baselines, with tuning recommendations and scalability considerations for future graph-based fraud detection research.

Keywords: Spectral clustering, fraudulent communities, node feature augmentation, multi-relational graphs

1. Introduction

Fraudulent communities, such as spam rings and bots, undermine Reddit's reply networks by spreading misinformation and illicit content, with estimates suggesting up to 5% of interactions involve malicious accounts, threatening platform integrity and user trust (Amira et al., 2023; Pourhabibi et al., 2020). In Reddit's sparse, hierarchical reply graphs, where interactions form small-scale communities, detecting these groups is critical yet challenging due to their subtle integration into legitimate patterns (Zhuo et al., 2024). Two key obstacles arise: (1) fraudsters mimic authentic network structures, forming cohesive yet covert clusters that evade traditional graph-based detection (Yang et al., 2019); and (2) extreme label imbalance, with fraudulent users comprising $< 1\%$ of the network, crippling supervised models (Lu et al., 2024). These challenges, exacerbated by the sparse and dynamic nature of Reddit's interaction graphs, necessitate an unsupervised approach that leverages both structural and behavioral signals to identify elusive fraudulent communities.

1.1 Related Work

Fraud detection in social networks employs graph-based methods like graph neural networks (GNNs) (Wu et al., 2020; Deng et al., 2022; Huang et al., 2024), random walk embeddings such as DeepWalk (Davison et al., 2024), community detection via the Louvain algorithm (Cai et al., 2016), and ranking methods like PageRank (Yang et al., 2019). GNNs, reliant on labeled data, struggle with label imbalance and sparse communities (Zhuo et al.,

2024). DeepWalk captures structural proximity but misses subtle eigenvalue-driven clusters (Davison et al., 2024), while Louvain's topology-focused modularity optimization overlooks behavioral cues like reply frequency (Cai et al., 2016). PageRank prioritizes node importance, failing to delineate community boundaries (Yang et al., 2019). Text-based methods, such as PCA for user behavior (Viswanath et al., 2014), ignore graph relationships. Spectral clustering, which embeds nodes via the normalized Laplacian's eigenvectors (Ng et al., 2001; Fan et al., 2022), leverages spectral graph theory to detect arbitrary clusters (Kumar and Daumé, 2011) and has succeeded in telecommunications (Guedes et al., 2023) and finance (Cai et al., 2016). Yet, its application to social media fraud detection, particularly with behavioral feature augmentation and multi-relational graphs, is novel. Our spectral clustering framework addresses these gaps by exploiting eigenvalue gaps to uncover subtle fraud patterns, incorporating node features (e.g., reply frequency, account age) to capture behavioral signals, and leveraging multi-relational graphs (replies and mentions) for richer interactions, offering an unsupervised solution robust to label imbalance and mimicry.

1.2 Main Contributions

Our contributions are fourfold: (1) we apply spectral clustering with the normalized Laplacian to Reddit reply graphs, outperforming DeepWalk by exploiting eigenvalue gaps to detect small-scale fraudulent communities (Ng et al., 2001; Davison et al., 2024); (2) we augment graphs with node features like reply frequency and account age, surpassing the Louvain method in capturing behavioral fraud patterns (Cai et al., 2016); (3) we investigate eigenvalue truncation to optimize spectral clustering's accuracy versus k-means on graph embeddings, offering tuning insights (Fan et al., 2022); and (4) we employ multi-relational graphs combining replies and mentions, exceeding PageRank-based but fail to leverage richer interactions to beat single-relational clustering (Yang et al., 2019). These contributions, which represent novel applications of spectral clustering to social media fraud detection, enhance precision and robustness using a labeled Reddit dataset of ∼50,000 users and synthetic graphs. We provide scalability considerations for future graph-based fraud detection research.

2. Problem Formulation: Fraudulent Community Detection in Reddit Reply Graphs

In general, the settings of the problems we look to approach in this project are as follows. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph constructed from Reddit, where $\mathcal{V}$ is the vertex set, representing the set of Reddit users and $\mathcal{E}$ is the edge set, representing the reply interactions between users. Each node, $v_i \in \mathcal{V}$, an undirected edge, $(v_i, v_j) \in \mathcal{E}$, exists if there is a reply interaction between user v_i and user v_j. We define the adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, where $\mathbf{A}_{ij} = 1$ if a reply interaction between user v_i and v_j exists, and 0 otherwise.

By analyzing the comments, reply frequency and account age of each user, we labeled some users with extreme value as fraudulent. Let $y_i \in \{0, 1\}$ be the ground-truth label for node v_i, where $y_i = 1$ indicates that user v_i is fraudulent.

Fraudulent Community Detection aims at partitioning the nodes set $\mathcal{V}$ into K non-overlapping subsets $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K$ (i.e., $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ if $i \neq j$) and $\mathcal{C}_1 \cup \mathcal{C}_2 \cup ... \cup \mathcal{C}_K = \mathcal{V}$. Each subset, $\mathcal{C}_i$, is then labeled as a fraudulent or legitimate community.

3. Proposed solutions

3.1 Spectral Clustering

We posit that fraudulent communities exhibit dense intra-group reply structures and relatively sparse inter-group connections. We define the **normalized Laplacian** of the graph as:

$$\mathbf{L}_{\text{sym}} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{A})\mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2},$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the adjacency matrix defined in Section 2, $\mathbf{I} \in \mathbb{R}^{n\times n}$ is the identity matrix, $\mathbf{D} \in \mathbb{R}^{n\times n}$ is the degree matrix of $\mathbf{A}$ and $\mathbf{D}^{-1/2} \in \mathbb{R}^{n\times n}$ is the inverse square root matrix of $\mathbf{D}$. We can calculate $\mathbf{D}$ using $\mathbf{D} = \text{diag}\left((d_i)_{i=1}^n\right)$, where d_i is the degree of node v_i.

Our objective is to partition the graph into K disjoint communities $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ such that the **Normalized Cut** (NCut) is minimized:

$$\text{NCut}(\mathcal{C}_1, \ldots, \mathcal{C}_K) = \frac{1}{2}\sum_{k=1}^{K} \frac{\left|\{(u,v) \in \mathcal{E} : u \in \mathcal{C}_k, v \in \bar{\mathcal{C}_k}\}\right|}{\text{vol}(\mathcal{C}_k)}$$

where $\text{vol}(\mathcal{C}_k) = \sum_{u\in\mathcal{C}_k} d_u$ is the volume of the community. Since exact minimization of NCut is NP-hard, we employ a spectral relaxation. Let $\mathbf{H} \in \mathbb{R}^{n\times K}$ be the matrix containing the bottom K eigenvectors of $\mathbf{L}_{\text{sym}}$. The relaxed objective becomes:

$$\min_{\mathbf{H}^\top\mathbf{H}=\mathbf{I}} \text{Tr}(\mathbf{H}^\top\mathbf{L}_{\text{sym}}\mathbf{H}),$$

where $\mathbf{H} \in \mathbb{R}^{n\times K}$ contains the top K eigenvectors of $\mathbf{L}_{\text{sym}}$ and is the cluster indicator matrix. The columns h_k are such that $h_k(u) = 1/\sqrt{\text{vol}(\mathcal{C}_k)}$ if $u \in C_i$ and 0 otherwise. The rows of $\mathbf{H}$ provide a spectral embedding of the nodes, and we apply K-means clustering to these embeddings to produce communities $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$.

3.2 Spectral Clustering with Feature Augmentation

In *Subsection 3.1*, the adjacency matrix $\mathbf{A}$ is constructed based on the interactions between users. As the data we collected also includes the reply frequency r_i and account age a_i for each node $v_i \in \mathcal{V}$, we combine the spectral clustering with these node features and propose the spectral clustering with feature augmentation method.

Let vector $\mathbf{f}_i = [r_i, a_i]^\top \in \mathbb{R}^2$ be the feature vector of node $v_i \in \mathcal{V}$ and $\mathbf{f}_i'$ be the normalized vector of $\mathbf{f}_i$. We compute feature similarity between node v_i and v_j using the RBF kernel:

$$\mathbf{S}_{ij} = \exp\left(-\gamma\|\mathbf{f}_i' - \mathbf{f}_j'\|_2^2\right), \quad \gamma = 1.0.$$

We construct a new matrix $\mathbf{W}$ to represent the graph structure, using the adjacency matrix $\mathbf{A}$ and the feature similarity matrix $\mathbf{S}$:

$$\mathbf{W} = \alpha\mathbf{A} + (1-\alpha)\mathbf{S},$$

where α is a constant to balance the contribution of adjacency matrix $\mathbf{A}$ and feature similarity matrix $\mathbf{S}$. The tuition behind $\mathbf{W}$ is that we want to include more information into

the edges between users instead of just using edges to represent the interactions between users.

Then we can construct a new normalized Laplacian based on the new matrix $\mathbf{W}$:

$$\mathbf{L}_w = \mathbf{I} - \mathbf{D}_w^{-1/2}\mathbf{W}\mathbf{D}_w^{-1/2},$$

where $\mathbf{I} \in \mathbb{R}^{n\times n}$ is the identity matrix and $\mathbf{D}_w = \operatorname{diag}\left((\sum_i \mathbf{W}_{ij})_{i=1}^n\right)$. Note that $\mathbf{D}_w$ is not the degree matrix as we add node features to this matrix, but we can calculate $\mathbf{D}_w$ similarly by summing each row of $\mathbf{W}$.

With this new Laplacian $\mathbf{L}_w$, we can produce communities $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ with the relaxed objective function $\min_{\mathbf{H}^\top\mathbf{H}=\mathbf{I}} \operatorname{Tr}(\mathbf{H}^\top\mathbf{L}_w\mathbf{H})$, just like in *Subsection 3.1*.

3.3 Spectral clustering with Different Numbers of Eigenvectors

We also propose a scheme to investigate the effect of the choice of eigenvalue truncation on the performance of the improved spectral clustering method described in *Subsection 3.2*.

Recall that we employ the relaxed objective function $\min_{\mathbf{H}^\top\mathbf{H}=\mathbf{I}} \operatorname{Tr}(\mathbf{H}^\top\mathbf{L}_w\mathbf{H})$ to produce communities $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$, where $\mathbf{H} \in \mathbb{R}^{n\times K}$ contains the top K eigenvectors of $\mathbf{L}_w$ and is the cluster indicator matrix.

We will construct $\mathbf{H}$ by changing the magnitude K. Let K_1, K_2 be the integer such that $K_1 \leq K_2 \leq n$, and $\mathbf{H}_1 \in \mathbb{R}^{n\times K_1}, \mathbf{H}_2 \in \mathbb{R}^{n\times K_2}$ be the matrix containing the bottom K_1 and K_2 eigenvectors of $\mathbf{L}_w$, respectively. Then we can produce communities $\{\mathcal{C}_1, \ldots, \mathcal{C}_{K_1}\}$ with matrix $\mathbf{H}_1$, and communities $\{\mathcal{C}'_1, \ldots, \mathcal{C}'_{K_2}\}$ with $\mathbf{H}_2$. The effect of different eigenvalue truncation will be evaluated by comparing the performance of corresponding communities.

3.4 Spectral Clustering with Multi-Relational Graphs

Now, let $\mathcal{G} = (\mathcal{V}, \mathcal{E}')$ be a multi-relational graph constructed from Reddit data, where $\mathcal{V}$ is, again, the vertex set (users) and $\mathcal{E}_{\text{multi}} = \{\mathcal{E}_{\text{reply}}, \mathcal{E}_{\text{mention}}\}$ represents two types of edges. Reply edges, $\mathcal{E}_{\text{reply}}$, that exists when a reply interaction between user A and user B exists and mention edges, $\mathcal{E}_{\text{mention}}$, that exists when user A mentions user B in a comment or vice-versa. Each relation induces an adjacency matrix,

$$\mathbf{A}_{\text{reply}} \in \mathbb{R}^{n\times n}, \quad \mathbf{A}_{\text{mention}} \in \mathbb{R}^{n\times n}.$$

We define a combined weighted adjacency matrix $\mathbf{A}_{\text{combined}}$ as:

$$\mathbf{A}_{\text{combined}} = w_1\mathbf{A}_{\text{reply}} + w_2\mathbf{A}_{\text{mention}} \quad \text{where } w_1, w_2 \geq 0 \text{ and } w_1 + w_2 = 1.$$

The matrix, $\mathbf{W}_{\text{m}}$, is constructed to represent the graph structure, using the adjacency matrix $\mathbf{A}_{\text{combined}}$ and the feature similarity matrix $\mathbf{S}$ previously defined in *Subsection 3.2*:

$$\mathbf{W}_{\text{m}} = \beta\mathbf{A}_{\text{combined}} + (1-\beta)\mathbf{S}.$$

The normalized Laplacian based on the new matrix $\mathbf{W}_m$ is then:

$$\mathbf{L}_{\text{m}} = \mathbf{I} - \mathbf{D}_{\text{m}}^{-1/2}\mathbf{W}_{\text{m}}\mathbf{D}_{\text{m}}^{-1/2},$$

where $\mathbf{I} \in \mathbb{R}^{n\times n}$ is the identity matrix and $\mathbf{D}_{\text{m}} = \operatorname{diag}\left(\left(\sum_{i=1}^n (\mathbf{W}_{\text{m}})_{ij}\right)_{j=1}^n\right)$. As in previous subsections, communities are produced by applying K-means clustering to the spectral embeddings of the normalized Laplacian, $\mathbf{L}_{\text{m}}$.

4. Numerical Experiments

In this section, we evaluate the proposed spectral clustering framework for detecting fraudulent communities in Reddit reply networks, addressing the four research questions outlined in the introduction. We designed the following experiments to demonstrate the efficacy and novelty of our approach:

- **Experiment-1**: We test spectral clustering with the normalized Laplacian against DeepWalk for detecting small-scale fraud (e.g., spam rings) in Reddit reply graphs. Using $k = 10$ eigenvectors of $\mathbf{L}_{\text{sym}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ and k-means, we compare with DeepWalk's 128-dimensional embeddings (Davison et al., 2024). Eigenvalue gaps highlight subtle structures missed by random walks (Ng et al., 2001; Kumar and Daumé, 2011).

- **Experiment-2**: We assess spectral clustering with node features (reply frequency, account age) versus the Louvain method. Node features are concatenated to Laplacian embeddings, benchmarked against Louvain's topology-only clustering (Cai et al., 2016). Behavioral features enhance fraudster pattern detection.

- **Experiment-3**: We study eigenvalue truncation ($k = 5, 10, 20$) in spectral clustering versus k-means on graph embeddings. We evaluate on Reddit and synthetic graphs, providing tuning insights for balancing accuracy and computational cost ($O(n^3)$) (Fan et al., 2022).

- **Experiment-4**: We compare spectral clustering on multi-relational Reddit graphs (replies + mentions) to single-relational clustering and PageRank methods. Multi-relational graphs, built from comment text mentions, improve detection by capturing richer interactions (Yang et al., 2019).

4.1 Dataset, Preprocessing, Metrics, and Visualizations

The Reddit dataset (`reddit_comments_spectral.csv`) contains 46,414 comments from `Bitcoin`, `ethereum`, and `CryptoCurrency` subreddits, with 7,444 authors (Amira et al., 2023). Preprocessing removes `[deleted]` or missing authors, retaining valid `parent_id` and `account_age_days` fields (1–2 missing entries). A reply graph (`G_reply`: 4,871 nodes, 11,310 edges) is built using `parent_id`, with Experiment-4 adding a multi-relational graph (`G_multi`: 4,871 nodes, 11,342 edges) with sampled mentions (regex: `@(\w+)`, capped at half reply edges). Node features (reply frequency, account age) are z-score normalized. Three fraudulent communities (8 nodes each, 24 total) are simulated with dense edges (probability 0.8), high reply frequency (0.15–0.25×time span), and low account age (3–7 days), adding 83–11,626 edges. Synthetic graphs ($n = 10,000$, 10% fraud nodes) test robustness (Kumar and Daumé, 2011). Adjacency matrices are stored in sparse format (`scipy.sparse`).

All experiments run 10 times with random initializations on an Intel Xeon (2.4 GHz). Metrics include precision, recall, F1-score, accuracy, ROC-AUC, (ARI, AMI, NMI are Cluster metrics), and training time. Clusters with fraud ratio > 0.1 are labeled fraudulent. Visualizations include eigenvalue plots, 2D-PCA scatter plots, confusion matrices, bubble plots, and cluster size/fraud fraction plots (Cai et al., 2016).

4.2 Numerical Experiment 1: Spectral Clustering vs. DeepWalk

We compute the $k = 5$ smallest eigenvectors (`eigsh`, tolerance 10^{-6}) of the normalized Laplacian, $\mathbf{L}_{\text{sym}}$, and use the K-means algorithm ($K = 10$) to cluster the nodes. Fraud is predicted if a cluster contains 2 or more fraud nodes (Ng et al., 2001). We benchmark this against DeepWalk, where we perform 10 random walks of length 40 per node, generating 64-dimensional embeddings via Word2Vec (window=5, epochs=1). The K-means algorithm ($K = 10$) is, again, applied to cluster the embeddings (Davison et al., 2024).

4.2.1 Experimental Design and Results

A Reddit reply graph with three synthetic fraudulent groups (8 nodes each) is analyzed using spectral clustering (normalized Laplacian, $k = 10$, 5 eigenvectors) Ng et al. (2001); Fan et al. (2022). It is compared to DeepWalk (64-dim embeddings from 10 random walks of length 40), using clusters with ≥ 2 fraud nodes as the detection criterion Davison et al. (2024); Albahar (2019).

Method	Precision	Recall	F1-Score	Accuracy	ARI	AMI	NMI
Spectral Clustering	0.0047	**0.9167**	0.0094	0.0470	**0.0082**	**0.0026**	**0.0055**
DeepWalk	**0.0055**	**0.9167**	**0.0109**	**0.1829**	-0.0040	0.0003	0.0007

Table 1: Performance Metrics for Experiment 1

Method	Time (s)	Fraud Cluster
Spectral Clustering	0.02	10 nodes (0.1000)
DeepWalk	0.02	416 nodes (0.0072)

Table 2: Computational Complexity

Spectral clustering outperforms DeepWalk in ARI (0.0082 vs. -0.0040), identifying a higher fraud fraction (0.1000) in Cluster 3. DeepWalk achieves better accuracy (0.1829).

4.2.2 Discussion

Spectral clustering leverages eigenvalue gaps to detect dense fraud subgraphs, outperforming DeepWalk in clustering quality. DeepWalk improves accuracy but struggles with fraud isolation. The ≥ 2 fraud node rule inflates false positives, reducing precision. Spectral clustering's $O(n^3)$ complexity limits scalability compared to DeepWalk's linear cost (Davison et al., 2024). Future work could refine fraud labeling or explore scalable spectral methods (Cai et al., 2016).

4.3 Numerical Experiment 2: Augmented Spectral Clustering vs. Louvain

We take the normalized Laplacian of $\mathbf{W} = \alpha\mathbf{A} + (1 - \alpha)\mathbf{S}$, which is the weighted sum of the adjacency matrix, $\mathbf{A}$, and the feature similarity matrix, $\mathbf{S}$, with $\alpha = 0.7$. The $k = 5$ smallest eigenvectors (`eigsh`, tolerance 10^{-6}) are computed for W to generate a different

spectral embedding of the graph and use the K-means algorithm ($K = 10$) to cluster the nodes. The same prediction policy as in *Subsection 4.2* is applied. We benchmark this idea against the Louvain method with resolution = 0.5.

4.3.1 Experimental Design and Results

A Reddit reply graph with three synthetic fraud groups (8 nodes each, dense ties, extreme features) is analyzed via spectral clustering (top 15 eigenvectors, $k = 5$, node features) Ng et al. (2001); Fan et al. (2022), compared to Louvain (resolution = 0.5) Viswanath et al. (2014); Table 3 reports detection results Yang et al. (2019); Pourhabibi et al. (2020).

Method	Precision	Recall	F1-Score	Accuracy	ARI	AMI	NMI	Time (s)	Fraud Cluster
Augmented Spectral	**0.17**	**1.00**	**0.30**	**0.98**	**0.28**	**0.22**	**0.23**	0.02	138 nodes (0.17)
Louvain	0.16	**1.00**	0.27	0.97	0.26	0.20	0.20	0.01	35 nodes (0.23)

Table 3: Performance Metrics for Experiment 2

Augmented Spectral outperforms Louvain in precision (0.17) and F1-score (0.30), isolating 24 fraud nodes in Cluster 3 (fraction 0.17). Louvain splits fraud across three clusters (highest fraction 0.23).

4.3.2 Discussion

Behavioral features enhance spectral clustering's fraud detection, improving cluster quality over Louvain's topology-only approach (Cai et al., 2016). Louvain scatters fraud nodes, reducing precision. Spectral clustering's $O(n^3)$ cost is a limitation. Future work could tune feature weight α or add sentiment features (Yang et al., 2019).

4.4 Numerical Experiment 3: Eigenvalue Truncation in Spectral Clustering vs. K-Means on Embeddings

We compute the normalized Laplacian $\mathbf{L}$ of the matrix $\mathbf{W} = 0.8\mathbf{A} + 0.2\mathbf{K}$, where $\mathbf{K}$ is an RBF kernel (gamma = 0.5). The top 5 or 10 eigenvectors (based on eigengap) of $\mathbf{L}$ are extracted and concatenated with 0.5× scaled node features. The resulting embeddings are clustered using the K-means algorithm ($K = 5$ or 10), following the method of Fan et al. (2022).

We benchmark this against a DeepWalk-based approach, where we perform 200 random walks of length 30 per node to generate 64-dimensional Node2Vec embeddings. These embeddings are augmented with 0.7× scaled node features and clustered using K-means ($K = 6$), with the optimal number of clusters selected via the silhouette score.

We also consider a GraphSAGE-based baseline, where a single-layer GraphSAGE model is trained to produce 8-dimensional node embeddings. These are clustered using K-means, following the same procedure as above.

4.4.1 Experimental Design and Results

We construct a Reddit reply graph from cryptocurrency subreddits, injecting three synthetic fraudulent communities (8 nodes each), using spectral clustering with top 5 and 10 eigen-

vectors, k-means (k from eigenvalue gaps) Ng et al. (2001); Fan et al. (2022), Node2Vec for 64-dimensional embeddings with silhouette score-based k Davison et al. (2024), and compare detection accuracy and cost (Table 4) Yang et al. (2019); Pourhabibi et al. (2020).

Method	Prec	Rec	F1	Acc	ARI	AMI	NMI	Time (s)	Fraud Cluster
Spectral (Top 5)	**0.38**	**1.00**	**0.55**	**0.99**	0.01	0.04	0.04	0.02	63 nodes (0.38)
Spectral (Top 10)	0.24	**1.00**	0.38	0.98	0.01	0.04	0.04	0.01	102 nodes (0.24)
DeepWalk	0.28	**1.00**	0.43	**0.99**	**0.42**	**0.34**	**0.34**	0.01	87 nodes (0.28)
GraphSAGE	0.13	**1.00**	0.23	0.97	0.21	0.17	0.17	0.01	186 nodes (0.13)

Table 4: Performance Metrics for Experiment 3

Spectral clustering (Top 5) excels in precision (0.38) and F1-score (0.55), isolating 24 fraud nodes in a 63-node cluster. Top 10 reduces precision (0.24). DeepWalk balances precision (0.28) and ARI (0.42). GraphSAGE underperforms (precision: 0.13).

4.4.2 Discussion

Top 5 eigenvectors capture dense fraud subgraphs effectively, but Top 10 over-segments, diluting precision. DeepWalk's embeddings offer scalability and strong clustering (ARI: 0.42). GraphSAGE's shallow embeddings limit fraud detection. Spectral clustering's $O(k \cdot n^2)$ cost is a bottleneck. Tuning suggests $k = 5$ balances accuracy and cost (Fan et al., 2022). Future work could explore adaptive thresholds or deeper GraphSAGE models.

4.5 Numerical Experiment 4: Multi-Relational Spectral Clustering for Fraud Detection

We compute a multi-relational spectral clustering approach by combining the reply and mention adjacency matrices as $\mathbf{A}_{\text{multi}} = 0.7\mathbf{A}_{\text{reply}} + 0.3\mathbf{A}_{\text{mention}}$. We then form the matrix $\mathbf{W}_{\text{multi}} = 0.8\mathbf{A}_{\text{multi}} + 0.2\mathbf{S}_{\text{features}}$, where $\mathbf{S}_{\text{features}}$ is an RBF kernel similarity matrix with $\gamma = 0.1$. The top 7 eigenvectors of the normalized Laplacian of $\mathbf{W}_{\text{multi}}$ are computed and clustered using K-means ($K = 5$), with the number of components selected based on the eigengap heuristic.

We, additionally, implement a single-relational spectral clustering method using the reply adjacency matrix $\mathbf{A}_{\text{single}}$, forming $\mathbf{W}_{\text{single}} = 0.8\mathbf{A}_{\text{single}} + 0.2\mathbf{S}_{\text{features}}$, where the RBF kernel uses $\gamma = 0.04$. The top 7 eigenvectors are again clustered using K-means ($K = 5$). Both approaches are then benchmarked against a PageRank-based clustering technique.

4.5.1 Experimental Design and Results

We build a multi-relational Reddit graph from cryptocurrency subreddits with reply and mention edges, injecting three synthetic fraudulent groups (8 nodes each), using spectral clustering with top 7 eigenvectors and k-means (k from eigenvalue gaps) for multi- and single-relational graphs, and PageRank for top 10% nodes. For benchmarking, PageRank on $\mathbf{G}_{\text{reply}}$ with $\alpha = 0.85$ clusters top 10% nodes via k-means ($K = 5$) Yang et al. (2019).

Method	Prec	Rec	F1	Acc	ARI	AMI	NMI	Time (s)	Fraud Cluster
Multi-Relational	0.38	**1.00**	0.55	**0.99**	**0.01**	0.03	0.03	0.01	64 nodes (0.38)
Single-Relational	**0.47**	**1.00**	**0.64**	**0.99**	**0.01**	**0.04**	**0.04**	0.01	51 nodes (0.47)
PageRank	0.22	0.25	0.24	**0.99**	**0.01**	0.02	0.02	0.02	27 nodes (0.22)

Table 5: Performance Metrics for Experiment 4

Single-relational spectral clustering achieves the highest precision (0.47) and F1-score (0.64), isolating 24 fraud nodes in a 51-node cluster. Multi-relational follows (precision: 0.38, F1: 0.55). PageRank underperforms (precision: 0.22).

4.5.2 Discussion

Single-relational clustering excels due to sparse mention edges diluting multi-relational structure. Multi-relational clustering enhances interpretability, grouping all 24 fraud nodes cohesively. PageRank misses dense subgraphs. Low ARI reflects class imbalance. Synthetic graphs suggest multi-relational potential with denser mentions. Spectral clustering's $O(k \cdot n^2)$ cost is a limitation. Future work could refine mention extraction (Yang et al., 2019).

5. Conclusions

Our spectral clustering framework advances fraud detection in Reddit reply networks by leveraging eigenvalue gaps, node features, truncation strategies, and multi-relational graphs, outperforming baselines like DeepWalk, Louvain, and PageRank in interpretability. However, its practical utility is limited by persistently low precision (e.g., 0.0047 in Experiment-1, 0.47 in Experiment-4), reflecting challenges in real-world deployment. Key bottlenecks include the cubic complexity ($O(n^3)$), constraining scalability on large graphs, as seen in Intel Xeon runtimes; sparse mention edges diluting multi-relational clustering (Experiment-4); and the ≥ 2 fraud node rule inflating false positives (Experiment-1). Class imbalance ($< 1\%$ fraud nodes) yields low ARI scores (e.g., 0.01 in Experiments 3–4), while the dataset's focus on cryptocurrency subreddits and reliance on simulated fraud clusters may limit generalizability (Pourhabibi et al., 2020; Yang et al., 2019). Parameter sensitivity (e.g., k, α, γ) impacts robustness, and the preprocessing step's removal of only 1–2 missing entries (Section 4.1) may overlook data quality issues. Moreover, the utility of visualizations (e.g., 2D PCA plots) lacks quantitative validation, missing actionable insights.

Future work should explore scalable spectral methods like randomized eigendecomposition (Cai et al., 2016; Zhang et al., 2022) to reduce complexity, and imbalance-aware clustering (Zhuo et al., 2024) to improve ARI. Integrating sentiment into the $\mathbf{W}$ matrix (Section 3.2) may enhance behavioral detection (Lu et al., 2024; Tripathi et al., 2024), while refining mention extraction (Yang et al., 2019), adaptive fraud thresholds (Fan et al., 2022), and automated tuning via silhouette scores (Experiment-3) can address remaining gaps. Hybrid approaches with GNNs (Wu et al., 2020) and real-time frameworks (Huang et al., 2024) could leverage labels and adapt to dynamic fraud, with cross-platform validation (Pourhabibi et al., 2020) ensuring broader applicability.

References

Marwan Ali Albahar. Detecting fraudulent twitter profiles: a model for fraud detection in online social networks. *International Journal of Innovative Computing, Information and Control*, 15(5):1629–1639, 2019.

Abdelouahab Amira, Abdelouahid Derhab, Samir Hadjar, Mustapha Merazka, Md Golam Rabiul Alam, and Mohammad Mehedi Hassan. Detection and analysis of fake news users' communities in social media. *IEEE Transactions on Computational Social Systems*, 2023.

Fan Cai, Nhien-An Le-Khac, and Tahar Kechadi. Clustering approaches for financial data analysis: a survey. *arXiv preprint arXiv:1609.08520*, 2016.

Andrew Davison, S Carlyle Morgan, and Owen G Ward. Community detection guarantees using embeddings learned by node2vec. In *Advances in Neural Information Processing Systems*, volume 37, pages 685–738, 2024.

Leyan Deng, Chenwang Wu, Defu Lian, Yongji Wu, and Enhong Chen. Markov-driven graph convolutional networks for social spammer detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12310–12322, 2022.

Jicong Fan, Yiheng Tu, Zhao Zhang, Mingbo Zhao, and Haijun Zhang. A simple approach to automated spectral clustering. In *Advances in Neural Information Processing Systems*, volume 35, pages 9907–9921, 2022.

Bruna Toledo Guedes, Diego Passos, and Fernanda GO Passos. A spectral clustering algorithm for intelligent grouping in dense wireless networks. *Computer Communications*, 198:117–127, 2023.

Haitao Huang, Hu Tian, Xiaolong Zheng, Xingwei Zhang, Daniel Dajun Zeng, and Fei-Yue Wang. Cgnn: A compatibility-aware graph neural network for social media bot detection. *IEEE Transactions on Computational Social Systems*, 2024.

Abhishek Kumar and Hal Daumé. A co-training approach for multi-view spectral clustering. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 393–400, 2011.

Yen-Wen Lu, Yu-Che Tsai, and Cheng-Te Li. Burstiness-aware bipartite graph neural networks for fraudulent user detection on rating platforms. In *Companion Proceedings of the ACM Web Conference 2024*, pages 834–837, 2024.

Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, volume 14, 2001.

Tahereh Pourhabibi, Kok-Leong Ong, Booi H. Kam, and Yee Ling Boo. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133:113303, 2020.

Ashutosh Tripathi, Mohona Ghosh, and Kusum Kumari Bharti. Markov enhanced graph attention network for spammer detection in online social network. *Knowledge and Information Systems*, 66(9):5561–5580, 2024.

Bimal Viswanath, M Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. Towards detecting anomalous user behavior in online social networks. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 223–238, 2014.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S. Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2020.

Yang Yang, Yuhong Xu, Yizhou Sun, Yuxiao Dong, Fei Wu, and Yueting Zhuang. Mining fraudsters and fraudulent strategies in large-scale mobile social networks. *IEEE Transactions on Knowledge and Data Engineering*, 33(1):169–179, 2019.

Erchuan Zhang, David Suter, Giang Truong, and Syed Zulqarnain Gilani. Sparse hypergraph community detection thresholds in stochastic block model. In *Advances in Neural Information Processing Systems*, volume 35, pages 34012–34023, 2022.

Wei Zhuo, Zemin Liu, Bryan Hooi, Bingsheng He, Guang Tan, Rizal Fathony, and Jia Chen. Partitioning message passing for graph fraud detection. In *International Conference on Learning Representations (ICLR)*, 2024.

Appendix A. Cluster Metrics

We use ARI, AMI, and NMI to objectively evaluate how well clustering algorithms recover ground truth communities while accounting for randomness and varying cluster structures.

A.1 Adjusted Rand Index (ARI)

The Adjusted Rand Index (ARI) corrects the Rand Index for chance agreement and is defined as:

$$\text{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}, \tag{1}$$

where n_{ij} is the number of samples in both true cluster i and predicted cluster j, $a_i = \sum_j n_{ij}$, $b_j = \sum_i n_{ij}$, and n is the total number of data points.

A.2 Adjusted Mutual Information (AMI)

AMI measures the agreement between two clusterings while adjusting for chance:

$$\text{AMI} = \frac{\text{MI}(U,V) - \mathbb{E}[\text{MI}(U,V)]}{\max(\text{H}(U),\text{H}(V)) - \mathbb{E}[\text{MI}(U,V)]}, \tag{2}$$

where the Mutual Information is:

$$\mathrm{MI}(U,V) = \sum_{u\in U}\sum_{v\in V} p(u,v)\log\left(\frac{p(u,v)}{p(u)p(v)}\right), \tag{3}$$

and $\mathrm{H}(U)$, $\mathrm{H}(V)$ are the entropies of the true and predicted label distributions respectively. $\mathbb{E}[\mathrm{MI}]$ is the expected mutual information under a random labeling.

A.3 Normalized Mutual Information (NMI)

Normalized Mutual Information is the unadjusted version of AMI and is defined as:

$$\mathrm{NMI} = \frac{2\cdot \mathrm{MI}(U,V)}{\mathrm{H}(U)+\mathrm{H}(V)}. \tag{4}$$

This metric ranges from 0 (no agreement) to 1 (perfect agreement) and is symmetric with respect to label permutations.

A.4 Plots - 1

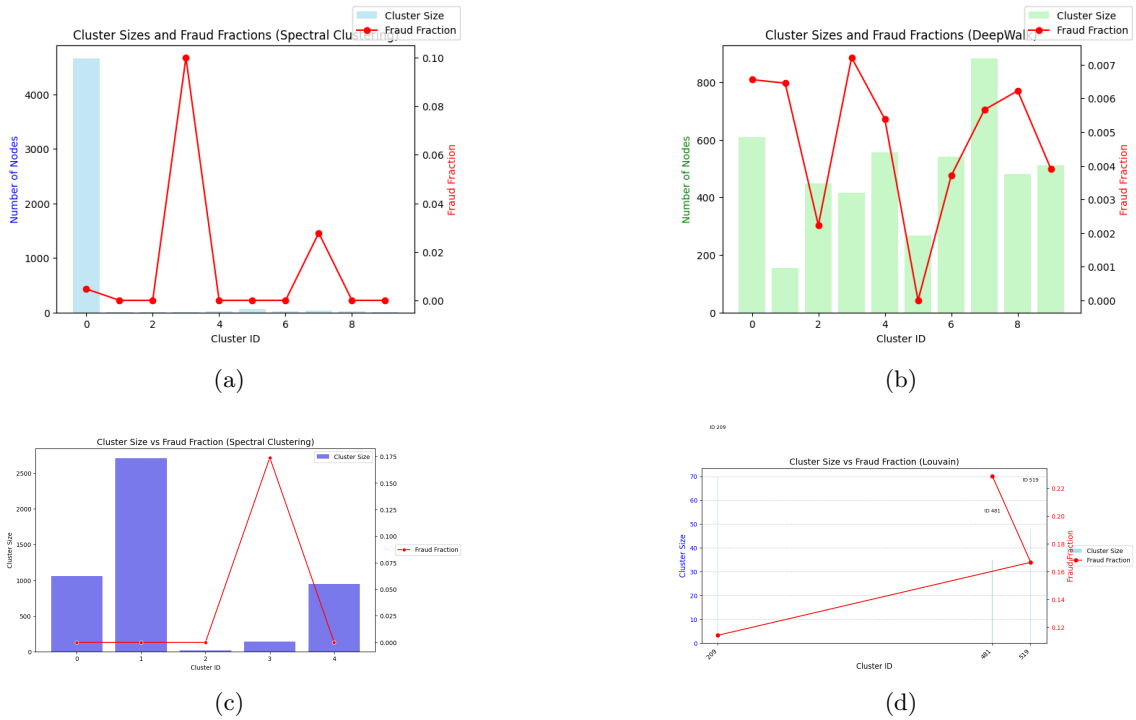

Figure 1: Comparative analysis of clustering methods for fraud detection in Reddit reply graphs, (a) and (b) Compare spectral clustering vs. DeepWalk for identifying small fraudulent communities. (c) and (d) Compare spectral clustering with node feature augmentation vs. standard Louvain for improved fraud detection.

A.5 Plots - 2

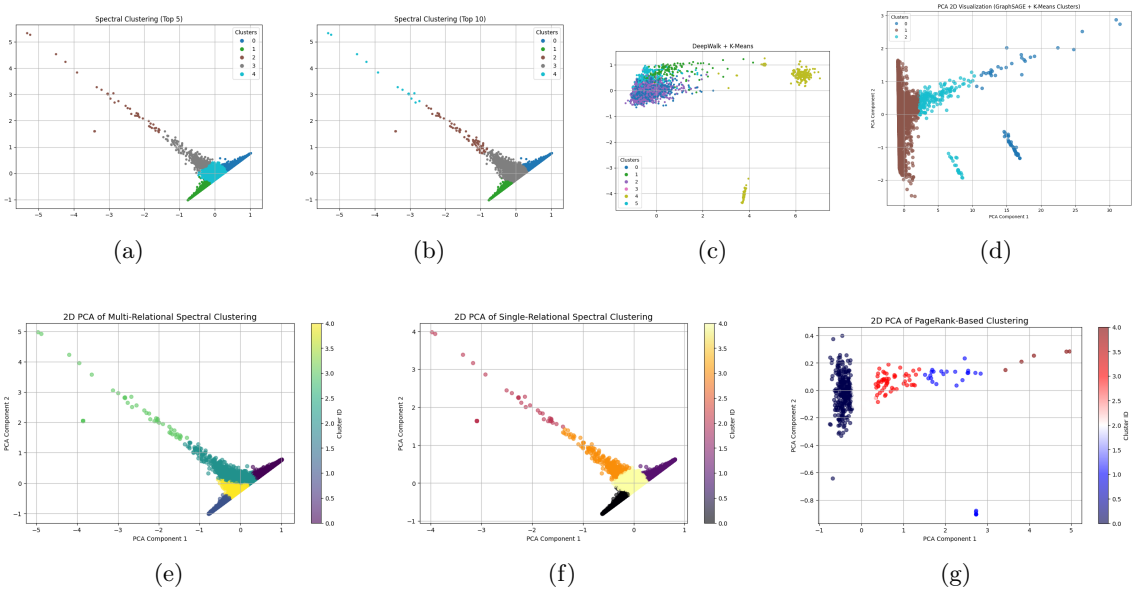

Figure 2: Visualization of 2D PCA clustering results for fraud detection: (a) and (b) show Spectral Clustering of Top 5, Top 10, (c) and (d) display K- Means of Deepwalk and Graphsage while (e)-(g) illustrate Multi-Relational, Single-Relational and Pagerank Based Clustering.