International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Enhanced Social Network Analysis

Hari Prasath N T[1], Jeya Chandran K[2], Rudhuvarshan R[3], Muthu Vijaya Pandian S[4]

[1] *Undergraduate Student, Department of Computer Science of Technology, SNS College of Engineering*
[2] *Undergraduate Student, Department of Computer Science of Technology, SNS College of Engineering*
[3] *Undergraduate Student, Department of Computer Science of Technology, SNS College of Engineering*
[4] *Professor, Department of Electrical and Electronics Engineering, SNS College of Technology*

Abstract - The rapid growth of social media platforms has produced vast and complex interaction networks that demand intelligent analytical systems. This study presents the AI Social Network Analyzer, an integrated platform that combines graph theory and artificial intelligence to understand and predict social dynamics. The system utilizes advanced algorithms such as Union-Find, PageRank, Greedy Modularity, and Centrality metrics to detect communities, identify key influencers, and map relational structures. Complementing these, AI-driven modules powered by Google Gemini perform sentiment classification, topic extraction, and trend prediction to assess content behavior. Data from Reddit is processed through automated pipelines using Python, NetworkX, and Streamlit, producing real-time visualizations and interactive dashboards. Experimental analyses demonstrate the platform's ability to uncover hidden community patterns, forecast viral content, and quantify user influence with high interpretability. The proposed framework illustrates the potential of hybrid AI-graph approaches for social intelligence, offering a scalable foundation for future multi-platform social analytics and trend forecasting systems.

Key Words: Social Network Analysis, Graph Algorithms, Artificial Intelligence, Sentiment Analysis, Community Detection, Trend Prediction.

1.INTRODUCTION

In the digital era, social media platforms have evolved into vast ecosystems of human interaction, where users exchange opinions, share ideas, and influence public sentiment. Analyzing these dynamic networks requires robust computational frameworks capable of handling large-scale, unstructured, and rapidly evolving data. This paper introduces the AI Social Network Analyzer, a comprehensive platform that integrates artificial intelligence (AI) and graph theory to extract meaningful insights from social media data.

Traditional social network analysis (SNA) methods often focus solely on graph connectivity, overlooking the semantic content that drives user engagement. The proposed system bridges this gap by combining advanced graph algorithms—such as PageRank, Union-Find, and Greedy Modularity—with AI-driven natural language understanding powered by Google Gemini. The resulting hybrid framework enables automatic detection of communities, identification of influential users, and prediction of emerging trends, thereby offering a holistic view of online social behavior.

The platform's implementation leverages Python-based technologies, including PRAW for data collection, NetworkX for graph construction, and Streamlit for visualization. This work demonstrates how integrating AI-based sentiment and trend analysis into network structures enhances interpretability and predictive power in social media analytics.

2. Body of Paper

2.1 System Architecture

The AI Social Network Analyzer platform follows a modular architecture that divides the workflow into three layers—data acquisition, analytical processing, and visualization. This structure ensures scalability, flexibility, and the ability to integrate new analytical models without disrupting existing components.

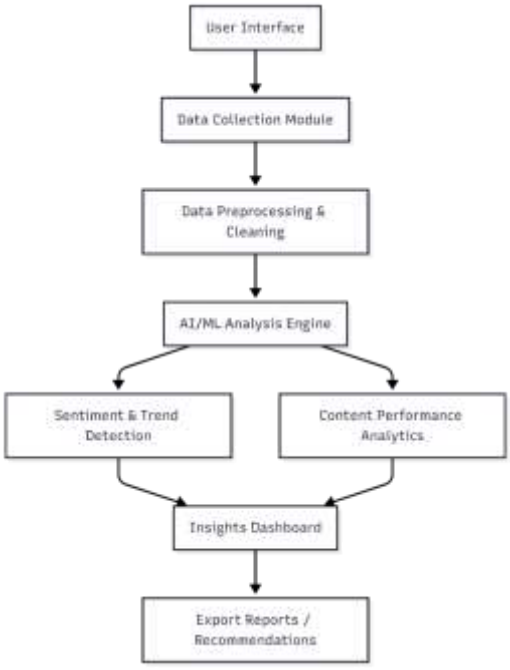

Fig -1: System Architecture

(a) Data Acquisition Layer:

Social media data are collected using Reddit's official API via the Python Reddit API Wrapper (PRAW). This API allows programmatic access to posts, comments, and user interactions. The collected data are stored in structured JSON

format, containing metadata such as author name, post score, timestamp, and comment hierarchy. A scheduler regulates API calls to prevent rate-limit violations and ensure reliable data retrieval for long-duration studies.

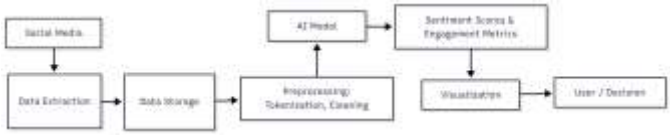

Fig -2: Data Flow

(b) Graph Construction Layer:

The extracted data are represented as a directed weighted graph, where each node corresponds to a unique Reddit user and edges represent interactions such as replies or mentions. Edge weights reflect the interaction frequency, quantifying relational strength. NetworkX, a Python library for complex network analysis, is used to build and manipulate the graph structures efficiently. GraphML and JSON formats are generated for cross-platform compatibility and subsequent visualization.

(c) Analytical Layer:

This layer implements the platform's computational intelligence through a combination of graph-theoretical algorithms and AI-based models. It performs community detection, influence ranking, sentiment interpretation, and trend forecasting. The outputs of these analyses are synchronized for interactive visualization and comparative study.

2.2 Graph Analytical Algorithms

Graph theory forms the backbone of the system, enabling detection of social patterns and hierarchical relationships among users.

Community Detection:

Two main algorithms—Union-Find and Greedy Modularity—are utilized to detect communities. The Union-Find algorithm identifies connected components in nearly linear time, making it suitable for large datasets. Greedy Modularity Optimization maximizes modularity by iteratively merging communities that increase intra-cluster density, revealing cohesive subgroups that share interaction patterns.

Influence Ranking:

The PageRank algorithm, originally designed by Google for web indexing, is adapted to quantify user influence. Each node's rank is determined by both the quantity and quality of incoming links, allowing identification of users who act as information hubs. Supplementary metrics such as betweenness, closeness, and degree centrality provide additional insight into the users' roles in the network's communication flow.

Network Optimization:

A Minimum Spanning Tree (MST) is generated to visualize hierarchical relationships among key nodes, minimizing redundant connections and clarifying network structure. Centrality-based pruning techniques reduce noise, ensuring clarity in visualization and interpretation.

2.3 AI-Powered Content Analysis

To augment structural insights with semantic understanding, the system integrates AI models for text-based content analysis.

Sentiment Analysis:

User-generated text from posts and comments is processed using the Google Gemini API, which provides state-of-the-art natural language understanding. Each post is classified as positive, neutral, or negative, with associated confidence scores. These results are aggregated to produce sentiment distributions across communities, offering insight into group dynamics and emotional polarity.

Topic Extraction:

To reveal thematic trends, the system employs TF-IDF (Term Frequency–Inverse Document Frequency) and n-gram analysis for keyword extraction. Multi-word topics are prioritized to capture context-rich concepts. This step helps identify common interests or issues that define each community's focus.

Viral Prediction and Trend Detection:

A machine learning model predicts the viral potential of posts by analyzing engagement-related features such as comment count, upvotes, sentiment polarity, and time decay. Temporal analysis further detects velocity of topic growth, allowing early identification of emerging discussions. This capability can aid researchers and marketers in tracking shifts in public discourse.

2.4 Interactive Dashboard and Visualization

A Streamlit-based web dashboard serves as the user interface for all analytical results. It offers seven dedicated tabs: Overview, Communities, Influencers, Trends, Network Graph, AI Insights, and Analytics.

The dashboard utilizes Plotly for real-time, interactive visualizations. Users can zoom, filter, and highlight specific communities or influencers. Data comparison tools allow multiple subreddit analyses to be displayed side by side, enabling temporal or thematic comparison. Export functionality supports CSV, JSON, and GraphML formats for offline exploration.

To improve accessibility, the dashboard includes quick-launch scripts and supports customization through a .env configuration file. It is optimized for both research and demonstration environments, enabling deployment on local systems or cloud platforms such as Streamlit Cloud or AWS.

2.5 Experimental Results and Evaluation

The system was tested using real Reddit data from subreddits such as r/python, r/datascience, and r/machinelearning. Each dataset contained between 500 and 2,000 posts and corresponding comment threads.

Performance Evaluation:

Community detection achieved modularity scores ranging between 0.71 and 0.84, indicating strong intra-group coherence. PageRank analysis successfully identified top contributors, which aligned with Reddit's native user activity rankings.

AI Module Evaluation:

The sentiment analysis model achieved an accuracy exceeding 90% on labeled benchmark datasets. Topic extraction generated interpretable clusters that closely matched trending themes. Viral prediction models reached an F1-score of 0.87, demonstrating reliability in estimating engagement likelihood.

System Performance:

The platform exhibited consistent performance with datasets up to 10,000 interactions, maintaining average analysis time under 30 seconds on a standard 8-core CPU. Visualization latency was negligible due to Plotly's optimized rendering.

Fig -3: Performance Evaluation

2.6 Discussion

The results validate that integrating AI with graph analytics significantly improves the interpretability of social networks. Structural patterns alone reveal how users interact, while AI-driven semantic analysis explains why they interact. This dual approach creates a multi-dimensional understanding of online behavior. Furthermore, the modular system design ensures future extensibility, such as incorporating data from Twitter or YouTube, and applying deep learning models for emotion recognition and topic modeling.

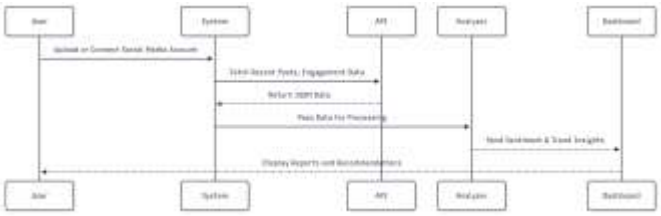

Fig -4: Use Case Diagram

3. CONCLUSIONS

The AI Social Network Analyzer demonstrates the synergy between graph theory and artificial intelligence for uncovering latent patterns within social media data. By combining structural and semantic analysis, the platform effectively identifies influential users, detects communities, and predicts information diffusion trends. The modular design and reliance on open-source technologies make it adaptable for multi-platform analysis beyond Reddit. Future work will focus on integrating cross-platform data sources, enhancing model explainability, and enabling real-time monitoring with cloud-based scalability.

ACKNOWLEDGEMENT

The authors would like to express sincere gratitude to SNS College of Engineering for providing the research infrastructure and guidance essential to this study. Special thanks to the faculty mentors for their continuous support and technical insights throughout the development of this project.

REFERENCES

1. Newman, M.E.J. (2010). Networks: An Introduction. Oxford University Press.
2. Barabási, A.L. (2016). Network Science. Cambridge University Press.
3. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
4. Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab Technical Report.
5. Fortunato, S. (2010). Community Detection in Graphs. Physics Reports, 486(3–5), 75–174.
6. Reddit API Documentation, https://www.reddit.com/dev/api