



Enhancing Social Network Prediction in Graph Neural Networks Using Graph Theoretical Approaches: Social Network Analysis on The Facebook Ego Dataset

Pharsana Parveen M¹, Sr. Stanis Arul Mary A²

Research Scholar, PG and Research Department of Mathematics, Nirmala College for Women, Coimbatore,
Tamil Nadu, India - 641 018¹

Assistant Professor, PG and Research Department of Mathematics, Nirmala College for Women, Coimbatore,
Tamil Nadu, India - 641 018²

Abstract: This paper focuses on social network analysis and prediction by incorporating Graph theoretical concepts into the architecture of Graph Neural Networks (GNNs), with a focus on Facebook Ego Dataset. We propose an algorithm to leverage graph theoretical principles such as graph coloring, mutual friends, and weighted edge traversal to enhance link prediction tasks in Graph Neural Network. This approach optimizes Graph Neural Network's performance by capturing local and global structural patterns within social networks. By using graph theoretical technique-based algorithm, the developed model aims on improved accuracy and diversity of friend recommendation in social networks. The study helps us to understand how integrating graph coloring helps in enhancing node embedding. The proposed algorithm generates improved social predictions with higher accuracy and meaningful insights. The results accentuate the importance of applying classical graph theoretical concepts with recent deep learning techniques, providing an efficient framework for social network prediction and analysis.

Keywords: Graph Neural Network, Product Recommendation System, Node Embeddings. Graph Coloring, User Product Graph.

I. INTRODUCTION

In the modern world, social platforms like Facebook, Twitter, etc., generate huge amount of data based on User interaction. There social platform gathers valuable information from user interaction such as user interacting with contents, forming connections, etc., which can be analyzed for delivering personalized recommendations. These recommendations help the social platform to provide tailored recommendations that best fits their choices and preferences.

Social network analysis (SNA) is playing an eminent role in providing valuable insights into human behaviours, patterns and connections. Using machine learning models, Social Network Analysis (SNA) allows us to predict the relationship of user preference and social behaviours, which has a great impact on application such as recommendation of friends, suggestions for contents, etc.

Traditional Machine learning models analyzes user data and their interactions and provides efficient social network predictions by using various methods, few includes,

1. Collaborative filtering: The historic behaviours of similar users is constructed as a matrix (use-item matrix) where rows represent uses, and columns represents items. This allows the system to predict unknown entries, resulting in prediction of potential friend.
2. Matrix factorization: The idea of decomposing of use-item interaction matrix into smaller matrix is used in 'Matrix Factorisation' method. It discloses the hidden relationship among the users and helps us in finding a potential user suggestion.
3. Link prediction: The objective of this method is to predict the possibility of a new connection between two users. It uses various similarity metrics to identity the recommendation.



4. Random walks: This particular graph traversal method helps in finding the relationship between users in a social network. Random walks i.e., random steps from one user to another user helps in predicting the possible user connection (i.e., friend prediction)

Graph theory helps us in handling relationship between entities to reduce the complexity of the network. In graph theory, social networks are represented as $G = (V, E)$ where,

- Vertices(V)-users in the social networks
- Edges(E)-represents the interaction between users (e.g., friendship)

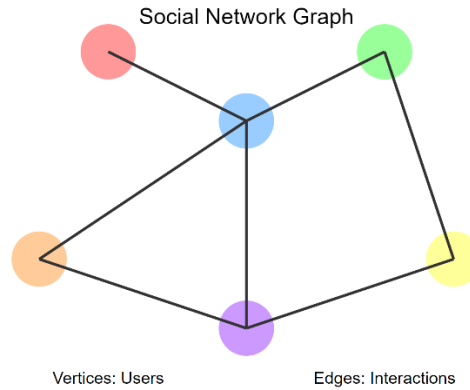


Fig. 1 Representation of Social Network as a Graph

This graph-based representations are highly effective in analyzing the structure of the social networks.

These are several other graph properties such as centrality of node, clustering and path length are used to identify influential users, community clustering and possible friend recommendations. For instance, degree centrality is used in identifying the influences. Betweenness centrality helps in identifying users who bridges various communities and choose them as recommendation for those communities.

These graph-based methods passed way for graph-neural network (GNNs) in social network analysis (SNA). GNNs extends power of graph theory by combining graph-based methods with deep learning. They learn from complex interactions and propagates the information across nodes and edges to equate node embedding. The node embedding in GNN represents position of each user in the network along with its associated information. Encoded with local and global information. The weight of the edge represents the strength of interactions. GNN-based nodes are effective in link-predictions to predict potential relationships based on graph structure and uses interactions.

II. LITERATURE REVIEW

The application of Graph theory in Social Network Analysis (SNA) is eminent due its representation of entities as vertices and relationship as edges in graph. Many researches and studies have been conducted in both structure and content-based analysis in Social Network Analysis (SNA).

- **Structure-Based Analysis:** Analyzing Social Network Analysis (SNA) from a structural perspective was established by Viswanath (2009) [1] and Humski et al (2019) [2]. The factors that were tailored are user relationships over time and synthetic expanded social graph. De Nooy (2012) [3] tailored centrality measures in detecting popular nodes. Ranking system based on likes, shares and comments, was introduced by, Satapathy et al (2016) [4]. The importance of degree distribution and clustering coefficient was studied by Wang, Krim and Viniotis [5], using Markov graph models. At 2016, Krid [6] introduced a new model for Online Social networks case of Facebook.
- **Content-Based Analysis:** The content-based analysis focuses on information based on user interactions, activities and posts. At 2013, Reza Hassazadeh [7] used semi-supervised graph algorithms to find outliers. At 2012, Catanese [8] proposed a new approach in visualizing social network data using geodistic distances and connected components. Anwar and Abulaish [9] used message similarity relationships in finding community interaction.
- **Combined Approach and Challenges:** Shtivastara and Pal (2009) [10] introduced a combined approach including both structure-based and content-based analysis. Even though, it suffered in Scalability in using large datasets.



There is a growing demand in integrating various frameworks which efficiently handles large and complex datasets, providing real-time, meaningful insights.

III. APPLICATION OF GRAPH THEORY IN SOCIAL NETWORK ANALYSIS (SNA)

Traditional methods in Social Network Analysis (SNA) depends on statistical methods. These statistical methods concentrate on isolated properties of users. For example, number of friends of a user or number of messages received for a user. The defect of the model is that it fails to capture the connections between the users and only complex relationships.

Graph theory helps in configuring a framework by considering the entire network instead of focusing on isolated properties of users. This property helps in deeper understanding of the global structure of the network, detects underlying pattern and future prediction of interactions. It plays a major role in identifying influence, community detection and nature of users.

Graph based algorithms efficiently handles large scale networks by effective computation with graph traversal and learning techniques when compared to traditional methods.

TABLE I SOLUTION FOR VARIOUS SNA USING GRAPH THEORY

Problem	Description	Solution
Detection of communities	The identification of closely connected subgroup with the network. It is used to be noted that users may have more interactions with each other.	<i>Clustering algorithms</i> (e.g. Spectral clustering) can be used in finding group of similar users.
Link Prediction	The prediction of future relationships between users based on current interactions.	<i>Link prediction metrics</i> such as Jaccard similarity and common neighbors can be used to analyze node relationships to predict new edges.
Centrality and Influence Detection	The identification of most influential users for marketing and advertisement purpose	<i>Centrality measures</i> including degree centrality, betweenness centrality detect the most impacted node in the network.
Friend Recommendation	Recommendation of new potential friends based on user-user interaction.	<i>Collaborative filtering</i> in combination with graph traversals techniques and coloring algorithms allows for efficient friend recommendation.
Information propagation	Modelling and maintains the spread of information across a network and identify the users who distributes the message more.	<i>Diffusion models</i> e.g. Linear threshold represents the flow of information across the nodes and helps in identification of key influences.

IV. PROBLEM DESCRIPTION

A. Facebook Friend Prediction in Social Network Analysis (SNA):

Facebook as a large social networking platform connects billions of people worldwide. The major challenge faced by Facebook is providing an effective friend recommendations for their users to increase user engagement and grow the network. In order to recommend a potential friend to a user, Facebook relies on Social Network Analysis (SNA) techniques.

B. Problem Statement:

Facebook's friend recommendation system needs to predict and suggest potential friendship for users. The parameters to be considered must include the consideration of their existing connections, common activities and mutual connections.

Despite the massive scale of the network, the system should identify potential connection with billions of users and connections, evaluating and prioritizing new recommendation for every user in real-time is expensive. This might also involve the users who lack dense populated network, which makes prediction tasks easier. It is difficult to handle the dynamic changes in the connections (new connection and breakage of old connections)



V. WEIGHTED GRAPH COLORING USER RECOMMENDER ALGORITHM (WGUR-ALGORITHM)

A. Construction of Graph with Nodes and Edge:

Construct an undirected graph from social network datasets by considering user as node and user-user interaction as edges $G = (V, E)$, where,

- i. V is the set of *Users* (nodes)
- ii. E representing the connection-based friendship (interaction) between the Users

Each node $U_i \in V$ represents a user, and each edge represents a friendship between users U_i and U_j .

Construct a graph from the considered dataset, where each edge represents the connections. Initially, the strength of interaction between users is set as equal.

B. Assignment of Graph Coloring:

- i. *Step 1*: This step involves graph coloring to assign colors to users.
 - a. Group users with mutual friends by assigning them the same colors.
 - b. Users in different groups with distinct friends are assigned different colors.
- ii. *Step 2*: Focusing on color Diversity in Recommendations: Prioritize friends from different color groups other than U_1 while calculating the common interaction counter for a candidate friend U_y .
Friends in different social clusters should be given more weights.
Let $C(U_1)$ be the color assigned to User U_1 .

C. Common Interaction Counts and Edge Weighting:

- i. *Step 1*: Identification of Candidate friends.
 - a. Let U_1 be the target User. The candidate friend of U_1 is U_y who are not directly connected to U_1 .
 - b. The traversal is to be prioritized for different color groups to ensure diversity in recommendation other than that of U_1 .
- ii. *Step 2*: Identification of Mutual Friends: For every candidate friend U_y , identify the set of mutual friends, i.e., the nodes shared between U_1 and U_y .

$$M(U_1, U_y) = \{M \in G : (U, M) \in E \text{ and } (M, U_y) \in E\} \quad (1)$$
- iii. *Step 3*: Computation of Weighted Common Interaction Counter: Consider candidate friend U_y , now let us compute "Weighted Common Interaction Counter" based on:
 - a. The total number of mutual friends between U_1 and U_y .
 - b. The sum of edge weights between mutual friends that connect U_1 and U_y .

$$C(U_y) = \left[\text{Number of Mutual Friends} + \sum_{i \in M(U_1, U_y)} (W(U_1, i) + W(i, U_y)) \right] \times \text{Color Bias}(U_1, U_y) \quad (2)$$

- Number of Mutual Friends: Count of nodes shared in common by U_1 and U_y .
- $W(U_1, i)$ – weight of edge between U_1 and i
- $W(i, U_y)$ – weight of edge between i and U_y
- $\text{Color Bias}(U_1, U_y) = \begin{cases} 1.5 & \text{if } C(U_1) \neq C(U_y) \\ 1 & \text{if } C(U_1) = C(U_y) \end{cases}$

This ensures recommendation from different color groups.

The weight may be binary in absence of any special interaction information

D. Recommendation Logic:

- i. *Step 1*: Scoring Based on color and Weighted Counter: The score $C(U_y)$ represents the counter value including the contributions of number of mutual friends, the strength of interactions and color bias.
- ii. *Step 2*: Recommend top k friends:
 - a. Once $C(U_y)$ is calculated, select top k users with the highest score.
 - b. Prioritize users from different color groups, whenever similar $C(U_y)$ is encountered.

E. Alternative Scenarios:

- i. *Scenario 1*: Whenever no users obtain a meaningful friendship score, recommend users with higher edge weights from diverse color groups.
- ii. *Scenario 2*: Whenever multiple users obtain same score, choose based on different color.



F. Optimization of Graph Traversal:

- i. *Color-Guided Traversal*: It prioritizes exploring users with different colors from U_1 ensuring diverse friend recommendations.
- ii. *Prioritize Strong Connection*: It prioritizes stronger mutual friends based on edge weights, ensuring meaningful recommendation.
- iii. *Skip Already Connected Friends*: Excluding user already connected to the target users avoids recommendation of existing friends.

VI. ILLUSTRATION OF ALGORITHM

To Understand the algorithm in a deeper sense, let us consider an illustration where six users are represented as nodes U_1, U_2, U_3, U_4, U_5 and U_6 . The friendship between any user is represented as an edge between them. The user-user interaction and the weight assigned is explained in the following table.

TABLE II USER-USER INTERACTION TABLE

User	Friends (Edges)
U_1	U_2, U_3
U_2	U_1, U_3, U_4
U_3	U_1, U_2, U_5
U_4	U_2, U_5, U_6
U_5	U_3, U_4, U_6
U_6	U_4, U_5

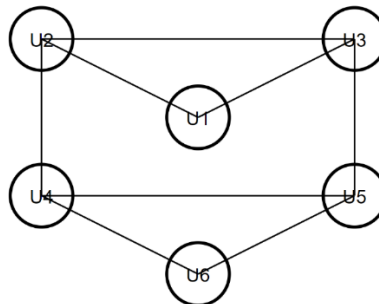


Fig.2 Representation of Users and User-User interaction as Node

All edges are weighted as 1, as here the interaction strength is ignored for scalability.

Next, we group users by allocating colors to each user based on their connections and ensuring two friends have no common color. Let the color of U_1 be color 1, Since U_2 , and U_3 are adjacent nodes, they are assigned a different color, similarly all nodes are colored.

TABLE III GRAPH COLORING

User	Assigned Color
U_1	Color 1
U_2	Color 2
U_3	Color 3
U_4	Color 1
U_5	Color 2
U_6	Color 3

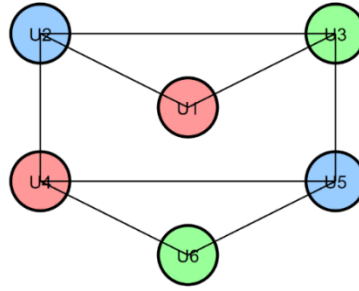


Fig. 3 User-User Interaction with Graph Coloring Representation

Next, Candidate Friends are identified for U_1 . Since, U_1 is already connected to U_2 and U_3 , the candidate friends for U_1 are U_4 , U_5 and U_6 .

TABLE IV IDENTIFY CANDIDATE FRIENDS FOR U_1

Already Connected Friends	Candidate Friends
U_2, U_3	U_4, U_5, U_6

Note we identify Mutual Friends and Common Interaction Counter for each candidate based on the number of mutual friends and their color bias.

TABLE V MUTUAL FRIEND IDENTIFICATION AND COMMON INTERACTION COUNTER

Candidate Friend	Mutual Friends with U_1	Edge Weights $w(U_1, \text{Mutual Friend}), w(\text{Mutual Friend}, U_y)$	Color Bias	Common Interaction Counter $C(U_y)$
U_4	U_2	$w(U_1, U_2) = 1, w(U_2, U_4) = 1$	1 (same color)	$1 + (1 + 1) \times 1 = 3$
U_5	U_3	$w(U_1, U_3) = 1, w(U_3, U_5) = 1$	1.5 (different color)	$1 + (1 + 1) \times 1.5 = 4$
U_6	U_4	$w(U_1, U_4) = 0$ (no direct edge), $w(U_4, U_6) = 1$	1.5 (different color)	$1 + (0 + 1) \times 1.5 = 2.5$

We now rank the candidate friends U_4 , U_5 and U_6 based on their Common Interaction Counter (U_y) and assign them Rank.

TABLE VI SORTING BASED ON COMMON INTERACTION COUNTER

Candidate Friend	Common Interaction Counter $C(U_y)$	Ranking
U_5	4	1
U_4	3	2
U_6	2.5	3

The top friend recommendation for U_1 is U_5 , followed by U_4 and U_6 based on their rank respectively.

VII. EXPERIMENTAL RESULTS

In the study, we developed a GNN model to enhance friend recommendation based on our proposed algorithm which primarily focuses on mutual friends and other graph theoretical approaches. The Facebook Ego Dataset was collected from the SNAP STANFORD EDU Database. The dataset includes information about user-user interaction. The Data was preprocessed and converted to a graph structure by representing Users as nodes and establishing an edge connection



whenever two Users are existing friends on Facebook. The feature engineering uses User connections, mutual graph coloring, and edge weights. The GNN was trained with an 80- 20 Train-Test Split dataset.

The architecture consists of 3 GNN layers, the input layer consisting of node features is fed through multiple Graph Convolutional Layers, which aggregate information from its adjacent nodes to enhance node representations. The aggregation includes the influence of mutual friends and weighted interaction scores between them. For non-linearity between layers, ReLu activations were used. The final layer consists of a soft-max classifier that provides the likelihood of a new connection as an output. The model ranks potential friends for each user and recommends the top N highest-ranking users as new friends. The proposed algorithm achieved significantly higher accuracy (90.08%) when compared to traditional methods, outperforming them. The proposed model promotes diversity based on user interaction patterns, making it stand out from traditional GNN models.

VIII. CONCLUSION

The study demonstrates the importance of integrating classical graph theoretical approaches with Graph Neural Networks in enhancing social network predictions. Our study primarily focuses on suggesting a possible friend for users within Facebook Ego Dataset. With concepts such as mutual friends, graph coloring and edge weighting incorporated into the model, the model outperforms the prediction of friends by an improved accuracy of 90%. Future work can extend to exploring dynamic datasets, integrating additional features, and refining model parameters to further enhance the predictive capabilities of GNNs in social network analysis (SNA). As a conclusion, the combination of graph theory with advanced Machine learning and Artificial Intelligence, provides valuable insights and recommendations than can impact user engagement in social network platforms.

REFERENCES

- [1]. B. Viswanath, A. Mislove, M. Cha, & K. P. Gummadi, *On the evolution of user interaction Workshops in Facebook*. SIGCOMM 2009 - Proceedings of the 2009 SIGCOMM Conference and Co-located Workshops, Proceedings of the 2nd ACM Workshop on Online Social Networks, WOSN 2009, pp. 37-42. (2009)
- [2]. L. Humski, D. Pintar, & M. Vranic, *Analysis of Facebook Interaction as Basis for Synthetic Expanded Social Graph Generation*. IEEE Access, 7(c), pp. 6622-6636. (2019)
- [3]. W. De Nooy, Social network analysis, graph theoretical approaches to. *Computational Complexity: Theory, Techniques, and Applications*, 9781461418, pp. 2864—2877. (2012)
- [4]. S. C. Satapathy, A. Joshi, N. Modi, & N. Pathak, Preface. *Advances in Intelligent Systems and Computing*, pp. 409, (2016)
- [5]. T. Wang, H. Krim, & Y. Viniotis, *A generalized Markov graph model: Application to social network analysis*. IEEE Journal on Selected Topics in Signal Processing, 7(2), pp. 318—332. (2013)
- [6]. D. K. Krid, N. B. S. Grati, & R. Robbana, *A new model for Online Social Networks case of Facebook*. Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, A/CCSA, 2016-July. (2016)
- [7]. R. N. Reza Hassanzadeh, *A Semi-Supervised Graph-Based Algorithm for Detecting Outliers in Online-Social-Networks*, pp. 577—582. (2013)
- [8]. S. Catanese, P. De Meo, E. Ferrara, & G Fiumara, *Analyzing the facebook friendship graph*. CEUR Workshop Proceedings, 685, pp. 14—19. (2010)
- [9]. T. Anwar, & M. Abulaish, *Mining an enriched social graph to model cross-thread community interactions and interests*. MSM '12 - Proceedings of 3rd International Workshop on Modeling Social Media, pp. 35-38. (2012)
- [10]. S. Shrivastava, & S. N. Pal, *Graph mining framework for finding and visualizing substructures using graph database*. Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009, pp. 379-380. (2009)