Traditional Graph-Based vs. AI-Driven Approaches in Social Network Analysis: A Comparative Study

Ayushi Chandrawal[1*] & Durgesh Dangi[2]

[1*,2] Academic Associate, Indian Institute of Management, Indore,
[1*]Email- ayushic@iimidr.ac.in, [2]durgeshd@iimidr.ac.in

HOW TO CITE:
Ayushi Chandrawal, Durgesh Dangi (2026). Traditional Graph-Based vs. AI-Driven Approaches in Social Network Analysis: A Comparative Study. International Journal of Special Education, 41(10s), 1043-1058

COPYRIGHT STATEMENT:
Copyright: © 2026 Authors.
Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:
Traditionally, graph-based data mining used social network analysis algorithms such as measurement and influence propagation. These models adopted an analytical and mathematical model that was well suited in analyzing static and semi static social networks (Tabassum et al., 2018). But with the adoption of neural networks and deep learning techniques (Qiu et al., 2018), newer approaches have been designed for analyzing complex and dynamic networks.

In this paper, a review of both the traditional as well as the artificial intelligence-based approaches is made with respect to their strengths, features, and complementary aspects. Various concepts such as community discovery, link prediction, influence evaluation, and anomaly detection have been discussed.

In brief, the use of graph-based data mining methods in social network analysis was based on conventional models that concentrated on measurement and influence detection. These approaches were designed using a mathematical and statistical approach that could be used to analyze both static and semi-static networks. However, with the current fast development in artificial intelligence and machine learning technology, new methods that are better suited for the analysis of complex data structures and dynamic networks have emerged. In this research, an analysis is conducted on traditional and artificial intelligence data mining methods and the advantages and dimensions of the two approaches will be highlighted.

Keywords: Social network analysis, Traditional graph-based methods, AI-driven approaches, Community detection, Link prediction, Influence analysis and anomaly detection.

1. INTRODUCTION

Nowadays, the data obtained from many of these practical applications is represented in the form of a network of interconnected entities (Tabassum et al., 2018). Social Network Analysis (SNA) has proven itself to be an important tool for analyzing communication and interaction and behavior within the modern world. Graph-based data mining is very useful, as it allows discovering trends in social networks, formation of communities, interconnections between people and influential individuals, and any anomalies. The analysis of such networks was based on graph-based methods which involved the application of mathematical and statistical models in order to discover interconnections, communities, and the mechanism of influence distribution. Although the traditional approach can help find much important information about

the analyzed social network, it cannot cope with dynamic networks and large data sets.

However, due to the appearance of artificial intelligence (AI) systems and machine learning (ML), it has become possible to develop the alternative computational methods (Zhou et al., 2020). AI allows processing huge amounts of dynamic data and predicting various phenomena in real-time. In this work, we discuss the comparison of the classic approach based on graphs with the one using AI and ML, and its applications in community detection, link prediction, influence diffusion, and anomaly detection (Daud et al., 2020).

Innovations in Artificial Intelligence and Machine Learning have opened up opportunities for Social Network Analysis. Artificial Intelligence-based methods such as Graph Neural Networks, deep neural networks, autoencoder and reinforcement learning have demonstrated their effectiveness in performing prediction tasks on graphs (Corso et al., 2024), they are highly scalable and adaptable in their nature (Aarthi et al., 2024). They can capture patterns from data right away, and work well with large and dynamic graphs (Hamilton, 2020). On the other hand, despite the strengths mentioned above, AI-based methods are usually computationally expensive (30), and hard to interpret in comparison to previous methods (Lundberg & Lee, 2017).

While a lot of research has been conducted on Social Network Analysis techniques such as community detection, link prediction, influence analysis, and anomaly detection, very little research has been done comparing the conventional graph-based approaches and the contemporary Artificial Intelligence approaches (Tabassum et al., 2018). In this paper, the two approaches will be compared based on four of the main Social Network Analysis tasks. To do this, a methodology will be adopted which considers several aspects of evaluating the benefits, drawbacks, and applicability of both approaches so as to facilitate a choice between them based on their use in research and practical cases. so that it becomes easier to choose among them according to their usage in practice.

2. METHODOLOGY

This paper performs a comparative literature review on the traditional graph-based approach and the advanced AI based approach for the analysis of social networks. This is done to understand the advantages, disadvantages, and applicability of both approaches for various purposes such as community detection, link prediction, influence assessment, and anomaly detection (Tabassum et al., 2018). For achieving this purpose, both approaches have been analyzed in terms of their performance based on the parameters of accuracy, scalability, interpretability, computational cost, and dynamic adaptation to changing networks.

2.1 Literature Collection

For selecting articles related to this research, four main databases were used, namely Google Scholar, Web of Science, Scopus, and the ACM Digital Library. Search queries were generated by putting together expressions related to various techniques (centrality methods, modularity, graph neural networks, deep learning, reinforcement learning, and graph-based methods), along with various aims (community discovery, link prediction, influence evaluation, and anomaly detection within social networks). The search took place during the time interval between 2005 and 2025, which means that both classical approaches and state-of-the-art developments in artificial intelligence were covered.

The inclusion criteria for the papers in the review process were as follows:

- These papers should suggest or study at least one technique for one of the four tasks of social network analysis.
- These papers should provide enough mathematics and conceptual knowledge about these techniques.
- These papers should be peer-reviewed or indexed journal papers available online.
- The language of the papers should be in English only.

2.2 Comparative Dimensions

This study evaluates both approaches over 29 dimensions; these dimensions fall into five broader groups.

- Performance
- Efficiency
- Usability
- Ability to adopt dynamic networks
- Characteristics of data

2.3 How the Comparison was Conducted

Firstly, we discussed and reviewed standard solutions for each Social Network Analysis (SNA) task from graph based traditional methods and AI driven methods in terms of working important of characteristics, principles and practical applications. These details were utilized to design a summary comparison table (Table 1) that evaluate both methods over several aspects such as accuracy, scalability, interpretability, computing cost and adaptability. This comparison was based on the findings of the reviewed literature, emphasizing the identification of the merits limitation and suitability of each method for various SNA tasks.

3. TRADITIONAL GRAPH-BASED APPROACH SOCIAL NETWORK ANALYSIS

3.1 Community Detection

Community detection is a concept in social network analysis that involves identifying groups or clusters within a network where nodes are connected to each other (Fortunato, 2010). Traditional community detection refers to the methods that were initially and widely used and depend on mathematical and graph-theoretical techniques to find communities in social networks. These methods' main objective is to divide a network into groups of nodes that are more closely connected within the group compared to the rest of the network. These communities usually represent real-world social groups, such as families, collaborative teams, or interest-based circles, e.g., on Facebook and Twitter. Traditional approaches do not make use of artificial intelligence or machine learning. Instead, they use graph properties and algorithms to identify the connections between nodes (people or objects). These are some of the most commonly used traditional techniques:

3.1.1 Graph Partitioning

This category of techniques divides each vertex into g predetermined parts, thereby reducing the connections among the detected units. The cutting size is defined as the number of vertices that cross the group clusters. If the number of clusters is unknown and an initial division with a small slice size is used, the solution obtained will be trivial. Community detection in networks, otherwise referred to as graph or network clustering (Zhang & Chen, 2018).

3.1.2 Hierarchical Clustering

The visualizations can represent a hierarchical organization, where each community can be composed of multiple small groups at various levels. In such cases, the multiple levels of community architecture in the data structure can be identified using hierarchical clustering methods. The vertex similarity measure is the foundation of hierarchical clustering algorithms. They do not need a fixed number and size of communities. Dendrograms are a more appropriate representation. Hierarchical clustering

3.1.3 Partitional Clustering

The network nodes are classified by minimizing or optimizing their loss product based on the proximity of (assumed) clusters. The following is a list of some partitional clustering methods and commonly used loss functions:

• Minimal K-clustering (Naderi et al., 2024)

• K-clustering sum

• K-center (Bateni et al., 2021)

• K-median (Fomin et al., 2024)

3.1.4 Spectral Clustering

This encompasses all methodologies employed to partition the graph into groups by implementing the eigenvectors of the provided data matrices. It

transforms a designated set of items into a collection of points in multiple-dimensional space, with parameters determined by the components of the eigenvectors. This process uncovers hidden properties of the actual dataset; therefore, the spectral clustering technique is useful for combining information that cannot be effectively clustered using k-means directly.

3.2 Link Prediction

Link identification is forecasting future associations across sets of unrelated links through using existing connections (Daud et al., 2020). Visualize the social networking platform Facebook using graphs, where each node represents a user from Facebook, Instagram, or Twitter, and an association denotes a relationship between two separate users or nodes. A connection may represent any type of interaction between two users, including posts, comments, reposts, and follows (Pócs et al., 2021). Link prediction helps uncover patterns and relationships in networks and is an essential tool for building smarter, more responsive systems in both research and real-world applications.

3.2.1 Similarity-Based Methods

Similarity methods offer a standard approach to link estimation, considered more accurate than alternative methods due to their simplicity and shorter processing time. In these approaches, a similarity score is computed for each unconnected node pair using specified link prediction metrics, including graph-based, content-driven, or a combination of both. The node pair with the higher similarity score is considered more likely to establish a connection in the future. Preliminary research into similarity-based methodologies primarily focused on graph-based metrics, which can be examined at both local and global levels. These parameters are divided into three types: Quasi-local Indices, Global Indices, and Local Indices (Kumar et al., 2020).

3.2.1.1 Common Neighbors

Within a specified network or graph, the total number of similar neighbors between two nodes (x and y) is expressed as the size of their intersection, indicating the shared portion of their corresponding neighborhoods (Kumar et al., 2020).

$$S(x, y) = |\Gamma(x) \cap \Gamma(y)|$$

3.2.1.2 Jaccard Index

This measurement method resembles Common Neighbor (Kumar et al., 2020). In addition, it simplifies the previously mentioned ratings, as represented below.

$$S(x, y) = |\Gamma(x) \cap \Gamma(y)| / |\Gamma(x) \cup \Gamma(y)|$$

3.2.1.3 Adamic/Adar index

The Adamic/Adar index, after slight modification by Liben-Nouvel et al., is a matrix of Adamic and Adar that calculates the level of similarity between two web pages based on common attributes and is employed in link prediction.

3.2.2 Statistical Model

Statistical models have a crucial role in understanding the structure and behavior of social networks. Such models employ probability theory and statistical inference to quantify patterns, predict relationships, and identify structures such as communities or influence paths.

3.2.2.1 Stochastic Block Models (SBM)

Hierarchical models are unlikely to represent the majority of networks. A simpler, more universal method for modeling such networks uses the block model, in which vertices are partitioned into blocks and groups. The probability of a connection between two vertices depends on the blocks they share. A new framework has been introduced that uses a stochastic block model to represent networks and identify missing and fraudulent links. The researchers measure the reliability of connectivity in an observation network, which is subsequently used to identify missing links (non-existent connections with higher reliability) and spurious connections (existent linkages with lower reliability).

3.2.2.2 Hierarchical Structure Model

The models described below assume that most real networks have hierarchically structured forms, with nodes aggregated into groups, which are, in turn, broken down into segments and so on over several levels. To construct an algorithm that statistically estimates parameters by statistical approaches, some sample work systematically imprints such a structure from network information. The link-generation probability is subsequently estimated using these factors.

3.3 Influence Analysis

Influence analysis of social network analysis (SNA) studies the extent to which a person or organization influences other entities in a network. Influence analysis is widely used to interpret information dissemination, attitude change, and decision-making across a broad range of areas, including marketing, public health, and politics.

3.3.1 Centrality Metrics

Centrality measures indicate the influence of units within a network. They are among the most widely adopted unit-level measures derived from network data (O'Malley & Marsden, 2008). A key idea in the study of social networks is centrality. From a conceptual standpoint, centrality quantifies a person's position as a key player in a social network. Network analysis and graph theory are often-utilized tools. Numerous centrality measures, degree centrality (Opsahl et al., 2010), closeness centrality (Borgatti & Everett, 2006), betweenness centrality (Zhang & Luo, 2017), eigenvector centrality (Rakaraddi & Pratama, 2021), and Katz centrality, are all proposed in studies for extensive usage in social influence analysis.

3.3.1.1 Degree Centrality

The level of centrality can be determined by the total count of linkages associated with a node, particularly the quantity of edges linked to it. In an oriented network, degree centrality is measured by two metrics: in-degree, the count of edges directed towards a node, and out-degree, the count of edges emanating from the node (Peng et al., 2018).

3.3.1.2 Betweenness Centrality

Betweenness centrality estimates the frequency with which a node serves as a conduit along the shortest path connecting two other nodes. Assuming that item transfer follows the shortest paths, a node with elevated betweenness centrality exerts a greater impact throughout its network. Betweenness centrality is widely applicable across fields such as biology, social networks, and transportation.

3.3.1.3 Closeness Centrality

The closeness of centrality determines the mean distance from a particular vertex to all the others across a network. Influence evaluation can serve as a metric for assessing each node's ability to propagate information throughout the network. A node's closeness centrality correlates positively with its advantageous position within the social network.

3.3.1.4 Eigenvector Centrality

Eigenvector centralization quantifies the influence of a particular node inside the network. Equivalent values for every node within a network are predicated on the assumption that linkages to high-scoring vertices are more likely to enhance their overall score than linkages to low-scoring nodes.

3.3.2 Influence Propagation Models

These models mimic the propagation of influence through a network.

3.3.2.1 Independent Cascade Model (ICM)

Each node that becomes activated (influenced) gets one opportunity to influence its neighbors. Influence propagates step by step, like a chain reaction.

3.3.2.2 Linear Threshold Model (LTM)

A node is influenced when a specified number (or percentage) of its neighbors are influenced. It also encompasses peer pressure or cumulative influence.

3.3.1 Page Rank Algorithm

Page and Brin developed the PageRank algorithm, which underlies Google's search engine. In PageRank, all hyperlinked pages are assigned numerical weights that estimate their importance. PageRank uses the web topology to indicate the value of any page. PageRank is so applicable that we could even use it for social networks, in the sense that the reciprocity relationship between social networks could be represented as links to microblogs, and nodes could even be treated as websites in the PageRank algorithm.

3.4 Anomaly Detection

An abnormality, an anomaly, involves an item of data whose activity markedly deviates from the standard pattern of the other data. Typical pattern of the remaining data. Such unusual trends or observations can also be called outliers, abnormalities, peculiarities, conflicting observations, contaminants, surprises, or exemptions, contingent upon their respective domains. Anomalies arise due to fraudulent activities, human mistakes, or system faults

Abnormal activity identification, also known as outlier detection, is a core problem in data mining that involves identifying aberrant patterns in a dataset. While other methods of data mining, such as categorization, grouping, and repeated pattern identification, determine common structures within a dataset, anomaly detection identifies a comparatively small number of elements that deviate from the normative behavior of the majority of data objects in the dataset (Jin et al., 2018).

3.4.1 Graph-based Anomaly Detection

Graph-based detection of abnormalities involves analyzing data in graph form and applying graph mining algorithms.

Graph-based identification of anomalies aims at recognizing abnormal entities such as edges, nodes, or subgraphs inside graph-structured data. Graph-based methods for detecting anomalies identify irregular structural components within graphs, including the emergence of unexpected nodes or edges, the absence of expected vertices or edges, or alterations in the properties of vertices or edges.

3.4.2 Distance-Based Methods

Distance-based abnormalities are termed "global anomalies." Typically, Euclidean or Mahalanobis distance serves as the basis for distance metrics.

The proximity of an object is analyzed depending on a certain distance criterion for each object. When the proximity of a data item o excludes a significant number of items across the total collection D, meaning that the specified neighbors have a limited number of components, while 'o' is considered an oddity. The explained approaches employ two universal parameters that are d and β

4. AI-DRIVEN TECHNIQUES OF SOCIAL NETWORK ANALYSIS

Emerging advancements in AI and ML introduced enhanced strategies that improve the limitations of traditional methods. Strategies including deep learning, neural networks, and probabilistic models, are applied in AI-based strategies to improve efficiency and accuracy.

4.1 Community Detection

Community detection is the method of discovering clusters of highly connected nodes within a social network. In traditional methods (Fortunato, 2010), algorithms like modularity optimization or hierarchical clustering were used. However, with the advent of Artificial Intelligence (AI), community detection has become more accurate, flexible, and scalable—especially in large-scale, dynamic networks. Utilizing the strength of AI can strongly improve the accuracy and value of community identification.

4.1.1 Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) (Corso et al., 2024) are strong, robust machine learning algorithms that process data directly in graph form. As regards community detection, it is possible to automatically discover the node representations that carry semantic meaning

through GNNs that use the information from the structure of the graph and node feature vectors. They outperform any other type of handcrafted features in detecting dense clusters of nodes and are very useful for managing large and complex network environments [Aarthi et al., 2024]. Through repeated combinations of the information of neighboring nodes, GNNs are capable of discovering the fine-grained structure of relational dependencies in social graphs (Fortunato, 2010).

4.1.2 Text and Topic Detection Technique Based on ML

Text and topic analysis using machine learning algorithms is an integral component in community detection in various social media applications. These approaches measure the semantic similarities among text content in order to discover latent communities that cannot easily be found in the structure alone. NLP tools and methodologies such as tokenizing, stop words removal, and lemmatization allow textual content to be transformed into numerical form by means of techniques such as TF-IDF, Word2Vec, or BERT. Topics within a dataset are extracted via topic modeling techniques including LDA and BER Topic, which enable user and document classification by topics of interest. This creates vectors of topics which can be then used by clustering algorithms like K-Means or DBSCAN to detect communities that have similarities within the themes of interest. Other techniques take the intersection between text similarity and interaction graph, making use of algorithms like Text-GCN or Graph SAGE for enhanced community discovery. Overall, topic- and text-based machine learning techniques provide a valuable complement or alternative to traditional structural community detection algorithms.

4.2 Link Prediction

4.2.1 Random Forest and Gradient Boosting

The Random Forest approach and gradient optimization are collaborative learning methods that combine multiple decision trees to improve forecast accuracy. They perform well in link prediction tasks by learning node attributes and structural features (Breiman, 2001).

Random Forest

The random forest method is a collaborative method that generates a collection of decision trees and integrates their forecasts to provide more accurate and stable predictions than any single tree. It works by using a technique known as bagging (bootstrap aggregating), in which a number of decision trees are learned separately, each from a distinct, unique portion of the data. Its overall output can be computed with combining (in regression) or voting (in classification) the predictions of the separate trees.

How Random Forest Works:

Bootstrap Sampling: Random samples of the data are generated through sampling with replacement.

Tree Construction: A tree is constructed on every subset of data, with a randomly chosen set of attributes at every division to ensure the diversity of the trees.

Aggregation: For classification, predictions are determined by a plurality vote across all trees. For the regression analysis, the predicted values from each tree are averaged.

Mathematical Representation:

Let T1, T2, …, Tn be the input trees in the forest, and let x be the input data instance. The prediction for classification is:

= mode(T1(x),T2(x),…,Tn(x))

For regression, the prediction is:

Gradient Boosting

Gradient boosting is a strong ensemble method, too, but rather than constructing independent trees, it constructs them sequentially. The new tree at each step corrects the errors of the past trees. Gradient boosting optimizes a loss function. In every iteration, the technique incorporates a fresh tree to reduce residual errors, which represent the discrepancy between the current predictions and the actual labels.

How Gradient Boosting Works:

Initial Prediction: Begin with a basic prediction, often the mean for regression or the most common class for classification.

Residual Calculation: Calculate residuals (errors) between actual values and predicted ones.

Tree Building: An additional tree is trained to predict the residuals of the current model.

Update: The forecasts are updated by incorporating the new tree's forecasts, scaled by a learning rate to regulate each tree's contribution.

Repeat: The procedure continues until the algorithm reaches a given number of trees or attains the required accuracy level.

Mathematical Representation:

Let F(x) be the existing model, and r indicates the residual of the Ith data point. The residual is computed as

In each iteration, a novel tree h(x) be constructed to optimize residuals. The new model becomes:

$F(x) \leftarrow F(x) + \eta h(x)$

Where η is the learning rate process, which determines the impact of each tree.

4.2.2 Autoencoders

Autoencoders, especially Graph Autoencoders (GAEs) and Variational Graph Autoencoders (VGAEs), are neural network architectures well-suited for unsupervised learning on graph data. They learn low-dimensional node representations that summarize the network's underlying structure, which proves useful for link prediction tasks (Fu et al., 2023).

Variational Graph Auto-encoder

Encouraged by advances in deep learning, a large number of VGAE-based methods for link prediction have been proposed in recent years. Drawing on the variational autoencoder (VAE), the most prominent effort is the variational graph auto-encoder (VGAE), a comprehensive architecture that utilizes a graphical convolutional network (GCN) encoder and a product in the decoder.

Disentangled Graph Encoder

The disentangled graph encoder (Fan & Gao, 2024) is designed to learn disentangled node embeddings, comprising three major modules: feature projection, factor-aware aggregation, and concatenation.

Recently, the tremendous success of disentangled representation learning in the computer vision community has drawn much attention to graph-structured data. Graph disentangled representation learning seeks to learn low-dimensional embeddings by factorizing the representations of entities (nodes, edges, subgraphs) in a graph into several explainable components, enabling embeddings to generalize to unseen nodes and be useful for downstream tasks.

4.3 Influence Analysis

4.3.1 Deep Neural Network (DNN)

The Deep Neural Network (DNN) (Qiu et al., 2018) structure plays a significant role in AI-driven influence analysis in social network analysis (SNA). It is more advanced than conventional graph metrics because it learns autonomously from intrinsic influence patterns in user behavior, interactions, and shared content within the network. DNNs can model complex, non-linear user dependence, enabling them to capture subtle influence flows that are often neglected by rule- or centrality-based approaches. By historical behavior patterns, DNNs will be able to predict future user behavior, i.e., content adoption, event attendance, or product purchases, as a function of being a peer-influenced behavior. This would be extremely useful for targeted advertising, tracking the diffusion of disinformation, and research on opinion dynamics in online forums. DNN relies on network topology and user attributes to discover which user attributes and relations most influence the spread Furthermore, its scalability and versatility make it well-suited for examining large, dynamic social networks in near real-time. Thus, DNNs offer a powerful, data-driven platform for detecting and leveraging social

influence in contemporary information-rich settings.

4.3.2 Deep Reinforcement Learning for Influence Maximization (DREIM)

Deep Reinforcement Learning (Li, 2023) for Influencer Identification in Big Networks proposes a new approach based on Deep Reinforcement Learning (DRL) to address the Influence Maximization challenge within social networks (Tan et al., 2019). Unlike the typical greedy or heuristic approaches, DREIM finds the optimal methods for selecting influential nodes by exploring the network. The approach frames the influencer selection problem as a sequential decision problem, where each node selection (action) determines subsequent influence propagation. With DRL, the agent learns to adapt to network dynamics and maximize long-term influence coverage. DREIM outperforms baseline methods for choosing influential and diverse influencers. It also scales well to massive and large-scale networks and is thus suitable for real-world social media or marketing use cases.

4.4 Anomaly Detection

4.4.1 Anomaly Detection Using GAN

As the algorithm that discriminates is trained to differentiate between conventional and produced data distributions, it then uses reconstruction error to compute the anomaly score. With the convergence of the approach, the discrimination algorithm should have gained the ability to reconstruct regular data from the original distribution X. The discriminative algorithm generally has the capacity to restore the original data after passing through its generator. However, the algorithm of discrimination might not be able to recover erroneous data. Then, the discriminator, after it has ascertained that the data of the variance generators was not following the standard distribution of data, will raise the reconstitution error. The formula used to calculate the suspicious value (, i) for every input, x, will be as follows:

Then set a level φ at which irregularities have been created. An evaluation input is considered unusual if φ>0. After optimizing how to train performance, their proposed methodology can be applied in order to determine all nodes in the test dataset are subsequently evaluated for their irregularity degree using their recovery error, which was computed through assistance from the discriminator loss.

4.4.2 Hybrid AI Models

Belavagi proposed a predictive model for intrusion detection using machine learning methods, including Gaussian Naive Bayes, Random Forest, Logistic Regression, and Support Vector Machines (SVMs). The hybrid detection method proved more effective than a single classifier (Rahman et al., 2021).

We develop an integrated method for anomaly identification in social networks. The algorithm we use is constructed by sequentially integrating various machine-learning techniques:

1) Decision tree (C5.0) (Rahi et al., 2024)

2) Support Vector Machine (SVM) (Kecman, 2005)

3) Naive Bayesian Classifier (NBC) (Zheng & Li, 2024)

Initially, this C5.0 decision tree serves to categorize those who use social networks under two classifications:

Abnormal users, along with typical members. This C5.0 decision tree was selected for the following purposes:

1) The C5.0 technique utilizes significantly smaller amounts of memory compared to conventional decision tree methods.

2) C5.0 minimizes inaccuracy through cutting and discarding unnecessary or irrelevant features associated with labeling classes in extensive datasets.

A Support Vector Machine (SVM) is applied in the subsequent phase to categorize atypical consumers as highly satisfied or dissatisfied. At the last stage, NBC will be used to categorize dissatisfied consumers into dangerous and non-dangerous groups. Individuals.

The investigation's results indicate that the suggested approach performs admirably in identifying abnormal as well as depressive users within social media networks.

Integrating k-means clustering as a technique along with the C5.0 decision tree and SVM will mitigate all inherent issues in in classification.

1) The matter of class preponderance and

2) The matter of involuntary allocation.

Decision Tree (c5.0)

The Decision Tree (C5) is a widely used classification technique. C5.0 produces more compact decision trees with fewer rules than C4.5. The precision of C5.0 is superior to that of C4.5. C5.0 inherently facilitates the elimination of useless properties. The C5.0 technique offers superior memory efficiency, cross-validation, knowledge discovery, reduced defect trimming, and enhanced feature selection capabilities compared to C4.5.

Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a supervised learning model used for regression and classification. SVM is primarily used for classification, differentiating between two distinct groups in a test dataset. The aim of SVM is to identify the optimal hyperplane for effectively classifying the given data. The support vectors, which are crucial in constructing the model, remain unaffected by additional data. In SVM, the Gaussian kernel is employed for both linear and non-linear spaces.

Naive Bayesian Classifier (NBC)

The Naive Bayesian approach identified individuals in the social network by computing the posterior probabilities. The naive Bayesian approach is used for labeling dissatisfied consumers as depressed or non-suicidal depending on self-harming behavior. The mathematical representation of Naive Bayes and the principles of Naive Bayes and the principles of Naive Bayesian inference are both articulated in an especially straightforward manner:

Table 01: Comparative Analysis of Traditional and AI-Driven Methods in Social Network Analysis

Different strengths and weaknesses that characterize each paradigm of SNA vary based on the purpose of analysis. Conventional techniques are often valued due to their interpretability and relative simplicity, while the use of artificial intelligence in social network analysis ensures higher accuracy, scalability, and flexibility; however, conventional techniques tend to be less computationally intensive than those based on AI. In order to analyze the strengths and weaknesses in more detail, the research explores the use of each paradigm on four key tasks in social network analysis: community detection, link prediction, impact assessment, and anomaly detection. The findings of the comparative study are presented in Table 1.

S. No	Aspect	Traditional Methods	AI-Driven Methods
1	Scalability	It resists well on small or mid-size networks but performs poorly on very large networks or with recent data.	Tuned to deal with huge, complex, and dynamically evolving networks efficiently through parallel processing and prudent system design (e.g., graph neural networks) (Aarthi et al., 2024).
2	Accuracy	Use predefined formulas or rules (such as centrality or modularity), which may make it difficult to see beyond surface-level patterns.	Learn implicit and nuanced relationships from data directly, typically making more accurate predictions (Tan et al., 2019).
3	Dynamic Adaptability	Mostly designed for static networks. If the network is dynamic, the analysis would typically need to be recreated.	Can see and respond to network changes over time, typically in real-time, using models such as RNNs, GNNs, or reinforcement learning (Hamilton, 2020).
4	Computational Cost	Efficient computation for small networks; simple algorithms like BFS have a polynomial-time complexity.	Requires more computing power, especially training, and typically demands GPUs or cloud-based setups (Zhou et al., 2020).

5	Interpretability	Simple to interpret—results can be described using intuitive metrics such as node degree or clustering.	Less straightforward to interpret ("black box" models), but something like SHAP or LIME can be used to explain what the model is doing (Lundberg & Lee, 2017).
6	Anomaly Detection	Uses simple rule-based methods or statistical methods to identify outliers.	Advanced models (e.g., autoencoders or hybrid AI (Rahman et al., 2021) methods) to recognize hidden or weak anomalies (Zhang, 2025).
7	Link Prediction	It holds only for observable network structures, as measured by similarity metrics such as Common Neighbors or the Jaccard Index (Kumar et al., 2020).	They use sophisticated methods such as GNNs (Breiman, 2001), node embeddings, and supervised learnin to identify latent or future links (Tahsien et al., 2020).
8	Community Detection	Average algorithms are Girvan-Newman or modularity-based clustering (Fortunato, 2010).	This approach uses deep graph and reinforcement learning to detect communities more dynamically and accurately (Rashnodi et al., 2024).
9	Handling Time-Based Data	Requires manual configuration to monitor changes over time (e.g., time snapshots); this can become less nuanced.	Applies time-aware models, such as Temporal GNNs, LSTMs, or attention-based models, to better capture how networks change over time (Kazemi et al., 2020).
10	Data Requirements	Can function even when only the network's structure is present; it does not require additional data.	Performs well with a lot of labeled data and high node/edge knowledge.
11	Robustness to Errors	Degrades when data is missing or noisy.	Used for dealing with missing data or noise by techniques like robust training and regularization.
12	Flexibility Across Tasks	Flexibility Across Tasks: Typically trained on a particular task, such as node ranking or community detection.	Artificial intelligence models can be optimized or trained to execute a set of tasks—such as prediction, classification, or ranking—within a single network (Wu et al., 2021).
13	Tools & Software	Utilizing software Gephi, Pajek, or NetworkX to graph and analyze.	High-level AI libraries such as PyTorch Geometric, DGL, TensorFlow, or Stellar Graph-based learning.
14	Human Effort Required	Human Effort Required It requires humans to manually adjust parameters and select the appropriate models.	Able to minimize manual tuning via AutoML or auto model selection (AutoML Solutions - Train Models without ML Expertise, n.d.) (He et al., 2021).
15	Real-Time Usage	Not ideal for real-time or stream data.	Is to be used for real-time monitoring, like fraud detection or observation of social media trends, based on online learning methods (Su et al., 2020).
16	Model Generalization	Tend to be task-specific and will not generalize outside the original design context.	Can generalize between different tasks (e.g., transfer learning from one network to another) (Kumar et al., 2023).
17	Learning Capability	Do not learn from data; learn by explicit programming and mathematical rules.	Self-improve constantly through learning patterns and feedback from training.
18	Data Fusion	Difficult to integrate multi-modal data (e.g., text, image, interaction logs).	combines multi-input models, attention mechanisms, and data types in a simple way.
19	Feature Engineering	involves hand-picking and building features.	Respect the latest trends in deep learning to automate feature extraction (Goodfellow et al., 2016).
20	Anomaly Resilience	An anomaly: adaptability, strict rules, or thresholds with a tendency to produce false positives.	Uses learned and probabilistic representations to identify subtle, hidden anomalies (Zhang, 2025).
21	Prediction Speed (after training)	Fast on small networks; performance degrades on complex systems.	Slow to train, but inference is extremely fast and scalable thereafter (Zhou et al., 2020).
22	Visualization Support	Direct and intuitive visualizations with centrality, clusters, etc.	Needs sophisticated visualization of learned embeddings or attention weights (Hamilton, 2020).
23	Noise Sensitivity	Very sensitive to noisy/incomplete links/nodes.	Can handle and adapt to noisy/incomplete data with strong training (Goodfellow et al., 2016).

24	Context Awareness	Does not see context (e.g., user activity, temporal patterns).	Embraces context in learning models for further research (Tan et al., 2019).
25	Personalization	Difficult to personalize results based on user or group behavior.	Allows personalized models on learned user interactions (e.g., recommendation systems) (Aarthi et al., 2024).
26	Reproducibility	Very high reproducibility based on deterministic algorithms.	More challenging to reproduce due to stochastic training procedures and hyperparameter sensitivity (Goodfellow et al., 2016).
27	Development Tools	Comprehensively available and mature analytical toolboxes.	Emerging ecosystem with high-frequency updates (e.g., PyG, TensorFlow GNN, DGL (Wu et al., 2021)
28	Training Data Needs	Can operate on small or synthetic data sets.	Requires large-scale labeled or semi-supervised data for optimal performance (Tan et al., 2019).
29	Performance Metrics	In terms of structural graph measures like modularity, density, and shortest paths (Fortunato, 2010).	Performance is assessed using precision, recall, AUC, and task-specific loss functions (Wu et al., 2021).

There are some distinctions revealed by the comparative research of traditional graph-based techniques vs. AI-based approaches in the case of social network analysis. It is noteworthy that in this case, there is no correlation between the level of interpretability and the accuracy of predictions. The literature indicates that the AI-based models and techniques, machine learning and Graph Neural Networks-based approaches outperform the traditional ones when it comes to such tasks as link prediction, community detection, or anomalous behaviour recognition.

However, it should be noted that in this case, the AI-based approach represents a black-box model that makes it rather challenging to comprehend the process. In comparison, traditional graph-based approaches provide high-level interpretability and transparency as the outcomes are connected with the characteristics of the graph under consideration.

In addition to the above discoveries, another significant discovery was made regarding scalability and adaptability. The classic graph-based approaches function best within the medium range of size networks using limited computing power, while they do not function well when dealing with large and highly-complex social networks.AI approaches perform better when handling big data. So, they are more adapted to the social network environments that are constantly growing and changing.

The data also shows that neither traditional nor AI-driven methods are always better for all evaluation parameters. The success of a method depends on things like the size of the network, the amount of data, the computational resources and the goals of the investigation.

AI-driven methods offer advantages like accuracy, scalability and adaptability while traditional methods are better in terms of interpretability, reproducibility and simplicity of implementation. Recently research has sparked interest in solutions that combine graph-theoretic characteristics with machine learning approaches. These methodologies can use the benefits of both AI-driven methods and are a potential path for future research, in social network analysis.

5. CHALLENGES AND FUTURE DIRECTIONS

Social network analysis is advancing in highly impactful and scalable analysis through its transformation emerging with the highly AI based technological developments. If technological advancements are one side of a coin, then its associated challenges are the other side of the same coin. Same goes with the AI Driven advancements in the SNA. In this scenario few challenges emerged that restricted the full utilisation of those AI based advancements in social network analysis. When these AI based analysis models are designed, trained and operationalized, computational complexity arises in the process as they require structured deep learning models for them to

provide efficient analysis results, such deep learning models are highly dependent on the large networks (Zhou et al., 2020). This inclusion of deep learning and large networks in constructing further require ample resources including computational power, memory and specialized equipment such as GPU or TPU (Aarthi et al., 2024). This computational complexity challenges the ease of access to such advanced AI based SNA for the researchers who were operating with limited resources.

Also, these AI based SNA models among such models, deep neural networks particularly are "black boxes", these inclusions make it difficult to understand or decode the process that AI model has undertaken for providing the analysis results or decisions (Lundberg & Lee, 2017). This uncomprehends process results in trust and accountability complexities due to lack of transparency. These can be exemplified with cases of fraud detection or monitoring of any misleading information and traceability of the fraud and misrepresentations (Zhang, 2025).

Along with the challenges discussed above, there are other issues that are related to the privacy concerns and the bias linked to its relative users. AI based models heavily based on data supplied by its users, as they depend on user data, it generates the possibilities of developing societal bias issues due to which they are bound to either unintentionally compromise user confidentiality (36). Since these AI based models provide logical assistance for the advanced SNA, this particular challenge can be addressed and resolved by giving proper considerations to the user's concerns. While adding justifiable and responsible approaches to the user consent along with ethical deployment of these models can resolve the challenges which gives fullest utilisation of advanced AI tech without getting interrupted by associated challenges concerning privacy (Goodfellow et al., 2016).

While introducing the AI based technologies and advancements in the SNA domain one of the other challenges they bring in along with the benefits and ease of analysis, is the training and updating such AI driven models to keep them up with the continuing advancements and developments. These AI based models are effective and efficient, only and only if they are trained and modified in due course of time. But concerns arise as such training data are difficult to access because either they are unavailable in real world or even if they are available, they are held in hands of the monotony that makes their accessibility too expensive to be afforded by the limited resource users (Tan et al., 2019).

Thus, for the full utilisation of the AI based advancements in social network analysis can be achieved only by resolving the such discussed challenges through future research to find the possible efficient solutions. In this aspect a new future updated AI model for SNA can be designed while integrating such efficient and advanced solutions to the above challenges.

With this new AI model for SNA will provide advanced analysis with reduced computational cost, more efficient architecture with explained procedures of arriving on decisions or providing the results (Lundberg & Lee, 2017), and incorporate the ethical consideration for user's consent and user privacy protocols in the model design (Goodfellow et al., 2016). For developing such modified AI model of SNA expertise of professionals from different core areas or subject field will be needed such core fields include computer science for technological developments (Kumar et al., 2023), sociology for user attributes and human interaction with technology, governance and ethics for user consent and privacy aspect along with other expertise as and when required throughout model designing and development process.

6. CONCLUSION

SNA, as an advanced method for understanding and analysing the communication analogies along with behaviours and interactions depicted by the people in present physical and virtual environments. Being with traditional graph-based SNA methods lacking in various aspects such as advanced validated accuracy, vast data and with efficient time managed results. Introduction of AI in SNA brought as a solution to such limitations of the traditional SNA methods.

While discussing the traditional methods and AI based SNA methods, this papers study comprehensively discuss comparative survey of both methods where traditional methods include statistical model, graph theory and similarity-measures that as founding stone provide knowledge for four core tasks of SNA uncovering the network analysis stream and structures, works on detecting communities, link predictions, influence analysis and predicting the outliers. The traditional methods were idealized for their interpretability, uncomplicated and ease of applications and direct applications, especially in case of small or specified network analysis.

These traditionally well-versed SNA methods became less effective in the present time which required more effective accurate results with highly strict timelines that too using the vast complex datasets and networks. This brought up the requirement of the AI driven advanced techniques and machine learning modules to be embedded in the SNA that holds considerable strengths in this context. Such integration of AI and machine learning modelling SNA will allow users to simulate high-dimensional, non-linear, and complex relationships. Such a complex analysis cannot be supported by conventional methods. Following these methods such as Graph neural networks (GNNs), transformer model, auto-encoders, and reinforcements leanings have been developed with embedded improved assured reliability, scalability, and adaptability in performing the basic tasks including detecting the unidentifiable, influence analysis, predicting the associations and detection and identification of the communities based on associations.

Though the advancements in SNA brought in by AI and machine learning provided the solutions to the limitation of the traditional methods and present requirements of the validated reliable and timely efficient results, it brings in some inbuilt limitations associated to its complex design, high processing based advanced system requirements, increased cost, and obscure procedure disclosure about the arrived conclusion and prediction made through the AI based SNA model. These AI based SNA also have ethical concerns as limitations regarding the data uses and its confidentiality. For building the optimal solution, a symbiotic combination of the merits of both traditional and AI based approaches to SNA can be utilized.

The above discussions of the study provide a guiding comprehension for the concerned researchers and practitioners in the process whether it's related to selecting the right model for the optimum utilisation as per their customised requirements, or if its concern to designing a customised SNA model.

REFERENCES

Tabassum, S., Pereira, F. S., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(5), e1256. https://doi.org/10.1002/widm.1256

Zhang, M., & Chen, Y. (2018). Link prediction based on graph neural networks. Advances in Neural Information Processing Systems, 31, 5165–5175.

Naderi, I., Rezapour, M., & Salavatipour, M. R. (2024). Approximation schemes for min-sum k-clustering. Discrete Optimization, 54, 100860. https://doi.org/10.1016/j.disopt.2024.100860

Bateni, M., Esfandiari, H., Fischer, M., & Mirrokni, V. (2021). Extreme k-center clustering. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6775–6783.

Fomin, F., Golovach, P., Inamdar, T., Saurabh, S., & Zehavi, M. (2024). Hybrid k-clustering: Blending k-median and k-center. ACM Transactions on Computation Theory. Advance online publication. https://doi.org/10.1145/3680534

Daud, N. N., Ab Hamid, S. H., Saadoon, M., Sahran, F., & Anuar, N. B. (2020). Applications of link prediction in social networks: A review. Journal of Network and Computer Applications, 166, 102716. https://doi.org/10.1016/j.jnca.2020.102716

Pócs, D., Adamovits, O., Watti, J., Kovács, R., & Kelemen, O. (2021). Facebook users' interactions, organic reach, and engagement in a smoking cessation intervention: Content analysis. Journal of Medical Internet Research, 23(6), e27853. https://doi.org/10.2196/27853

Kumar, A., Singh, S. S., Singh, K., & Biswas, B. (2020). Link prediction techniques, applications, and performance: A survey. Physica A: Statistical Mechanics and its Applications, 553, 124289. https://doi.org/10.1016/j.physa.2020.124289

O'Malley, A. J., & Marsden, P. V. (2008). The analysis of social networks. Health Services and Outcomes Research Methodology, 8(4), 222–269. https://doi.org/10.1007/s10742-008-0041-z

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. Social Networks, 28(4), 466–484. https://doi.org/10.1016/j.socnet.2005.11.005

Zhang, J., & Luo, Y. (2017). Degree centrality, betweenness centrality, and closeness centrality in social network. In Proceedings of the 2nd International Conference on Modelling, Simulation and Applied Mathematics (pp. 300–303). Atlantis Press. https://doi.org/10.2991/msam-17.2017.68

Rakaraddi, A., & Pratama, M. (2021). Unsupervised learning for identifying high eigenvector centrality nodes: A graph neural network approach. In Proceedings of the 2021 IEEE International Conference on Big Data (pp. 4945–4954). IEEE. https://doi.org/10.1109/BigData52589.2021.9671766

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., & Jia, W. (2018). Influence analysis in social networks: A survey. Journal of Network and Computer Applications, 106, 17–32. https://doi.org/10.1016/j.jnca.2018.01.005

Jin, X., Cui, B., Li, D., Cheng, Z., & Yin, C. (2018). An improved payload-based anomaly detector for web applications. Journal of Network and Computer Applications, 106, 111–116. https://doi.org/10.1016/j.jnca.2017.09.004

Corso, G., Stark, H., Jegelka, S., Jaakkola, T., & Barzilay, R. (2024). Graph neural networks. Nature Reviews Methods Primers, 4(1), Article 17. https://doi.org/10.1038/s43586-024-00294-7

Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3–5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Fu, J., Zhang, X., Li, S., & Chen, D. (2023). Variational disentangled graph auto-encoders for link prediction. arXiv. https://arxiv.org/abs/2306.11315

Fan, D., & Gao, C. (2024). Learning network representations with disentangled graph auto-encoder. arXiv. https://arxiv.org/abs/2402.01143

Qiu, J., Tang, J., Ma, H., Dong, Y., Wang, K., & Tang, J. (2018). DeepInf: Social influence prediction with deep learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2110–2119). ACM. https://doi.org/10.1145/3219819.3220077

Li, S. E. (2023). Deep reinforcement learning. In Reinforcement learning for sequential decision and optimal control (pp. 365–402). Springer Nature. https://doi.org/10.1007/978-981-19-7784-8_10

Rahman, M. S., Halder, S., Uddin, M. A., & Acharjee, U. K. (2021). An efficient hybrid system for anomaly detection in social networks. Cybersecurity, 4(1), Article 10. https://doi.org/10.1186/s42400-021-00073-5

Rahi, M., Ebrahimnejad, A., & Motameni, H. (2024). Evaluation of predicted fault tolerance based on C5.0 decision tree algorithm in irrigation system of paddy fields. International Journal of Intelligent Computing and Cybernetics, 17(2), 253–305. https://doi.org/10.1108/IJICC-06-2023-0154

Kecman, V. (2005). Support vector machines — An introduction. In L. Wang (Ed.), Support vector machines: Theory and applications (pp. 1–47). Springer. https://doi.org/10.1007/10984697_1

Zheng, X., & Li, C. (2024). Predicting students' academic performance through machine learning classifiers: A study employing the Naive Bayes Classifier (NBC). International Journal of Advanced Computer Science & Applications, 15(1). https://doi.org/10.14569/IJACSA.2024.0150121

Aarthi, E., Sheela, M. S., Vasantharaj, A., Saravanan, T., Rama, R. S., & Sujaritha, M. (2024). Integrating neural network-driven customization, scalability, and cloud computing for enhanced accuracy and responsiveness for social network modelling. Social Network Analysis and Mining, 14(1), Article 139. https://doi.org/10.1007/s13278-024-01295-2

Tan, Q., Liu, N., & Hu, X. (2019). Deep representation learning for social network analysis. Frontiers in Big Data, 2, Article 2. https://doi.org/10.3389/fdata.2019.00002

Hamilton, W. L. (2020). Graph representation learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 14(3), 1–159. https://doi.org/10.2200/S01045ED2V01Y202009AIM046

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Zhang, Y. (2025). Social network user profiling for anomaly detection based on graph neural networks. arXiv. https://arxiv.org/abs/2503.19380

Tahsien, S. M., Karimipour, H., & Spachos, P. (2020). Machine learning based solutions for security of Internet of Things (IoT): A survey. Journal of Network and Computer Applications, 161, 102630. https://doi.org/10.1016/j.jnca.2020.102630

Rashnodi, O., Rastegarpour, M., Moradi, P., & Zamanifar, A. (2024). Community detection in attributed social networks using deep learning. The Journal of Supercomputing, 80(18), 25933–25973. https://doi.org/10.1007/s11227-024-06338-3

Kazemi, S. M., Goel, R., Jain, K., Kobyzev, I., Sethi, A., Forsyth, P., & Poupart, P. (2020). Representation learning for dynamic graphs: A survey. Journal of Machine Learning Research, 21(70), 1–73.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4–24. https://doi.org/10.1109/TNNLS.2020.2978386

He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. Knowledge-Based Systems, 212, 106622. https://doi.org/10.1016/j.knosys.2020.106622

Su, L., Wang, W., & Zhang, Y. (2019). Strong consistency of spectral clustering for stochastic block models. IEEE Transactions on Information Theory, 66(1), 324–338. https://doi.org/10.1109/TIT.2019.2934157

Kumar, S., Mallik, A., & Panda, B. S. (2023). Influence maximization in social networks using transfer learning via graph-based LSTM. Expert Systems with Applications, 212, 118770. https://doi.org/10.1016/j.eswa.2022.118770

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.