SOCIAL NETWORK ANALYSIS USING MACHINE LEARNING APPROACHES

A Thesis Submitted to

DELHI TECHNOLOGICAL UNIVERSITY

for the Award of the degree of

DOCTOR OF PHILOSOPHY

In

DEPARTMENT OF INFORMATION AND TECHNOLOGY

by

SONIA

(2K18/Ph.D/IT/01)

Under the Joint Supervision of

Prof. KAPIL SHARMA
Professor
Department of Information Technology
Delhi Technological University

Dr. MONIKA BAJAJ
Assistant Professor
Computer Science Department
Deshbandhu College, University of Delhi

Department of Information and Technology

Delhi Technological University

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042 (India)

July 2023

DECLARATION

I , Sonia ,Ph.D. student (Roll No. 2k18/PhD/IT/01), hereby declare that the thesis entitled "**Social Network Analysis Using Machine Learning Approaches"** which is being submitted for the award of the degree of *Doctor of Philosophy* in Information Technology, is a record of bonafide research work carried out by me in the Department of Information Technology, Delhi Technological University. I further declare that this work is based on original research and has not been submitted to any university or institution for any degree or diploma

Date:------------------------

Place: New Delhi

SONIA

2K18/Ph.D./IT/01

Department of Information and Technology

Delhi Technological University

New Delhi-11000422

DELHI TECHNOLOGICAL UNIVERSITY
(GOVT. OF NATIONAL CAPITAL TERRITORY OF DELHI)
BAWANA ROAD, DELHI-110042

CERTIFICATE

This is to certify that the work embodied in the thesis entitled **"Social Network Analysis Using Machine Learning Approaches"** has been completed by Ms. Sonia (Reg. No.: 2K18/Ph.D./IT/01) under our supervision and guidance towards fulfillment of the requirements for the degree of Doctor of Philosophy to the Delhi Technological University,Delhi.This work is based on the original research and has been submitted in full or in part for any other diploma or degree of any university.

Date:-----------------

Supervisors

Prof. Kapil Sharma	Dr Monika Bajaj
Professor	Assistant Professor
Dept. of Information Technology,	Dept. of Computer Science,
Delhi Technological University	Deshbandhu College, University of Delhi

ACKNOWLEDGMENT

I am grateful to my supervisor, **Prof. Kapil Sharma**, and **Dr. Monika Bajaj** without whom this achievement would not have been realized. It was their valuable guidance and consistent encouragement through my research period, which helped me to overcome the challenges that came my way. I consider it a great opportunity to do my doctoral program under their guidance and to learn from his research experience. My sincere regards to **Prof. S. Indu, Vice-Chancellor, Delhi Technological University** for providing a platform for pursuing my Ph.D. work. I express my gratitude to **Prof. Dinesh Kumar Vishwakarma**, Head, Department of Information Technology, Delhi Technological University for their kind support and for providing the necessary facilities to undertake this research. I would also like to thank the senior staff members and lab technicians for helping me with my requirement during the research work. I would like to thank my fellow research scholars, friends, and hostel mates who offered me their time, help, and moral support whenever needed. I am grateful to my daughter (Ms.Purvi), my parents (Mr.Narendra Singh Choudhary and Mrs. Rameshwari), and my family for being an inspiration and strong support for me always and enlightening my path in life. Last, but not least, I am grateful to God for always giving me the strong support, the best path, and direction to make things successful.

Sonia
2k18/Phd/IT/01
Department of Information Technology
Delhi Technological University, Delhi

LIST OF PUBLICATIONS

- Sonia, Kapil Sharma, Monika Bajaj (2023), DWIM: Influence Maximization Using Deep Learning", Intelligent Automation & Soft Computing, IASC, 2023, vol.35, no.1, DOI: 10.32604/iasc.2023.026134 (Published SCIE: Impact Factor 1.632, Tech Science Press).
- Sonia, Kapil Sharma, Monika Bajaj(2023) Effectual Seed Pick Framework focusing on Maximizing Influence in Social Networks", in Wireless Communications and Mobile Computing Volume 2023, Article ID 3185391, 19 pages https://doi.org/10.1155/2023/3185391(Published, SCIE: Impact Factor 2.164, HINDAWI)
- Sonia, Kapil Sharma, Monika Bajaj(2023), Multi Neighbour Seed Selection for Influence Maximization, Sadhana Journal. ISSN :0973-7677(Communicated)(SCIE)
- Sonia,"Opinion Mining Techniques and Applications: Review" in Springer Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019), 2020, Volume 121.
- Sonia, Kapil Sharma, Monika Bajaj, “A Review On Opinion Leader Detection and It’s Applications” in Proceedings of IEEE 7th International Conference on Communication and Electronics Systems(ICCES-2022).
- Sonia, Kapil Sharma, Monika Bajaj,” Social Networks Analysis and Machine Learning: An Overview of Approaches and Applications” IEEE International Conference on Sustainable Computing and Smart System,(ICSCSS 2023)(accepted).
- Sonia, Kapil Sharma, Monika Bajaj, “Techniques of Community Detection: A Review” IEEE International Conference on Sustainable Computing and Smart System,(ICSCSS 2023)(Communicated).

ABSTRACT

Social network analysis is vital for uncovering meaningful patterns, structures, and dynamics within social networks, with applications spanning health, marketing, and finance. One prominent application is viral marketing, which harnesses social relationships and interactions on social media platforms to sway consumer behavior. Social media allow users to connect, share opinions about the product and services. In such a scenario Opinion leaders (OLs), individuals with expertise in specific subjects and a substantial number of followers, play a pivotal role in shaping others' opinions on social media provide a boost to viral marketing. Therefore the success of viral marketing campaigns is crucial to identifying opinion leaders and individuals with substantial influence within their networks.

Machine learning techniques have significantly advanced the development of accurate algorithms for identifying and evaluating these influential users, thereby maximizing the impact of marketing efforts. This work explores the feasibility of leveraging deep learning to approximate user influence. DeepWalk-based Influence Maximization (DWIM) algorithm is proposed that employs graph embedding techniques to identify the most influential nodes within the network.

The work introduces a seed selection framework for maximizing influence in pervasive healthcare, utilizing machine learning approaches to investigate the bidirectional effects of influence and trust. The proposed framework addresses challenges associated with a large number of patients, ultimately enhancing influence maximization through strategic seed selection. Specifically, the Fuzzy-VIKOR algorithm is proposed to identify target nodes that facilitate the rapid dissemination of information. By effectively tackling issues inherent in large patient populations, the framework proves beneficial for pervasive healthcare applications.

Furthermore, the thesis presents a Multi-Neighbor seed selection approach to enhance influence maximization. This approach accounts for the network's memory effect or social reinforcement effect and employs the Neighbor Degree Value (NDV) to estimate

the influence strength of selected seeds. It addresses challenges related to seed selection, such as limited coverage area and inadequate discriminatory power. The proposed Judgment Leader Pick Weighting Grade method, incorporating Judgment Grade Value for leader selection, effectively resolves trust management issues in ubiquitous services, ultimately amplifying influence maximization. This thesis explores the potential of Ant Colony Optimization (ACO), a nature-inspired optimization algorithm based on ants' foraging behavior, for influence maximization in social networks. It presents the framework for maximize influence spread by identifying seeds in the network for specific situations and issues.

Experiments are conducted to evaluate the proposed algorithms, and centrality measures are employed for result comparison. The experimental findings demonstrate that the proposed frameworks exhibit high precision, accuracy, F1-score, and recall compared to existing algorithms for influence maximization seed selection. Furthermore, the proposed methods efficiently identify a set of influential nodes within a computable timeframe, facilitating viral marketing efforts and enabling targeted recommendations for various products and services,

In summary, this thesis underscores the importance of social network analysis and machine learning techniques for identifying influential individuals in social networks, enabling more effective viral marketing and recommendations. The research introduces the DeepWalk Based Influence Maximization (DWIM) algorithm, which surpasses traditional centrality measures by integrating topical and topological features. DWIM holds promise for applications in diverse fields such as online marketing and outlier detection.

The thesis presents two algorithms for identifying influential opinion leaders that incorporates user interaction and trust management and can be potentially applied in viral marketing and outlier detection across various services. The effectiveness of the algorithms has been studied on pervasive healthcare, politics, product promotion, and service promotions. Notably, in healthcare, it enhances targeted service delivery, demonstrating superior performance compared to existing algorithms in terms of accuracy, recall, f1-score, and precision. Overall, this work contributes significantly to the advancement of social network analysis by providing a comprehensive framework for influence maximization and trust management.

Contents

List of Figures

List of Tables

ABBREVIATIONS

ANN	:	Artificial Neural Network
AI	:	Artificial Intelligence
BNS	:	Best Node Selection algorithm
BCT	:	Billion Scale Cost Aware Targeted
CELF	:	Cost Effective Lazy Forward
CIM	:	Community Based Influence Maximization
CoFIM	:	Community based Framework for Influence Maximization
DL	:	Deep Learning
DWIM	:	DeepWalk Based Influence Maximization
EEA	:	Exact Extraction Algorithm
GCN	:	Graph Convolutional Network
IM	:	Influence Maximization
IRIE	:	Impact Ranking and Impact Estimation
IC	:	Independent Cascade Model
KNN	:	K Nearest Neighbour
LSTM	:	Long Short Term Memory
LT	:	Linear Threshold
ML	:	Machine Learning
MIA	:	Maximum Influence Arborescence
OL	:	Opinion Leader

SKIM	:	Sketch Basec Influence Maximization
SKIM	:	Sketch Based Influence Maximization
SNA	:	Social Network Analysis
SIM	:	Simple Influence Maximization
TVM	:	Topic Aware Viral Marketing
TIM	:	Two phase Influence Maximization
VIKOR	:	VIseKriterijumskaOptimizacija I KompromisnoResenje

Chapter 1 INTRODUCTION

This chapter introduces the concept of social network analysis, machine learning, approaches, and applications of social network analysis.This chapter introduces about Opinion Leaders used for Influence Maximization The objectives of the research work are presented. In the end, the chapter-wise thesis coverage has been summarized.

1.1 Background of the study

Currently, social networking sites have become an essential part of human life. Social networking websites have allowed users to share information globally and provided an online communication platform.Social Networking sites are not restricted to Twitter, Facebook, LinkedIn, Instagram, WhatsApp, or Snapchat. The interactions among likeminded users over these sites boost social networking, representing groups with similar interests and likings. As a result, social networking has provided a powerful platform for marketers seeking to connect the customers.

With the emergence of online social networks and user-generated content, there are great opportunities to analyze social networks. Social networks have significantly helped web users communicate globally. Social network analysis (SNA) (J.Tang, 2009) uses networks and graph theory to analyze social structures. Social network analysis (SNA) is a research area that uses network and graph theory to explore the social arrangements and relationships between nodes (individuals or objects) and edges (links or connections) within a network. Network structures are characterized by nodes (individual actors, people, or things in the network) and the edges, links, or ties (interactions or relationships) which connect the nodes in the network. The social network analysis helps summarize users' opinions (nodes in the network) and their interests in specific brands, products, or services, discovering the links patterns between users and mining online platforms' events. This information provides insights into various applications. For example, a critical application of SNA is in identifying opinion leaders.

The individuals who significantly affect the views, opinions, and actions of others in their network are known as opinion leaders. Marketers have used the influencing power of opinion leaders for their online product promotion. However, the identification of opinion leaders is a complex exercise. It needs deep analysis of patterns in the data and identifying key features that distinguish opinion leaders from other network members. Researchers applied various machine learning techniques to automate the identification of opinion leaders. However, these techniques have shown promise in different application areas, including marketing, political campaigning, and public health.

The thesis studies and explores the viability of both topological and topical behavior while identifying opinion leaders. The work begins with an overview of social network analysis and its significance and proceeds with studying existing machine learning techniques used for SNA. Through SNA, researchers can extract user behavior patterns, opinions, and interests, enabling them to gain insights into the users' interactions on online platforms. This chapter discusses the techniques of social network analysis, its approaches, and applications to various domains related to the research area.

1.1.1 Techniques for Social Network Analysis

Artificial Intelligence (AI) is the technique that enables a machine to imitate human behavior. The study of artificial intelligence (P., 2012) (Huang, 2010) is a subfield of computer science that aims to build intelligent machines capable of learning, problem-solving, making decisions, and interpreting natural language tasks that generally call for human-level ability. Numerous sectors, including healthcare, banking, manufacturing, and transportation, stand to benefit from AI.

Machine learning (ML), (T. Veeramakali, 2022) (A.Kanan, 2022) (L. M. Norz, 2022) (Á. MacDermott, 2022) (Gawali, 2022). (Murphy, 2012) (Bharadwaj, 2021) (Géron, 2019) (Domingos P. , 2012) is a subset of artificial intelligence where models got trained through data, and it is considered supervised learning, unsupervised learning, and reinforcement learning. It includes creating statistical models and algorithms that let computers learn from data rather than being explicitly programmed. To learn hierarchical representations of data, deep learning (DL) (Won, 2018), a subset of machine learning entails training neural networks with several layers. Natural language processing, speech recognition, and computer vision are just a few of the fields where Deep Learning has excelled. Deep learning utilises the human brain structure, and this structure is known as

Artificial neural network. We required a much higher volume of data for deep learning to train our machine. A neural network consists of neurons where the information processing takes place.

1.1.1.1 Machine Learning

Machine learning is considered as supervised learning, unsupervised learning, and reinforcement learning as in Figure 1.1 and all machine learning algorithms overview in Figure 1.2 is given:

Supervised Machine Learning: It is broadly classified into two types based on the problem it solves: classification and regression. The algorithms under supervised learning are linear regression, logistic regression, ridge and lasso, decision tree, ada boost, random forest, gradient boosting, xgboost, naive bayes, support vector machine and KNN (*k*- nearest neighbor).

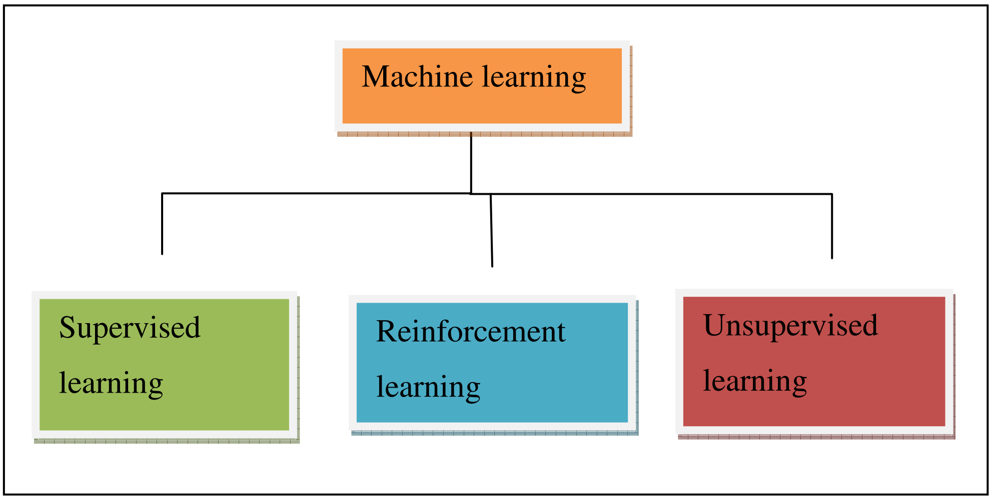


Figure 1.1 Types of Machine Learning Algorithm

Unsupervised Machine Learning: It is further classified in two types based on the problems it solves, like clustering or grouping and dimensionality reduction. The algorithms used for unsupervised machine learning to solve clustering problems are k-means clustering, hierarchical clustering, k-nearest neighbour clustering, dbscan clustering, and for dimensionality reduction iterated distillation and amplification (IDA) and principle component analysis (PCA) algorithms are used.

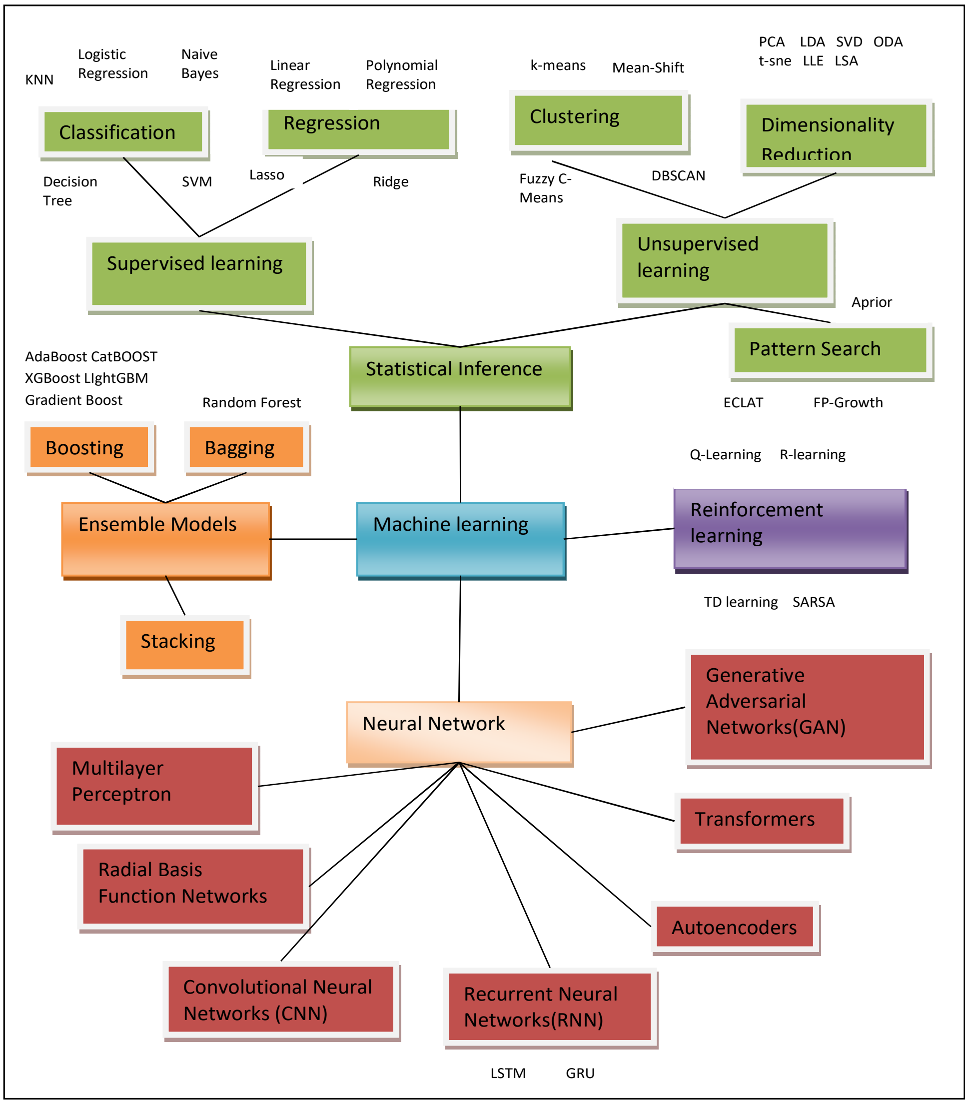


Figure 1.2 Machine Learning Overview

Reinforcement Machine Learning (D. Posner, 2015) (Richert, 2015) (Dey, 2016) (Qiu, 2016): It is a machine learning technique that focuses on the activities that software agents should perform in a given environment. Two categories of reinforcement learning techniques exist: Positive and Negative.

Statistical Inference involves drawing conclusions from data by making inferences about underlying populations or distributions. It's used to test hypotheses, estimate parameters, and make predictions based on statistical models. Techniques like hypothesis testing, confidence intervals, and Bayesian inference are used for this purpose.

Neural Networks are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected nodes (neurons) organized in layers and are particularly suited for tasks involving complex patterns and large datasets. Deep neural networks, including convolutional neural networks (CNNs) for image analysis and recurrent neural networks (RNNs) for sequence data, have been highly successful in various domains.

Ensemble Algorithms combine the predictions of multiple machine learning models to improve overall performance and reduce overfitting. Common ensemble techniques include bagging (e.g., Random Forest), boosting (e.g., AdaBoost, Gradient Boosting), and stacking.

1.1.2 Approaches of Social Network Analysis

Social network analysis (SNA) is an interdisciplinary field that examines social structures, relationships, and interactions between individuals, groups, or organizations. SNA can be applied to different application areas, including sociology, psychology, anthropology, business, and computer science. There are several approaches to conducting social network analysis, each with its focus and set of methods (Crossley, 2017). Some of the most commonly used approaches are:

Structural Approach: The structural approach focuses on the overall structure of the network, including measures of centrality (i.e., the importance of nodes in the network), density (i.e., the extent to which nodes are connected to each other), and clustering (i.e., the tendency of nodes to form groups). This approach is useful for understanding the global properties of a network and identifying key players within it (Wasserman, 1994).

Actor-Centered Approach: The actor-centered approach emphasizes the characteristics of individual actors, including their attributes (such as age, gender, occupation) and their relationships with other actors in the network. This approach often involves examining patterns of ties or relationships between actors, such as the strength or frequency of their interactions (April, 2009).

Egocentric Approach: The egocentric approach focuses on the perspective of individual actors, such as their social networks, social support, and the resources they receive from their network. This approach typically involves interviewing individuals about their social connections and analyzing the resulting data. The egocentric approach is useful for understanding how individuals are embedded within larger social networks and how they interact with other actors (Mccarty, 2001).

Dynamic Approach: The dynamic approach examines how social networks change over time, including the formation of new ties, the dissolution of existing ties, and the evolution of network structure. This approach often involves longitudinal data collection and analysis and can reveal how social networks respond to changes in the environment (Snijders, 2010).

Mixed Method Approach: By integrating various data sources and analytical techniques, the mixed methods approach aims to offer a more comprehensive understanding of social networks. For example, it might combine survey data with qualitative interviews or use network visualizations alongside statistical analysis. This approach can provide a more nuanced view of social networks and help identify unexpected relationships or patterns (Creswell, 2018).

Overall, social network analysis provides a powerful set of tools for understanding the complex structures that emerge from social interactions. By combining different approaches and methods, researchers can gain a more comprehensive understanding of how social networks function and evolve over time. Following are the key Research areas related SNA:

Identifying key players or influencers: SNA is helpful in identifying individuals or groups that play important roles in a social network, such as key players or influencers (Borgatti S. P., 2016). This information can be used in marketing, politics, or other areas to target specific individuals or groups.

Studying the diffusion of information: SNA can be used to study how information spreads through a social network (Katherine, 2012). By examining how information moves from one person or group to another, scholars and researchers can better understand the factors that affect the transmission of data.

Analyzing the impact of social network interventions: SNA can be used to evaluate the effects of social network interventions (Reingen, 1994), such as introducing a new policy or program. Researchers can determine whether the intervention significantly impacted the network by analyzing the network structure before and after the intervention.

Studying the formation of social networks: SNA can be used to study how social networks form and evolve. By analyzing the connections/links between individuals or groups, researchers can gain insights into the factors that drive the formation and evolution of social networks (Reingen, 1994) .

Predicting behavior or outcomes: By examining the structure of a social network and the attributes of its constituent individuals or groups (Borgatti S. P., 2016)., Social Network Analysis (SNA) can be employed to predict future behavior or outcomes within the network. Researchers can use this technique to predict how the network will evolve based on its current state and underlying patterns.

1.1.3 Applications of Social Network Analysis

SNA has various uses in various fields, and its applications continue to grow as more data becomes available, and new analysis techniques are developed. As a result, SNA has multiple applications in different fields, including:

Sociology and Anthropology: SNA is used to analyze social structures such as communities, social groups, and organizations (Scott J. W., 1996). It helps in understanding the patterns of social interactions, identifying key actors, and studying the dynamics of social networks.

Marketing: SNA analyzes customer relationships and identifies key influencers in social networks. It can also be used to study the diffusion of information and product adoption (J., 2001).

Politics: SNA is used to study political networks, such as political parties, interest groups, and lobbying organizations (Scott J. , 2012). It helps understand the power dynamics within these networks and identify influential actors.

Health and Medicine: SNA is used to study the spread of diseases and the diffusion of health information (Celentano, 2010). Social Network Analysis (SNA) can also be utilized to scrutinize the social networks of healthcare providers and how they affect patient outcomes.

Business and Management: SNA is used to study the structure of organizations, including communication patterns, collaboration, and influence (Borgatti S. P., 2016). It can also identify influencers and their impact on organizational performance.

Crime and Terrorism: SNA is used to study criminal networks, such as drug trafficking organizations and terrorist cells (McAdam, 1982). It helps identify key actors, their roles, and their relationships.

Education: SNA is used to study student relationships, teacher-student interactions, and peer learning networks (Mao, 2014). It helps identify influential students and teachers and understand the learning process dynamics.

Sports: SNA is used to study the relationships among athletes, coaches, and teams (Sarmento, 2014). It helps in identifying influential players and coaches and understanding the dynamics of team performance

These are just some examples of the diverse applications of SNA. However, the work focuses on the most prominent application, i.e., online marketing.

1.1.4 Online Marketing

Social Network Analysis these days is mostly prevalent for the use of online marketing, opinion mining, and recommendation system for any product and services but the concept of opinion leader mining can be applicable in different application areas like online marketing, healthcare system, politics, agricultural field. Social network analysis (SNA) can be a valuable tool in online marketing, as it can help identify and analyze relationships between individuals, groups, or organizations online. Here are some ways in which SNA can be applied in online marketing as discussed below:

1. Influencer Identification: SNA can help identify influential individuals or groups in a social network with a high degree of connectivity and who can potentially amplify marketing messages to a broader audience (Zhao Y. K., 2018). By analyzing different centrality measures of the network, marketers can predict the key opinion leaders and influencers.

2. Brand Advocacy: SNA can predict brand advocates or individuals who are highly connected in a social network and have a positive attitude towards a brand (Brown, 2007). Marketers can leverage these advocates to promote the brand and increase its reach and engagement.

3. Social Listening: SNA can be used to monitor social media conversations and track how information flows through a network (Kim A. J., 2012). By analyzing network topology, marketers can identify patterns of information diffusion and track the spread of marketing messages.

4. Targeted Advertising: SNA can be used to find out the subgroups within a social network based on shared characteristics, interests, or behaviors. By targeting these subgroups with customized marketing messages, marketers can increase the effectiveness of their advertising campaigns (Borgatti S. P., 2009).

5. Community Building: SNA can be used to find out the groups of individuals who are highly connected within a social network and who share common interests or behaviors. By fostering engagement and building relationships with these communities (McAlexander, 2002), marketers can increase brand loyalty and advocacy

Social Network Analysis has some applications related to online marketing , such as follows and in Table 1.1 :

- Recommendations of any services, products etc.
- Viral marketing techniques and implementations
- Opinion mining of customers, whether it's positive, negative or neutral
- Sentiment analysis of customers for any product or service
- Finding opinion leader for viral marketing
- Maximization of influence
- Use of different soft computing algorithms for maximum influence problem

Table 1.1 Applications of SNA for Online Marketing

S. No.	Applications	Author
1.	Online targeting for advertisement	(Li Y. Z., 2015)
2.	Information flow driven personalized Recommendation	(Song, 2006)

3.	Viral Marketing in SN	(Leskovec J. A., 2007) (Chen W., 2010)
4.	User Ranking in SN related to Healthcare	(Tang, 2012)
5.	Social Networks Analysis for influence/diffusion	(S. Peng, 2017)
6.	Analysis of Social Network related to academics	(Dietz, 2007) (Guo Zhen, 2014)

Overall, SNA can provide understanding of the structure and dynamics of social networks in the online space, which can inform marketing strategies and tactics. However, it is important to use SNA in conjunction with other marketing research methods and to interpret the results in context.

1.2 Research Objectives

The proposed research work is aimed at :

1. To identify the correlation between opinion leadership and other measures other than the centrality measures which are used to measure eminence and centrality in networks.

2. To identify more specific context-based rules and to construct a comprehensive ‘‘reputation and trust-based system’’ after identifying bidirectional effects between influence and trust.

3. To identify the performance of evolutionary algorithms that maximize influence spread compared to other existing algorithms.

1.3 Contributions of Research Work

The significant contribution of the research work is summarized as follows:

- The DeepWalk Based Influence Maximization (DWIM) algorithm is superior to other centrality measures used to find the influential seed as this algorithm combines topical and topological features of a node in a network. Hence, its application has potential in diverse domains like online marketing and outlier detection.

- Effectual seed pick algorithm identifies the opinion leaders who are highly influencing potent. Its applicability is not only in pervasive healthcare but can be applied in other domains like politics, product promotion, and service promotion. In pervasive healthcare, the algorithm facilitates the targeted delivery of healthcare services, improving overall effectiveness. Comparative analysis with existing algorithms confirms the superior performance of the proposed approach, as evidenced by higher accuracy, recall, f1-score, and precision in selecting seeds for maximization of influence.
- The multi-neighbor seed selection introduces a seed-dependent influence maximization algorithm that addresses network effects and social reinforcement effects .The algorithm maximizes the acquiring probability by considering the number of neighbours of the seed and utilizing the Neighbour Degree Value(NDV). Incorporating user interaction in seed selection,the framework enhances trust management using opinion leaders.The proposed algorithm proves effective in viral marketing and offers potential for outlier detection in ubiquitous services.This work provides a comprehensive framework for influence maximization and trust management, contributing to advancing social network analysis.

1.4 Outline of the Thesis

- Chapter 1: This chapter provides an overview of Social Network Analysis and the significance of machine learning for SNA. It describes the approaches used in the literature and highlights the prominent key research areas of SNA. Finally, the chapter states the application areas of SNA and discusses various methods applied to the most renowned online marketing application.
- Chapter 2: This chapter provides an overview of the Influence Maximization Problem in the context of viral marketing and its associated challenges. It offers state-of-the-art literature on the Influence Maximization Problem and Identification of Opinion Leaders and describes the various diffusion models proposed in the literature.

- Chapter 3: This chapter explores the viability of deep learning for the influence approximation of users. The proposed method combines topical and topological aspects of a user in the network using collaborative filtering. Then, DeepWalk an algorithm of deep learning, is used to identify the most influential nodes using graph embedding.
- Chapter 4:In this chapter, Effectual Seed Pick Framework for Influence Maximization has been proposed to study the bidirectional effects between influence and trust. The proposed structure effectively addresses different issues related to a large number of patients, and thus, increased influence maximization using seed nodes is helpful for pervasive healthcare.
- Chapter 5:This chapter proposes a Multi Neighbour seed selection approach to enhance the efficacy of Influence Maximization. The proposed method considers the network's memory effect or social reinforcement effect. It presents the Multi neighbor seed selection approach, in which neighbor degree value (NDV) is determined to estimate the influence strength of the seeds.
- Chapter 6: This chapter discusses about use of Evolutionary algorithm to find out Opinion Leader in a social network .
- Chapter 7: This chapter discusses the conclusion and future scope .

Chapter 2 LITERATURE REVIEW

Social networks as unique information or commercial dissemination channels attracted marketers for viral marketing, and in turn, viral marketing as a successful marketing technique piques the curiosity of many researchers. Viral Marketing involves selecting a limited group of early adopters according to mutual belief between local public circles of relatives or companions to optimize the social network's impact expansion. Identifying a fixed fraction of users that can increase the role's transmission is described as an influence maximization problem. The fraction of users is referred to as Opinion Leaders or influencers. A plethora of algorithms have been presented to address this problem. This chapter presented a detailed summary of all existing approaches for influence maximization and identifying opinion leaders. Finally, it examines current research trends as well as prospective innovative avenues.

2.1 Influence Maximization

Nowadays, Social media ought to be an indispensable element in everyone's daily lives (Kempe David, 2003) (Richardson Matthew, 2002). A social network with interconnections is constructed from a group of actors for public social communications (Liu B. , 2011). Social media platforms play a crucial role in disseminating information, opinions, ideas, innovation, rumors, and so on (Centola, 2010) (Nekovee, 2007). This spreading mechanism is advantageous in viral marketing (Leskovec J. A., 2007) (Chen W., 2010), personalized suggestion (Song, 2006), feed ranking (Ienco, 2010), an advertisement with a specific target (Y. Li, 2015), choosing influential twitters (J. Weng, 2010) (E. Bakshy, 2011), selecting interesting weblog (J. Leskovec A. K., 2007), and so on.

Nowadays, people are connected through social media, so the influence maximization problem is prevalent among researchers. For example, consider the scenario of a commercial house's aggressive marketing, which aims to entice consumers to buy a specific product. The most straightforward approach to accomplish vigorous marketing is to choose a group of very prominent people and offer free trials to them. If they like the product, they will tell their friends about it. Then, because of their clout, many of their neighbors will try the product and spread the word to their neighbors.

This flowing progression will carry on, and a significant proportion of people will eventually intend to use that product.

This diffusion of information through the network causes social influence. With the advent of social media, this behavior in a networked system has received much attention (Jonard, 2004) (Banerjee, 2020). People interact in diverse social media and exchange information and views with their peers through word-of-mouth (WOM) communication. Therefore, it is vital for marketing on a large scale to target a chosen collection among those finest influential individuals from relevant public connecting media also spread its clout among numerous prospective individual consumers.

However, in viral marketing, the number of goods sent as special offer products are limited due to financial constraints or budget. As a result, if special offers are available to very noteworthy people, this challenge is to choose significant individuals from the network, and the method will be beneficial. With a restricted budget, Some brand-new agency's marketing staff may choose any limited amount of choices from early consumers besides allowing those to trial their merchandise while using social networks for commercial promotion. This is known adequately as influence maximization. Influence maximization can be defined by locating a K-sized subset of users (commonly referred to as beginnings, when K is a possible value determined ahead of time), leading to the most significant possible dispersal of influence in social networks. Whereas most marketing researchers investigate WOM diffusion progressions by employing theoretical approaches, such as approximate methodologies, experts and others in adjacent domains strive to create influence-diffusion replicas to increase and anticipate the spread of impact.

2.1.1 Basics of Influence Maximization

Social media are real-global networks that are modeled like one diagram plot G (V, E, W), where V denotes this group with vertices (nodules) that represent people from this social media, E means a group of edges that indicates the link between two individuals, also $E \subseteq V \times V$, as well as W Scales are symbolized by related with every edge. Based on this notion Influence Maximization Problem on social networks can also be modeled as an algorithmic representation in which the diffusion model influences the significance and calculation of W in the network.

2.2 Influence Maximization in Social Network

The concept of Influence Maximization was introduced by Pedro and Matt (Domingos P. R., 2001)in 2001. Kempe et al. formalized this as a problem in 2003 (Kempe, 2015). Many researchers have conducted numerous follow-up studies to maximize influence, such as the IRIE model (K. Jung, 2012) and the IPA model (J. Kim, 2013). The aim of the Influence Maximization issue was intended to discover the k most influential seed manipulators. However, the identification of k seeds is a very challenging task. It involves calculating each user's influencing power in the network by analyzing its relationships and interactions with other users in the network.

2.2.1 Challenges of Influence Maximization Problem

A short explanation of the primary research issues associated with the Social Influence Maximization Problem is presented below, which helps determine the solution approach to tackle the problem.

1. *The trade-off between exactness and computing period*: Generally, the Social Influence Maximization Problem (SIM) is calculatingly difficult from a classical and parameterized complexity-theoretic standpoint (Schölkopf, 2012). The obvious way would be to utilize a heuristic method to identify seed nodes in a reasonable amount of time. As a result, seed set creation will take less time. On the other hand, the number of impacted nodes formed by the seed nodes might be arbitrarily low. In this case, it is critical to building algorithms that will run in a reasonable amount of time while also keeping the slit amongst this ideal binge also these binges caused by these seeds are usually chosen via the process as minimal by way of easily.
2. *Breaking over the obstacles of submodularity*: The social impact parameter may not be submodular in many actual scenarios, such as opinion and subject-specific influence maximization. This occurs as a point could change whose condition has shifted between affirmative into unfavorable while conversely. Because those social impact functions lack the submodularity property within the circumstance (A. Goyal, 2011), resolving this Problem is possibly more difficult.
3. *The realism of the issue* (Walker, 2012): Basically, the issue makes numerous conventions, for example, that every chosen seed will behave as expected in the

spreading process, that impacting every network node is equally significant, and so on. In some cases, these assumptions may be impractical. For example, consider the scenario of target advertisement, in which a set of target nodes is chosen instead of all nodes, and its purpose is to increase impact inside those goal nodes. Alternatively, a seed node may underperform in the impact-spreading process because of the diffusion's probabilistic nature. If we relax these assumptions, this may grow increasingly difficult to solve this or related variations.

4. *Scalability*: In real-global social networks, there is a substantial percentage of nodes and a billion edges. Because of this, when it comes to tackling Influence Maximization (IM) and related issues for real-world social media, extensibility should be a key consideration for any solution method (J. Leskovec D. C., 2010).

5. *Theoretical challenges*: Any solution methodology for a calculation issue involves two factors. The first is the calculation period. When the procedure is applied to actual-world difficult situations, It was assessed in terms of compiled code. This computational complexity was the second factor. These were considered along with methodologies limited monotonically. Theoretical study over several computer difficulties was constantly preoccupied with that later part of such issue (Kempe, 2015). As a result, designing algorithms with adequate asymptotic bounds is some hypothetical difficulty for some Influence Maximization issues.

Because of the intrinsic difficulty of the IM Problem, researchers have devised diffusion models to attain close to ideal influence spread across time. The diffusion model determines how an energetic node's influence spreads. The following section explains the diffusion models applied to the IM problem.

2.3 Diffusion Model for Influence Maximization

Various diffusion models have been studied in the literature to predict the influence in the digital world. The below section briefly explained the existing models studied to depict the influence spread in social networks.

There has recently been a large volume of literary works on developing dissemination theories in fields like epidemiology, information systems, and protocols, including data mining. Because the primary goal is to examine the hierarchical components of IM, the part also looks the models often employed in IM.

Different models employ various strategies to represent how a user transitions according to a dormant to an excited phase; it was impacted considering their surroundings. The part concentrations were sole with four typical designs that were often employed over IM issues, viz., the Independent Cascade (IC) design, the Linear Threshold (LT) design, the Triggering (TR) design, and the Time Aware design.

2.3.1 The Independent Cascade Model

The Independent Cascade (IC) (Kempe David, 2003) diffusion model was a well-known and thorough-examined diffusion design as in Figure 2.1. It takes this into account a user v to be stimulated independently, such as everyone their new companions through inserting this impact likelihood $p_{u,v}$ near every verge e = (u, v During sampling interval 0, the IC type dispersion example develops over consecutive stages based just on impact possibilities as well as a seed collection S. With probability pu;v, in step t, each existing user u activates any one their aggressive peers v who are passive now step t - 1. The activation process is analogous to spinning one penny with a likelihood of a head $p_{u,v}$: When this outcome is just a head, v is engaged; alternatively, it's dormant. Remember that each u gets a single opportunity to engage the incoming companions. Then, u remains vigorous so that such stimulation is ended. When no more nodes can be triggered, the diffusion instance ends. If S was equal starting an activated point was chosen, then a random something that which described below was used. Utilized, the S is the estimated amount of individuals who will have an impact on the prevalence of the seed batch-activated nodes, according to this IC model.

Some of the earlier works relied on heuristic probability assignment to compute the influence possibilities where the Weighted cascade is a popular one in which the On edge e = (u, v), it allocates $p_{u,v}$ as $1/d_{in,v}$; thus $d_{in,v}$ was comprehensive in of v. Few recent research suggests learning stimulus probabilities from information, such as transmission acts in social networks. The task of discovering rim chances using previous dissemination activities is formalized, like an example likelihood acceleration issue. In the presence of a chart G = (V, E) also, the collection in actions for propagation by itself uses this Expectation Maximization (EM) technique can calculate

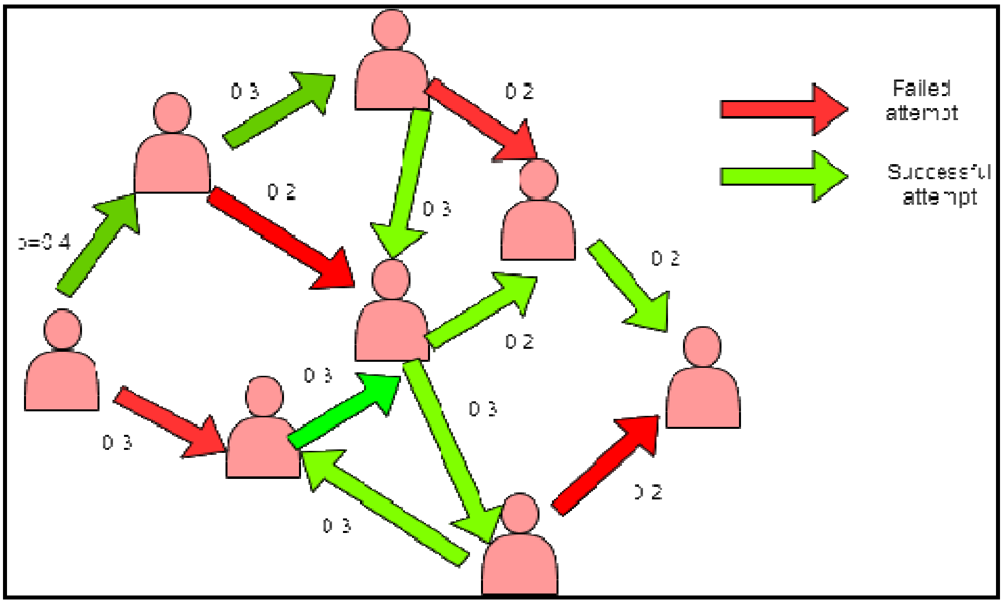


Figure 2.1 Independent Cascade

transmission probability continuously for every $e \in E$ to maximize the total chance of all acts. Following that, the SPINE algorithm was proposed to simultaneously learn the social graph structure and the propagation probability, in conjunction with ideal constraints maximizing its journal chance because of their spread acts being generated by this interpersonal show's architecture. It likewise studied this issue by determining edge probabilities, proposed more adaptable algorithms, and considered this problem in data streams. According to the review below, the authors describe efficient algorithms for estimating probabilities with a single go out of every activity. There are several ongoing research comprehending pattern readability, such as the Probably Approximately Correct (PAC) framework, which analyzes the learning ability of diffusion models and establishes a reduced constraint on the number of rounds involved in learning an IC paradigm based on knowledge technology

2.3.2 The Linear Threshold Model

Linear Threshold Model (Granovetter, 1978) is another inspirational dispersion ideal. The essential principle According to LT, a person's position may get altered from being dormant to proactive provided a "significant" percentage of their arriving companions were engaged as in

Figure 2.2 .Every edge $e = (u, v) \in E$ with LT design is connected with encumbrance $b_{u,v}$. Here N_I (v) is referred to as such group persons v's arriving

neighbors and must satisfy $\Sigma_{u}\ _{\epsilon}N_{I\ of\ v\ ,}\ b_{u,v} \leq 1$. Furthermore, Every client v is then assigned some limits v. When taking another look at a case of molecular spreading, the LT design initially uniformly Every recipient's v variable is selected within that region [0, 1]. The process continues with separate phases. Step 0 marks Members under S are marked proactive, while some are marked dormant. It then successively changes each user's status: All members who were busy in step t-1 remain busy in step t, and other members v are dormant in step t - 1 will be active when the sum quantity of busy companions in N_I (v) is greater than θ_v.

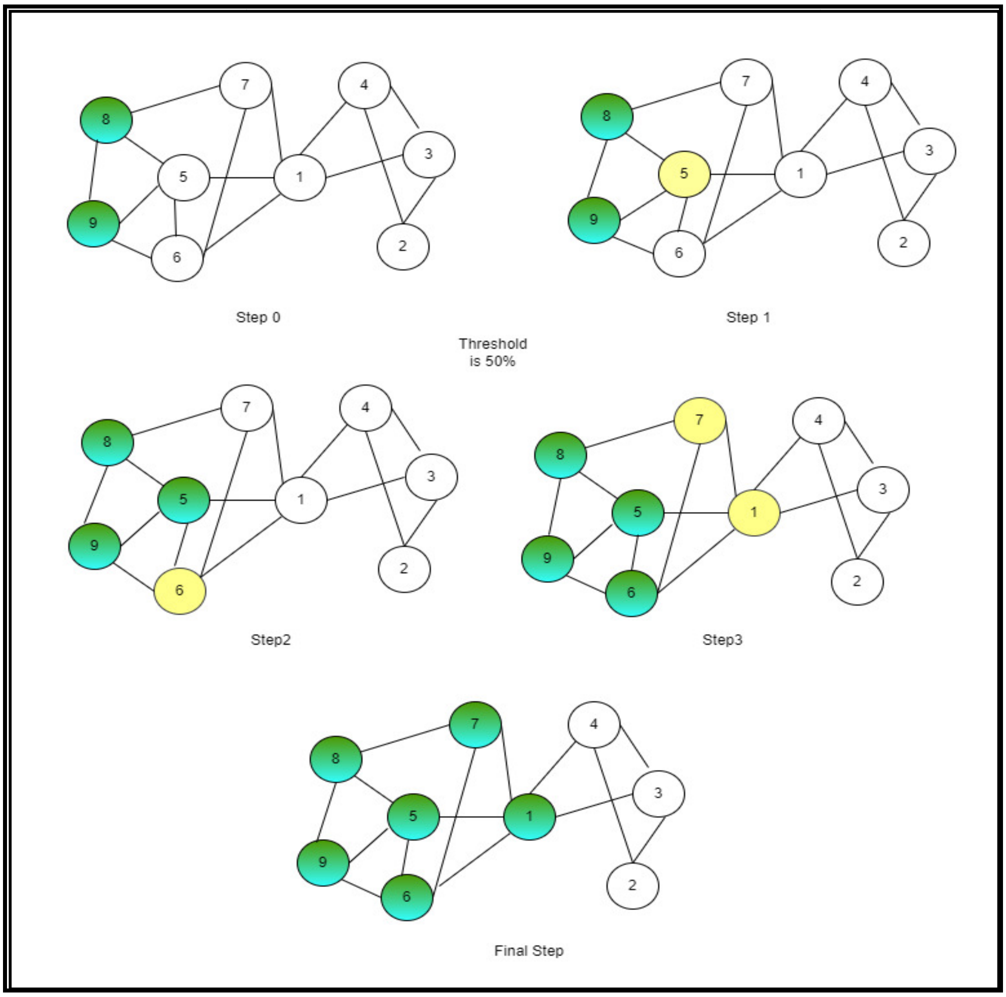


Figure 2.2 Linear Threshold Model

When no more users are triggered, the diffusion instance ends as in .Given numerous examples for drift dynamics, these LT models affect the dispersion on the seeds cluster

S, i.e., σ (S), denotes as predicted the limited connections triggered when S is initially started. Many IM techniques use approximations to determine its mass $b_{u;v}$ on every verge e = (u,v) E, including equally allocating e at random with a likelihood from the collection (0.1,0.01, 0.001). However, here state of the art, there is no information technique enough for the LT paradigm.

2.3.3 The Triggering Model

The triggering model (Cheng, 2013) was proposed to be homogenized with the IC and LT models. The TR model states distribution that transfers v's colleagues are a subcategory of v's acquaintances. In the chance that a neighbor subset may impact v given any user v. The TR version selects a haphazard "activation array" TV for user v based on the previously stated dissemination across selections of v's peers for each iteration of the diffusion process. The process then proceeds in separate steps. A seed set S is used to re-initialize the diffusion instance. Following the acceleration step, a dormant junction v evolves into busy Whether it contains a companion including the selected activating collection, then proceeds to stage t, $T_{v,}$ which was started along the step t - 1. Like IC and LT, its impact on S during TR is likewise the anticipated amount of nodes engaged. It is demonstrated that IC and LT designs have unique instances of the TR model. Several universal models expand IC and LT more than TR. According to the review described below, the IC model is extended to a Decreasing Cascade Model. Like pv, DC denotes the possibility of impact from junction w to junction v for a share Of the total of v's active neighbors (w, S). To represent the declining value situation, DC applies pv(w, S) pv(w; T) aimed at S T.Kempe et al. (Kempe, 2015) offer some General Threshold (GT) designs encompasses IC also the LT designs. GT defines this threshold factor v as $f_v(S)$,here S denotes some set of peers who are engaged of v. When $f_v(S)$ surpasses some assessment v, v enters such dispersion progress. Even though GT design is inapplicable to IM during several cases, it shows that the impact product with GT possesses ability identical monotonic as well as perhaps sub modular characteristics from once your limited processing is harmonic low cost solution as well as the limits were set randomly.

2.3.4 Time-aware Diffusion Models

IC, LT, then TR were period-insensitive designs in which those dispersion ends solitary after no nodes can be triggered. Nevertheless, public relations efforts are frequently duration-sensitive, necessitating the maximization of effect spread within a time limit (Carley, 2015). To address this desire, time-alert models are proposed, and current research is used roughly grouped into the following divisions: 1) discrete-time models, in which diffusion occurs in discrete stages, also 2) continuous-time designs, in which another's procedure person modifying additional occurs throughout time.

Discrete-time models enhance IC by simulating the practice of spreading through one site towards another as a discrete random variable across duration steps. The model is fundamentally related to IC and LT because diffusion occurs in discrete phases. One node's procedure impacting another is fundamentally constant in real-world circumstances. Continuous-time diffusion models are presented can elicit its qualities. The Continuous-Time IC model considers the Probability of paired spreading across junctions as an unceasing time distribution. Known a node u's activation time tu, the probability of u triggering the nearest companion v during all time moment $t_v > t_u$ was described by $p(t_v|t_u; u,v)$, where u,v are the moment impact dissemination attribute used should figure out the degree of swaying between u till v. T > 0 in the presence of a predetermined stopping time , every case of propagation of CT comes to a halt when no further nodes are triggered before T. The exponential model is a common choice for the time-aware influence distribution, with $p(t_v|t_u;\alpha_{u,v}) = \alpha_{u,v}.\ e^{\alpha_{u,v}(t_v - t_u)}$ if $t_v > t_u$ then 0 otherwise. Thus Dyna Diffuse design assumes that node propagation rates drop with duration progressively. In the presence of a node u that is stimulated on duration t also a significant competitive advantage *(u→ v)* along the propagation level r(u, v), then the possibility of diffusion via *u* towards *v* duration t_0 as$1\text{-}e^{r(u,v)(t_0\text{-}t)}$ where $(t_0 > t)$. Dyna Diffuse, and on the other hand, limits the diffusion duration by adjusting toward a predetermined value threshold $T > 0$. The equivalence of Dyna Diffuse and CT is demonstrated. Some studies have also been done on simulating the temporal changing aspects of the dissemination procedure. Though the main objective of such studies seems to be to explain Monitoring evidence with a timing impact on behavior, neither IM techniques are created on those models towards the finest ability. Because the study focuses on the hierarchical side, both IM and CT are very extensively used moment systematic IM methodical investigation, further will compare several IM methodologies using CT design.

2.3.5 Non-Progressive Diffusion Representations

Non-progressive diffusion models (Schölkop, 2012) (J.Tang, 2009) (M. Gomez-Rodriguez, 2011) include several diffusion models. The primary distinction between progressing with non-progressing systems is that effective model nodes can be deactivated. The SIR/SIS system and Votersystem are two examples of non-progressing diffusion systems. Other IM procedures were planned in both models. This questionnaire's goal was to evaluate the computational facets of IM and that generally known systems within that field include LT, IC, TR, and CT.

Bhagat et al. (Bhagat, 2012) presented the LTC model, a propagation system that considered these variables. Then dignified and proved that acceptance maximization, by way of opposed to influence maximization, remains NP-hard. In addition, the predicted propagation factor of adaptation below LTC system remains monotone and submodular, allowing the conventional greedy technique to be employed toward obtaining such an approximation answer.

Goyal et al. (Goyal A. B., 2010) investigated ways to learn those probabilities from the history of previous propagations. They offered stationary as well as duration-reliant models for snatching, and algorithms for learning and validating the models' parameters. Goyal et al. approach is improved to reduce scans through a list of activities, an essential source for the interpretation issue influence possibilities. That must be because the operation report may became relatively big. Here executed a large number of trials to put our learning algorithms to the test.

Chen et al. (W.Chen, 2011) presented a comparison to the autonomous cascade design that considers the emergence and the spread of views. A fresh design includes some explicit parameter called quality factor for simulating real behavior of individuals being dissatisfied with a material due to flaws in the source, as well as prejudice against negativity (Critical comments typically outnumber good thoughts). Chen et al approach is well-known in this social psychology literature, and a robust approach designed method calculated the impact on graph architectures famous in such social psychology literature. Moreover, this algorithm served as the foundation for developing a heuristic algorithm for influence maximization.

He et al. (X. He, 2012)investigated the competitive linear threshold model with an influence-blocking maximization issue (IBM). The greedy approximation algorithm is available because of the optimal solution for IBM. The problem was a low-cost solution under this CLT model. Then construct an effective program CLDAG to defeat the optimization technique leisurely. CLDAG also improves other optimization methods for the IBM issue, such as the closeness algorithm, which chooses straight neighbors with noxious seeds, demonstrating that CLDAG is a reliable and stable methodology.

Gayraud et al. (N. T. H. Gayraud, 2015) examined the influence issue maximization on dynamic networks, where the network evolves while the diffusion process is taking place and creates being challenged Evolving Independent Cascade (EIC) as well as Evolving Linear Threshold (ELT) diffusion systems. Gayraud et al. work demonstrated significant variations between diffusion in static and changing graphs, both in theory and practice and that neglecting the network's dynamic nature is incorrect. This dynamic model, which considered the relevance of timing in dissemination, resulted in a fundamentally different diffusion process.

Kimura et al. (Kimura, 2006) developed the SPM and SP1M, different natural scenarios for the spread of data in a social media platform, in which the outcome *(A)* of an objective set *A* can be easily evaluated in a plausible setting. The natural greedy method's accuracy is promised for the influence maximization challenges. In this paradigm, a dormant node can only become proactive by following the quickest route from the initially active nodes. The SP1 design was then somewhat this SP Version has been changed, developed by the same researcher, and asserts how a dormant node has a probability of triggering at a given period.

Chen et al. (Chen W. W., 2009) explored efficient maximum influence via two opposing viewpoints. The major is to present new Greedy as an improvement to the basic optimization technique, while the next is to provide Probabilistic reasoning for Degree Discount as an improvement to influence spread. To calculate the most significant impact established during the IC as well as LT systems, two approaches for extrapolation, PMIA, and LDAG, were planned (Goyal A. B., 2010) (Chen W. W., 2009) (Leskovec J. K., 2007). Although it has been demonstrated in LDAG that according to its LT paradigm, calculating impact distribution over the DAG has a proportional computation period, also the technique about the regional DAG building remains offered towards further minimizing the calculation period.

Chen et al. (Chen W. W., 2010) along with Liu et al. (Chen W. Y., 2010), presented the moment impact profit-maximizing issue, wherein its framework acts takes into account period constraints, and demonstrated its repetition of submodularity this same principle of impact dispersion with a limited time, as well as suggested techniques for resolving the issue. Wang et al. (W. Chen et al., 2012) develop IMIC-OC, an autonomous sequence impact maximizing approach to compute positive influence.

Game theory is a profit-maximizing strategy. The research on game strategy is restricted to several persons; otherwise, organizations must adhere to rigorous guidelines. It always takes advantage of the adversary's plan to increase gain. Because among economic factors, rewards, and decision-making strategy, a piece of knowledge is either distributed or not spread. For many years, the game strategy was employed in typical network investigations. Li,et al. (Li J. X., 2012) suggested a game-based network expected utility design group behaviors. They argue that individual information behavior's micro-level characteristics are far more intricate than macro-level features.

Due to their sociality and unpredictability, group behaviors in a network frequently exhibits high uncertainty. The evolution game model is well suited to dealing with the data-adaptive issue dissemination in social media. The overall analysis has been shown in Table 2.1. Our study of IC, LT, and GT systems concludes that the three systems were employed to influence spread or maximization. For instance, with maximizing, the scientist's viewpoints might class as macro or micro. This macro perspective examines coming dissemination through the ocular of a network protocol, whereas the close look examines upcoming diffusion through the ocular of user behavior.

Table 2.1 Different Models Used to Maximize the Influence on Social Networks

Basic model	Advantages	Disadvantages
Independent Cascade Model (Kempe David, 2003)	• The likelihood of dissemination projection • Influence maximization • Add up the execution hours and storage usage. • The likelihood of information diffusion episodes	• Cannot affect factors • Not focused on the prediction of information spread • Only focused information content, not user profile
Linear Threshold Model (Granovetter, 1978)	• Focused information content, as well as the user profile • Can focus based on topics • Select the most significant influence nodes • Select the nodes based on the activation threshold • Predict the topic distribution	• Not predicting the information spread • Not predict the information diffusion in a dynamic network
Game theory Model (Li J. X., 2012)	• Focused relationship and cost • Focused individual information behavior • Predict the information spread and diffusion in a dynamic network • Also, predict the relationship between the users.	• Robust • Sender-center model

We know through a review of the writings. When those three systems are compared, IC models are sender-centered, concentrating exclusively on data senders, while LT systems are receiver-centered. GT systems were highly impartial that considered the entire network gain.

2.4 Approximation Based Models

These techniques provided the influence spread's lamest limit. However, most had technical challenges; as the network area spreads, so does proper running time. Nevertheless, several methods in exponential growth limits existed for that domain that was close to optimum.

Kempe et al. (Kempe, 2015) were initial members for addressing every subject of social impact. Therefore as a feasible solution, maximizing and investigating the calculational challenges using both diffusion systems, LT and IC. They assumed using their investigations because the overall societal situation impact performs σ (.), was both bland and submodular, so a greedy technique for seed set selection was presented.

In most cases, Kempe et al. studies' main goal was to alleviate the adaptability issue caused by using the Basic Greedy algorithmic methodology. Unfortunately, some landed on heuristics, which resulted in a solution far from optimal. However, some experiments substantially decreased the limitation, not losing the approach rate. Researchers also included techniques that could offer estimation certainty in this section.

Leskovec et al. (Leskovec J. K., 2007) suggested the Cost-Effective Lazy Forward (CELF) approach that benefited from this social influence function's submodularity. The core notion of their research was that a node's present iteration's modest improvement in impact dissemination could not be more significant than its small benefit in earlier versions. With this concept, Those who drastically decrease the number of evaluations of such influence estimate terms (σ (.)), significantly enhancing processing duration while maintaining similar asymptotic difficulty as the Basic Greedy methodology.

Goyal et al (Goyal A. L., 2011) accessible CELF++, the optimized version of CELF that used the submodularity capability of the social influence parameter. The vital principle in CELF++ is that if the node has the most significant modest improvement from scanned users so far in. The present iteration is included field crop with perhaps its most recent version, and u's optimal profit in (u) concerns S's previous best, and u should not need to be mathematical within the following iterative process. Although Its exponential difficulty does not change, the results demonstrated that CELF++ is 35–55 percent quicker than CELF.

Cheng et al. (S. Cheng, 2013) created a stationary Greedy set of rules for addressing SIM Problems that guarantees assured accuracy and excellent scalability. Cheng et al algorithm is divided into two steps. In the first stage, R Monte Carlo snapshots of the social network are obtained, with all edge (uv) chosen due to the corresponding diffusion possibility p_{uv}. It begins with a null seed set, a node having the highest mid-range boundary profit in influence spread across every tested print must choose like seed node in the second stage. This technique will be repeated until k nodes were determined.

Chen et al. (Chen W. W., 2010) along with Wang et al. (C. Wang, 2012) the maximum influence arborescence (MIA) also Prefix excluding MIA (PMIA) impacted network topologies, and also the dissemination possibility estimated out about donor cluster toward a module that is not a seeding link via combining some impact possibilities among rims along the quickest route, were presented. The Maximum Influence Path has a very high likelihood of propagation, as well as everyone anticipated that influence spreads only by local arborescence. Because of this, the model is known as MIA. At PMIA (Prefix without MIA) archetypal, Aimed at every source si, the greatest impact route to neighboring nodes must exclude all seeds that came before si. In this basis of these two diffusion scenarios, they suggested greedy methods for seed set selection. According to the results, together MIA as well as PMIA can attain a great rate with adaptability

Borgs et al. (C. Borgs, 2014) offered a completely new strategy to solve the SIM system due to the LT system by oppositely employing sequencing. Aside from the MCS proceeds, it was a novel method to assess impact diffusion. The technique is close to ideal. It operates in two stages. The input social network is used to build a hypergraph (H) stochastically during the first phase. The seed set selection is the focus of the second phase. That was accomplished by continuously selecting this node using the most significant degree of H and eliminating it and its incidence edges from H. The seed group for diffusion is the k-element set obtained in this manner.

Cohen et al. (E. Cohen, 2014) planned the Sketch-Based Influence Maximization (SKIM) algorithm. This makes your Simple Optimization technique better while assuring that Any cluster you select again for seed production provides a residual benefit that comes near an optimum for each repetition, with a suitable enough likelihood of expectancy.

The results revealed that SKIM outperforms, Two-Phase Influence Maximization (TIM), Impact Ranking and Impact Estimation (IRIE), and others regarding scalability without sacrificing influence spread.

Tang et al. (Y. Tang, 2014) premeditated this Two-phase Influence Maximization (TIM) procedure. This algorithm contains two steps, as the name implies. First, TIM calculates a bottom limit on the most significant anticipated diffusion of influence throughout every k-sized set in the first phase. It then estimates a component using these bounds. During the following sections, a sample of reverse reachability (RR) sets was selected in the social media. Next, as a result, That generates a k-sized seed set that includes as the most incredible range of RR sets possible. A heuristic is presented for enhancing the processing duration of TIM by taking every RR sets formed in quite an initial stage of phase 2 of TIM when input data. To reduce the running time of TIM, a heuristic was presented that grosses entirely of such RR collections formed provides stimuli through the interim solution of synthesis second stage in TIM and utilizes a ruthless attitude to this highest attention issue to select the seed set. TIM+ is the name given to this enhanced version of TIM. According to the results, TIM+ is twice as quick as TIM.

For the Topic-aware Viral Marketing (TVM) challenge, Nguyen et al. (H. T. Nguyen, Stop-and-Stare: Optimal sampling algorithms for viral marketing in billion-scale networks) created this Stop-and-Stare Algorithm (SSA) as well as their volatility variation DSSA. With slight modifications, this solution methodology may have been used to answer the SIM issue, and it was demonstrated because their methods use the fewest RR set sample approximation. Nguyen et al. (H. T. Nguyen, 2017) projected the Billion-scale Cost-award Targeted (BCT) set of rules to solve the cost-aware targeted viral marketing (CTVM) problem that they presented. The result Technique can be used to solve the SIM issue.

As a result of this discussion, it is crucial to take notice of the adaptability issue. Subsequent research has minimized the problems found with the Basic Greedy method. But conversely, as the online social data set has grown, creating scalable algorithms has remained a significant focus.

2.5 Community Based Models

The majority of a society pattern emerged in actual social media. The society is essentially a subset of nodes heavily linked to one another but weakly linked with remaining network nodes. As a result, a community based solution framework (CBSF) for SIM Problems has been established.

Wang et al. (Chen W. W., 2010) presented a greedy approach based on the community to solve the SIM issue. The technique is divided into dual steps: identifying organizations based on knowledge dissemination and then choosing organizations to discover connected components. In addition, the method performed level reduction and arbitrary optimization.

Chen et al. (Y. C. Chen, 2014) invented CIM, a CBSF for tackling the SIM Problem. They picked several individual seed sets in every group by utilizing the

Table 2.2 The Overview of the Different Algorithms Used

Methods	Significance	Limitation
SimpleGreedy (Kempe David, 2003)	• Guarantees of precision	• Large processing duration with restricted adaptability
CELF (J. Leskovec A. K., 2007)	• Problems that are simple to implement • Provides a guarantee of approximation	• Not usable for practical problems • Computational time is still high, and there is insufficient scalability.
CELF++ (A. Goyal, 2011)	• Provides a guarantee of approximation • The processing duration is smaller for Simple Greedy and CELF.	• Computational time is insufficient for working on real-world social networks.
MIA, PMIA (Chen W. W., 2010) (C. Wang, 2012)	• Provides a guarantee of approximation • The first methodology centered upon this MIA diffusion model	• Because the algorithm is parameter sensitive, an acceptable result must be obtained. • The parameter must be properly calibrated.
Static Greedy. (S. Cheng, 2013)	• Provides a guarantee of approximation • To a certain level, it solves that scalability-accuracy dilemma.	• Dealing on massive data needs more than just adaptability. The capacity needed is higher than in Basic Greedy. • Working on huge files needs more than flexibility.

Borg et al.'s Method (C. Borgs, 2014)	• Offers assurance of approximation • This technique represents a methodological advance in the research on impact optimization.	• Three factors govern the approach. • Lack of practical experimentation
SKIM (E. Cohen, 2014)	• Provides approximation bound • Experimentation with large datasets	• These techniques have properties. So, proper tuning of these parameters may be an issue.
TIM+, IMM (Y. Tang, 2014)	• As opposed to previous techniques, it is easily customizable.	• The Techniques consist of Factors. So, factor tuning may be an issue
Stop-and-Stare (H. T. Nguyen, 2016)	• Provides the approximation smallest number of observations • IMM is even quicker.	• The methods are made up of factors. As a result, parameter tweaking might be a difficulty.
BCT (H. T. Nguyen, 2017)	• A generalized framework for SIM and related issues is presented.	• Many features are included in the suggested technique. To get good result, characteristics must be adjusted.

community structure; Researchers chose the finalized seed set for dispersion among the prospective seed sets. As a result, CIM outperformed various cutting in terms of effect spread, edge computational approaches such as CDH-Kcut, CDH-SHRINK, and to the greatest extent possible, was superior.

Rahim Khan et al. (K. Rahimkhani, 2015) devised and termed ComPath, a CBSF for addressing SIM problems using the LT Model. First, they detected communities using the Speaker-Listener Label Propagation Algorithm (SLPA) introduced by Xie et al. (J. Xie, 2011), after which the most influential groups were determined and possible seed nodes. Finally, they chose from the final seed set with the candidate seed set according to the intra-distance between the nodes in the individual seed set. As a result, ComPath has the potential to outpace CELF, CELF++, and LDGA, the most significant level criterion and the most incredible Relevance optimization method.

SLPA program. They proposed a seed selection approach that computes the influence of Goyal et al. (A. Goyal, 2011) and devised a strategy for spreading. INCIM may outpace various cutting-edge approaches such as LDGA, SIMPATH, IPA (striking similarity approach aimed at the SIM Problems stated through (Kim J. K., 2013), lofty Sitelinks, also an elevated optimization were all examples of algorithms.

Shang et al. (Shang, 2017) devised and termed CoFIMa to resolve the SIM issue. The study used a diffusion system that operates in 2 steps. The initial stage seed set S was prolonged to include S's neighbor nodes, which are often assigned to various communities. This spread of the impact within groups was then calculated in the second phase. As a result, CoFIM could achieve a more remarkable influence spread than IPA, TIM+, MDH, and IMM.

Li et al. (X. Li, 2018)offered some communal-oriented solutions to solve such SIM issues in which operators are geographically restricted. They created a social influence-based organization discovery technique using harmonic segmentation and a seed-choosing methodology considering the society influence index. According to the findings, the method is much as effective than various state techniques while achieving roughly the same impact distribution. Table 2.3 Community Based Problem SolvingTable 2.3 shows the overview of Community based problem-solving.

Table 2.3 Community Based Problem Solving

Technique	Significance	Limitation
Local Arborescence structures method. (Chen W. W., 2010)	• First, look for community-based solutions to the SIM problem.	• There is no guarantee of approximation.
CIM (Y. C. Chen, 2014) i	• When contrasted with several relative importance strategies, adaptability seems superior.	• There is no guarantee of approximation. • There are a few parameters. • The effectiveness of this method is dependent on the parameters being properly tuned.
ComPath (K. Rahimkhani, 2015)	• Sustainability is higher when relative to other relative importance strategies, CELF and LDGA.	• There is no guarantee of approximation.
INCIM (A. Bozorgi, 2016)	• Compared to numerous Probabilistic reasoning predicated on prominences, LDGA, and Simpath, adaptability had been improved.	• There is no guarantee of approximation.
CoFIM (Shang, 2017)	• High scalability • Compared with several similar group initiatives, this method	• There is no guarantee of approximation.

	spreads influence more effectively.	
Community-based Seeds Selection (CSS) Algorithm (X. Li, 2018)	• In inquiries about impact enhancement, ‘geographical preference' is considered.	• There is no guarantee of approximation.

There has recently been numerous research on the influence issue by utilizing the community structure while considering that the network's links may be unknown at first, non-submodular influence function, and so on. It is vital to highlight that the methodology is essentially a set of heuristics. However, these solutions leverage the underlying social network's community detection to reduce the SIM Dilemma to the grassroots layer.

2.6 Heuristics Based Models

Influence spread is not limited in any way by algorithms in this category as discussed in Table 2.4. On the other hand, most of them are more scalable and run faster than the algorithms in the previous group.

Preeminence is a very well metric of a node's significance in a system in analysis tools (A. Landherr, 2010). Many centrality-based strategies for the SIM Problem have been suggested in the writings, such as the Maximum Degree Heuristic (MDH) (consider K nodes with the most significant level like a starter node), the High Clustering Coefficient Heuristic (HCH) (consider K vertices having greatest agglomeration correlation coefficients) (B. M. Tabak, 2014), and the Elevated Search Ranking Algorithmic (choose those institution or organization having that most incredible search ranking score).

Using Impact ranking (IR) and Impact estimating (IE), Jung et al. (K. Jung, 2012) suggested this method to resolve the SIM Issue in IC itself and Expansion, IC-N (autonomous destructive comment avalanche). Like belief propagation, they used an overall effect standing strategy to build their method. Each node's effect spread will overlap if we select the top k. A simple influence approximation procedure was used to forecast surplus influence control of seed on another node in this network, thereby bypassing this issue altogether. Analyzing with MDH, Page Rank, PMIA, and other assumptions, IRIE has a better influence distribution.

For the SIM Problem, Galhotra et al. (Galhotra S. A., 2015) developed an extremely extensible algorithm. Every node u V (G) receives a rating benefit due to this method (the normalized total is all set most basic routes of frame size d beginning at a certain point). In contrast to SIMPATH, ASIM has a duration of O (kd(m + n)). A similar influence spread may be achieved with ASIM, despite requiring less computing time and memory.

In their OCI model, Galhotra et al. (Galhotra S. A., 2016) took into account unfavorable opinions. Using OCI as a starting point, they presented two fast heuristics for this problem: Opinion Spread Influence Maximization (OSIM) and EaSyIm, both of which run in O(kD(m + n)). In all cases, there are two phases to the algorithms. For each node in the first phase, a score is assigned predicated upon their distribution of effect across everyone pathway that originate at that node. The node processing phase ,next phase in the procedure. They are chosen as seed nodes because they have the highest score.

A heuristic strategy was presented by Cordasco et al. for picking the target set in an undirected social network. Trees, cycles and entire graphs can be solved in the most efficient manner possible using this technique. On the other hand, this heuristic performs substantially better for real-life social networks than the different techniques in the literature.

Using Metaheuristic optimization methods fall under this category, and Several of these were founded upon techniques towards evolutionary computing. Bucur et al. (Squillero, 2016) (Bucur, 2016)used a genetic method to solve the SIM

Problem. This study revealed the feasibility of finding an approximate solution to the influence spread problem using a primary genetic operator.

Jiang et al. (Jiang Q. S., 2011)developed a replicated strengthening-oriented technique for addressing the SIM issue based on the IC system. Compared to existing heuristic methods, the suggested techniques perform quickly, based on the information.

A new greedy approach devised by Tsai et al. (Tsai, 2016)should address this SIM issue according to its IC paradigm by merging some the evolutionary program with the new suggested methodology. According to their findings, the GNA can disseminate influence 10 percent more than the genetic algorithm.

A discrete particle swarm algorithmic technique to tackle the SIM Problem was suggested by Gong et al (Gong, 2016). For establishing the seed set, The level reduction algorithm was employed, and the local influence estimation (LIE) parameter was used can simulate the two-hop affect It was also designed to accelerate

Table 2.4 Heuristic Based Problem Solving

Algorithm	Advantages	Disadvantages
Centrality-Based Heuristics (A. Landherr, 2010)	• Highly scalable	• Poor solution quality
Degree Discount Heuristic (DDH) (Chen W. W., 2010)	• Simple to put into action • Seed set efficiency outperforms arbitrary and relative importance algorithms.	• Its seed-setting grade was insufficient sufficient achieve the appropriate degree or impact.
SIMPATH (C. Borgs, 2014)	• Great strides have been made in terms of capacity.	• Only works in LT, the diffusion Algorithm model comprises parameters, and its practical

	• The seed set quality has also improved.	performance is determined by how well the parameters are tuned.
LDGA (J.Tang, 2009)	• Substantial progress has been achieved in terms of scalability.	Performs only on LT Model diffusion
IRIE (Xuanhao Chen, 2021)	• Functions in the face of a negative impact as well. • Significant progress has been made in terms of scalability.	• There is no guarantee of approximation.
ASIM (Chen W. W., 2010)	• Scalability has been improved significantly	• Will neither give an assurance on accuracy
EaSyIm107 (Wang C. C., 2012)	• Considers happening presence of a negative impact	• Will neither give an assurance on accuracy
Evangelism in SN (G. Cordasco, 2018)	• Ensures the best loops with tree lings. • Working with massive data needs more than just scalability.	• There is no guarantee of approximation.

the computation completion. It is concluded from the data that this methodology outperforms CELF++ while requiring less calculation time.

The studies on heuristic approaches will now be shortened in this section Because centrality-based heuristics (CBHs) primarily evaluate network structure, generated impact spread in most circumstances is far lower than that of other cutting-edge methods. DDH, on the other hand, performs marginally better than other CBHs because it limits the selection of two nearby nodes. SIMPATH is helpful for seed selection since This features a consumer option to allow you to regulate the exchange among precision and total time. It is quicker since LDAG is rooted in the concept of calculation of influence spread in DAGs.

2.7 Context Aware Models

Context aware IM research has emerged in current trends. Spreading on the conventional IM difficulty, context-aware IM analysis takes into account ambient characteristics, including subject, timing, as well as region, to "personalize" the IM solution for its purposes. Each part investigates IM procedures in any number more circumstances.

Topic-aware Influence Maximization (TAIM) expanded the general IM issue, considering into account of the subjects of the item being disseminated. To solemnize the notion, TAIM incorporated themes to symbolize together item features and consumer considerations and preferences, the influence to be dependent upon both the seed set S and the topics. Next, a query of matters seeks the ideal seed collections which optimized topic-aware impact. It divided all current TAIM investigations into two groups. Initial With particular matter users, the class includes IM. Assumes protuberances to be particular matters also seek to optimize impact upon a subgroup or people pertinent for the inquiry subjects. This latter was IM, which systematizes this notion because all borders were particular matters tried en route to exploiting impact using a novel topic-dependent modeling approach.

Some TAIM research (Guo J. Z., 2013) (Y. Li, 2015) (Nguyen, 2016) concentrated on exerting maximum impact on individuals who seem to be connected to the inquiry topics, i.e., particular matter prospects. Technically, these investigations established the idea of benefit to distinguish the users. The influence is urged to be calculated as the predicted sum of the benefits of the stimulated operators, which is called a battered influence. According to that, here present approaches for determining the seed collections which optimized such battered influence in analytical models for fundamental.

Recent studies (Aslay, 2014) (Shuo Chen, 2015) (Chen W. L., 2016) (Fan, 2018) (Li Y. T., 2017)concentrated on a Particular matter diffusion system on IM. The current design is based on the idea that every edge e = (u; v) exists somewhere amongst multiple individuals u and v topic dependant. That was due to the given notion that v can be triggered via u in certain subjects while staying dormant in another. The main difficulty here is the vast quantity of possible inquiries, everyone correlates towards certain subject dissemination and yields various probability charts. Such a simplistic approach would first calculate $P_{u,v}$ at every periphery based just on request, & subsequently apply as previously stated IM techniques to such encouraged chart as previously stated. The method would be too luxurious, necessitating the development of cost-effective strategies to sustain operational topic-aware IM questions.

Classical IM methods presume every diffusion instance finishes merely once there are no more nodes to affect. So the diffusion process might take a long time to conclude, and the hypothesis is irrational. To establish a this influence constrained by duration, moment influence maximization is presented. First, distinct moment scattering designs (Lee, 2012) use the distinct distribution stage also as moment assessment &limit this diffusion process's most significant stage. Liu et al. (Liu B. C., 2012) (Lee, 2012) (B. Liu, 2014) then Lee et al. (Lee, 2012) independently suggest this LAIC as well as CT-IC simultaneously, although both designs are remarkably similar.

Because of the increasing prominence of site social media, this current technique using place term advertising has sparked academic interest in local-aware impact maximization (LAIM). This prior knowledge behind LAIM can increase the overall effect among regionally related individuals rather than certain customers around standard IM parameters. En route to address this issue, many ways to combine general location element systems programs have been developed (Guo L. Z., 2017) (Wang X. D., 2018) (Wang X. Z., 2016) (Zhou T. C., 2015). (Agarwal, 2020) Models to estimate, social influence namely Time Decay Features Cascade Model (TDF-C) and Time Decay Features Cascade Threshold Model (TDF-CT) have also been proposed in literature.

2.8 Evolutionary Algorithm Based Solution for Influence Maximization

Ma and Liu (Liu N. M., 2014)analyzed the characteristics of 3 seed networks and identified OL in the Fukushima nuclear crisis using the SuperedgeRank algorithm. Jiang et al. (Jiang L. G., 2013) used MapReduce to develop and deploy a BBS OL mining solution founded on an adaptive Page Rank algorithm. (C. C. Huang, 2020) took the main procedure of Internet browsing–the Page Rank model–and proposed an approach to determine Internet opinion leaders by integrating the understanding of the effect of linguistic data and emotional preferences; they also tested the strategy with an empirical investigation. C.C.Huang et al. (C. C. Huang, 2020) used the IS Rank method, which integrates the PR values of the Page Rank algorithm with appreciation degree, to merge impact with SA depending on the content of postings and filtered OL. Deng et al. (X. Deng, 2013) proposed a node closeness approximate calculation method for discovering core OL on Micro Blog graphs, which offers increased accurateness and moves better, and created a Micro Blog crawling and analytics tool based on SINA Micro Blog APIs. Page Rank is an efficient sorting algorithm, but its performance degrades dramatically as the number of available nodes grows. Jing and Jing et al. (Jing, 2014) presented a hybrid data mining technique depending on user characteristics and conversation networks that consists of three parts: a technique for analyzing users' authority, activity, and impacts, a method for considering sentiment polarity in an interaction network, and a consolidated approach for recognizing microblog OL based on the HITS algorithm.

OLs on Twitter usually have a large following and significantly influence their followers. Their blog is usually full of useful data and breaking news for everyone else (Ghosh, 2021). Followers enjoy sharing their opinion leaders' tweets data because they believe it is essential for others to know. All of these qualities of Twitter OL reinforce the notion that knowing who the OLs are on Twitter is valuable. On the other hand, existing options do not focus on specific opinion leaders but on general ones, as we have described. Researcher (Banati, 2013) compared the effectiveness of evolutionary algorithms to the current greedy technique for this application. It evaluates the appropriateness of two popular EA's, Differential Evolution (DE) and Firefly (FA), to the k-Max-Influence problem. Epinions, Wiki-Vote, Slashdot datasets were used in the research. In terms of maximal influence experienced as well as advantage obtained by boosting the value of k, the results demonstrated that both approaches DE and FA outperform the Greedy strategy. FA consistently outperforms DE among evolutionary techniques. The data demonstrate that FA keeps its outcomes consistent and has a larger chance of scoring than DE and Greedy. Table 2.5 discusses different evolutionary algorithm based Influence Maximization problems.

Online SN has become more integrated into people's work and daily lives as social media and activities evolve. OLs are activists who use media to spread information and filter statements in mass media. As a result, effective monitoring of OL may impact the dissemination and evolution of public sentiment to some level. Several traditional techniques of OL mining rely primarily on the user's network structure, ignoring the importance and relevance of contextual information in the development of opinion leaders. Furthermore, these methodologies only rank users' impact on a global scale and are ineffective at discovering low-influence local opinion leaders.

Table 2.5 Evolutionary Algorithm Based Approaches

Author	Method	Evaluation Results	Ref.
Junwei Wang and Dingwei Wang [2008]	Particle swarm optimization	Goose team optimization attempts to incorporate several animal aggregating patterns into the optimization process. Thanks to the goose team flight technique, the program can effectively optimize four difficult nonlinear functions.	(Wang J. &., 2008)

W.R.M.U.K. Wickramasinghe and X. Li [2008]	Differential Evolution	MDEPSO is a simple and reliable EMO method that performs similarly to NSGA-II.	(Li W. R.)
Yiyi Zhao et al. [2018]	confidence-based opinion dynamics model	The trust factor that opinion followers have for OL uniquely affects their ability to influence others.	(Zhao Y. K., 2018)
sMichałWeskida et al. [2018]	Evolutionary Algorithm parameters	They found that the sets of ranking parameters employed in the evolutionary process had a strong correlation, often around 0.9.	(Weskida, 2019)
M.R. Martínez-Torresa et al. [2018]	Particle Swarm Optimization (PSO)	The findings can aid in a better understanding of the authentication mechanism in eWOM groups and the avoidance of reputation manipulations by fraudulent accounts.	(Martínez-Torres, 2018)
Lokesh Jain and Rahul Katarya [2018]	Firefly Algorithm	The suggested methodology is superior and ideal for identifying the opinion leader in a community and across the SN at large.	(Katarya, 2019)

Sofya A. Semenkovich et al. [2019]	SVM	When comparing the proposed algorithm to current ones, it was discovered that they all consider certain nodes prospective leaders.	(Tsukanova, 2019)
Lokesh Jain et al. [2019]	whale optimization algorithms	Depending on factors and computational time, they also concluded that the communal partition approach is superior to the other communities' detection techniques.	(L. Jain, 2019)
S. Mohammad Aghadam et al. [2021]	Grey Wolf Optimizer (GWO) algorithm	The number of known OL discovered by this algorithm is large, and the proposed method has the advantage of being compatible with various criteria and producing long-term results in multiple methods.	(Aghadam, 2021)
AfsanehRazghandi and Yasser Elmi [2020]	Leach algorithm	With a correlation value larger than 0.7, the algorithm has a significant association with the classic centrality indexes.	(Razghandi, 2020)
Dr.K.Saraswath et al. [2020]	Multi-Objective function based Ant Colony Optimization (MOFACO)	The classification performance of IG-SVM RBF (100,0.1) outperforms IG-CART by 17.26%, according to the results.	(Saraswathi, 2020)
Xu Qiang et al.	Ant Colony	This research contributes	(Qiang,

[2021]	Optimization	significantly by helping select an appropriate OL based on the criteria qualities.	2021)

Despite these crucial findings, properly comprehending opinion dynamics via social processes is never easy. Individual information obtained from various avenues is important in decision-making and opinion formation. It lays the groundwork for forecasting the outcomes of multiple courses of action. However, not everyone has the same knowledge foundation. When information is difficult to absorb or obtain owing to incompleteness, people often shape their beliefs by interacting with others who share their viewpoints on the same subjects. A basic example is that it is easy for someone to generate views on places they have never seen or foods they have never tried because decisions can be formed based on information from others. Popular computational approaches have been concentrating on how opinion emerges in a social setting for a few decades, in which opinion is either condensed to an arbitrary number or analyzed relevant to a particular situation.

2.9 Inferences of Literature Review

- Current research is centered on this technical basis of the Simple Influence Maximization (SIM) issue. Viral marketing is a prominent application of social influence maximization. Therefore, appealing to a customer would be useful in that case, especially when they have many additional system-contributing members. In addition to the node selection cost, recent research incorporated benefit as an additional SIM Situation element .just; the utility-driven influence diffusion model was used to handle the influence maximization challenge. Because this, is an entirely novel direction, numerous interesting works can be created.
- Scalability is still a concern in this field. Instead of employing a Monte Carlo simulation-based spread estimate, reverse reachable set-based spread estimation was recently presented to address the scalability issue. Following this work, all of the primary SIM Problem techniques, like TIM, IMM, and TIM+, take advantage

of its approach when an impression radiates estimation procedure to improve versatility.

- The influence function's submodularity was critical for building efficient and theoretically constrained IM solutions. In some cases, the influence function's submodularity criterion was overly tight. Non-sub modularity emerged because each node in the influence network may alternate both pro and con viewpoints. In such cases, the greedy framework was rendered ineffective. A proposed future approach for describing the Interaction parameters uses greater complex parameters to offer a finer answer over basic algorithms. Furthermore, the theoretical features that were poorly effective model ensure that the extended IM issue might be roughly estimated.
- Present IM diffusion models concentrated on this basis of impact that occurs involving a set of links connected by an edge. In actual society, individuals are impacted mainly through relationships or by social standards. A forerunner in the adherence IM issue is proposed in many existing papers. The conformity-aware diffusion model, on the other hand, focuses on extracting conformity features based on the uniform grid and hence excludes various social categories that are represented whereby individual features might be deduced. One potential step is incorporating access to individual identities into the conformance IM issues.

2.10 Major Findings and Limitations of Existing Research Work

- Although different measures of centrality are used, still there are other constructs that can be used to measure prominence and centrality in networks .It would be necessary to study the correlation between these measures and opinion leadership.
- The performance of evolutionary algorithms are superior and have higher probability to incur maximum influence spread within the given budget as compared to Greedy approach used for influence maximization problem.
- Future research can focus on developing more precise context-based rules. Measuring the tie strengths based on various variables is another intriguing direction. Additionally, we have discovered that in a real social network, influential people are typically viewed as more trustworthy than other users, and

vice versa. Determining the reciprocal relationships between influence and trust and building a comprehensive "reputation and trust-based system" are therefore important tasks.

2.11 Conclusion

A detailed survey on Influence maximization with general media has been conducted here. Following section 2.2, this work discussed the influence maximization issues in social media and research challenges on finding the effect seed in the network to increase the spread of information. The following sections explain the different types of model that has been used to increase the influence in social networks and the algorithms used in that models to overcome the challenges in diffusion models. Further, section 3.3 also explained the influence maximization based on the time, topic, and location in the social networks. Finally, the last section summarizes the methods for disseminating knowledge through social media. The chapter discusses five aspects of issue-solving approaches are offered to promote information dissemination, which will assist researchers and practitioners in having a thorough knowledge of the problem and greater exposure in this sector.

However, all these Influence Maximization methods promised excellent seed identification results but did not consider the network's memory or social reinforcement effect. The work proposes the seed-dependent influence maximization algorithm, which controls the memory and social reinforcement effects, is used to obtain the maximum acquiring probability when the number of neighbors of the Seed at a particular time increases. Furthermore, the work proposes the Effectual seed pick framework that studies the bidirectional effects between influence and trust to construct a comprehensive reputation and trust-based system. It also explores the viability of deep learning-based DeepWalk algorithm that considers the various topological measures for Influence maximization.

Chapter 3 DEEP LEARNING FOR INFLUENCE MAXIMIZATION

The proposed method in this chapter combines topical and topological aspects of a user in a social network to identify the most influential nodes using graph embedding. The technique uses collaborative filtering and cosine similarity to find similar users and then applies DeepWalk for graph embedding. The resulting algorithm, called DeepWalk-based Influence Maximization (DWIM), uses deep learning to identify influential users in computable time.

3.1 Introduction

Influence maximization (IM) in social network analysis is one of the central research areas. IM aims to identify a set of influential seeds in the network and to maximize the diffusion of the information or influence. Unfortunately, inefficient computational algorithms are frequently used in conventional influence-maximizing techniques, which do not scale well for vast networks. To address this challenge, this chapter introduces DWIM, a novel approach for influence maximization based on DeepWalk, a state-of-the-art node embedding technique. To effectively capture a network's structural properties and identify influential nodes, DWIM makes use of DeepWalk's strength.

A user's centrality inside the network, which can be measured in several ways, defines how significant they are in propagating information in a social network. The three most common metrics are Closeness Centrality (CC), which gauges a node's proximity to every other node in the network, Betweenness Centrality (BC), which measures a node's importance in information transmission from one point to another; and Degree Centrality(DC), which counts the number of connections a node has.

Various variants of classical centrality measures such as Katz centrality (Katz, 1953), Google's Page Rank (Langville, 2011), alpha centrality (Zelen, 1989), and Subgraph centrality (Estrada, 2005) have been proposed to identify the most critical nodes. Studies also focused on methods that include the degree of trust and its propagation, game theory, etc. Moreover, it has been observed that user's activities also influenced their acquaintances (Trusov, 2010). moreover, to identify opinion leaders, the user's actions have been analyzed to learn the frequent pattern (Goyal A. B., 2008). However, not only the network topologies, text mining is also considered for identifying opinion leaders (Liu N. M., SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork, 2014) (Bodendorf, 2010). Studies also focused on topic or domain-specific opinion leaders (Van Der Merwe, 2009) using a novel frequent pattern mining approach, leaders are also discovered by analyzing the community and their actions (Van Der Merwe, 2009). Buyers are also ranked based on their past purchase behavior on eBay for efficient marketing (Lam, 2009). Various heuristic models based on classical Linear Threshold (LT) propagation have been developed (Bonchi, 2011) (Chen W. W., 2010). Rating behavior also maximizes product adoption (MNaik, 2013). Various existing studies related to viral marketing focused on influence maximization by using k number of seeds. Kempe et al. (Kempe David, 2003) proposed a greedy approach based on Cascade Models (Kim H. B., 2014).

While previous efforts in identifying influential nodes in social networks have yielded impressive results, they often require extensive knowledge of the entire network, which can be time-consuming and unfeasible for many applications. In contrast, the presented work utilizes graph embedding techniques to leverage the social network, using asymmetric similarities for node recommendation (Zhou C. L.) and random walks for node representation, which applies to large-scale networks. The suggested approach for identifying the influence of nodes in the network based on random walks is described in the following section.

3.2 Background

A modest initial selection of nodes in a network is chosen to maximize the dissemination of influence or information. The aim is to find people who, when triggered, can start a chain reaction of influence that affects a significant section of the network. The scalability and accuracy of conventional influence maximization techniques like the greedy algorithm and Monte Carlo simulations are constrained. These techniques are frequently computationally intensive, making them unsuitable for large-scale networks. DeepWalk is a popular unsupervised learning method for producing network node embeddings. It uses the SkipGram model notion from natural language processing to extract structural information from graphs. Through random walks on the network that DeepWalk treats as sentences, SkipGram models can be used to develop low-dimensional representations of nodes. These node embeddings capture the structural characteristics of the network and enable various downstream tasks, including node classification, link prediction, and community detection.

3.2.1 Framework

This section discusses the underlying idea of graph embedding and investigates its application to the influence maximization problem. Two critical modules in the architecture combine social network user's topological and topical characteristics. Collaborative filtering is utilized to do this, and cosine similarity is used to find comparable users for each social network user. After that, graph embedding is carried out by randomly traversing the network built utilizing the similarities between individuals. These resemblances serve as probabilities to pique the interest of nearby users.

3.2.2 Preliminaries

This section formally explains Graph Embedding on social network data and its most well-known technique, DeepWalk (Perozzi, 2014). Graph Embedding involves learning a mapping function representing each word in the corpus as a low-dimensional vector equation, as shown in Equation 3.1

$$\boldsymbol{\varphi}: \mathbf{v} \in \mathbf{V} \rightarrow \mathbf{R^d} \text{ , } \mathbf{where\ d} \ll |\mathbf{V}| \tag{3.1}$$

φ: This represents the mapping function or embedding function. It takes a word v (belonging to the vocabulary V) as input and maps it to a low-dimensional vector in R^d .

v: This variable stands for a word from the vocabulary V. The goal of graph embedding is to represent each word in the corpus as a low-dimensional vector in R^d .

V: This is the vocabulary of the corpus. It consists of all the unique words or nodes in the dataset.

R^d : This represents the d-dimensional Euclidean space, where each word in the vocabulary V is mapped as a low-dimensional vector. The value of d is significantly smaller than the size of the vocabulary, denoted as |V|.

The basic concept used in the literature for embedding is Word2vec (Mikolov T. C., 2013). It transforms each word of the corpus into embedding vectors. For Instance, consider two sentences for instance: *“Hi I am here”, and “Hi I am there”*. These sentences can be represented as a set of sequences of nodes as in Figure 3.1 and construct a vocabulary called corpus *V={Hi, I, am, here, there}*.

Now each vocabulary word is represented as a vector of dimension V as in the example Hi = [0,0,1,0,0]; I=[0,0,0,1,0]; am=[1,0,0,0,0]; here=[0,1,0,0,0]; there=[0,0,0,0,1].

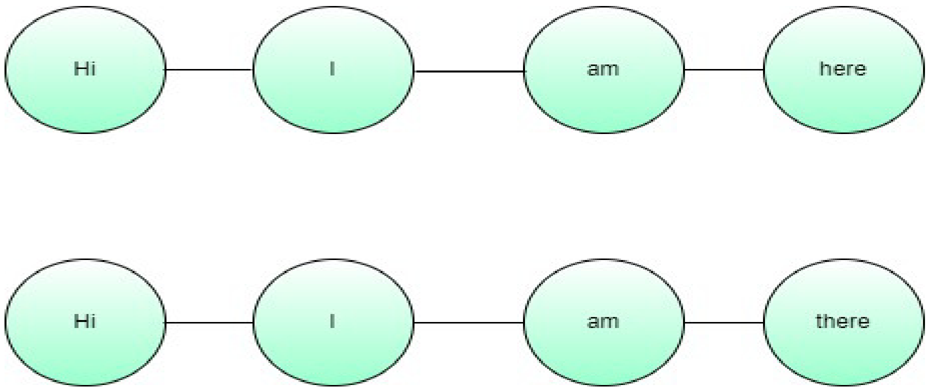


Figure 3.1 Sequence of Words in a Sentence

The main concern of this representation is to visualize each word with one dimension. In our example, "here" and "there" are closer to each other as compared to "Hi" and "am." So it is a great form to represent the nodes in low-dimensional vectors that facilitates the depiction of their relationship with other nodes in the network.

The SkipGram (Mikolov T. C., 2013) neural network is used by word2vec to generate these embeddings. In this network, the input layer takes an encoded vector of

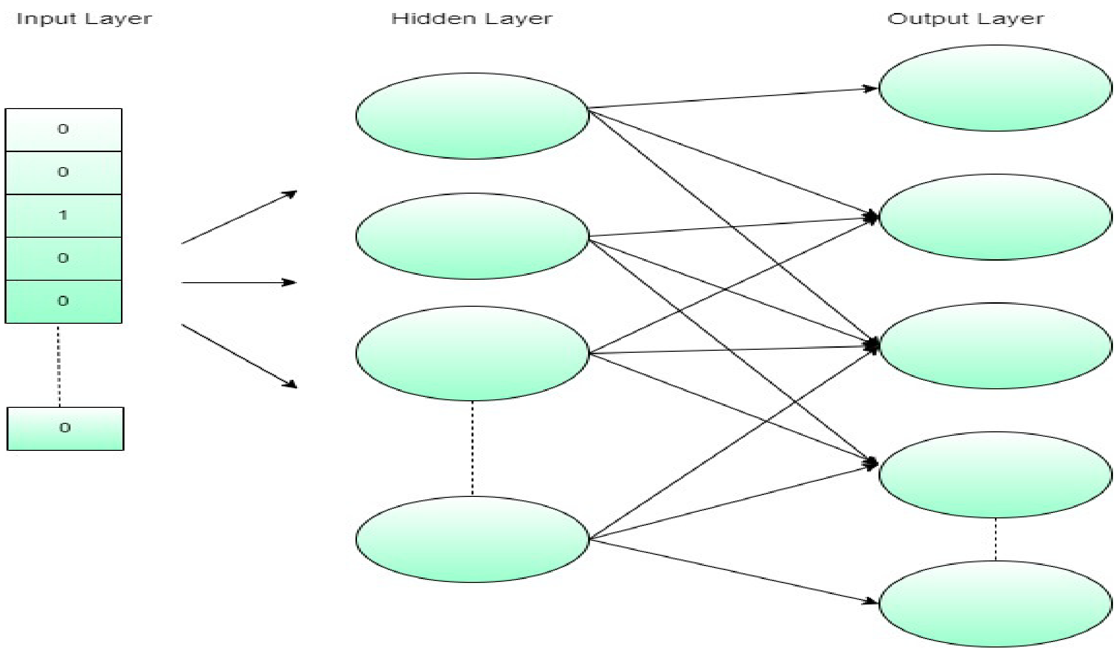


Figure 3.2 Neural Network for Skipgram

dimension *d*, which is equal to the words in the dictionary, and the hidden layer presents its embedding. The output layer predicts the neighborhood words using a softmax classifier, as shown in Figure 3.2.This transforms each word of the corpus into embedding vectors such that similar words occupy close spatial positions. The current paper exploits this property of graph embedding to identify the node with the maximum influence. The following section describes its applicability in more detail.

3.3 Proposed Framework

This section formally explains the applicability of graph embedding to social networks for estimating the influence of nodes.

A graph G (*V, E*) can be used to illustrate a social network that has nodes that are connected. The nodes in networks and user social interactions are represented by *V* and *E*, respectively.

The presented DeepWalk based influence maximization algorithm has two essential components, first a DeepWalk algorithm (Perozzi, 2014) and a second SkipGram algorithm (Mikolov T. C., 2013) to learn the embedding of each node in a

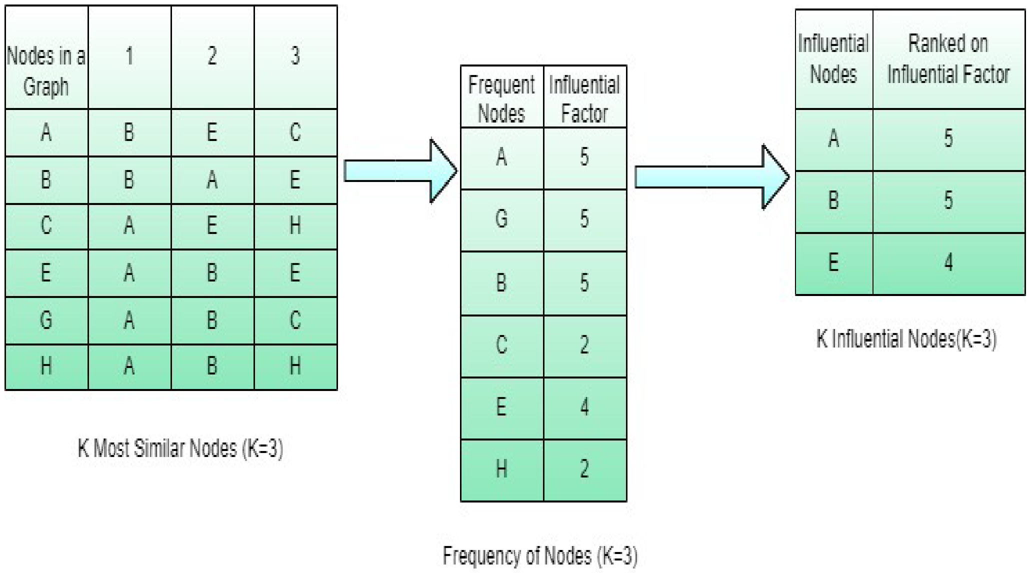


Figure 3.3 Computing Influential Factor of a Social Network

social network. After learning the embeddings, the algorithm explores each node's k most similar nodes and generates a pool of nodes of size $V \times k$. Then compute the influential factor of each node based on its frequency, as shown in Figure 3.3. We used the DeepWalk algorithm and SkipGram algorithm as a part of our proposed algorithm DWIM which are used frequently for graph embedding as follows.

DeepWalk Algorithm (Perozzi, 2014) : DeepWalk(G,w,d,ϒ,t)

Input: Graph G(V,E)

Window_size: *w*

Embedding size: *d*

walk per vertex: ϒ

walk length: *t*

Output: Matrix of vertex representations

$\Phi \epsilon R^{|V| x d}$

1. *Initialization: Sample Φ from* $u^{|V| x d}$

2. *Build a binary Tree T from V*
3. *for i=0 to ϒ do*
4. *O = Shuffle(V)*
5. *for each $v_i \epsilon$ O do*
6. *W_{vi}=Random Walk(G, $v_{i,t}$)*
7. *SkipGram(Φ,$W_{vi,}$w)*
8. *end for*

end for

The DeepWalk algorithm utilizes random walks on a given social network graph as input, where different paths are traversed and given to the SkipGram model for learning. The SkipGram model is a basic deep learning model used for predicting word and context words, and in this case, it uses the random walks generated on the graph as the "bag of words".

SkipGram Algorithm (Mikolov T. C., 2013) : SkipGram(Φ,$W_{vi,}$w)

---

1. *for each $v_j \epsilon$ W_{vi} do*
2. *for each $u_k \epsilon$ W_{vi}[j-w : j+w] do*
3. *$J(\Phi) = -\log Pr(u_k \mid \Phi(v_j))$*
4. *$\Phi = \Phi - \alpha * \frac{\partial J}{\partial \Phi}$*
5. *end for*

end for

Our proposed algorithm utilizes DeepWalk and SkipGram models to identify the k most influential nodes in a social network. The process involves predicting the most similar nodes to any given node, and the node with the highest similarity to others is deemed the most influential. To calculate the similarity between users, we employ cosine similarity via collaborative filtering, which yields user-to-user similarity, used as a probability to activate other users. When two user nodes are highly similar, the chances of getting activated are greater. We then obtain various paths by random walks on the graph and employ SkipGram to predict the most similar nodes in the social network. In the following part, we will examine how the suggested method is applied to actual big data from social networks as in flowchart given

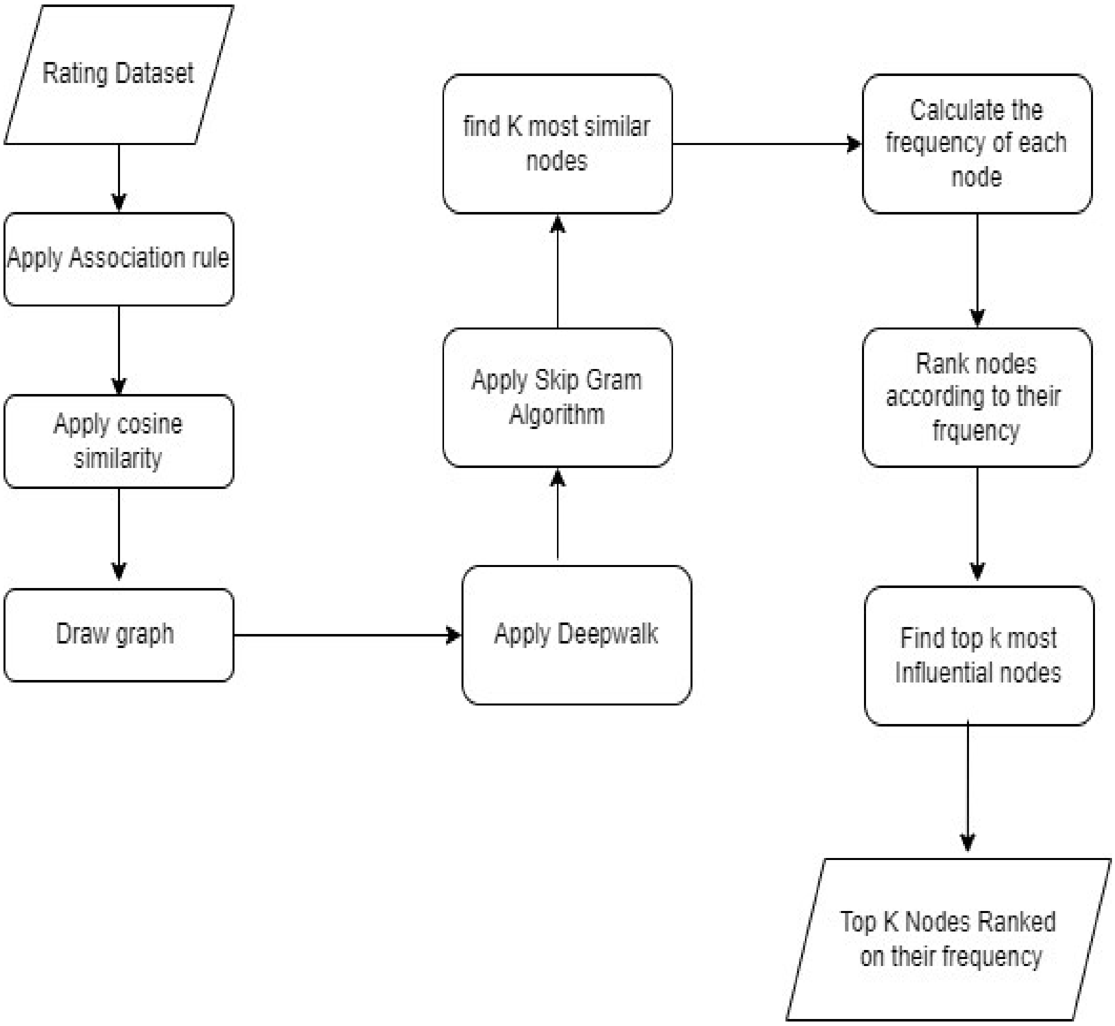


Figure 3.4 Deepwalk Based Influence Maximization

Figure 3.4 . First, the top k nodes determined by our method are contrasted with those determined by several centrality metrics, such as degree centrality, closeness centrality, Page Rank, and eigenvector centrality.

Proposed Algorithm :DeepWalk based influence maximization (Sonia, 2023)

Input: *DWIM(G,k,t,l,w,d)*

No. of nodes k *Walk per node: t*

Length of walk: l

Window size: w

Dimensions: d

Output: k = top k most influential (MI) nodes ranked on their frequency

1. NodeSequence = GenerateNodeSequence(G, t, l) // Generate node sequence using DEEPWALK algorithm

2. Embeddings = LearnEmbeddings(NodeSequence, d) // Learn embeddings using SkipGram algorithm

3. SimilarNodes = FindSimilarNodes(Embeddings, k) // Find k most similar nodes for each node

4. Frequency = CalculateFrequency(SimilarNodes) // Calculate the frequency of each node

5. RankedNodes = RankNodesByFrequency(Frequency) // Rank each node according to its frequency

6. MI = SelectTopKMostInfluentialNodes(RankedNodes, k) // Find top k most influential (MI) nodes

Return MI

3.4 Experimental Study

The experiments seek to determine whether the suggested method effectively locates the most influential nodes (MI) in social networks. Comparisons are made between the performance of the MI found by the proposed model and traditional centrality metrics such as Degree Centrality, Closeness Centrality, Page Rank, and Eigenvector Centrality. The studies make use of two real-world datasets with different sizes and densities, namely as in Table 3.1:

(i) Zachary's karate club (Zachary, 1977): It is a friendship network of 34 university-based karate club members.

(ii) Epinions: It is the who-trust-whom of Epinions.com. Each vertex is a site member, and the link between them represents who trusts whom (Meyffret, 2014).

Table 3.1 Summarization of Dataset

Dataset Name	Total Nodes	Total Edges
Zachary's karate club	34	78
Epinion	75888	508837

These datasets are collected from Stanford Large Network Dataset collections. The features of the datasets are shown in Table 3.2. Experiments are conducted with k=10 for the karate club data set and k=20 for the Epinion dataset. However, the other parameters are the same for all experiments, such as w =4,d=64,t=10,l=100.

Table 3.2 Details of Top-10 MI for Karate Club

Rank	Node_id	DC	CC	PR	EC
1	34	1.030303	0.55	0.1034546	0.3733712
2	1	0.969697	0.5689655	0.0992321	0.3554835
3	33	0.7272727	0.515625	0.0733062	0.308651
4	2	0.5454545	0.4852941	0.0543403	0.2659539

5	3	0.6060606	0.559322	0.059197	0.3171894
6	14	0.3030303	0.515625	0.0302543	0.2264697
7	4	0.3636364	0.4647887	0.0366125	0.2111741
8	9	0.3030303	0.515625	0.0304115	0.2274051
9	31	0.2424242	0.4583333	0.0248246	0.1747603
10	24	0.3030303	0.3928571	0.0316533	0.1501233

Table 3.2 presents the Karate data set findings from the proposed model's performance evaluation on the provided data sets. The suggested approach was used to retrieve the top 10 most influential (MI) nodes from this data set. The outcomes show that the proposed approach successfully finds the group of critical nodes with the most significant influence as assessed by DC, CC, PR, and EC. The centrality measures of each node ranked by the proposed approach are illustrated in Figure 3.5, Figure 3.6, Figure 3.7, Figure 3.8, Figure 3.9 and Figure 3.10.

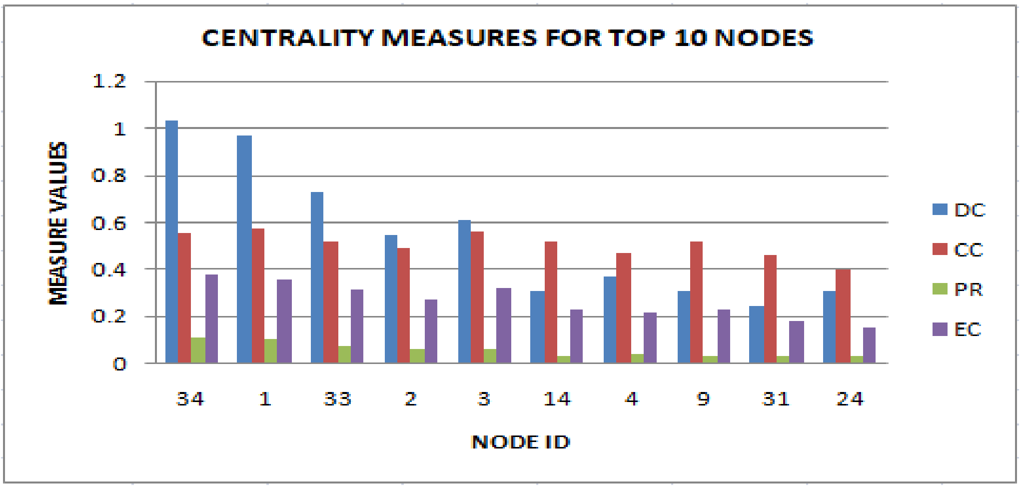


Figure 3.5 Centrality Measures for Karate Club

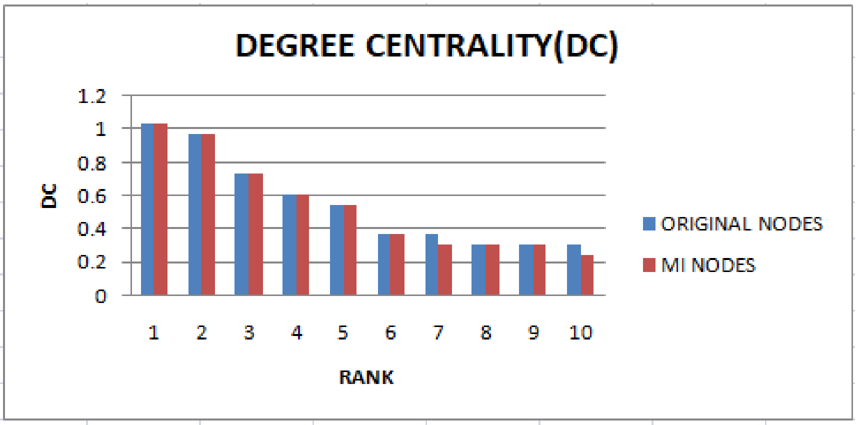


Figure 3.6 DC Comparison of Top 10 MI for Karate Club

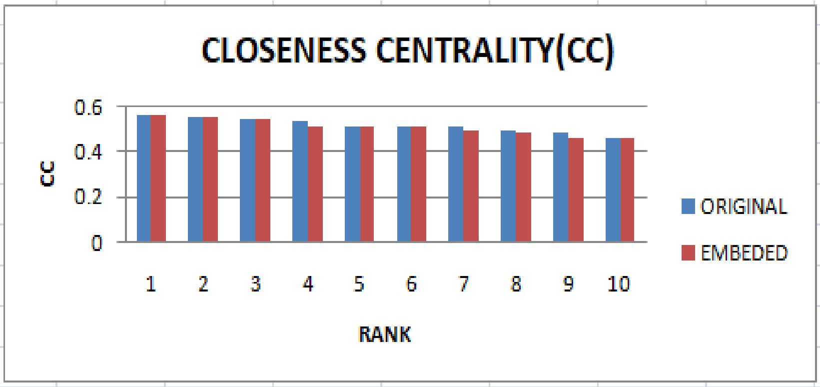


Figure 3.7 CC Comparison of Top 10 MI for Karate Club

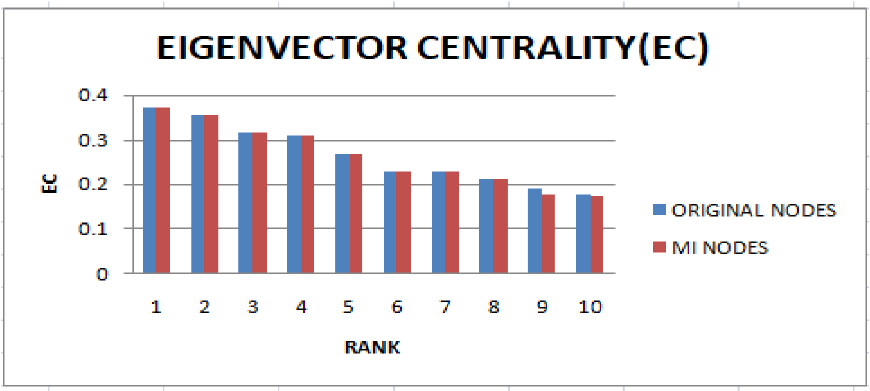


Figure 3.8 Comparison of Top 10 MI for Karate Club

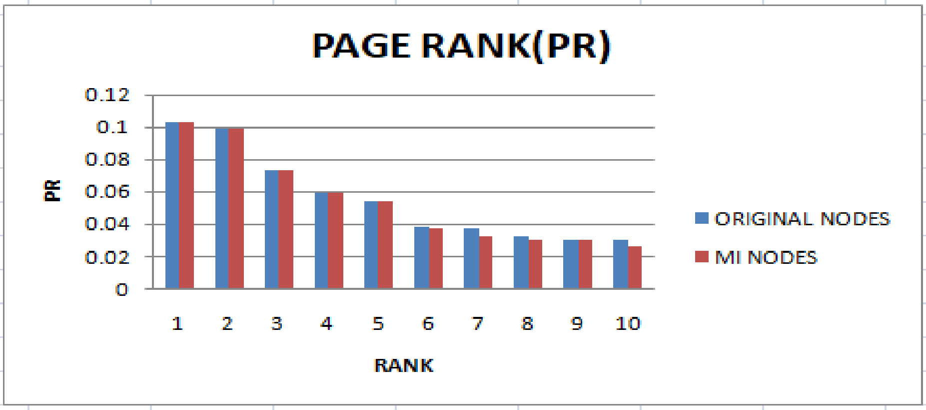


Figure 3.9 PR Comparison of Top 10 MI for Karate Club

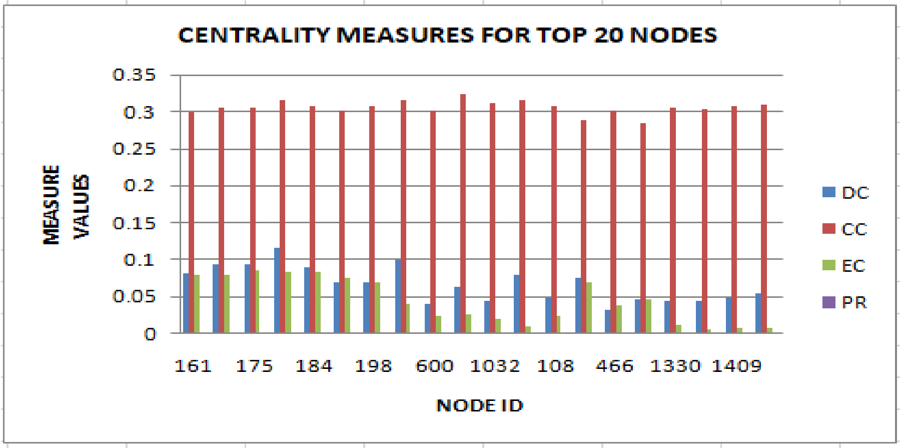


Figure 3.10 Centrality Measures for Karate Club

The evaluation of the proposed algorithm on the two data sets has been presented in Table 3.2 for the karate data set and Table 3.3 for the Epinion network. The proposed algorithm was applied to extract the most influential nodes based on the topological and topical features of the social network users. In both data sets, the proposed method identified the set of vital nodes with maximum influence, as measured by the classical centrality measures of DC, CC, PR, and EC.

In Figure 3.5, Figure 3.6, Figure 3.7, Figure 3.8, Figure 3.9 and Figure 3.10 and Figure 3.11 depict the performance of the algorithm for different network centrality measures such as degree centrality, closeness centrality, eigen centrality and page rank centrality of each node ranked by the proposed approach. These results were compared with the top 20 nodes in the original network concerning the given criteria. It was observed that the proposed method efficiently extracted the nodes with values of centrality measures that were near to or proportionate to the existing values in the graph. Overall, the evaluation results demonstrate the effectiveness and applicability of the proposed approach in identifying the most influential nodes in social networks.

Table 3.3 Details of Top-20 Rank Nodes for Epinion

Rank	NODE_IDS	DC	CC	PR	EC
1	161	0.0811619	0.2989117	0.0011499	0.0789562
2	174	0.0922683	0.305522	0.001174	0.0783467
3	175	0.0922683	0.305522	0.001174	0.0851888
4	180	0.114481	0.314919	0.00008417	0.0819235
5	184	0.0888509	0.3061248	0.0010054	0.0828359
6	185	0.0679197	0.3002913	0.0012395	0.0734336
7	198	0.0683469	0.3061248	0.0010312	0.0687223
8	568	0.0986758	0.3150666	0.0008793	0.0383735
9	600	0.0384451	0.301278	0.0006445	0.0225858
10	745	0.061085	0.3235573	0.0008882	0.0238793
11	1032	0.0435711	0.3107945	0.0008413	0.0185931
12	1418	0.0781717	0.3148207	0.0005255	0.0077826
13	108	0.0465613	0.3063108	0.0004147	0.0225176
14	172	0.0747544	0.2888111	0.0010853	0.0680583
15	466	0.0303289	0.301233	0.0004886	0.0378859
16	560	0.0452798	0.2836929	0.0004869	0.045139
17	1330	0.0435711	0.3053833	0.0003236	0.011254

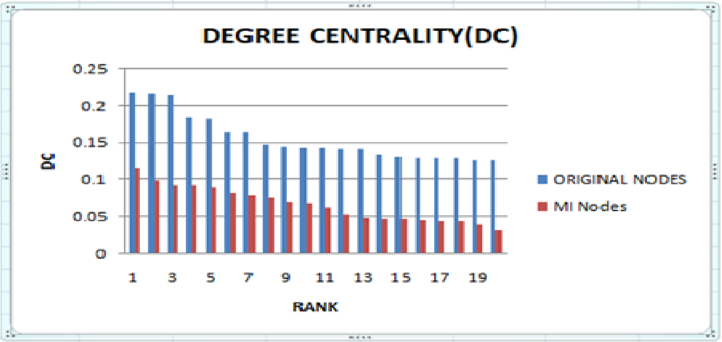
18	1408	0.0439983	0.3021355	0.0005156	0.0045623
19	1409	0.0469885	0.3074781	0.0004485	0.0061307
20	1453	0.0525416	0.308277	0.0004301	0.0069624

The effectiveness of the suggested strategy is assessed by locating the top k influential nodes using conventional centrality measures as DC, CC, EC, and PR. The collected findings are then contrasted with the full k nodes in the network that have the most influence, according to each centrality metric. According to the analysis, the suggested method is effective and quick in locating the k most important nodes.

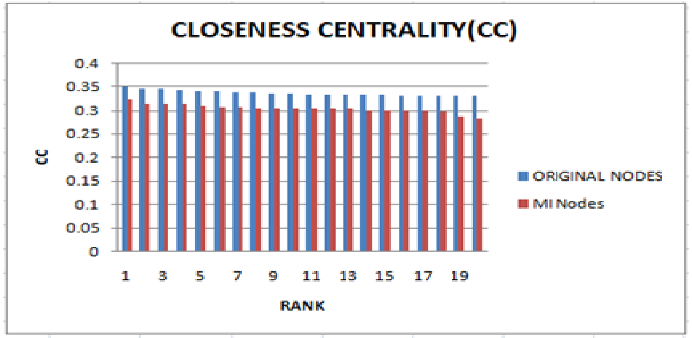
3.5 Conclusion

In this chapter, we introduced DWIM, a novel approach for influence maximization based on DeepWalk. By leveraging the power of node embeddings, DWIM effectively captures the structural characteristics of a network and identifies influential nodes for maximizing influence spread. The experimental evaluation demonstrated the effectiveness of DWIM in large-scale networks and showcased its scalability and accuracy compared to existing methods. The performance of the proposed method was evaluated through experiments and compared with centrality measures. The results showed that DWIM could find k influential nodes and help with viral marketing, outlier detection, and recommendations for different products and services. Additionally, the chapter introduces Graph Embedding and its most prominent method, DeepWalk, which is used in the proposed method. The chapter also explains the concept of word embedding using SkipGram neural network. It provides an example of representing a vocabulary as low-dimensional vectors to depict the relationship between nodes in the network.

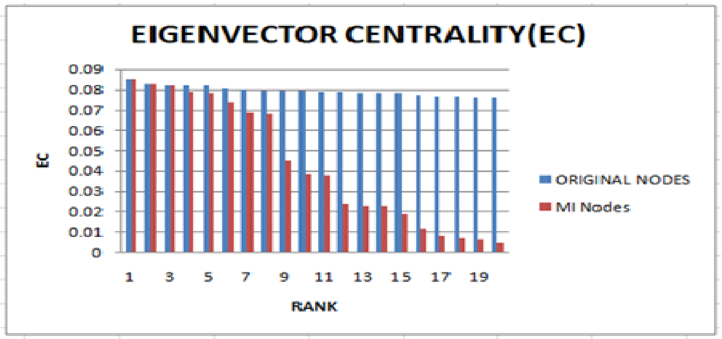
DWIM opens up new possibilities for efficient influence maximization in various domains, including social networks, marketing, and information diffusion. Future research can explore extensions of DWIM to incorporate additional network features and develop more advanced seed selection strategies for improved influence maximization.



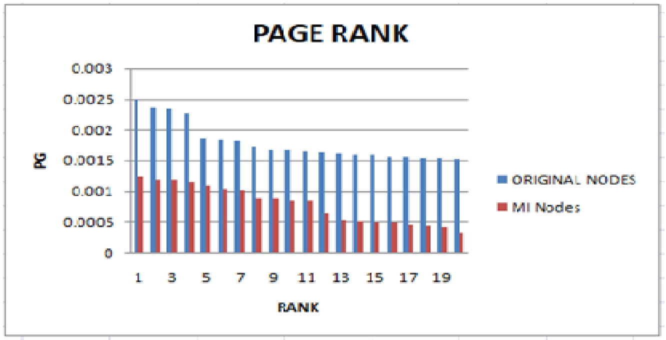
(a) DC Comparison of Top 20 MI Nodes for Epinion Dataset



(b) CC Comparison of Top 20 MI Nodes for Epinion Dataset



(c) EC Comparison of Top 20 MI Nodes for Epinion Dataset



(d) PR Comparison of Top 20 MI Nodes for Epinion Dataset

Figure 3.11 Comparison of Centrality for Top 20 MI Nodes for Epinion

Chapter 4 EFFECTUAL SEED PICK FRAMEWORK FOCUSING ON MAXIMIZING INFLUENCE IN SOCIAL NETWORKS

*This chapter introduces specific context-based rules and constructs a comprehensive ''reputation and trust-based system'' by identifying bidirectional effects between influence and trust. Here effectual seed pick framework for pervasive healthcare is discussed, which provides maximum dispersal of any information or services. The proposed algorithm Fuzzy-VIKOR (*VIse Kriterijumska Optimizacija I Kompromisno Resenje) *helps identify the targeted node to spread information rapidly. Consequently, the proposed structure effectively addresses different issues related to a large number of patients, and thus increased influence maximization using seed nodes is helpful for pervasive healthcare.*

4.1 Introduction

Today, the emergence of social media is helpful for the healthcare system, where everyone is closely connected. Large numbers of people can be reached by using seed nodes to provide medical advice, facilities, new changes in treatment, and any health ministry guidelines. As today's world is dealing with COVID-19 so the main objective is to provide healthcare services to many people irrespective of time and locality. As people suffering from corona are coping with mental health issues, the seed pick framework using machine learning for the Influence Maximization technique is proposed to help provide pervasive healthcare. For pervasive healthcare, an effectual seed pick framework is required focusing on influence maximization using machine learning. The proposed algorithm Fuzzy-VIKOR helps identify the targeted node to spread information rapidly. Consequently, the proposed structure effectively addresses different issues related to a large number of patients, and thus increased influence maximization using seed nodes is helpful for pervasive healthcare. The experiments show that the proposed framework has high precision, accuracy, F1-score, and Recall compared to other algorithms employed to find influence maximization seeds.

4.2 Effectual Seed Pick Framework for Pervasive Healthcare

4.2.1 Proposed Method

Social media is an interactive, computer-mediated platform that enables people to create and share information. Its use in healthcare has been widely debated. Despite the increasing number of healthcare professionals who use it, the advantages of social media remain unclear. In the pandemic scenario, the utilization of social media in healthcare was touchy. There is still considerable debate about the payback of social media in actual learning and improving the quality of treatment offered. Every COVID-19-related topic to summarize the conversation around the viral eruption on the various platforms has been extracted and evaluated. The SkipGram neural network is used to create distributed representations of words for the corpus of the text from each social media platform after applying the Exact Extract Algorithm (EEA) to the textual content on each platform. The term is represented as a vector to consider the themes around which the COVID-19 debate is focused, and the Partitioning Around Medoids (PAM) technique is used to cluster the words whose vector representations use the cosine distance matrix as a closeness measure. To optimize the influence across various social media platforms, we must first separate the content that is linked to Covid-19. The prior strategies concentrated on single qualities and node attributes for opinion leader selection to achieve the presumptive communities. However, social media success is often based on choosing qualities of the target audience with variable weights, such as age, gender, or location, depending on how important they are in terms of the campaign's effectiveness. The proposed novel Multi-Criteria Based Effectual Seed Pick Framework with a unique opinion value for each criterion employed to choose the first seeds to successfully attain the network's targeted nodes by Best Node Selection Method (BNS) in which the network users/nodes are distinguished not only by their centrality relationships with other nodes/users but also by a set of unique parameters addresses this intriguing research gap as in Figure 4.1. After reaching the targeted multi-attribute nodes, the information spread in a social network proceeds through the Effective Seed Pick Process (ESP) in which multiple attributes are taken into account to choose the seeds with the best chance of eventually propagating the information to the targeted node. To obtain a network node ranking with the highest to reach the destination nodes in the social networks, the multi-criteria evaluation of the nodes has been done. Fuzzy- VIKOR is used based on their relative distance to their solution. As a result, the proposed approach gives exact information about Covid-19 from various social networks and improves the node selection to increase the influence

4.2.2 Process Flow

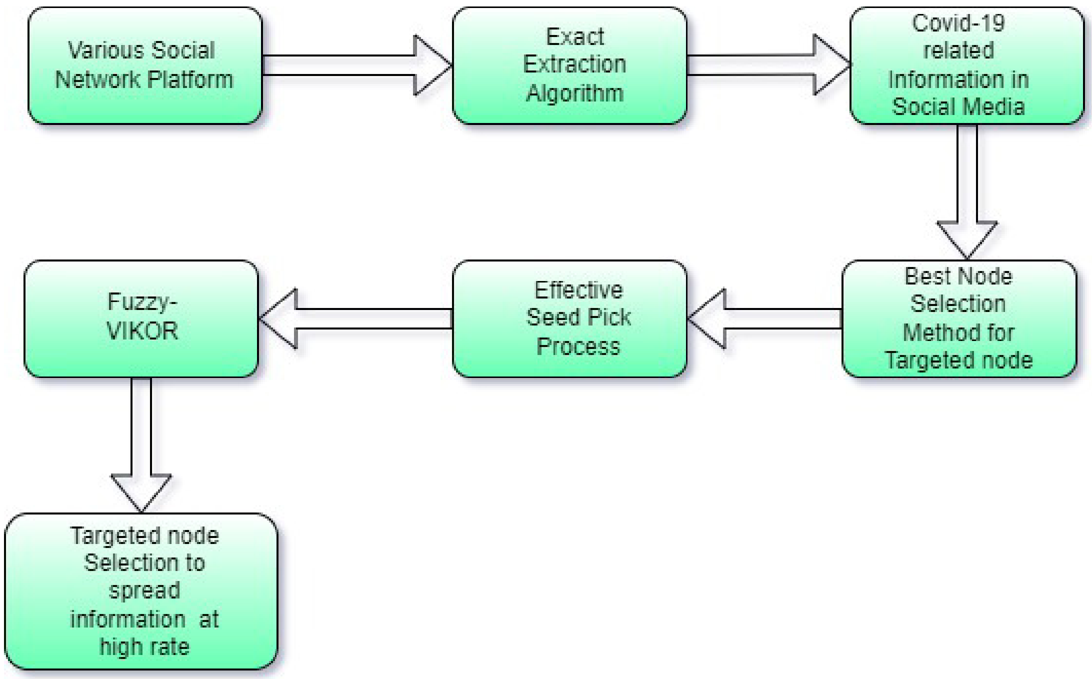


Figure 4.1 Effectual Seed Pick Framework for Pervasive Healthcare

4.2.3 Find Various Social Network Platform

In this work covid 19 twitter comments dataset which contains tweets having hashtags: #coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #covid19, #covid_19 from 17 March onwards have been used. The dataset also consists of the other hashtags: #epitwitter, and #ihavecorona. The different variables of the dataset consist of the text tweeted by the tweeter user, the location of the user account, the country code, and the hash tags used by the users. As the volume of the tweets was extensive, there can be some gaps for some hashtags as all the tweets with the hashtags cannot be captured. To make the dataset file manageable, the data is split into half month data in one file and the rest in the other file. Some hashtags are frequently used, and some are used less frequently. As a result, less regularly used hashtags will remain for a longer time.

The most frequently used hashtag is #coronavirus, which seems to be the most popular. Despite scraping over 500,000 Tweets, there were still gaps between the number of retweets and the hashtag. This dataset excludes retweets since the retweets option has been set to FALSE (although a count of retweets is provided as a variable)

4.2.4 For Analyzing User Engagement in Social Networks

This work introduced the SkipGram neural network to learn word embedding for the text of social media platforms. Furthermore, it learns word embedding for topics related to the corona. We apply the SkipGram neural network to represent social media platforms' text corpus. To maximize the average log probability of a sequence of words represented by *w1, w2, ..., wT*, we can apply stochastic gradient descent and compute the gradient through the backpropagation rule as in Equation **4.1**

$$\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=-k}^{k} \log p(w_t + j/w_t)\right] \tag{4.1}$$

Where k is the size of the training window, as a result, during training, comparable word vector representations are close to one another. Each word in the SkipGram model is connected to its input and output vectors, *uw* and *vw,* in some way. Given the word w_j, the likelihood that the word w_i would be accurately predicted is given by Equation 4.2

$$p(w_i/w_j) = \frac{exp\left(u_{w_i}^T v_{w_j}\right)}{\sum_{i=1}^{V} exp\left(u_l^T v_{w_j}\right)} \tag{4.2}$$

One corpus vocabulary contains V words. The effectiveness of training can be influenced by two critical factors: the dimensionality of word vectors and the size of the window surrounding the words. Here, we have utilized six words for the window and 200 as the vector dimension, which is a conventional choice for training large datasets.

Where $p(w_i/w_j)$ is the probability of correctly predicting a word

V is no. of words in the corpus vocabulary

u_w is input vector

v_w is output vector

4.2.5 Clustering the Words

The Cluster of the words is done by executing the Partitioning Around Medoids(PAM) algorithm, which uses the proximity metric, the cosine matrix in the particular vector representation. To evaluate the topics related to CORONA, we used PAM to cluster words and the cosine distance matrix as the proximity metric of words' vector representations using Equation 4.3, Equation 4.4, Equation 4.5, Equation 4.6. After locating the k clusters, figure out the typical silhouette width for each value of k. We must compute the average pairwise Jaccard similarity between clusters using 90% of the data to establish the cluster's stability. The subject of each cluster is then established by creating word clusters. To examine the discussions related to the global outbreak of Covid-19, how can we analyze the conversations taking place, The distribution of word clusters for a particular content c can be defined as the distribution of its words over themes c. To evaluate the importance of each topic within a corpus , we limit our analysis to contents c with a maximum value of c greater than 0.5 and consider each of these contents as an individual topic.

$$\boldsymbol{I} = \left[\frac{\boldsymbol{R_0}}{(\mathbf{1}+\boldsymbol{d})^{\boldsymbol{t}}}\right]^{\boldsymbol{t}} \qquad 4.3$$

Where I is an incidence

t is add up of days

R_0 is a basic reproduction number

d is a damping factor

$$\partial_t S = -\beta S \cdot {}^{I}/_{N} \tag{4.4}$$

$$\partial_t I = \beta S \cdot I/N - \gamma I \tag{4.5}$$

$$\partial_t R = \gamma I \tag{4.6}$$

Where S is the count of susceptible

I is adding up of infected

R is add up of recovered

4.2.6 Best Node Selection Method(BNS) for Distinguishing the Network Nodes

Best Node Selection Method distinguished the network nodes and users by various distinctive characteristics, such as user friends, followers, and favorites, in addition to their centrality interactions with other nodes and users. Selecting the seed in every attribute based on their highest potential of spreading the information to the targeted node chosen in the BNS method.

4.2.7 Selection of Targeted Node Using Fuzzy-VIKOR

Next, we have to select the targeted node with the highest ranking in the social networks by the Fuzzy- VIKOR method based on their relative distance to their solution for spreading the information to the maximum number of users.

The criteria values of all the network's vertices are then used to construct a decision matrix, D[x_{ij}] in Equation 4.7, where the vertices are represented by the m rows and the criteria are represented by the n columns.

$$DD[x_{ij}] = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{pmatrix} \tag{4.7}$$

Where $D[x_{ij}]$ is decision matrix

m row represents the vertices

n columns represent the criteria

In the algorithm the second step is to normalize the decision matrix. Different formulas are employed for the cost and benefit criteria, respectively: The ability to alter the weights of various decision criteria makes the Fuzzy-VIKOR-based techniques more versatile than the conventional aggregating procedures.

The analyst adjusts the weights of each choice criterion following the decision maker's preferences. To maximize the chance of reaching the targeted network nodes through the seeded network nodes as much as is practical, the marketer modifies the weights of several criteria in the case of the seed selection problem under consideration. The weights are established depending on the analyst's expertise, knowledge, and experience.

The ideal solutions for the algorithm's fourth phase that is Fuzzy-Vikor in next section 4.3 Proposed Algorithm, V_j and V_j^+, respectively, are computed using Equation 4.8, Equation 4.9, Equation 4.10. The positive ideal solution for the seed selection problem under study would be a vertex with the highest possible values for each criterion. The negative ideal solution, in contrast, would be a vertex with the lowest values for each criterion.

$$\boldsymbol{V_{ij} = w_j \cdot r_{ij}} \qquad 4.8$$

$$\boldsymbol{V_j^+ = \{v_1^+, v_2^+, v_3^+, \ldots, v_n^+\}} \qquad 4.9$$

$$\boldsymbol{V_j^- = \{v_1^-, v_2^-, v_3^-, \ldots, v_n^-\}} \qquad 4.10$$

Where V_j^+ and V_j^- are positive and negative ideal solution

$$D_i^+ = \sqrt{\sum_{j=1}^{n}\left(V_{ij} - V_j^+\right)^2} \qquad 4.11$$

$$D_i^- = \sqrt{\sum_{j=1}^{n}\left(V_{ij} - V_j^-\right)^2} \qquad 4.12$$

Where D_i^+ and D_i^- are Euclidian distance in Equation 4.11, Equation 4.12 between each network vertex

$$CC_i = \frac{D_i^-}{D_i^- + D_i^+} \qquad 4.13$$

The vertices are then ranked using the acquired CCi scores to create the final ranking as in Equation 4.13, which can subsequently be used to choose the vertices for the first network.

4.3 Proposed Algorithm

Proposed Algorithm **(Sonia, Effectual Seed Pick Framework focusing on Maximizing Influence in Social Networks, 2023)**

1. FindVariousSocialNetworkPlatforms():

- Retrieve a list of various social network platforms

2. AnalyzeUserEngagementInSocialNetworks(platform):

- Input: social network platform

- Apply EEA (Extended Exponential Algorithm) to the platform

- Use SkipGram neural network to build the word distribution for the text corpus

- Represent words as vector representations

- Assess the COVID-19 topics in the platform

3. ClusterWords(particular_vector_representation):

- Input: vector representation of words

- Apply the Partitioning Around Medoids (PAM) algorithm

- Use the cosine matrix's closeness metric to cluster the words

4. SeparateCOVID19Contents(clustered_words):

- Input: clustered words

- Identify and separate the contents related to COVID-19

5. DistinguishNetworkNodesUsingBNS(centrality_relationship, unique_attributes):

- Input: centrality relationship with other nodes, unique attributes (e.g., age, sex, gender)

- Apply the BNS (Bayesian Network Structure) method to distinguish network nodes

6. SelectBestTargetNodesInEachAttribute():

- For each attribute:

- Select the best target nodes based on the BNS method

7. ESPForSeedSelection(attribute, target_nodes):

- Input: attribute, target nodes

- Apply the ESP (Efficient Seed Pick) algorithm to select seeds

- Consider the highest potential of spreading information to the targeted nodes

8. SelectTargetedNodeWithHighestRanking():

- For each social network:

- Apply the Fuzzy-VIKOR method to rank the nodes based on their relative distance to the solution

- Select the targeted node with the highest ranking

9. Repeat steps 4-8 whenever new information diffuses in the social networks

4.3.1 Extract Covid -19 Related Words:

Functions to ExtractCovid19RelatedWords(TwitterDataSet)

Input: TwitterDataSet - Dataset containing Twitter data

Output: Covid19RelatedWords - List of COVID-19 related words

1. CleanStopWordsSpecialCharHashtags(TwitterDataSet) // Clean the Twitter dataset by removing stop words, special characters, and hashtags

2. BuildWordEmbedding(TwitterDataSet) // Build word embeddings for the cleaned Twitter corpus

3. PAM = CalculatePointwiseMutualInformation( ) // Calculate Pointwise Mutual Information (PAM) scores for word pairs using the word embeddings

- Iterate over each word pair (wi, wj) in the corpus

- Compute the PAM score for the word pair based on the formula

$$P(wi/wj) = \frac{e_{xp\left(u_{wi}^{T} v_{wj}\right)}}{\sum_{l=1}^{v} exp\left(u_{l}^{T} v_{wj}\right)}$$

4. DistributedRepresentation = TrainSkipGramModel(TwitterDataSet) // Train the SkipGram model to obtain distributed representations of words

- Use stochastic gradient descent with the gradient formula:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=-k}^{K} log\, P\left({w_{t+j}}/{w_t}\right)\right]$$

5. RestrictContents(DistributedRepresentation) // Restrict the contents by using an importance formula

- Calculate the importance I for each word based on the formula: $I = \left[\frac{R\,O}{(1+d)^t}\right]^t$

6. Covid19RelatedWords = SelectCovid19RelatedWords(RestrictedContents) // Select COVID-19 related words based on predefined criteria

Return Covid19RelatedWords

4.3.2 Best Node Selection Method:

Function to SelectBestNode(SkipGramOutput, Attributes)

Input: SkipGramOutput - List of Covid-19 related words from the SkipGram model

Attributes - Set of attributes (e.g., Followers, Favourite, Friends)

Output: BestNode - The selected best node

1. AssignNumericalValues(Attributes) // Assign numerical values to the attributes

- Assign numerical values to each attribute based on a predefined scale or criteria

2. CentralityMeasures = ComputeCentralityMeasures(Attributes) // Compute additional centrality measures between the attributes

- Compute various centrality measures, such as betweenness centrality or eigenvector centrality, based on the attributes

3. SumValues = AddValues(SkipGramOutput, CentralityMeasures) // Add the numerical values of SkipGram output and centrality measures

- Iterate over each word in SkipGramOutput

- Add the assigned numerical value of the word with the corresponding centrality measure value

4. BestNode = FindMaxNode(SumValues) // Find the node with the maximum summed value

- Find the word in SumValues with the highest value

5. Return BestNode

4.3.3 Effective Seed Pick Process :

Function to SelectSeedWithHighestPotential(NetworkG, Attributes)

Input: NetworkG - Network represented by SkipGram model

Attributes - Set of attributes (e.g., Followers, Friends, Favourite)

Output: SeedNode - The selected seed node for maximum information propagation

1. CentralityMeasures = ComputeCentralityMeasures(NetworkG) // Compute centrality measures for the network

- Compute various centrality measures, such as betweenness centrality or eigenvector centrality, for the nodes in NetworkG

2. AssignNumericalValues(Attributes) // Assign numerical values to the attributes

- Assign numerical values to each attribute based on a predefined scale or criteria

3. HighestPotential = 0

4. SeedNode = None

5. for each Node in NetworkG do

6. NodePotential = 0

7. for each Attribute in Attributes do

*8. NodePotential += CentralityMeasures[Node] * NumericalValues[Node][Attribute] // Calculate the potential for each attribute*

9. end for

10. if NodePotential > HighestPotential then

11. HighestPotential = NodePotential

12. SeedNode = Node

13. end if

14. end for

15. Return SeedNode

4.3.4 Fuzzy – Vikor: to Select the Best Node from the Multiple Attributes:

Function to SelectBestNodeFromAttributes(SeedNodes, Attributes)

Input: SeedNodes - List of seed nodes picked in the effective seed pick process

Attributes - Set of attributes

Output: BestNode - The selected best node for information propagation

1. ConstructNodeMarkers(SeedNodes)

- Initialize an empty dictionary NodeMarkers

- Iterate over each attribute in Attributes

- Calculate the node marker value for each node based on the formula:

*NodeMarkers[node] = 1/k * sum(($\tilde{xij}$)^m) for m = 1 to k*

2. DetermineBestAndWorst(NodeMarkers)

- Initialize variables f_1^*= = -infinity and $f_1^\wedge$= infinity

- Iterate over each node marker value in NodeMarkers

- Update , f_1^*= max(, f_1^*, NodeMarkers[node])

- Update $f_1^\wedge$ = min($f_1^\wedge$, NodeMarkers[node])

3. CalculateSAndR(NodeMarkers, f_1^*, $f_1^\wedge$)

- Initialize empty lists S and R

- Iterate over each node marker value in NodeMarkers

- Calculate S_j = sum$\left[(wi(f_i^* - fij)) \;/\; (f_1^* - \;\; f_1^\wedge)\right]$ for each node

- Calculate R_j = max$\left[(wi(f_1^* - fij))\ /\ (f_1^k - \ f_1^{\wedge})\right]$*for each node*

- Append S_j and R_j to the lists S and R, respectively

4. ComputeValues(S, R)

- Initialize variables $s^* = min(sj),\ s^{\wedge} = max(sj),\ R^* = min(Rj),\ R^{\wedge} = max\ (Rj)$

- Calculate $Q_j = v(sj - s^*/(s^{\wedge} - s^*) + (1\text{-}v)(Rj - R^*/(R^{\wedge} - R^*)$ *where v = (n + 1) / (2n), n is the number of nodes*

5. RankAlternatives(S, R, Q)

- Sort the nodes based on the values of S, R, and Q in ascending order

6. C1 = CheckCondition1(Q(A(2)) - Q(A(1)) >= DQ)

- Set DQ = 1 / (J - 1), J is the total number of nodes

- If the condition is satisfied, proceed to the next step; otherwise, go to Step 9

7. C2 = CheckCondition2(A(1) is best ranked by S or R/S and R)

- If A(1) is the best ranked by S or R/S and R, proceed to the next step; otherwise, go to Step 9

8. BestNode = A(1)

- Select the best node as A(1)

9. Return BestNode Selected from the multiple attributes to Propagate the information in a selected social network pervasive health care

4.4 Result and Discussion:

This section provides a thorough analysis of the implementation outcomes, highlights the effectiveness of our suggested system, and concludes with a comparison to ensure that our proposed approach outperforms the alternatives.

4.4.1 Experimental Setup:

The system requirements for this work were implemented in Python 3.9.6, and the simulation results are given below. The experimental design for using an efficient seed select framework for influence maximization is described in Table 4.1

Table 4.1 Experimental Setup

Tool	Python 3.9.6
OS	Windows 7(64 - bit)
Processor	Intel premium
Ram	8 GB RAM
Dataset	Social Network data, Twitter Data

4.4.2 Evaluation Metrics and Simulation Output:

This section presents the network simulation and the resulting outputs of implementations. The results are as follows in the following figures: Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7 shows the frequency of tweets concerning the location from which tweeter users are tweeting related to corona.

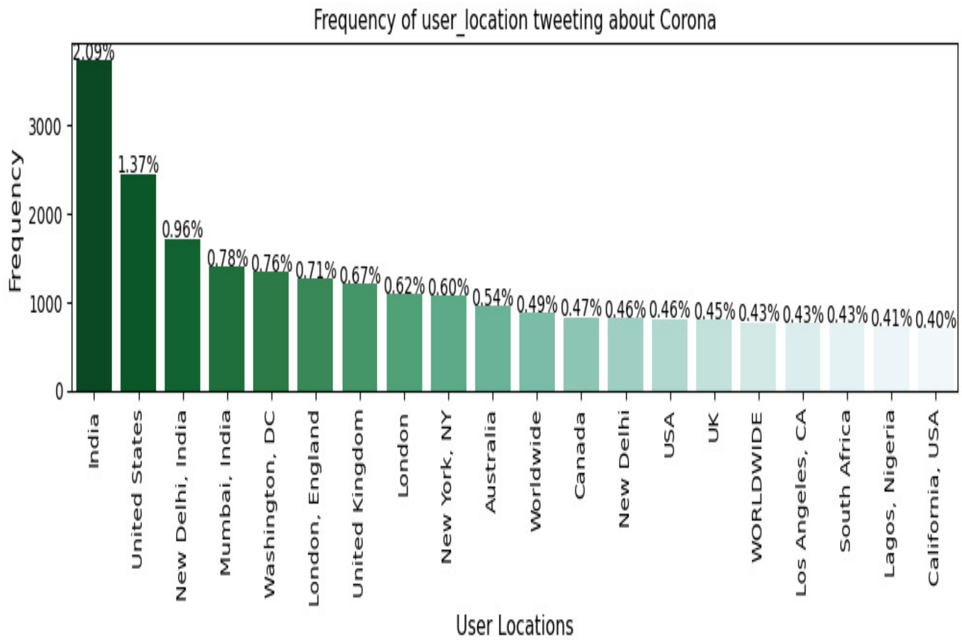


Figure 4.2 Frequency of Tweets Versus the Location of Users

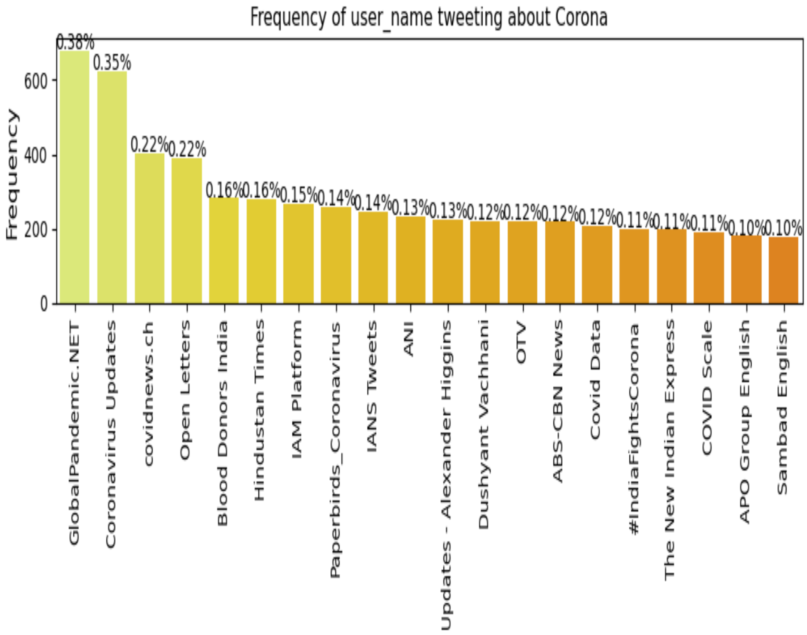


Figure 4.3 Frequency of user_name Tweeting About Corona

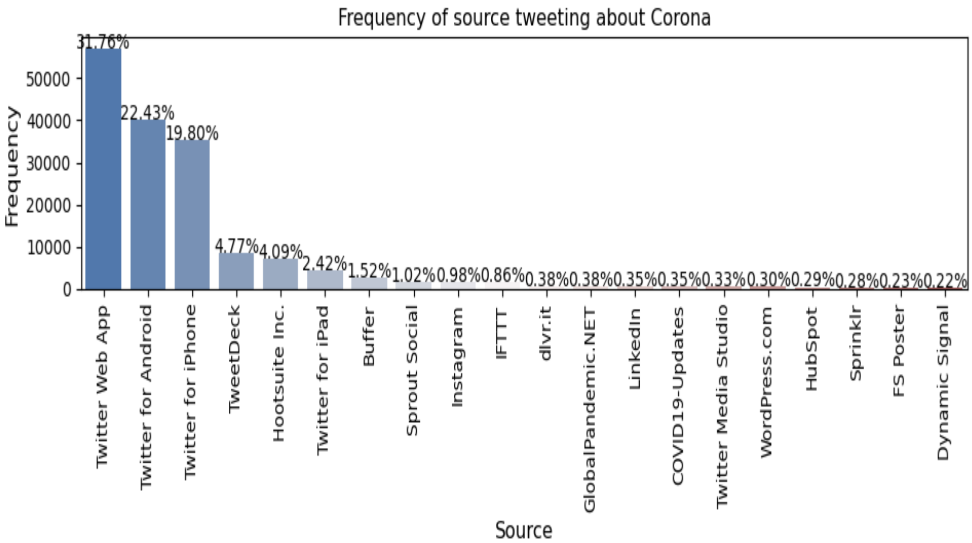


Figure 4.4 Frequency of Source Tweeting About Corona

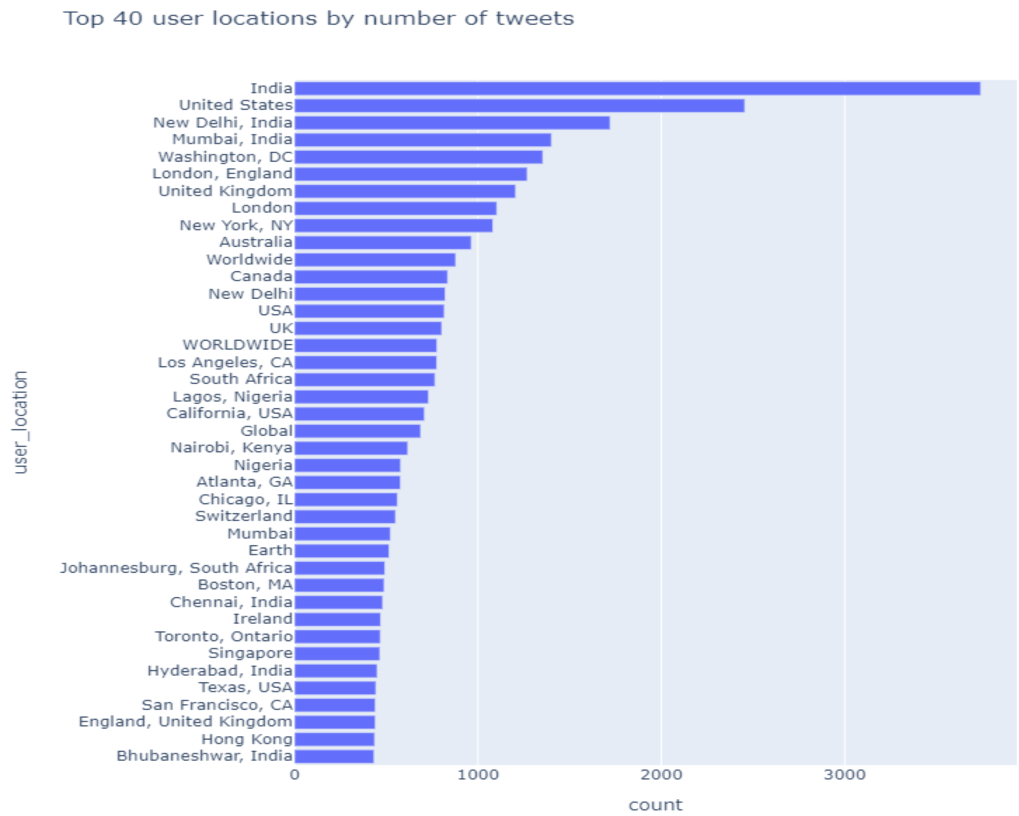


Figure 4.5 Graph Between user_location and the Count of Tweets Related To Corona

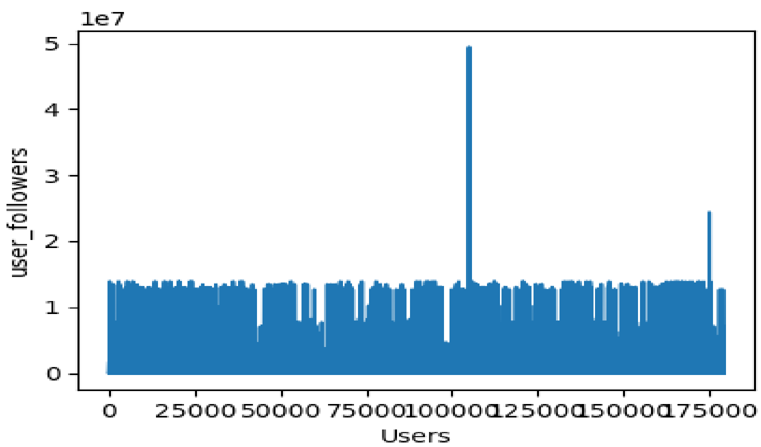


Figure 4.6 6 Graph Between users and user_followers

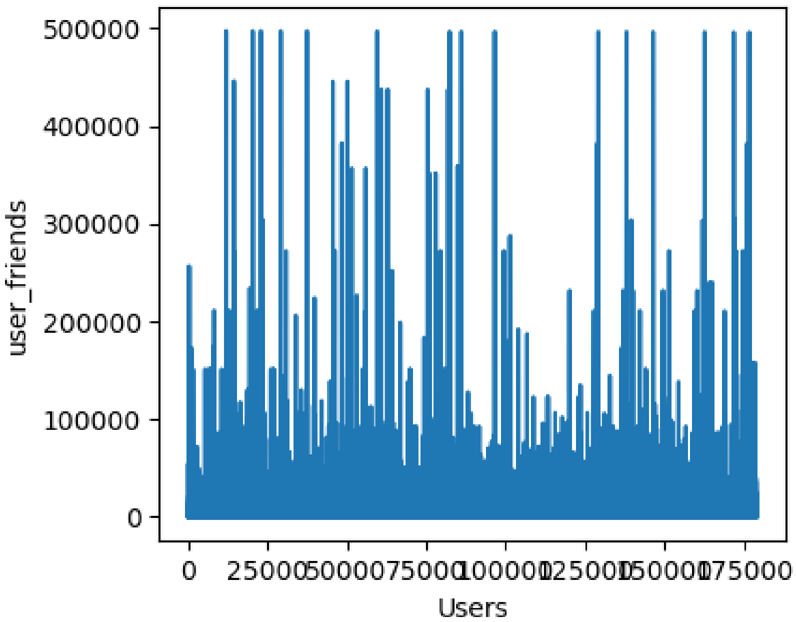


Figure 4.7 Graph Between users and user_friends

The following equation detects the accuracy of the proposed methodology and the obtained image.

The accuracy of the clinical text data is calculated using Equation 4.14 ,

$$\mathbf{Accuracy} = \left[\frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}}\right] * \mathbf{100} \qquad 4.14$$

TP- True Positive Value

TN- True Negative Value

FP- False Positive Value

FN- False Negative Value

F1 Score is defined as in Equation 4.15,

$$\mathbf{F1} = \frac{\mathbf{2 \times (Precision * Recall)}}{\mathbf{Precision + Recall}} \qquad 4.15$$

Where,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision is defined as how closely two or more measurements match one another. The equation is displayed as Equation 4.16,

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} \tag{4.16}$$

Where,

TP - True Positive

FP –False Positive

The capacity of a model to correctly predict an outcome is known as recall. The definition of the recall formula is as Equation 4.17,

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \tag{4.17}$$

where,

TP - True Positive

FN –False Negative

The Twitter dataset's accuracy is compared with the proposed system and other existing techniques like Rand, LDAG, MC-CELF, CEF++, CIM, ComPath, and Pas as in Table 4.2. The graph shows the accuracy of the proposed work; the proposed technique accuracy seems to be 97.9%, and existing techniques rely upon below proposed accuracy as in Figure 4.8. For example, existing technique Rand is 82.17%,LDAG is 81.08%, MC-

CELF is 64.9%, CEF++ is 82.87%, CIM is 81.4%, Compath is 78.1%, and Pas is 81%. From this result, it is determined that the proposed technique is more efficient.

The precision of the Twitter dataset is taken for comparison with the proposed system and other existing techniques. The graph shows the precision of the proposed work; the proposed technique precision seems to be 98.9%, and existing techniques rely

below proposed precision as in Figure 4.9. Existing technique Rand is 81.17%, LDAG is 82.08%, MC-CELF is 63.09%, CEF++ is 84.87%, CIM is 85.4%, Compath is 76.1%, and Pas is 87% as in Table 4.3. From these results, it is determined that the proposed technique is more efficient.

The Recalls of the Twitter dataset are compared with the proposed system and other existing techniques, as in The graph shows the recalls of the proposed work. Proposed technique recall seems to be 98.7%, and existing techniques rely below the proposed precision. Existing technique Rand is 82.17%, LDAG is 83.08%, MC-CELF is 66.9%, CEF++ is 84.67%, CIM is 85.8%, Compath is 73.1%, and Pas is 86%, as in Table 4.4. From these results it is determined that the proposed technique is more efficient as in Figure 4.10. The F1-Score of the Twitter dataset is taken for comparison with the proposed system and other existing techniques, as in Table 4.5. The graph shows in Figure 4.11 the F1-Score of the proposed technique F1-Score seems to be 98.7%, and existing techniques rely below the proposed precision. Existing technique Rand is 80.17%, LDAG is 81.08%, MC-CELF is 76.9%, CEF++ is 84.67%, CIM is 83.8%, Compath is 73.1%, and Pas is 85%. From these results, it is determined that the proposed technique is more efficient, as in Figure 4.11.

The Accuracy and F1-Score of the Gab dataset is taken for comparison with the proposed system and other existing techniques, as in Figure 4.12 The graph shows the accuracy and F1-Score of the proposed work, the proposed technique accuracy seems to be 97.7%, and F1-Score appears to be 98%, and existing techniques rely below proposed accuracy and F1-Score. These results show that the proposed technique is more efficient, as in Figure 4.12. Proposed work; proposed technique precision seems to be 97.8%, and recall appears to be 98%, and existing techniques rely below proposed precision and recall as in Figure 4.13. From this results, it is determined that the proposed technique is more efficient

Table 4.2 Accuracy of the Proposed Framework and Another Previous Algorithm

Methods	Accuracy(%)
Rand	82.17
LDAG	81.08
MC-CELF	64.9
CEF++	82.87
CIM	81.4
ComPath	78.1
PaS	81
Proposed	97.9

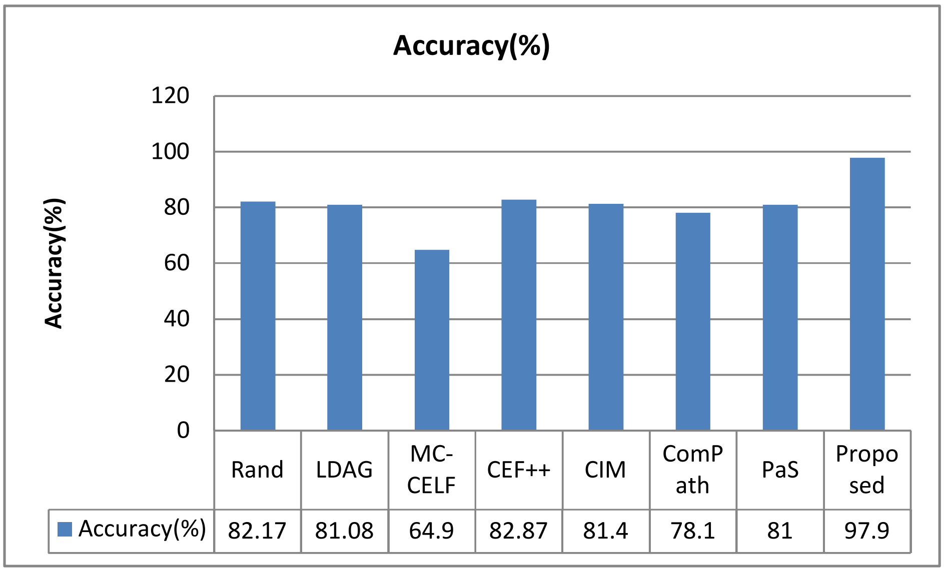


Figure 4.8 Graph Showing the Accuracy of the Proposed Framework

Table 4.3 Precision of the Proposed Framework and the Previous Algorithm

Methods	Precision(%)
Rand	81.17
LDAG	82.08
MC-CELF	63.09
CEF++	84.87
CIM	85.4
ComPath	76.1
PaS	87
Proposed	98.9

.

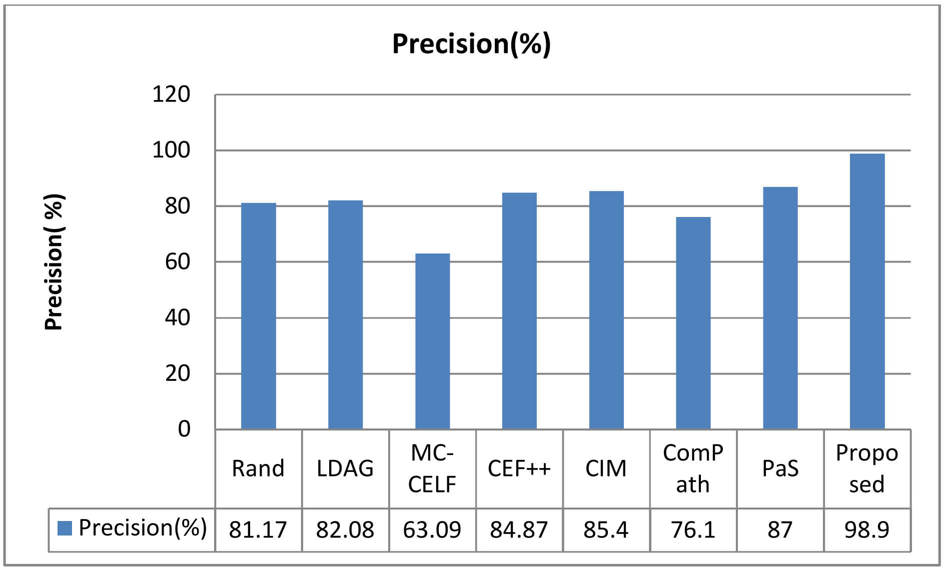


Figure 4.9 Precision of the Previous Algorithm and the Proposed Framework

Table 4.4 Recall Value of the Proposed as well as Other Previous Algorithms

Methods	Recall(%)
Rand	82.17
LDAG	83.08
MC-CELF	66.9
CEF++	84.67
CIM	85.8
ComPath	73.1
PaS	86
Proposed	98.7

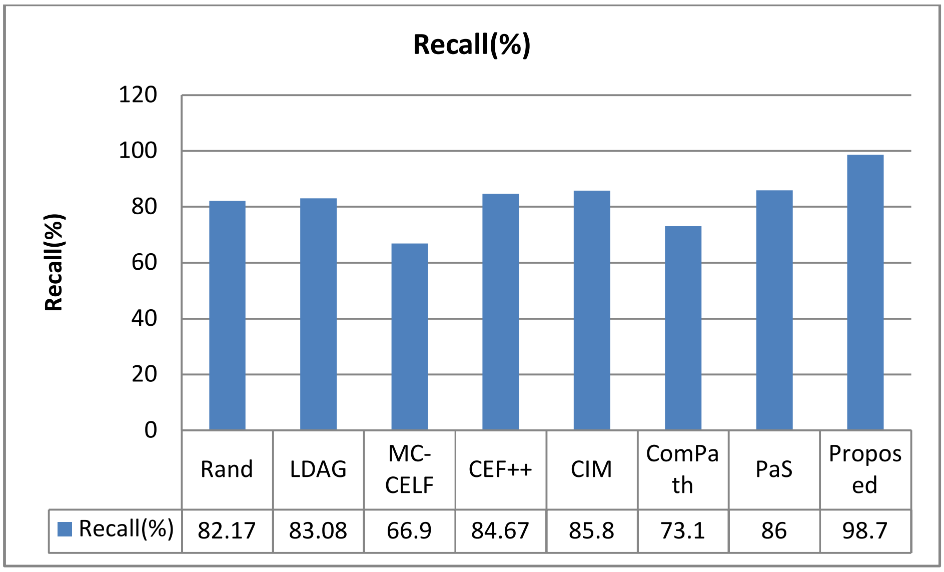


Figure 4.10 Graph Showing a Recall of the Proposed Framework Compared to Previous Algorithms

Table 4.5 F1-score for the Proposed as well as Other Existing Algorithms

Methods	F1-Scorel(%)
Rand	80.17
LDAG	81.08
MC-CELF	76.9
CEF++	83.67
CIM	83.8
ComPath	73.1
PaS	85
Proposed	98.7

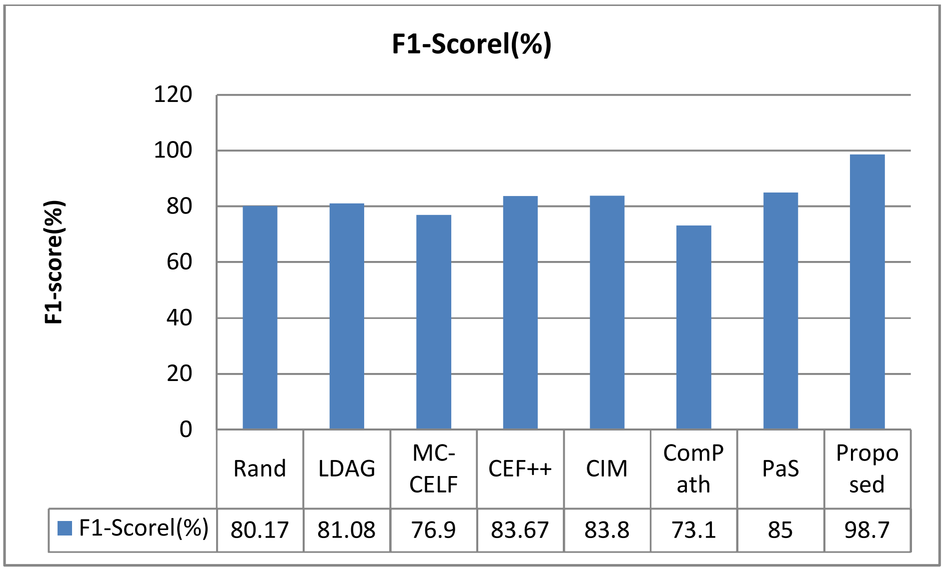


Figure 4.11 Graph Showing F1-Score for the Proposed as well as Existing Algorithms

The Gab dataset's precision and recall are compared with the proposed system and other existing techniques, as in Table 4.7. The graph shows the precision and recall of the Gab Datassset. The precision and recall of the Reddit dataset is taken for comparison with the proposed system and other existing techniques as in Table 4.8. The graph in Figure 4.14 shows the precision and recall of the proposed work; proposed technique precision seems to be 97.8%, and recall appears to be 98%, and existing techniques rely below proposed precision and recall. From this results it is determined that proposed technique is more efficient .

The Accuracy and F1-Score of the Reddit dataset is taken for comparison with the proposed system and other existing techniques as in Table 4.9. The graph as in Figure 4.15 shows the accuracy and F1-Score of the proposed work, proposed technique accuracy seems to be 97.8% and F1-Score seems to be 98%, and existing techniques rely below the proposed accuracy and F1-Score. From these results, it is determined that the proposed method is more efficient.

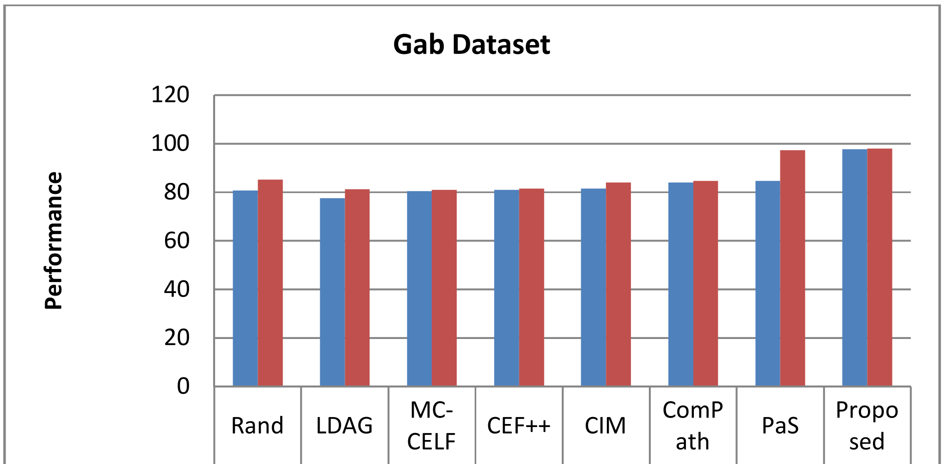
The precision and recall of the Instagram dataset is taken for comparison with the proposed system and other existing techniques as in Table 4.10. The graph shows the precision and recall of the proposed work; the proposed technique precision seems to be 97.8%, and recall seems to be 98%, and existing techniques rely below the proposed precision and recall. From these results, it is determined that the proposed technique is more efficient, as in Figure 4.16.

The Accuracy and F1-Score of the Instagram dataset are taken for comparison with the proposed system and other existing techniques as in. The graph in Figure 4.17 shows the accuracy and F1-Score of the proposed work, proposed technique accuracy seems to be 97.8% and F1-Score seems to be 98%, and existing techniques rely below the proposed accuracy and F1-Score. From this results, it is determined that the proposed approach is more efficient

The Youtube dataset's precision and recall are compared with the proposed system and other existing techniques, as in Figure 4.18. The graph shows the precision and recall of the proposed work; the proposed technique precision seems to be 97.8%, and recall seems to be 98%, and existing techniques rely below the proposed precision and recall as in Table 4.12. From this results, it is determined that the proposed approach is more efficient.

Table 4.6 Accuracy and F1-Score for Proposed as well as Other Algorithms

Methods	Accuracy(%)	F1-Score
Rand	80.82	85.31
LDAG	77.56	81.32
MC-CELF	80.45	81
CEF++	81	81.5
CIM	81.5	84
ComPath	84	84.6
PaS	84.6	97.34
Proposed	97.7	98



	Rand	LDAG	MC-CELF	CEF++	CIM	ComPath	PaS	Proposed
Accuracy(%)	80.82	77.56	80.45	81	81.5	84	84.6	97.7
F1-Score	85.31	81.32	81	81.5	84	84.6	97.34	98

Figure 4.12 Graph of Accuracy and F1-Score for Proposed and Existing Algorithm for Gab Dataset

Table 4.7 Precision and Recall of Proposed as well as Previous Algorithm for Gab Dataset

Methods	Precision(%)	Recall(%)
Rand	72.62	80.12
LDAG	71.78	78.30
MC-CELF	75.78	82.01
CEF++	81	81
CIM	81	82.4
ComPath	72	84
PaS	84.8	83.2
Proposed	97.8	98

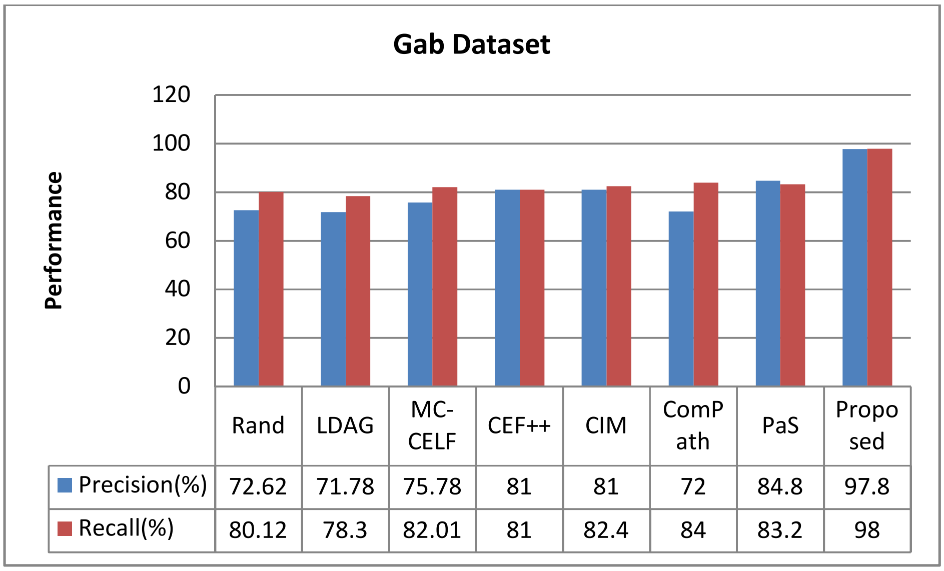


Figure 4.13 Graph For Precision and Recall of Proposed and Existing Algorithms for the Gab Dataset

Table 4.8 Precision and Recall of Proposed and Existing Algorithm for the Reddit Dataset

Methods	Precision(%)	Recall
Rand	71.82	81.22
LDAG	71.88	77.40
MC-CELF	73.78	81.01
CEF++	81	81
CIM	82	82.5
ComPath	84.4	84
PaS	84.6	83.6
Proposed	97.8	98

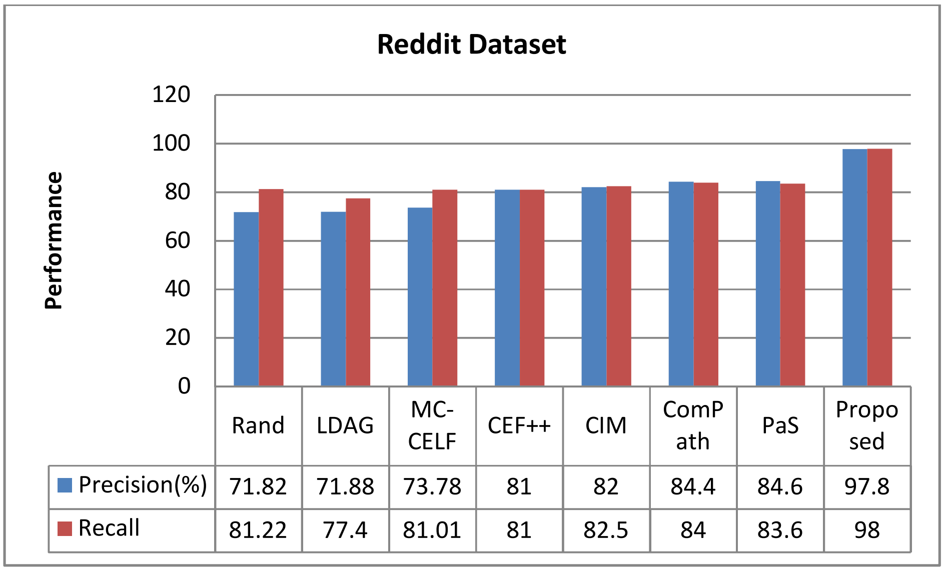


Figure 4.14 Graph For Precision and Recall of Proposed and Existing Algorithms for the Reddit Dataset

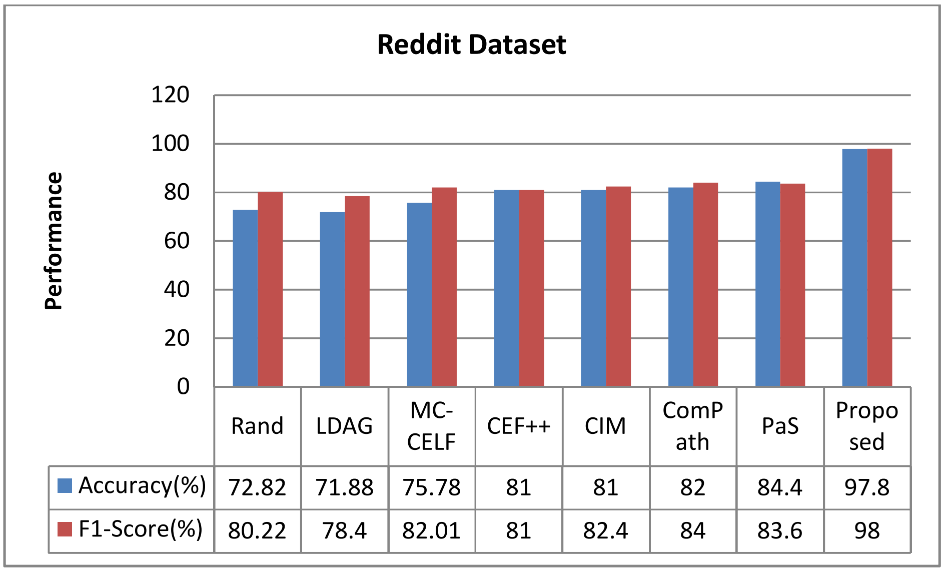


Figure 4.15 Graph of Accuracy and F1-Score of Proposed and Existing Algorithm for Reddit Dataset

Table 4.9 Accuracy and F1-Score of Proposed and Existing Algorithm for Reddit Dataset

Methods	Accuracy(%)	F1-Score(%)
Rand	72.82	80.22
LDAG	71.88	78.40
MC-CELF	75.78	82.01
CEF++	81	81
CIM	81	82.4
ComPath	82	84
PaS	84.4	83.6
Proposed	97.8	98

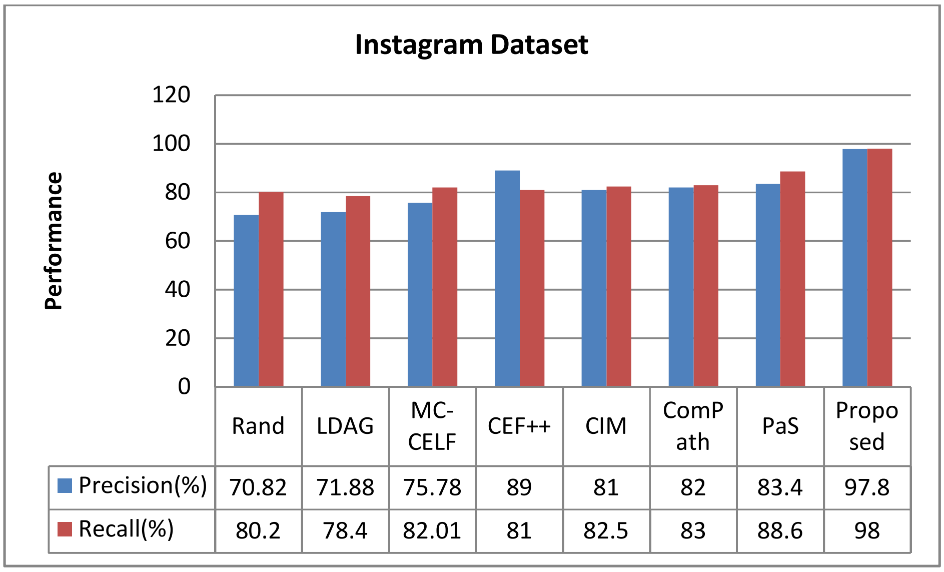


Figure 4.16 Graph of Precision and Recall of Proposed and Existing Algorithms for the Instagram Dataset

Table 4.10 Precision and Recall of Proposed and Existing Algorithm for the Instagram Dataset

Methods	Precision(%)	Recall(%)
Rand	70.82	80.2
LDAG	71.88	78.40
MC-CELF	75.78	82.01
CEF++	89	81
CIM	81	82.5
ComPath	82	83
PaS	83.4	88.6
Proposed	97.8	98

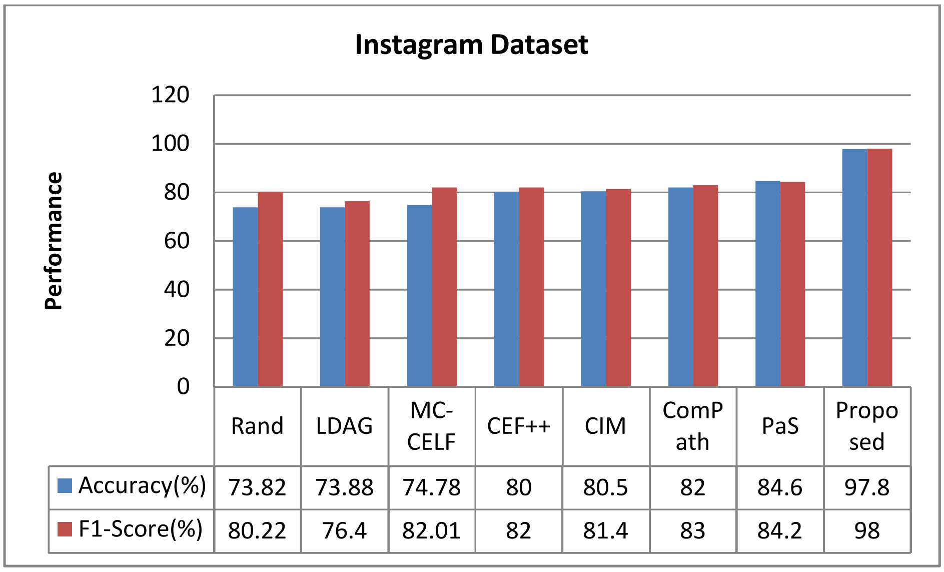


Figure 4.17 Graph of Accuracy and F1-Score of Proposed and Existing Algorithm for Instagram Dataset

Table 4.11 Accuracy and F1-Score of the Proposed and Existing Algorithm for the Instagram Dataset

Methods	Accuracy(%)	F1-Score(%)
Rand	73.82	80.22
LDAG	73.88	76.40
MC-CELF	74.78	82.01
CEF++	80	82
CIM	80.5	81.4
ComPath	82	83
PaS	84.6	84.2
Proposed	97.8	98

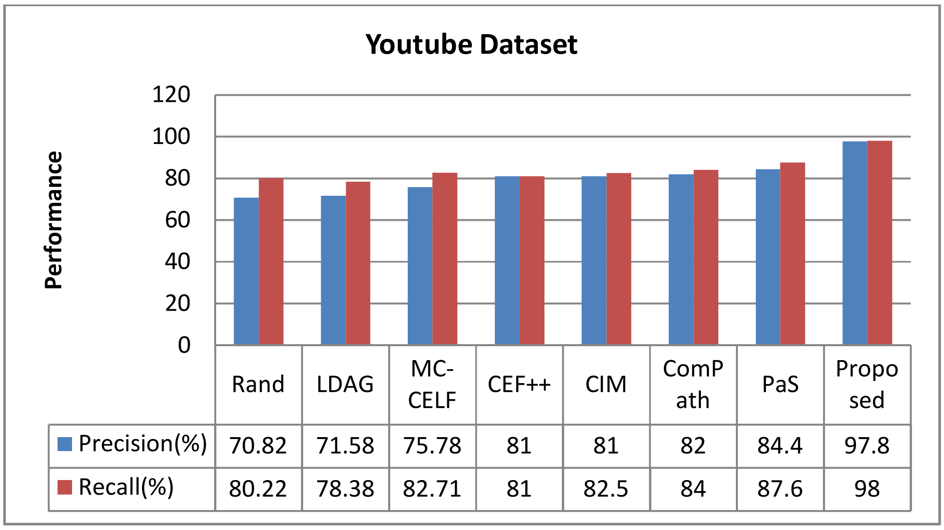


Figure 4.18 Graph of Precision and Recall of the Proposed and Previous Algorithms for the Youtube Dataset

Table 4.12 Precision and Recall of the Proposed and Existing Algorithm for the Youtube Dataset

Methods	Precision(%)	Recall(%)
Rand	70.82	80.22
LDAG	71.58	78.38
MC-CELF	75.78	82.71
CEF++	81	81
CIM	81	82.5
ComPath	82	84
PaS	84.4	87.6
Proposed	97.8	98

4.5 Conclusion

This chapter has proposed algorithm which is having maximum influence maximization in the network. The proposed work can be applied in any area or field like politics, product promotion, or service promotion. In this chapter, we have applied our proposed algorithm in the field of pervasive healthcare. The proposed framework is efficient to provide pervasive healthcare. When for different datasets like YouTube, Facebook, Reddit, Instagram, Gab,and Twitter, the proposed framework is compared with the existing algorithms to find out the seeds for maximum influence maximization, and the accuracy, recall, F1-score, and precision are high as compared to other existing algorithms.The next chapter discusses the consideration of net work's memory effect or social reinforcement effect on the Influence Maximization Problem.

Chapter 5 MULTI NEIGHBOR SEED SELECTION FOR INFLUENCE MAXIMIZATION

This chapter further extends specific context-based rules and the construction of a comprehensive ''reputation and trust-based system'' by identifying bidirectional effects between influence and trust. Here multi neighbor seed selection is used for influence maximization. The proposed approach here considers the social reinforcement effect and uses Neighbour Degree Value(NDV), the Judgement leader pick for Opinion Leader selection in the network.

5.1 Introduction

To enhance the efficacy of Influence Maximization, a Multi Neighbour seed selection approach is presented. The proposed method considers the network's memory effect or social reinforcement effect. It offers the Multi neighbor seed selection approach, in which neighbor degree value (NDV) is determined to estimate the influence strength of the seeds and addresses the issues of seed selection due to limited coverage area and lack of discriminatory power. The Judgment leader picks a weighting grade, in which the Judgment grade value is considered for leader selection. Consequently, the proposed structure effectively addresses the trust management issues in ubiquitous services and thus increases influence maximization.

5.1.1 Convolutional Neral Network Model for Social media marketing

Machine learning and deep learning are assisting us in gaining a deeper understanding of data and the knowledge that it can provide. A Convolutional Neural Network (CNN) that processes image data by assigning significance, through learnable weights and biases, to different features and objects present in the image is a deep learning algorithm. This enables the CNN to differentiate and identify various aspects and objects within the image. CNN requires less preprocessing of the data. Though filters in rudimentary methods are hand-engineered, CNN can learn these filters/characteristics with enough experience. CNN's job is to translate the pictures into a format that's easier to process while maintaining the functionality that makes a forecast popular. This is critical when designing a system that can learn features while still robust to massive datasets.

The Convolutional Neural Network has four layers:

1. The convolutional layer is the first layer, and it is here where the activity begins. It aims to distinguish image features. In nature, it usually progresses from wide (i.e., shapes) to narrow (i.e., colors) (i.e., identifying elements of an object, the face of a certain man, etc.).
2. The Rectified Linear Unit layer is the next layer (aka ReLu). This layer is a convolutional layer that has been extended. ReLu aims to improve the picture's non-linearity. It is a method for removing excess fat from an image to aid attribute extraction.
3. The pooling layer aims to reduce the number of input parameters, allowing regression to be performed. In other words, it focuses on the most critical elements of the collected data.
4. A typical feed-forward neural network is used in the linked layer. It's the last straight line before the finish line, and all is in sight. It'll only take a few days for the reports to be verified.

The proposed method as in Figure 5.1 uses the Graph Convolution Network (GCN). A graph convolutional neural network (GCN) is a neural network designed to process data represented as a graph structure. It can utilize the inherent structural information in graphs to perform convolution operations and learn useful features from the data. In other words, GCN is a neural network architecture that is tailored to handle graph-structured data and leverage the graph topology for improved performance. It solves classifying nodes in a graph when only a subset of nodes has labels (semi-supervised learning). The basic idea behind GCN is that it obtains feature information from all of its neighbors and the feature of the node itself for each node. Assume we're going to use the average equation. It will repeat this process for all nodes. Finally, these average values are fed into a neural network. The convolution process in graph convolutional neural networks (GCNs) is similar to that of convolutional neural networks (CNNs). It involves the multiplication of the input features by filters, which are essentially sets of weights. These filters act as sliding windows that traverse the input graph and extract features from adjacent nodes. In GCNs and CNNs, the filters are shared across the input data, allowing the network to learn and extract similar features across the entire dataset. This process of weight sharing enables the network to efficiently learn features from large datasets and generalize well to unseen data.

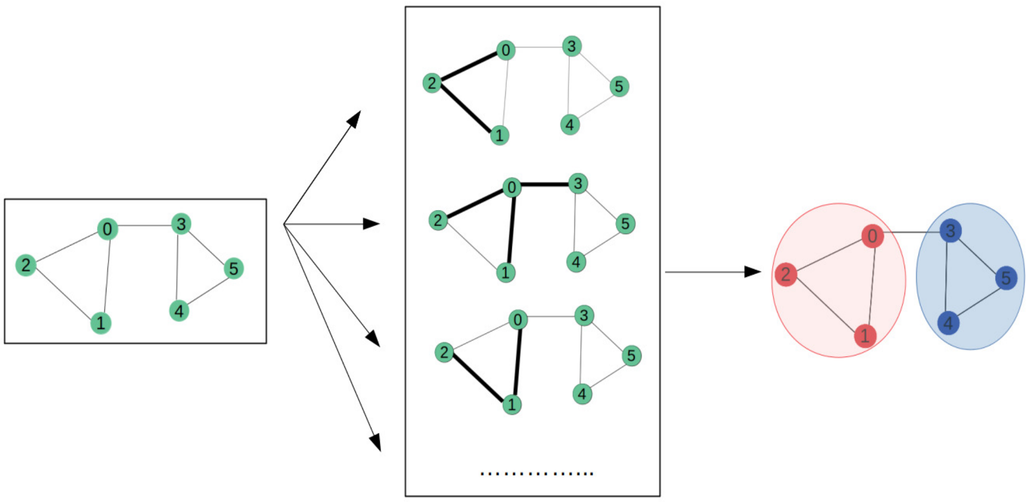


Figure 5.1 Illustration of Graph Convolution Networks

The GCN's fundamental concept in the proposed method is the weighted average of all neighbors' node features (including itself): lower-degree nodes have higher weights. The resulting feature vectors are then fed into a neural network for training. More layers can be added to render GCNs deeper. This process is repeated to find the seeds in the social network data.

5.2 Proposed Framework

5.2.1 An Effectual Seed Pick Framework

Social networks have long been a hot spot for effective viral marketing. In modern times, the concept of influence maximization has become more concentrated to meet the requirements of social network analysis in viral marketing. It aims to select influential users and Judgment leaders to spread as much information as possible through the social network. Several influence maximization approaches considering time and budget constraints have been introduced to achieve effective viral marketing. Despite their benefits, existing techniques had some limitations, which can be overcome by implementing an effective seed pick influence maximization framework based on a machine learning approach Figure 5.2 illustrates the proposed system.

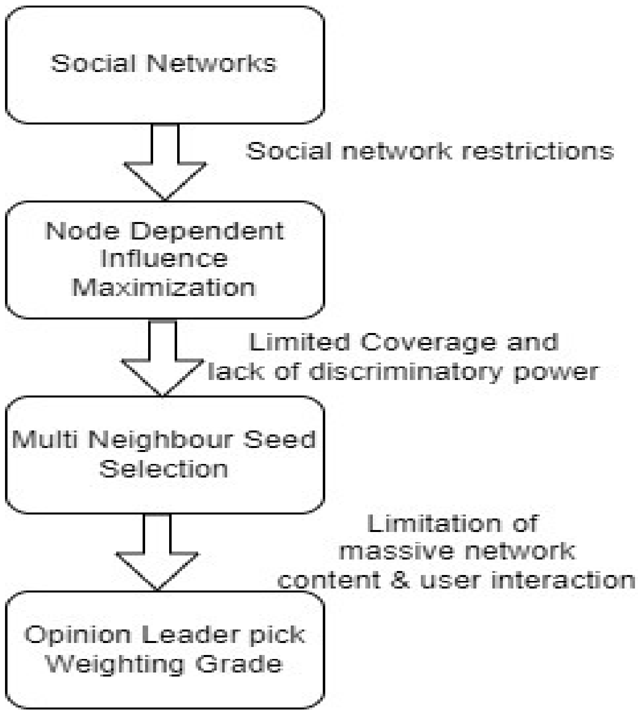


Figure 5.2 Schematic Diagram of the Proposed Approach

However, existing approaches did not take into account the network's memory effect or social reinforcement effect. As a result, previously used connections in the social network can impact the current information used in real-time. Furthermore, if at least one neighbor agrees and transfers information to the source node, the source node is compelled to do so. This necessitates a novel approach examining memory variations and the social reinforcement effect. Seed-dependent influence maximization is used to address this issue. The seed potentiality algorithm produces an optimal solution by sampling and verifying a specific number of seeds. Every Seed has the opportunity to increase its potentiality with each iteration, and the seeds with the highest potentiality are chosen to control the memory and social reinforcement effects.

Furthermore, when selecting seeds in Public systems, their influence strength, referred to as degree, is considered. However, when considering the influence strength, degree-based techniques face accuracy issues due to a limited consideration scale, resulting in a limited coverage area which is solved by the Multi neighbor seed selection method. The neighbor degree value (NDV) is evaluated to estimate the strength of the seeds' influence. NDV is the sum of the lowest path lengths among the Seed and its neighbor. The fixed NDV is chosen as the second-highest rank to address runtime inefficiency and application area constraints.

Furthermore, these techniques lack discriminatory power because many seeds have the same rank, resulting in an unstable situation when selecting the seeds. As a result, influence maximization is required, considering multiple neighbors' levels to calculate influence strength. The existing methods for determining Judgment leaders consider only the current network structure, ignoring temporal characteristics. As a result, the massive content of public systems and user interaction is removed, which can jeopardize the accuracy of the Judgment leaders' monitoring process, which is addressed by the weighting grade of the Judgment leader's pick, here the total of the multiplication of the social metric and its priority is referred to as the significant grade value in this case, and the average of all significant grade values obtained for updates in the social network is used as the threshold. Significant grade values more than or equivalent to the threshold will be considered to select Judgment leaders.

Furthermore, the number of outputs of the updates that are influenced among the overall sum of inputs of all updates is considered for Judgment leader selection. The Judgment leader with the most social values, as well as Judgment values in the updates, is chosen. The detailed description of each novel method has been explained in the upcoming sections.

5.2.2 Seed Dependent Influence Maximization

Influence Maximization (Kempe, 2015) (Li J. Y., 2016) is a complex process of selecting seed nodes in a social network to maximize influence spread. The Independent Cascade Model (ICM) (Goldenberg, 2001)is the most commonly used propagation method. In contrast, ICM recognizes only two states: active and inactive. Furthermore, the process's experience does not affect a node's ability to transition from inactive to active. This isn't enough to address the real issue of viral marketing. In this article, we propose Seed Dependent for Viral Marketing, a modern viral marketing propagation model that is best tailored to the issue of viral marketing and may also be used for viral distribution and the dissemination of some other living being. Each node is assigned one of five states as in Figure 5.3 untold, told, responsive, conforming, and outdated . The node has by no means heard of the advertisement of the manufactured goods; informed indicates that the node has heard of the publicizing; and unclear indicates that the node has never heard of the publicizing; sensitive indicates that the node is concerned in the manufactured goods but hasn't determined whether or not to buy it; and the conforming node is the one which purchases the services or product after getting interested. Since it has outlived the commodity, the node is deemed outdated and fails.

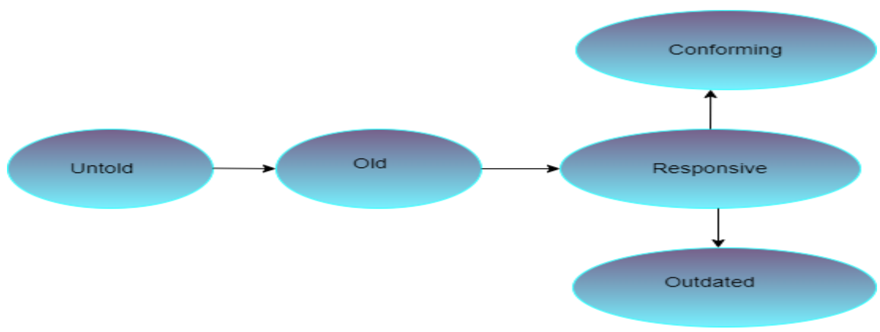


Figure 5.3 Word of Mouth Marketing Seed Based Cascade Structure

Figure 5.3 depicts the state transition process. Maximum nodes are initially unaware of the product. In addition, a small subset of initially told nodes attempts to tell their neighbors about the product. Once the information is said, with chance $P_{u,A,}$ a node A is transformed into receptive under it. Very certainly, P_A, the responsive node, will disseminate the knowledge to its neighbors and thus turn out to be conforming. The interested node becomes conforming or outdated after spreading the information and will no longer spread it.

It has been discovered that whether or not a person will purchase a product is determined by two types of factors: core reasons and exterior reasons. Core reasons indicate an individual's innate favorite meant for the merchandise. That is to say, it is the likelihood that an individual will buy the product, not at all the influence of their friends. The effect of the external environment on consumer buying habits is reflected in external factors. For example, external factors such as the impact of social strengthening and memory are essential. In other words, the more friends a person has, the more likely the individual would buy the product. As shown in Figure 5.3, the customer is initially unfamiliar with the merchandise. After knowing about it, the customer may become attracted, and the customer may then agree to take the merchandise.

Properly, the probability $pp_A(t)$ as in Equation 5.1 that a given node A purchases merchandise at time t are calculated as follows:

$$\boldsymbol{pp_A(t) = \left(pp_A^{min} - \ pp_A^{max}\right) exp^{-sr1me1(A,t)-sr2me2(A,t)} + pp_A^{max}} \qquad 5.1$$

Function (2)'s components are summarized as follows:

The probability of a node A making a purchase, denoted as pp_A^{min}, is influenced by internal factors. The node's memory effect is captured by two variables, *me1(A,t)* and *me2(A,t),* which represent the number of neighbors in responsive and conforming states at time t. The social affirmation effect is reflected by parameters *sr1* and *sr2* for receptive and conforming states. The total probability of purchase, pp_A^{max}, is the best estimate used to offset the impact of the social affirmation effect and memory effect. As *me1(A,t)* and *me2(A,t)* increase, $pp_A(t)$ approaches pp_A^{max} at speed specified by parameters *sr1* and *sr2*, both of which are positive.

5.2.3 Seed Potentiality Algorithm:

First, this algorithm focuses on accepted nodes rather than all influenced (attracted or acknowledged) nodes. Following this, we must recalculate $pp_A(t)$ after a round of propagation simulation.

In this algorithm, only a fixed number of candidate nodes are sampled in each round (Sampling). The guess (Verification) is verified once the candidates are produced. Specifically, each node is assigned a Prospective Possibility PP that will be improved or reduced based on efficiency.

Each node's possibility PP is initially determined by its out-rank as in Equation 5.2.

$$\boldsymbol{PP(A) \propto \log\left(deg_{out}(A) + 1\right)} \tag{5.2}$$

Where, $deg_{out}(A)$ is the node A's out-rank .

Formerly, *PP* was re-evaluated based on the outcome in every round using Equation 5.3, Equation 5.4, Equation 5.5, Equation 5.6

$$\boldsymbol{a = a_{tar}(seed \cup \{A\}) - a_{tar}(seed \cup \{u\})} \tag{5.3}$$

$$\boldsymbol{recal = \frac{|a|}{a_{tar}(seed \cup \{A\})}} \tag{5.4}$$

$$\boldsymbol{\alpha = \frac{1}{2}\left|log\frac{1 - recal}{recal}\right|} \tag{5.5}$$

$$\boldsymbol{PP(A) \propto PP(A)e^{-\alpha}} \quad ; \textbf{ if } \boldsymbol{a} < 0 \ \boldsymbol{and\ a} > 0 \tag{5.6}$$

Where $a = a_{tar}(seed \cup \{A\}) - a_{tar}(seed \cup \{u\})$is Conforming nodes that exchange *u* with *A* are improved or reduced as in Equation 5.3. The more capable a node is, the more likely it will be picked. After a reasonable number of iterations, each node has the potential to change the result. This is analogous to the theory of advancement of machine learning, in which the last classification increases the weight of incorrectly classified data while reducing the weights of properly classified data in which the network graph S and initial seeds with the size of k from social network data set has been given to the input of the neural network in machine learning approach. The inputs given to the neural network are processed with Seed Potentiality Algorithm 5.1. The potential depend Seed has been selected to avoid the memory and social enforcement effect.

Algorithm 5.1 Seed potential algorithm:

Input: network graph S, initial seeds size t, number of iterations p, number of nodes sampled each time q

Output: conforming seeds Seed

1.Set Seed ← φ R = 10000, c = 0

2. Initiate Seed with t nodes using heuristic algorithms;

*3. Initialize R*C according to (5);*

4. while a_{tar}(Seed) not converge and c < p do

5. pick one node u out of Seed with simply by order;

6. Seed = Seed \{u};

7. Sampling stage:

*8. sample |R *Seed| = q out of R*C randomly;*

9. Verification stage:

*10. for each node A∈R *Seed do*

11. $a_A = 0$;

12. For i = 1 to R do

13.a_A+= a_{tar}(S, Seed ∪{A})

14. end

15. $a_A = a_A/R$;

*16. MordernizeR*C according to (6);*

17.end

18. $v^+ = argmax_{v\in A\setminus Seed}\{a_A\}$;

19. Seed = Seed ∪{A^+ } ;

*20. remove A^+ from R*C;*

21.c++

22.end

23.return Seed;

Even though the above algothirm can rectify the issues of memory and social enforcement effect, some issues occur during seed selection due to limited coverage area and lack of discriminatory power, which is addressed by the Multi Neighbour seed selection approach and detailed explanation described in the next section.

5.3 Multi Neighbour Seed Selection Method

After addressing the memory and social enforcement effect need to address discriminatory power and coverage area issues in seed selection by the Multi Neighbour seed selection method.

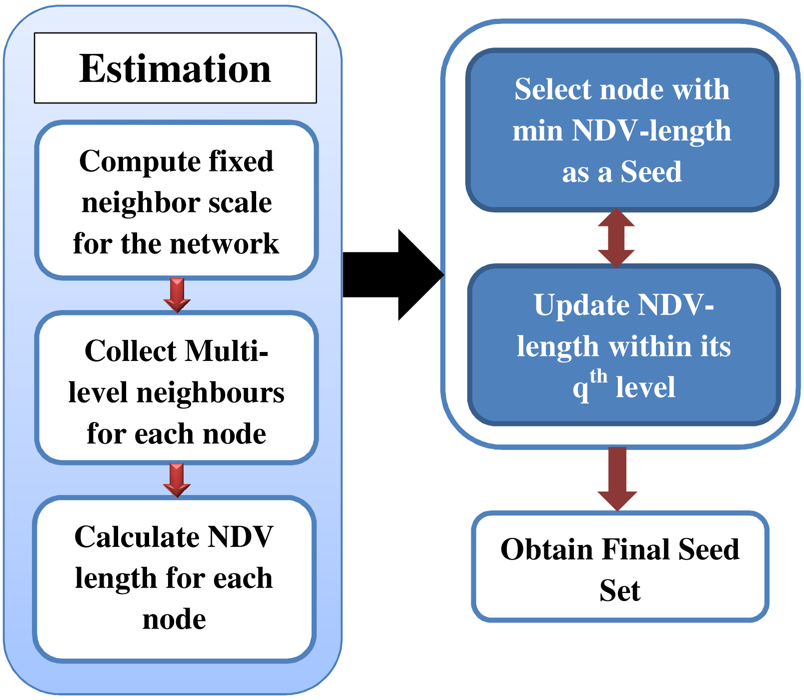


Figure 5.4 Overview of Multi Neighbour Seed Selection

The proposed algorithm finds the seed nodes that have very effective way of dealing with the rich-club phenomenon.

The neural network process the algorithm which is used to find out the NDV - length of each node in a particular node set from the output of the seed selection method and select the node which has the min NDV length as a seed in the output of the neural network in machine learning approach. Figure 5.4 depicts a high-level overview of our proposed method. The framework is divided into two sections: estimation of influence strength and selection of seed nodes. The proposed method begins by estimating each node's influence strength using the proposed metric NDV-length. Then it selects seed nodes with effective strategies for dealing with the rich-club phenomenon.

A fixed number of neighbors near the goal node are considered in the proposed system. This fixed number is called the rank value of the neighbor (also known as NDV), and it is the same for all nodes in the same network. When using NDV, only count the first-level neighbors of the evaluated node seed nodes at the beginning since the top-grade nodes normally differ in rank noticeably. However, because Nodes of greater grades are selected, choosing seed nodes solely established on first-level neighbors is inappropriate because many nodes may share the leftover maximum rank. Then, to distinguish the remaining nodes, the second-level neighbors should be considered.

To make the NDV method more user-friendly, propose a new metric called NDV-length to evaluate the intensity of a node's influence rather than the rank. The number of the weights allocated to its multi-level neighbors based on their distances from this node is revealed as the node's NDV length. This novel metric assesses a node's impact power by considering its multi-level neighbors and additional topological structure information, providing more discrimination than rank. Figure 5.5 shows a model of multi-level network neighbors of a node. The target node is Node1 in the center of the figure, which multi-level neighbors surround. The target node's first-level neighbors are Nodes 2–5, its second-level neighbors are Nodes 6–12, and its third-level neighbors are nodes 13-21.

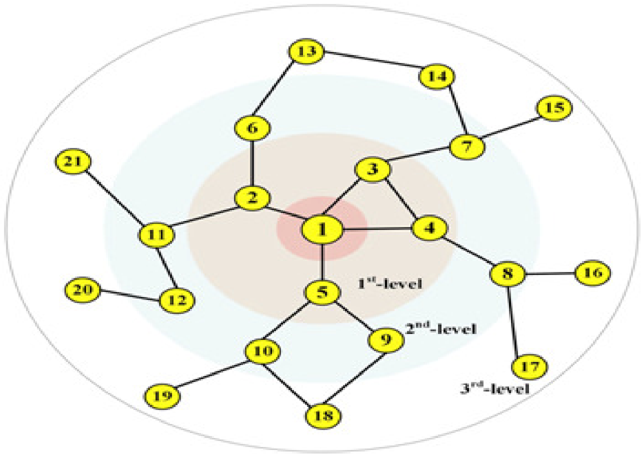


Figure 5.5 Neighbours on Several Levels of a Communication Link (Rui, 2020)

The proposed approach involves assigning distinct weights to each rank while treating all nodes at the same level equally in terms of their influence. For example, where the scale for a fixed neighbor is 15, consideration should be given to 15 nodes nearest to Node 1 such that all first- and second-level nodes, and any four-level third-level nodes, use the NDV longitude of the target node to be measured. The greater the influence intensity of a node, the shorter its NDV-length, which is determined using the following Equations Equation 5.7, Equation 5.8.

$$\boldsymbol{NDV\_length}(\boldsymbol{A}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{level-1}} (\boldsymbol{i} * |\boldsymbol{\Gamma_i}(\boldsymbol{A})|) + \boldsymbol{level} * \left(\boldsymbol{NDV} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{level-1}} |\boldsymbol{\Gamma_i}(\boldsymbol{A})|\right) \tag{5.7}$$

where

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{level-1}} |\boldsymbol{\Gamma_i}(\boldsymbol{A})| < \boldsymbol{NDV} \leq \sum_{\boldsymbol{i=1}}^{\boldsymbol{level-1}} |\boldsymbol{\Gamma_i}(\boldsymbol{A})| \tag{5.8}$$

Since the likelihood of the target node activating an ith-level neighbor lowers even as level index rises, neighbors at lower levels provide better gains to the given destination regarding influence intensity than neighbors at higher levels. It should be observed that a shorter NDV length implies a more substantial impact. As a result, if the level is lower, a lower weight will be assigned. Equation **5.7** uses a weight that is directly proportional to the neighbor's distance from a particular node, and all nodes at the same distance are given the same weight. The term "NDV length(A)" in the equation denotes the count of all the shortest paths between node A and its neighboring nodes that fall within a specified range of distances. These paths are considered across multiple levels in the graph.

According to Equation 5.8, the neighbor rank value determines the number of acceptable levels with each link; in other words, the node in the outer surface inside the neighbor rank value represents the total number of levels.

In general, the neighbor degree value (NDV) equals the network's second-largest rank plus one. If NDV is too high, each node's neighbors are calculated at too many levels, making the proposed scheme ineffective at runtime. Since NDV is restricted to first-level neighbors, if NDV is excessively lower, those high-rank nodes would have the same NDV length. NDV has been placed second to +1 – the lowest value to balance knowledge and performance. This means that the NDV length between the node with the highest rank and the node with the second highest rank can be separated. As a consequence, for all nodes except the one with the highest rank, in addition to the first-level neighbors, the second-level neighbors are also considered when calculating NDV length using algorithm 5.2.

Algorithm 5.2: Computing the neighbor rank value of a Network (Rui, 2020):

Input: Network S

Output: The neighbor rank value NDV

1: NDV= 0;

2: $maximum\_A$= argmax$A$ $(l\ (A))$;

3: $maximum\_l$= $l\ (temp\_A)$;

4: $B\leftarrow B\backslash \{maximum\_A\}$;

5: $temp\_A$= $maximum\_A$;

6: while $l\ (temp\_A) == maximum\_l$ do

7: $temp\_A$= argmax$A\{l(A)|A\in B\}$

8: $B\leftarrow B\backslash \{temp\_A\}$;

9: end while

10: $NDV = l(temp\_A) + 1$;

Once the neighbor rank value for a specified network is obtained, the NDV length of all nodes can be computed. First, a node's first-level neighbors are verified and compared with NDV. If $|\Gamma_i(A)|$ is less than NDV,the neighbors of the succeeding floor will be scanned iteratively. When $\sum_{i=1}^{level-1}|\Gamma_i(A)|$Equals or exceeds NDV, indicating that there are enough neighbors closest to the target node, the NDV-length(A) is calculated using Equation **5.7**.

5.3.1 Seed Selection

NDV length can be considered better accuracy and discriminative centrality than the rank. Picking seed nodes solely on NDV length would result in the rich-club phenomenon. Rise a node's NDV length until its qth-level neighbor is chosen as a seed node to solve this problem as in algorithm 5.3.

Algorithm 5.3: Computing the NDV-length of Nodes (Rui, 2020):

Input: Network S, fixed neighbor scale NDV

Output: The NDV-length

1. Set mark[i] to 0 for all nodes i in the network

2. Set NDV-length[i] to 0 for all nodes i in the network

3. For each node i in B:

a. Set mark[i] to 1

b. Create a queue Q containing i and its neighboring nodes

c. Set level to 1

d. Set sentinel to the end of the queue

e. Set nbs to the set of neighbors of i at level 1

f. While length(Q) < NDV + 1:

i. If length(nbs) < NDV + 1 - length(Q)

1. Add all unmarked neighbors of nodes in nbs to the queue

*2. Update NDV-length[i] by adding level * length(nbs)*

3. Increment level

4. Set mark[nbs] to 1

5. Set delta_sentinel to length(nbs)

ii. Else

1. Break out of the loop

g. For each node j in Q from sentinel to the end:

1. Set nbstemp to the set of unmarked neighbors of j

2. Add nbstemp to nbs

h. Set NDV-length[i] to the current NDV-length value

4. Return the NDV length for each node i in B and its neighboring nodes within the predefined range

---

The increase is determined by the length of the seed node and the likelihood of spreading between the seed node and the neighbor in the q^{th} step. In the case of q = 1 and q = 2, the likelihood of expanded distribution is reduced exponentially as the shorter path between two nodes increases using algorithm 5.4 and algorithm 5.5.

Algorithm 5.4: NDV1 (Rui, 2020)

Input: network *S*, NDV-length, seed number *x*

1: $SNDV1 \leftarrow \emptyset$;

2: for *i*= 1 to *x* do

3: *seed_A*= argmin*A*

{*NDV-length(A)*|*A*∈*B*};

4: $SNDV1 \leftarrow SNDV1 \cup \{seed\_A\}$;

5: for all *u* ∈*Γ (seed_A)* do

6: *NDV-length(u) = NDV-length(u) + NDV-length(seed_A) ∗p(seed_A, u)*;

7: end for

8: $B \leftarrow B \setminus \{seed\_A\}$;

9: end for

Output: The seed set *SNDV*1

Algorithm 5.5: NDV2 (Rui, 2020)

Input: network S, NDV-length, seed number x

1: $SNDV2 \leftarrow \emptyset$;

2: for i= 1 to x do

3: $seed\_A$= argmin$A$

$\{NDV\text{-}length(A)|A \in B\}$;

4: $SNDV2 \leftarrow SNDV2 \cup \{seed\_A\}$;

5: for all $u \in \Gamma\,(seed\_A)$ do

6: $NDV\text{-}length(u) = NDV\text{-}length(u) + NDV\text{-}length(seed\_A) * p(seed\_A, u)$;

7: for all $w \in \Gamma\,(u)$ do

8: $NDV\text{-}length(w) = NDV\text{-}length(w) + NDV\text{-}length(seed\_A) * p(u,w) *$

$p(seed\_A, u)$;

9: end for

10: end for

Output: The seed set *SNDV*2

The spreading probability increases over a wide area, resulting in a high discriminatory power from the above-proposed method. The interaction of users was overlooked in selecting Judgment leaders can be tackled by the Weighting grade of the judgment mediator's pick and will be well explained in the next section.

5.4 Weighting Grade of the Judgement Mediator,S Pick

The min-NDV length of the Seed from the seed set has been selected from the above method and has been given to the input of the neural network to pick the Judgment Mediator. Judgment leaders have a significant impression on the circulation of ideas, which affects the entire active progression of ideas. Judgment leaders are "campaigners" with extraordinary impact and appeal in the system's social network who frequently deliver facts or judgments to some people and apply influence over them. Judgment leaders can inspire the creation of collection judgment and important decisions and the development of available community Judgment circulation. Traditionally, some individuals are randomly assigned as Judgment leaders in the simulation system. As we all remember, judgment spreads in the public system. The features of social connection topology also affect circulation, and the main node is the Judgment leader with the most influences in Public systems; thus, it is challenging to precisely reproduce the purpose realism in the public system using casually establishing Judgment leader nodes. This work makes an effort to use social network research better to categorize the crowd's judgment leadter. Since a popular media forum is a series of social actors and their interactions, each actor's relationship is emphasized among others; this is the case. It provides the tool to depart from the social system and determine the function and properties of the social interactions depending on the dynamic features and the network environment. The three indices of the central category, centralness, and centrality of the social network are commonly used as keynote indicators in social networking research. The total number of direct links between nodes is used to calculate degree centrality. The formula is as follows in Equation 5.9:

$$\boldsymbol{C_x} = \frac{\boldsymbol{n_c}}{(\boldsymbol{n_m} - \mathbf{1})} \tag{5.9}$$

The variable *"nc"* represents the count of nodes that are linked to a particular mediator *"x",* while the variable "nm" indicates the total count of Mediators present in the network.

The capacity of a Mediator to be uncontrolled by other Mediators is defined by closeness centrality, which can be measured using the following equation.

$$\boldsymbol{C_x} = \frac{\boldsymbol{n_m} - \mathbf{1}}{\sum_{y \neq x} \boldsymbol{Avgd_{xy}}} \tag{5.10}$$

Where $Avgd_{xy}$ is the average distance from node x to y.

The Equation 5.11 provided can be utilized to evaluate the betweenness centrality of a node, which is a measure used to determine the significance of a particular node's role as a "mediator" in a given network.

$$C_x = \sum_{x \neq y \neq k} \frac{g_{yk}(x)}{g_{yk}} \qquad 5.11$$

The variable "gyk(t)" represents the count of shortest paths that connect nodes "y" and "k" in a network and specifically pass through the intermediary node "x". On the other hand, the variable "*gyk*" signifies the total number of shortest paths that connect nodes "*y*" and "*k*" in the same network.

The existing HK model, on the other hand, focused on the occurrence of uniformity and variation in the steady state of judgment circulation according to clear guidelines, as well as taking the inspiration of single communication on judgments into account. However, it quite reflects and discriminates if judgment mediators or usual personalities influence nodes. Generally, the critical node with the highest centrality in Public systems is the judgment mediator, which is assured to influence public judgment absorption, incorporation, and split-up. Even though the role of the judgment mediator is important in the evolution of Judgment *s*, they chose a judgment mediator node randomly from a social networking site. However, judgment mediators should be the recognized nodes in the actual situation, which necessitates accurate identification. Otherwise, ensuring the truth of judgment growth is difficult. Outdated judgment dynamic models assumed that the entire structure's perspectives were neither diminished nor improved, which is insufficient to describe Judgment regular setbacks. This is because while the revolutionary HK and its extensive models recognized the effect on decision transmission of individual participation, participants failed to consider the probability that their own views could spontaneously change. Personalities will broaden their understanding and develop their views using surfing sites on the internet and collecting information, automatically influencing their thinking of a particular event and their judgment right to decide. This is an observational hypothesis in which a participant's decision evolves over time despite the absence of any significant relationship between neighboring nodes, but little is understood about the parameter best fits this pattern of judgment evolution over time. The preferred approach uses the SNA centrality index to classify judgment mediators. It applies a revised system for natural judgment reversal based on the HK restricted based rank model to allow for natural reversal features induced by participant experience in the network context. It's used to emphasize the natural reversal of judgment and its internal mechanism in the evolution of public judgment. Individuals all upgrade their individual judgments in line with the critical graded rules, without acknowledging other considerations in judgment systems, such as the role of judgment mediators, according to the model mentioned above, and the mediators in this model all have the same conduct.

(Zhao Y. K., 2018) (Watts, 2007)The model being developed involves nm Mediators in a structure labeled as A, where A is defined as 1, 2, ..., n. At any given time t, the decision of a Mediator in A is denoted by a value xi(t) which falls in the range of 0 to 1. A vector x(t) represents the decisions made by all Mediators in A at time t. This vector x(t) comprises of individual decision values x1(t), x2(t), ..., xn(t) for each of the Mediators. Initially, at t = 0, each Mediator has a starting decision denoted as x(0). The HK model suggests that:

$$\boldsymbol{I(i, x(t)) = \{1 \leq j \leq n_m \mid |x_i(t) - x_j(t)| < \varepsilon_i\}} \qquad 5.12$$

Where in Equation 5.12 |x I (t)-x j (t)| is the difference between mediator i's judgment and that of its neighbor j, I(i,x(t)) in Equation 5.12 is a list of all individuals beyond Mediator i's significant grade at time t, I is the Mediator i's Judgment grade value, and |x I (t)-x j (t)| is a set of all individuals within Mediator. If the difference is less than the threshold value, the Mediator can choose to interact with its neighbors. All Mediators inside the graded boundary are assigned the same weight, while Mediators outside the boundary are not. This makes the issue easy to understand.

$$\boldsymbol{x_i(t+1) = |I(i, x(t)|^{-1} \sum_{j \epsilon I(i,x(t))} x_j(t), \qquad t \epsilon T} \tag{5.13}$$

The HK model examines the progress of ideas in discrete time intervals, in Equation 5.13 T = 0, 1, 2,... The complex system's final state changes as I adjust. By connecting Judgment Mediators to the HK model, this model is proposed to investigate the complex process of natural judgments reversal. Since a Judgment leader's social network status may affect ordinary groups, and a Judgment leader's view is seldom impartial, a Judgment leader's impact on Judgment diffusion processes varies from those of regular society. Judgment representatives may be categorized using the SNA centrality index, with the remainder being classified as ordinary Mediators.

Each Mediator interacts with other similar Mediators over time, and its interaction rules are based on the HK model. It has an impact if the neighbor Mediator is a Judgment leader or a normal Mediator. As a result, the following diffusion dynamics models are created, considering the heterogeneity between Judgment leaders and ordinary Mediators (Mirtabatabaei, 2012).

The relationship matrix *R* is an *n x n* matrix that represents the connections or relationships between individuals in a social network. *Rij* demonstrates if I have an individual connection with *j* for any individual I and j in A. When *Rij = 0*, individual *I* is *j*; not related to individual when *Rij = 1*, they are. When a network has m decision representatives, *m B = 1, 2,..., m.* As a result, there are n m ordinary men, with *C = m+ 1,m+ 2,..., n.* Without a doubt, A = BC. Below are the conditions for updating one's judgment. If Mediators is a normal person, IC. According to practice, a mediator is influenced by its neighbors, who may be either Judgment agents or ordinary people. To completely account for multiple results, the linear weighted rules of the joint variables for these two groups are used to model standard array Judgment dynamic progression:

1. At time t, compute the sets of common people and Judgment leaders that fall under Mediator i's Significant threshold:

$$\boldsymbol{OL}(\boldsymbol{i}, \boldsymbol{x}(\boldsymbol{t})) = \{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{m} | |\boldsymbol{x}_{\boldsymbol{i}}(\boldsymbol{t}) - \boldsymbol{x}_{\boldsymbol{j}}(\boldsymbol{t})| < \boldsymbol{\eta}_{\boldsymbol{i}}, \boldsymbol{j} \in \boldsymbol{B}\}, \boldsymbol{OP}(\boldsymbol{i}, \boldsymbol{x}(\boldsymbol{t})) = \{\boldsymbol{m} + \mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{n} | |\boldsymbol{x}_{\boldsymbol{i}}(\boldsymbol{t}) - \boldsymbol{x}_{\boldsymbol{j}}(\boldsymbol{t})| < \boldsymbol{\varepsilon}_{\boldsymbol{i}}, \boldsymbol{j} \epsilon \boldsymbol{C}\} \quad 5.14$$

Where η_i denotes Mediator i's significant level with judges, and *I* denotes Mediator i's significant level with regular citizens. The sets of Judgment leaders and common citizens under Mediator i's relevant threshold at time t are *OL(I x (t))* and *OP(I x (t)),* respectively as in Equation 5.14.

2. The median i's judgment social at time *t + 1* shall be influenced by the mediator j, which is in the corresponded meaningful thresholds, labeled bij(t) and cij(t), respectively, in relation to the node relationship in a network in determining the specific mediator j linked to a mediator I in the sets B and C.

$$\boldsymbol{b}_{\boldsymbol{ij}}(\boldsymbol{t}) = \begin{Bmatrix} \mathbf{1}, \boldsymbol{if}\ \boldsymbol{R}_{\boldsymbol{ij}} = \mathbf{1}, \boldsymbol{j}\epsilon\boldsymbol{OL}(\boldsymbol{i}, \boldsymbol{x}(\boldsymbol{t})), \\ \mathbf{0}, \boldsymbol{otherwise} \end{Bmatrix} \quad 5.15$$

$$c_{ij(t)=}\begin{Bmatrix} 1, if\ R_{ij}=1, j\epsilon OP(i,x(t))), \\ 0, otherwise \end{Bmatrix} \quad 5.16$$

Bij(t) as in Equation 5.16 shows whether Judgment *j'*s leader can affect the Mediator's viewpoint above the important threshold i. In OL I x(t) and Rij = 1 and bij(t) = 1, the judgement chief j would impact mediator s views. There will be no effect if this is not achieved. Similarly, the definition of cij(t) shows when an average individual j can affect the viewpoint of Mediator I during the trust interval i. If OP I x (t) and Rij = 1 and cij (t) = 1 the median j would affect i. If cij(t) = 0, Mediator j has no Mediator I influence.

The following equation Equation 5.17 will be used to update Mediator i's Judgment based on grade set and matrix R:

$$x(t+1) = \theta \frac{1}{\sum_{j\epsilon L(i,x(t))} b_{ij}(t)} \sum_{j\epsilon OL(i,x(t))} b_{ij}(t)\, x_j(t) + (1 - \theta)\, \frac{1}{\sum_{j\epsilon I(i,x(t))} c_{ij}(t)} \sum_{j\epsilon OP(i,x(t))} c_{ij}(t)\, x_j(t) \quad 5.17$$

Where 1 is the force of authority of the general population and is used to calculate the power of the members of the judgment. If there is no master of judgment, so the pattern is the HK model. All mediators belong to the usual community, and the HK paradigm is the basis for their decisions. The entire scheme includes all ordinary and Judgement Leader groups with greater value than zero.

If there are no changes in the judgments made using Equation 5.18, or if a specific condition is satisfied, then the process of upgrading will be halted:

$$\sum_{i=1}^{n} [x_i(t+1) - x_i(t)]^2 \leq \delta, \quad 5.18$$

If Mediator I is a judge, then OP B applies. Since the leader of the Judgment Mediators are more likely than ordinary Mediators to insist on their points of view, and they can be affected in the short term by other Judgment representatives with a certain degree of authority. As a result, the initial attitudes of Judgment leaders can affect the diffusion dynamics process. The initial attitudes of judgment leaders can be divided into two categories: positive and negative. Consequently, Equation 5.19 describes the evolution of positive-attitude Judgment leaders' viewpoints, while Equation 5.20 describes the evolution of negative attitude Judgment leaders' viewpoints.:

$$x_i(t+1) = \theta |OL(i, x(t))|^{-1} \sum_{j\epsilon OL(i,x(t))} x_j(t) + (1-\theta) \times 1 \quad 5.19$$

$$x_i(t+1) = \theta |OL(i, x(t)|^{-1} \sum_{j\epsilon OL(i,x(t))} x_j\,(t) + (1-\theta) \times (-1) \quad 5.20$$

Where x_j (t) denotes a Judgment leader's point of view.

To illustrate how individual judgment can obviously change over time even if individual neighbours have no collaborative control, a parameter must be included to modify the HK model, which varies with the time of evolution. Individual challenges and actions Psychological cognition capacity improve with increased awareness and interactions, according to trait theories. This means that the natural reversal parameter is a monotonically increasing element rather than a fixed constant. The parameter value needs to be small because individuals have a restricted understanding during the initial phases of evolution. However, as awareness and interactions expand, and an individual's judgment option cannot be consistent, the value of this parameter rises. The Sigmoid function is like this parameter because it has the following characteristics as in Equation 5.21:

$$\boldsymbol{s(x) = \frac{1}{1 + e^{-x}}} \qquad 5.21$$

Where x is an actual number and s(x)(0,1) is an interval-mapping function (0, 1). When a natural reversal parameter is built with Sigmoid function properties, a microscopic instantaneous transition cannot be articulated. The derivative is then measured as follows in Equation 5.22:

$$\boldsymbol{s'(t) = \frac{e^{-t}}{(1 + e^{-t})^2}} \qquad 5.22$$

The natural reversal parameter *(t)* is specially constructed based on equation Equation 5.21, taking into account the basic structure and essential controlling factors of natural reversal of Judgment, as seen in the following equation Equation 5.22:

$$\propto \boldsymbol{(t) = \frac{\mu e^{1/\beta(t-\gamma)}}{(1 + e^{1/\beta(t-\gamma)})^2}} \qquad 5.23$$

where (t) grows with time t and 0 1 is the effective index of (t). [1, 4] denotes the number of generations and the rate of evolution. *c* represents the time it takes for judgment to decay, which ranges from 10 to 30 seconds.

As a consequence of the inclusion of parameter, the decision modification modeling for the ordinary Mediator I is as follows in Equation 5.24:

$$\boldsymbol{x_i(t+1) = \propto (t)(\theta \frac{1}{\sum_{j\epsilon OL(i,x(t))} b_{ij}(t)} \sum_{j\epsilon OL(i,x(t))} b_{ij}(t) x_j(t) + (1-\theta) \frac{1}{\sum_{j\epsilon OP(i,x(t))} c_{ij}(t)} \sum_{j\epsilon OP(i,x(t))} c_{ij}(t) x_j(t)} \quad 5.24$$

The modification of the Judgment Equations for Judgment Representatives in positive and negative attitudes explains Equation 5.25 and Equation 5.26:

$$\boldsymbol{x_i(t+1) = \propto (t)(\theta |OL(i,x(t))|^{-1} \sum_{j\epsilon OL(i,x(t)} x_j(t) + (1-\theta) \times 1)} \quad 5.25$$

$$\boldsymbol{x_i(t+1) = \propto (t)(\theta |OL(i,x(t))|^{-1} \sum_{j\epsilon OL(i,x(t)} x_j(t) + (1-\theta) \times (-1))} \quad 5.26$$

where various parameter values represent a difference in the impacts on supporting or opposing groups, be it usual or Judgment leaders, and where distinct values in this parameter reflect an influence on supportive or opposing groups, be they usual or Judgment leaders.

The following is the method framework:

Step 1(initialization): Select the evolutionary time and parameters, including numbers of nodes, an interval of confidence, convergence and simulation measures.

Step 2 (Social Network Setup): Create the network needed.

Step 3 (assignment of a preliminary Judgment value): The preliminary Judgment xi(0 shall be delegated to each mediator and the preliminary Judgements of the system shall also or usually be assigned.

Step 4 (Judgment leader identification): Use the social network analysis tool to classify Judgment leaders in the created network. If Period = 1, a constructive Mediator can be considered a Judgment master. This Judgment leader can be appointed as a negative Mediator if Period = 2.
Step 5 (natural reversal parameter computation): Calculate each node's normal reversal parameter.
Step 6 (update of viewpoint): update the verdict value based on the model complexities of different network node forms (Judgment leaders and ordinary individuals).

As a result, the ultimate system solves viral marketing challenges, and the proposed architecture maximizes effects. The output of the neural network which gives the output as a number of seeds that rectify the issues in viral marketing and maximize the influences.

5.5 Result and Discussion

This section includes a comprehensive summary of the implementation results, our proposed system's success, and a comparison section to ensure that our proposed approach outperforms the competition.Table 5.1 describes the dataset used for the implementation of the proposed framework. We have used two datasets, Ca-HepTH collaboration network data and Facebook data.

5.5.1 Evaluation Metrics and Simulation Output:

This section presents the network simulation and the resulting outputs of implementations . Figure 5.6 illustrates the network simulation of a proposed method.

From Figure 5.6, Visualization of a Link provides an effective and cost-effective strategy for deciding how the network will behave under different operating conditions. Before deployment, the results of simulations can be used to evaluate network efficiency, identify possible problems, determine the root cause, and fix issues. Link visualization can also help link machinists determine network strength and resilience to cyber threats.

Table 5.1 Dataset Description

Dataset	Information
Ca-HepTH(Miscellaneous Networks)	The Arxiv HEP-TH (High Energy Physics - Theory) collaboration network is based on the e-print arXiv and covers research collaborations between authors of papers submitted to the High Energy Physics - Theory community. The graph has an undirected edge from i to j if author I and author j collaborated on a paper. If the paper has k co-authors, it produces a completely connected (sub)graph with k nodes. There are 9877 nodes and 51971 edges in this graph.
Facebook	This exploratory research provides insights from a Facebook dataset by identifying users who can be targeted more to increase business. These valuable observations can assist Facebook in making informed decisions about identifying useful users and making appropriate suggestions to them. This dataset has 99903 rows and 15 columns.

Simulating a network makes it possible to assess its susceptibility to cyber-attacks and the effectiveness of defensive measures in preventing or mitigating such threats. This evaluation can be done without actually exposing the physical network to real attacks, thus minimizing the risk of vulnerabilities being exploited in a live environment. Network simulation offers a cost-effective, low-risk, and straightforward method for predicting network performance and identifying potential problems before deployment.

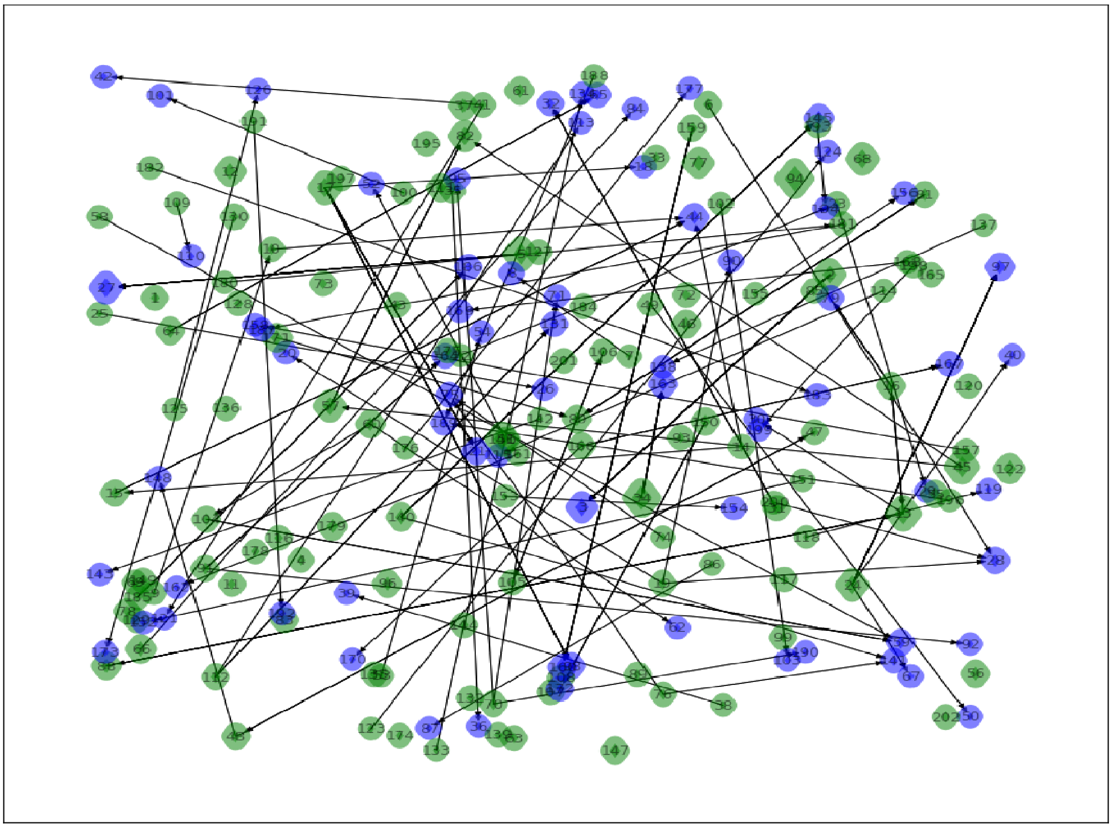


Figure 5.6 Visualization of a Link

Figure 5.7 describes the most important properties that determine a proposed method's network's structure (or topology). The degree distribution (or neighbour distribution) is as follows: i's the chance that a node picked at random has links (or neighbours). The number of connections a node has with other network nodes determines its rank. From the above graph, the proposed method will have a better connection with neighbours. Different centrality measures are used to identify the trusted opinion leader like In-Degree, Out-Degree, Degree, Betweenness Centrality, Closeness Centrality.

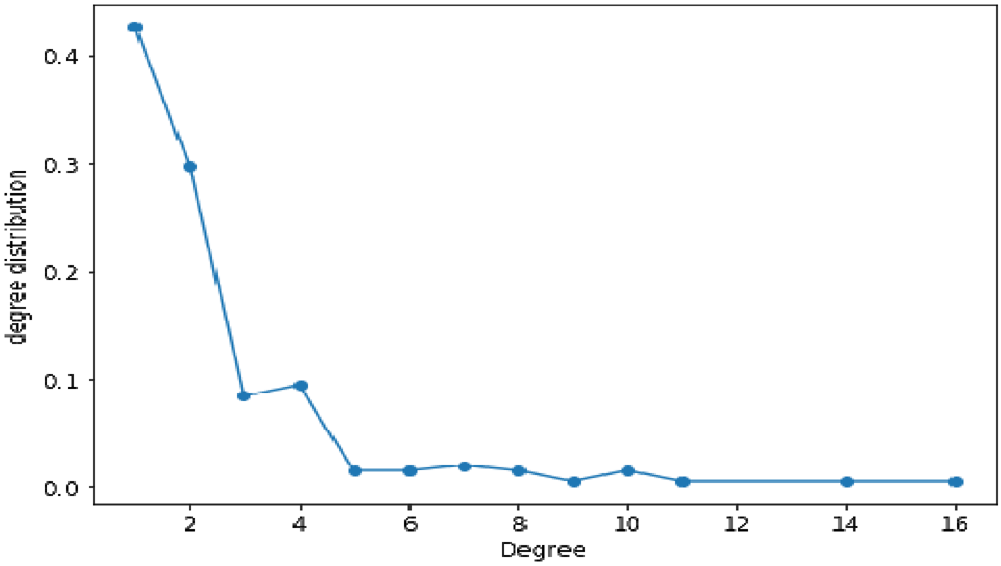


Figure 5.7 Degree Distribution of a Network

5.5.2 In-Degree Distribution:

The number of edges in the directed graph G that bind to user u is called u's In-degree, abbreviated as Idg (u) as in Equation **5.27**.

$$\boldsymbol{Idg(u)} = \sum_{\boldsymbol{A \in B \neq u}} \boldsymbol{\delta_{Au}} \qquad 5.27$$

When A is a follower of u, $\delta_{Au} = 1$; otherwise, $\delta_{Au} = 0$.

The Figure 5.8 describes the distribution of In-degree in the proposed method. It clearly explains that the proposed method has a better In-degree ie, the incoming nodes coming is more than existing methods.

5.5.3 Out-Degree Distribution:

In a directed graph, Figure 5.9 depicts the number of edges that emerge from a vertex.

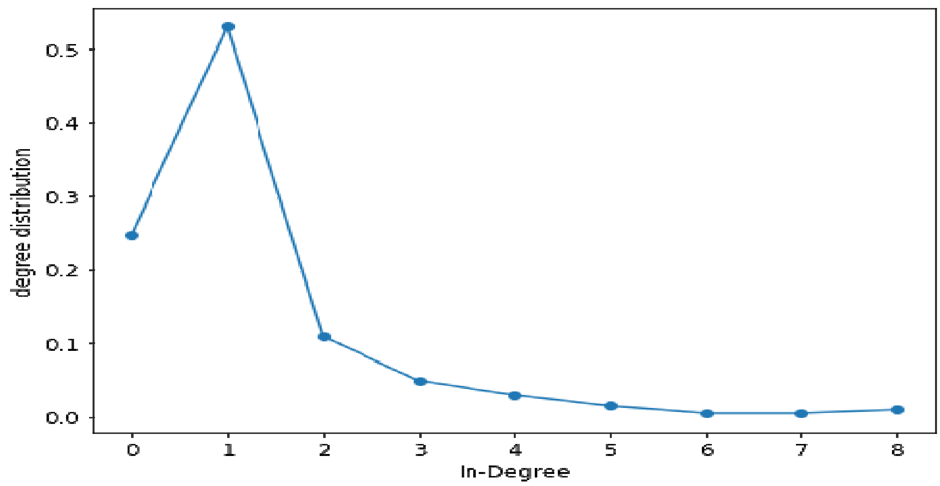


Figure 5.8 In-Degree Distribution

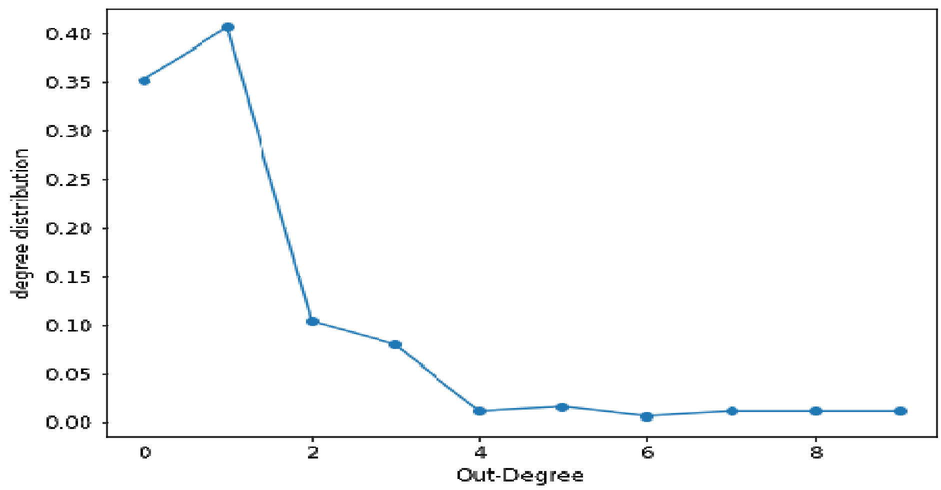


Figure 5.9 Out-Degree Distribution

5.5.4 Degree Centrality:

According to degree centrality, the value of each node is solely determined by the number of connections it contains. It means how many direct links each node has to the rest of the network in a single hop. For example, in a real social network, as more people join a person, the person communicates more intelligence, making them more likely to be a Judgment leader.

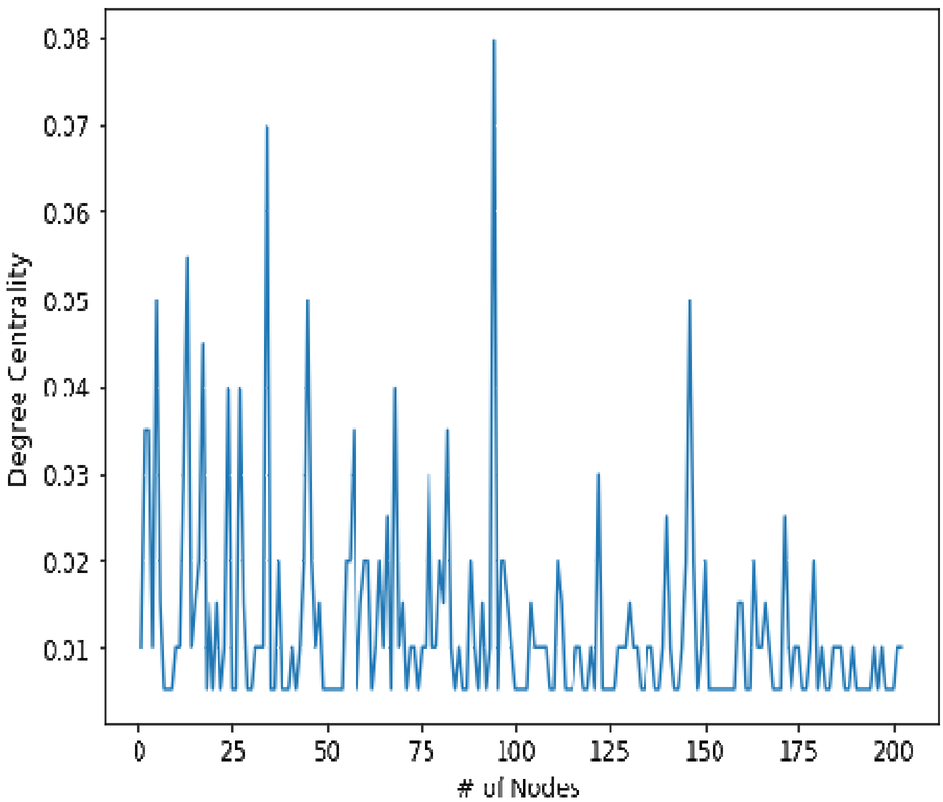


Figure 5.10 Degree Centrality

The relationship between the number of nodes and the degree of centrality is depicted in Figure 5.10. As the number of nodes in the graph grows, the proposed approach has a higher centrality value.

5.5.5 Between Centrality :

The number of times a node is on the shortest path between other nodes is measured by betweenness centrality.The graph in Figure 5.11 depicts which nodes in a network act as "bridges" to other nodes. It accomplishes this by identifying all of the shortest paths and calculating how many times each node intersects with one of them; it is also used to categorize entities that impact the flow of knowledge through a system.

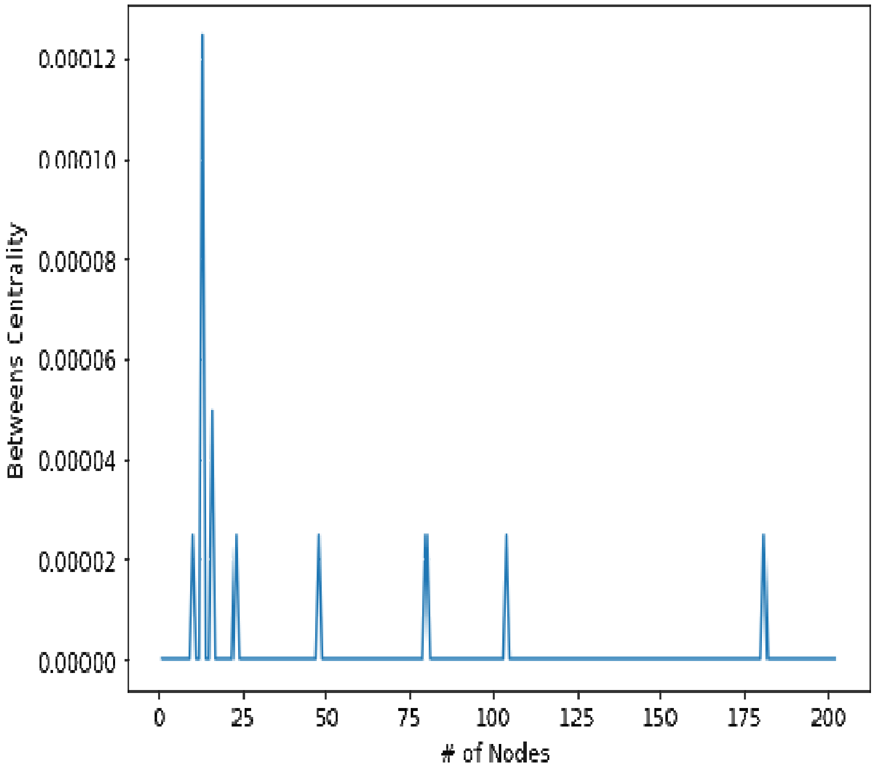


Figure 5.11 Betweenness Centrality

5.5.6 Closeness Centrality:

Identifying successful 'broadcasters' would be aided by closeness centrality, but all nodes will have identical ratings in a heavily connected network. Closeness can be more useful when looking for influencers in a single cluster.The graph in Figure 5.12 measures the shortest paths between all nodes before assigning a score to each node depending on its number of shortest paths. It's used to identify individuals who will most effect the whole network in the shortest time.

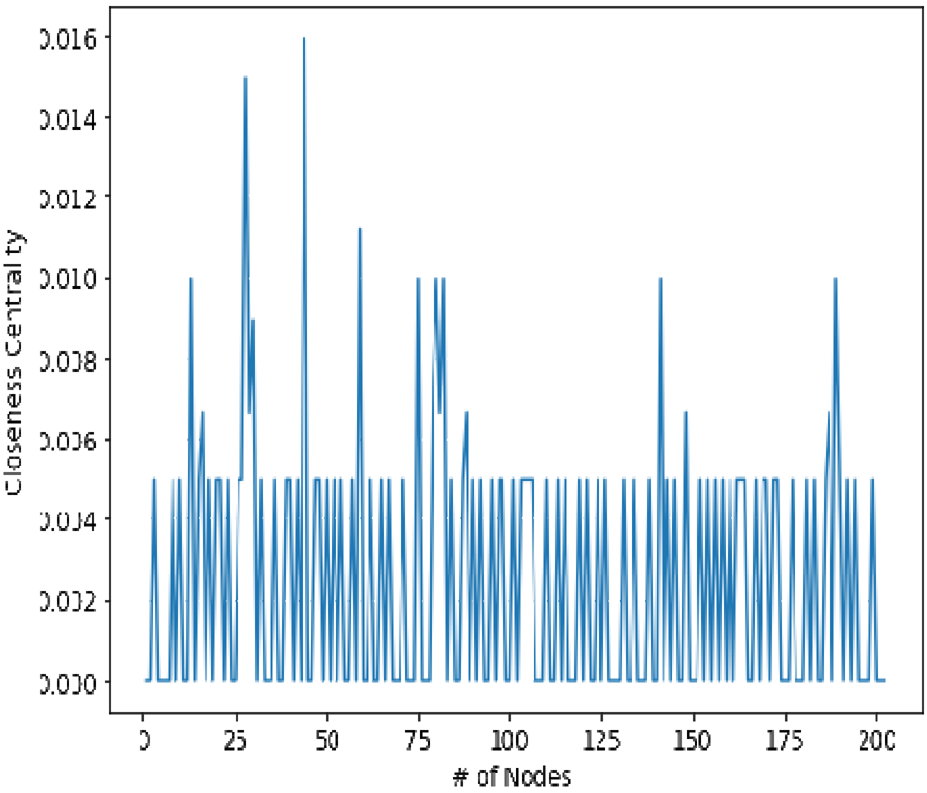


Figure 5.12 Closeness Centrality

As seen in the empirical result in Figure 5.13 , the spread against the seed set grows in proportion to the size of the seed set. It illustrates that as the seed set rises in size, the distribution of the proposed method increases, resulting in more viral marketing influencers.

The graph in Figure 5.14 shows the computation time for the proposed method. The computation time increases while increasing the size of the seed set.

5.5.7 Comparison Strategies:

The proposed method's results have been compared to existing approaches presented in this section.

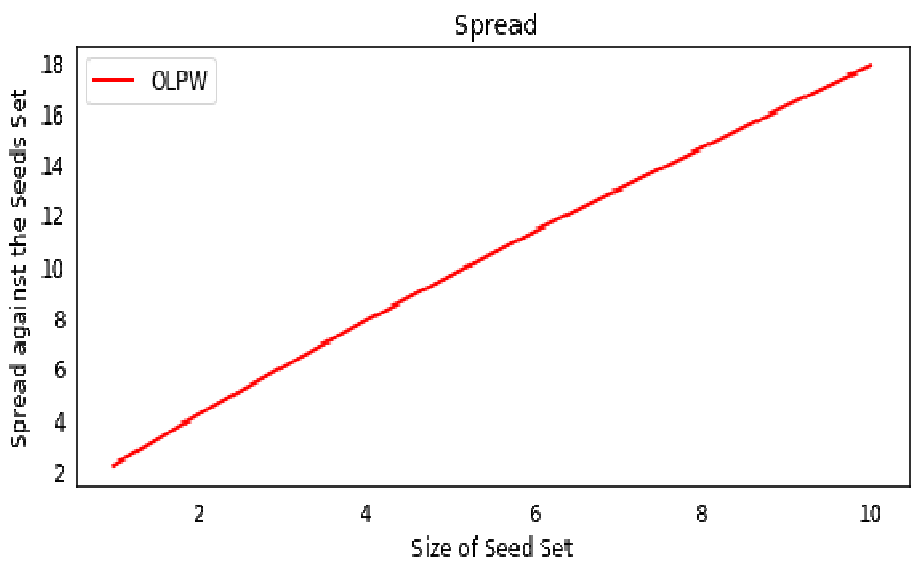


Figure 5.13 Expected Spread of Seed Set

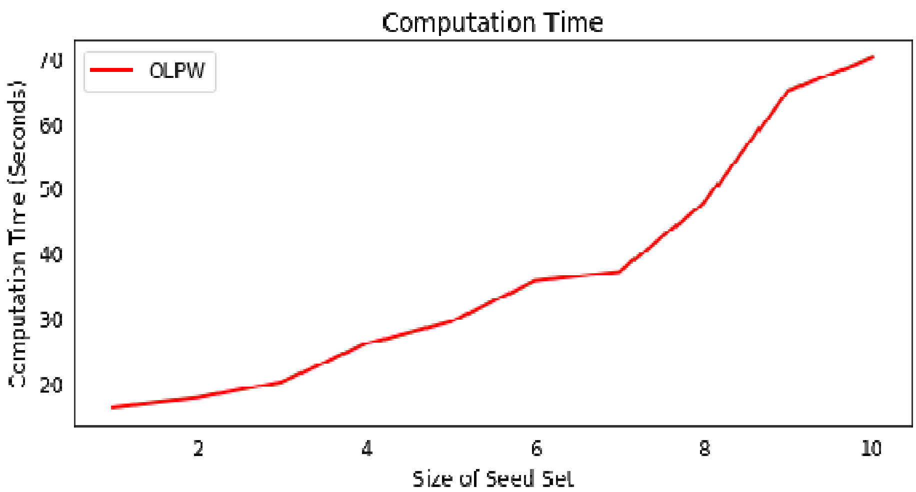


Figure 5.14 Computation Time

The spread of influence in the proposed method has been compared with existing methods in the above graph in Figure 5.15. From the graph, the greedy method has an influence spread of nearly 100 seeds while increasing the seed size to 28, the CI method has an influence spread of nearly 60 in the seed size of 28, TI-SC method has an influence spread of 150 in the seed size of 28. But in the case of the proposed method, the influence has been spread above 150 in the seed size of 28. It shows that the proposed method has a better influence spread while increasing the seed size compared to other existing methods.

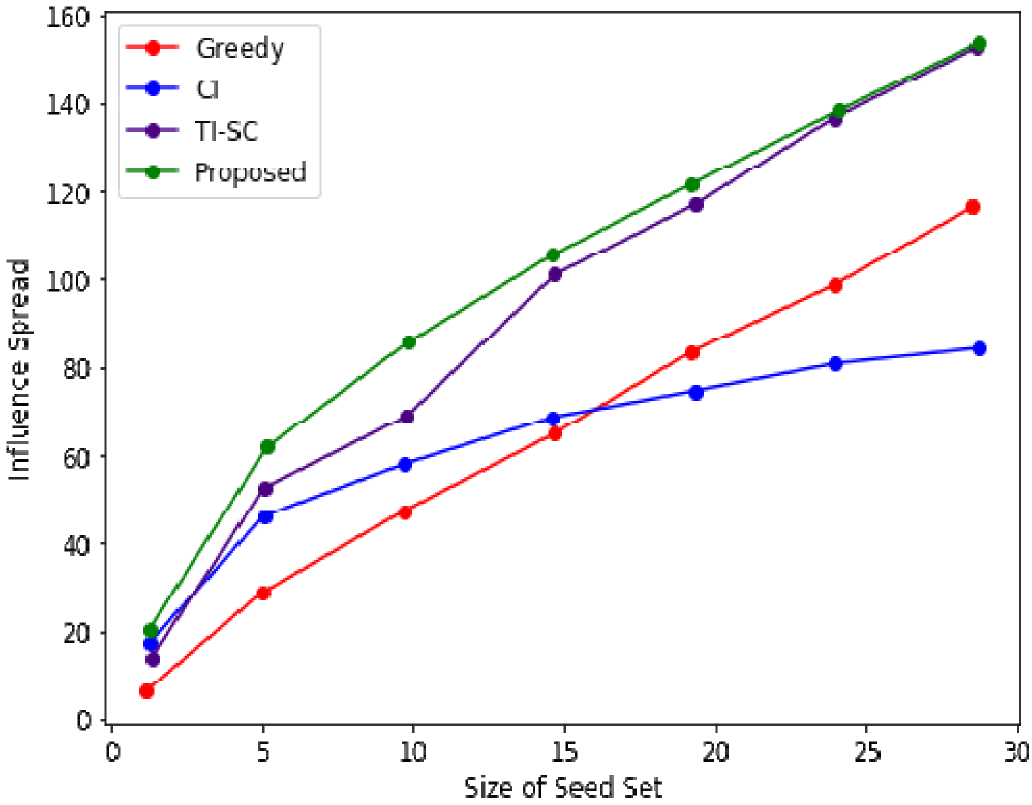


Figure 5.15 Comparison of Influence Spread

The computation time of various methods has been compared with existing methods in Figure 5.16. While comparing the existing methods like, greedy has a computation time of above 2000 s in a seed size of 10, the CELF method has a computation time of above 450s in a seed size of 10, the DSFLA method has a computation time of 750s in the seed size of 10, but the proposed method has a minimal computation time in the seed size of 10. The less computation time reduces the delay of influence spread and solves the memory and social enforcement issues.

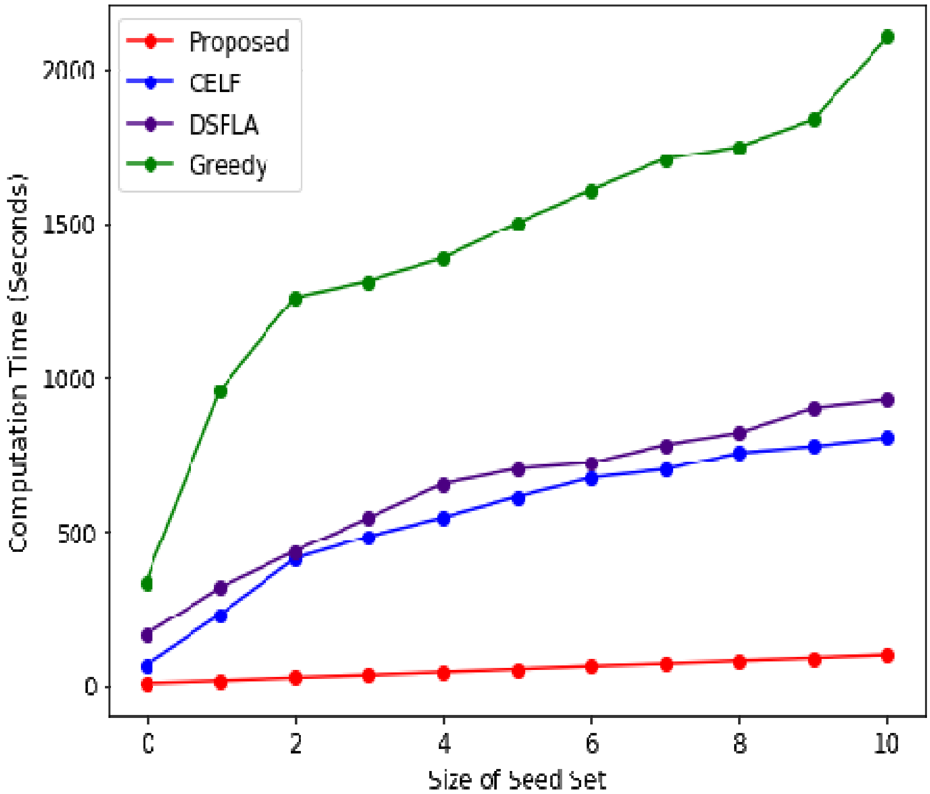


Figure 5.16 Comparison of Computation Time

5.6 Conclusion and Future Scope

As a result, the seed-dependent influence maximization algorithm, which addresses issues in the network's memory effect or social reinforcement effect, is used to obtain the maximum acquiring probability as the number of neighbors of the seed increases at a given time. Furthermore, the Multi neighbor seed selection approach is used to overcome problems during seed selection due to a lack of discriminatory power, in which the neighbor degree value (NDV) is calculated to estimate the influence strength of the seeds. Furthermore, the user interaction was ignored in the selection of Judgment leaders, which is addressed by the Judgment leader pick weighting grade. As a result, the proposed framework effectively addresses the issues in viral marketing and maximizes its influence. The proposed framework is useful for trust management using opinion leaders in ubiquitous services. The future scope of the proposed work is that it can be used for outlier detection in ubiquitous services as it is significant to find out the node which is having different behavior as compared to the trusted one. It is useful to identify the trusted node in the social network. For trust management, the trusted nodes in the network need to be identified and nontrusted nodes can also be identified, which are responsible for any promiscuous behavior.

Chapter 6 SEED SELECTION USING EVOLUTIONARY ALGORITHMS FOR INFLUENCE MAXIMIZATION

In today's era internet has evolved as a means for users to connect and share their opinions. So internet users can easily access information, put across their opinions, and connect with others through a wide range of social media and services. A powerful individual who is an expert in a given subject and has a significant number of followers whose opinions influence others on social media is known as an opinion leader (OL). In recent years, nature-inspired optimization algorithms have gained popularity for solving complex optimization problems. One such algorithm is Ant Colony Optimization (ACO), which is based on the foraging behavior of ants. This chapter explores the potential of ACO for influence maximization in social networks. This chapter considers the performance of evolutionary algorithms that maximize influence spread compared to other existing algorithms by identifying seeds in the network. In practice, however, the purpose is to detect leaders on specific issues and in definite situations, necessitating using the most effective, specially adjusted methodologies. The methods offered in this paper reflect these considerations in choosing a suitable method or approach for a given circumstance.

6.1 Introduction

Most research on online public perception (Chen T. S., 2020), comprising online opinion mining and distribution patterns, primarily focuses on the transmission of online communal thought on social networks (SN). Users' standing in public communities is uneven; those in the middle perform a significant and propelling part in the progression of online public opinions, whereas others on the perimeter are simply swayed by additional causes (Ramakrishnan, 2020). Online opinion leaders (HosseiniBamakanaIldarNurgalievabQiangQ, 2019) are frequently the core, and their engagement can quickly propel particular measures to the Online front OL are frequently at the middle, and their engagement can quickly propel specific events to the forefront of community belief. Public OL has the power to both vigorously steer social life and elicit various unpleasant feelings (S, 2021) (S. S. Lokesh Jain, 2020). Consequently, identifying and mining public OL is critical for directing the correct public opinion and monitoring network stability.

The technique of grouping public operators into closely linked and highly significant categories so that each type may be fine isolated from the others is known as community detection. The domains of SN research, spatial information technology, and smart grids all benefit from community detection (Chien, 2018) (B. Cai, 2019). This study employs communal exposure to create a communal opinion loop of individuals in social networks who share similar ideas and perspectives. We utilize the semantic cohesion (Cao, 2022) and structural compactness of the society to additionally boost the implication of OL in the situation rather than merely recognizing opinion communities by leveraging links among nodes (Bu, 2020).

6.2 Problem Definition

These methods have several disadvantages: On the one hand, self-reported data is subjective by definition, and self-claimed relevance is likely a reflection of self-assurance. An objective evaluation of opinion leaders would result in more dependable techniques for identifying leaders. On either hand, sociological studies are restricted by the resources required to conduct them; having scalable algorithms that can properly identify OL enables us to utilize the ever-increasing amounts of social data enabled by the Internet. Consequently, tools based on network science and data mining have been developed to identify OL(HosseiniBamakanaIldarNurgalievabQiangQ, 2019); this work focuses on these strategies.

6.3.1 Theoretical Basis and Concept Definition of OL

6.3.1.1 Network Opinion Leader

Individuals who have the potential to lead and play a vital role on the Digital platform are known as network thought leaders. They are activists who frequently supply information, thoughts, remarks, and influence to individuals in interpersonal interaction channels and mediators in the information chain formed by interpersonal communication

(Meißner, 2014). Compared to other groups, network opinion leaders are more informed, fluently skilled, politically savvy, and engaged (Sugawara, 2014). They build social interaction linkages in various ways, including self-identity, information involvement, and mutuality (Y. Xiong, 2018). OL mining is an essential matter in virtual network analysis that has a lot of implications for controlling public opinion on the internet and disseminating data. To find and study opinion leaders in the system, several scholars have used information sources, k-clique grouping (He., 2013), Page Rank index (Liu N. M., 2014), network analysis (Liu B. J., 2017), and other methods.

Unlike other social networks, members in the knowledge Q&A forum primarily gather material or expertise from satisfied developers through research, inquiry, and conversation. As a result, we focused on interrelatedness and conversation between users when recognizing online OL in the knowledge Q&A forum, such as whether the viewer frequently provides data, opinions, and remarks to other workers in the discussion community and whether the user has a consequence on other users.

6.2.1.1 Knowledge Q&A Community

Although many different types of modern popular social networks exist, knowledge Q&A groups have distinct qualities. These groups are classified into two types: knowledge groups and social groups. Knowledge groups share ideas and insights through queries. In contrast, social groups are mainly formed through interactivity and offline salon events (C. S. Sanghee Oh, 2008), and these types of groups have gained acceptance in recent years. In the United States, Quora and Stack Overflow are samples of knowledge Q&A groups, Knowledge-IN is an example, while in China, Baidu Know, Himalaya FM, and Zhuihu are examples. According to the findings obtained, the community has appreciated and is actively involved in the flow of information in Q&A groups. Secondly, knowledge Q&A groups, unlike other forms of media, foster a user-centered data flow environment (Y. Chen, 2021), in which users gather data or knowledge from content providers primarily through researching, inquiring, and connecting (Gazan, 2010). Users can also learn more about their interests by participating in a back-and-forth Q&A with further viewers or content providers in the forum under the postings they're involved in (M. L. G. Zhenya Tang, 2018). Operators can also express their problems and queries in other groups to get responses.

6.2.1.2 Emotional Cognition Theory (ECT)

It is a component of the emotional theory that is inspired by sociology. According to the hypothesis, once the body's data from outside commences the perception scheme, it'll be systematized and assembled by the perception scheme on the one side, as well as the data will activate the mind's positive or negative emotional feelings, which will then trigger some behavior propensities on either hand. In earlier studies, the extent of knowledge transfer (T. Guan, 2018) and social exchange theory (Jin, 2015), social capital theory (Johnson, 2007), and other theories or approaches were used to analyze operator data exchange in the knowledge Q&A community. Some experts attribute knowledge discovery and critical thought to communities' users' information behavior. Individuals, on the other hand, are emotional beings who are easily influenced by their surroundings. The material they observe and their own emotions influence how they disseminate information.

Furthermore, in the procedure of information analysis, the user's cognition and emotions will perform a significant part in the information processing (Berger, 2012). Operators in the group will display clear, expressive inclinations when post due to the body's observation, that is in line with the main features of the ECT. As a result, this study integrates ECT with communaldata dissemination behavioral science, which somewhat broadens the context of the theory's implementation but likewise aims to comprehend the emotive transformation of operators in the knowledge Q&A community throughout the pandemic by examining their temporal and spatial sentimental progression.

6.2.2 Approaches to OL Detection

6.2.2.1 Measures Based on Network Structure

This segment discusses methods for evaluating headship based on the network's structure as in Figure 6.1, with topological variables serving as markers for influential nodes. Because the societal place is a significant aspect of OL, heuristics based on vertex level, network pathways, or communal composition are employed to recognize possible leaders by their positioning inside the system. Figure 6.1 shows the different approaches to OL detection.

(a) Centrality measures

Calculating a vertex's centrality is a simple way of determining its importance. The centrality of vertices is a measurement of its significance in the network from a particular viewpoint (SergeyBrin, 2000). It can be stated in various ways, resulting in a variation of centrality measures. The centrality measure is appropriate in a given situation and is determined by which feature of the network topology the metric reflects. The degree of the vertex, often known as centrality measures, is the most basic centrality metric. According to this metric, the more interconnected a vertex is, the more relevant it is in the network. From the standpoint of opinion leaders, it appears evident that those with more connections have more opportunities for communication and impact. Eigen Vector centrality is an initiative to convey the qualitative feature of a vertex's connections. The metric additionally considers the neighbors' centrality, based on the notion that links to more prominent vertices are more relevant than linkages to less prominent vertices. Page Rank, which measures the significance of webpages in a search engine, is similar to eigen vector centrality (Freeman, 1977).

(b) Community-driven measures

Several methods for estimating an individual's significance in a network exist, but most of them—for example, the standard centrality measures —do not consider the channel's community composition. A community is defined as a network region with a greater connection density and is commonly created by vertex that shares mutual traits

(Fortunato, 2010) persons who regularly interact with one other or clusters of compatible persons. Internal and external connections are essential in terms of social position as an influence for OL. The positioning of people within a network's communal structure is an essential attribute, and it seems appropriate to consider their significance while judging their relevance. Embeddedness is one of the most critical metrics for identifying relevant vertices in a channel's community composition. It's the ratio of the overall degree to the internal degree. This metric measures how closely a vertex is linked to its community, yet, many vertices in network systems have no links to other societies and have strong integrated theoretical values (A. Lancichinetti, 2010) Embeddedness' capacity to discriminate between significant and irrelevant vertices is harmed in such instances.

Figure 6.1 Different Approaches to OL Detection (HosseiniBamakanaIldarNurgalievabQiangQ, 2019)

(c). Methods Based on Interaction

The connections between persons at a given time are captured by overall network information. On the other hand, interaction information records dynamic interactions between persons, often taking into account the liberal dimensions of these interactions. Monitoring user communications can assist in determining the movement of influence in a network, which provides valuable information about the network's most influential individuals. The flow of impact in a system is enormously responsible for the information diffusion mechanism. A node's local or global impact may be deduced in general if the dispersion procedure is observed or demonstrated correctly. All of the techniques in this area rely on interaction data, and the majority also examine structure network data; the algorithms vary from greedy search tactics to mining algorithms or clustering.

(d) Methods Based on Content Mining

This type of strategy is based on network construction and data gathered from the content that numerous nodes in the network expose and share. Subject information (because OL is topic reliant on), explicit user attributes, and sentiment are all extracted from the text. Although not all tactics are created expressly for identifying leaders, the approaches can be used in this situation.

(e) Methods Based on Content and Interaction

The techniques in this area use interaction information, content, and a network structure to find OL. The approaches take advantage of temporal cues including feedback flow, specific data, and exact composition, retrieved such as subject recognition and user-related aspects. Nearly methods in this area account for network topology, whereas others rely on content and temporal data (P. Parau, 2017).

6.3 A General Methodology for Ol Selection Using Ant Colony Optimization (Aco) Approach-

OLs are individuals in an SN who have the most impact on the acceptance or enjoyment of products/services by other persons. Formally and informally, OL controls the distribution and adoption of a new item, as well as the decisions of others, through diverse communication networks. Consequently, corporations' interest in OL has improved, making them a required primary marketing objective. Although there are so many distinct sorts of OL around the globe, selecting the appropriate one can be challenging. Several criteria will be used to identify an OL. These qualities were determined by reviewing the relevant literature. Due to the presence of diverse leaders, the feature set of these leaders will be expansive, and there may be redundant or irrelevant characteristics that impact selection. These were screened to eliminate undesirable characteristics.

6.3.1 Feature-Based Selection

Characteristics selection is performed to eliminate incorrect features to limit the number of OL. The dimensionality curse is caused by the fact that a diverse set of features makes processing difficult and exacerbates the complexity of computations. The features-based selection strategy is used to pick a subset of the original set while preserving the precision of the original set. This is accomplished by deleting unnecessary characteristics from the collection and sorting out suitable features. Features, effectiveness, and scalability can all be enhanced. The choice of appropriate leaders is based on their distinct characteristics and is required when working with a large variety of leaders (characteristics) datasets. Real-world issues necessitate the use of feature-based selection due to the abundance of obnoxious, inappropriate, or deceptive elements. The suggested method would identify suitable opinion leaders for assigning various organizational activities based on these criterion features.

6.3.2 Ants Generation

Ant Colony Optimization: Ant Colony Optimization is a metaheuristic algorithm inspired by the behavior of ants searching for food. Ants communicate through pheromone trails, which deposit on the ground to mark paths to food sources. These pheromone trails evaporate over time, and ants tend to follow paths with stronger pheromone concentrations. By following these trails, ants collectively find the shortest paths to food sources. This mechanism can be adapted to solve optimization problems, including influence maximization

ACO was suggested as a swarm intelligence technique. Ants have inspired a wide range of approaches and procedures, with ACO, a general optimization technique, being the most well-known and frequently employed. Various ant species' foraging behavior is utilized to enhance ant colonies. When ants migrate from one site to another, they deposit pheromones (chemicals) on the ground to guide other ants in the colony (members). Similarly, ACO solves optimization problems. Typically, the ACO technique is employed to determine the ideal solution. To solve difficulties, ants utilize two variables:

heuristic knowledge and the frequency of pheromones. Artificial ants that converse with one another can potentially provide superior results. The pheromone trail parameters are achieved by indirect interaction (sensing of the pheromone) between different ants. Ants do not transform themselves; instead, they affect how other ants perceive and understand the situation. ACOs can be utilized to address a range of difficulties.

6.3.3 Opinion Leader Selection

The procedure of picking a leader begins with forming an AC, which travels along several paths (edges) and selects a leader based on the pheromone value of each edge. If stopping requirements for ant visits are met, the ants will leave (traverse) and the best subclass of OLs will be identified, which will then be utilized in a variety of marketing or industrial contexts. If the traversal of ants during the visit does not fulfill the halting conditions, the pheromone values are modified and the cycle is repeated. Figure 2 shows the general flow for opinion leader detection with the Ants colony optimization algorithm. The work starts with the ant's generation with OL selection. If the criteria of the selection meet, then the program may end. And if the criteria don't meet then the new values should be modified and initialized again and the process continues till it finds the best opinion leader.

Applying Ant Colony Optimization to Influence Maximization The application of ACO to influence maximization involves mapping the problem to a graph representation, where nodes represent individuals in the social network, and edges indicate relationships or connections between them. The goal is to find a subset of nodes to maximize the spread of influence throughout the network. The following steps outline the application of ACO to influence maximization:

Step1: Initialization

- Initialize a population of ants and set pheromone levels on each edge of the graph.
- Randomly select a set of initial nodes as potential, influential individuals.

Step 2: Ant Behaviour

- Each ant traverses the graph, starting from one of the initial nodes.

- At each node, the ant probabilistically selects the next node to visit based on the pheromone levels and heuristic information (e.g., node degree, centrality).
- Both pheromone trails and heuristic factors influence the probability of selecting a particular node.
- The ant repeats this process until it visits a predefined number of nodes or exhausts the pheromone trails.

Step 3: Pheromone Updates

- After all ants have completed their tours, update the pheromone levels on each edge.
- Evaporate the existing pheromone trails to allow the exploration of new paths.
- Deposit pheromone on edges visited by the ants based on the quality of the solution found.

Step 4: Termination Condition:

Repeat Steps 2 and 3 for a specified number of iterations or until a termination criterion is met (e.g., a maximum number of iterations without improvement.

Step 5: Solution Extraction:

- Extract the influential individuals based on the pheromone levels on the graph's edges
- Higher pheromone levels indicate a higher probability of being influential
- Rank the nodes according to their pheromone levels and select the top-k nodes as the influential individuals.

6.3.4 Formation of Results

The selection of OL based on their characteristics begins with creating a random number of fake ants as in Figure 6.2 . The number of ants on the graph resembles the number of linked OLs. Each ant's traverse initiates the procedure of creating a graph from a single node. Ant moves probabilistically from a starting place, traversing multiple nodes before process completion. The resulting group of OLs is collected and examined to determine the ideal subset. When the best OLs have been identified, and the results are made public.

People's daily life have become increasingly reliant on Web 2.0. Several virtual communities, such as blogs and websites, are growing more engaged with time. People can readily voice their views publicly and observe what others say. The initial difficulty in identifying OLs is determining how a leader is characterized in various circumstances. If we define thought leaders as persons who can significantly influence others, then leadership can be regarded from various viewpoints. A leader's potential to convey persuasion is intuitively represented in the structural qualities of the SN. A typical definition of the leader notion is the origin of an action that spreads further and quicker than other activities.A comprehensive solution can be attained by integrating structural features, influence flow, and diffusion processes in networks. Essential to determine OL is their inherent qualities, which, in conjunction with other variables, determine their standing.

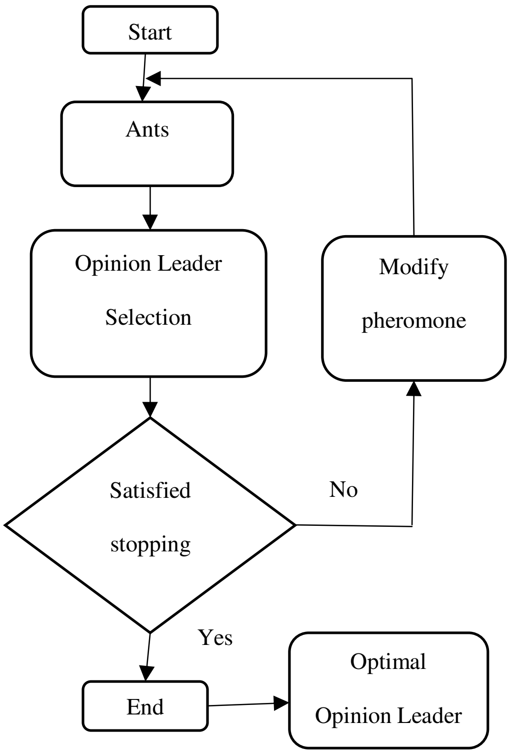


Figure 6.2 – A General Framework for Opinion Leader Detection (X. Qiang, 2021)

6.4 Discussion and Applications

By evaluating personal attributes and the information generated by leaders, including online posts, blog entries, comments on news stories, and even endorsements of others' work through sharing, one can indirectly analyze these intrinsic qualities. In actuality, all of these factors contribute to OL status, and ideally, methodologies for identifying OLs would consider all of these factors. Nevertheless, it is challenging to build broadly applicable procedures that consider all of these issues. Consequently, the most suitable approach to a particular problem relies on its environment.

6.4.1 Applications

When OLs are addressed, information can be conveyed more successfully in various areas. For instance, identifying OLs has been utilized in the medical industry to promote novel and effective therapies, such as HIV prevention and child health development. To successfully communicate information, climate change awareness efforts have utilized opinion leaders. In the middle to late years of the first decade of this century, Al Gore coordinated two initiatives that selected OL to give demonstrations and educate the general public on climate change issues. Detection of OLs is especially important in advertising, promotion, and product adoption, where organizations want to attract potential customers to their products as efficiently as feasible. Political campaigns can also profit from recognizing and contacting OL, as seen by George W. Bush's 2004 presidential election campaign, in which OL was chosen to endorse the campaign. Significant who the OL is can be advantageous in several areas, making their accurate and efficient recognition a crucial responsibility.

6.5 Conclusion

Nowadays, networks are the most popular medium among the general public, and practically everyone is connected to them, resulting in many societal difficulties as virtual networks spill over into the real world. OLs play a critical role in the distribution of information on a variety of topics. Previous studies used quantitative clustering approaches to recognize OL or used social assessment approaches based on interview self-reports or reviews. Some researchers have examined online posts to establish OL's social tendencies and interactions with their followers. As a result, this research must suggest a way of analyzing the features of social platform posts and their interactions. It is important to identify OL for an opinion leadership intervention to establish whether or not the dissemination of innovation succeeds. Few analyses have provided a comprehensive framework for researchers to determine opinion leaders. The goal of the current study was to present a synthesis of the methodologies for identifying OL using SN analysis techniques, based on this need. This research combined communication network data gathering methods with social network analysis to determine opinion leaders, and it presented how SN data were obtained in two different social systems. In social network analysis, three network centrality metrics are commonly used: degree, betweenness, and closeness. It will be exciting to see if other amalgamation approaches can be used to generate leaders in addition to Differential Evolution. We'd also like to investigate if this hybrid model can address rotational multi-objective situations, which are difficult to solve with current EMO methods.

Chapter 7 CONCLUSION AND FUTURE SCOPE

This chapter highlights the main conclusion and explains the possibilities of future scope based on the proposed algorithm and methodology for finding Opinion Leaders for Influence Maximization.

7.1 Introduction

The aim of powerful online marketing strategy is to target wide range of audiences with minimal cost. These online strategies capture thoughts, opinions and experiences shared by online users that have higher influential power than the other users in the network. These influential users or termed as opinion leaders are being targeted by marketers to promote their product and services.

The study proposed an approach of Influence Maximization using Deep learning that combined topological and topical features to achieve the best results. The work proposed Effectual Seed Pick framework that can be applied in any area or field like politics, product promotion or service promotion, or healthcare where people's opinion is involved. The seed-dependent influence maximization algorithm, which addresses issues in the network's memory effect or social reinforcement effect, is used to obtain the maximum acquiring probability as the number of neighbors of the seed increases at a given time. Furthermore, the Multi neighbor seed selection approach is used to overcome problems during seed selection due to a lack of discriminatory power, in which the neighbor degree value (NDV) is calculated to estimate the influence strength of the seeds. Furthermore, the user interaction was ignored in the selection of Judgment leaders, which is addressed by the Judgment leader pick weighting grade. As a result, the proposed framework effectively addresses the issues in viral marketing and maximizes its influence.

7.2 Major Results

The significant finding of the research work undertaken are:

1. Improved scalability: DeepWalk can effectively capture structural information in large-scale graphs, allowing influence maximization algorithms to scale better. By learning compact representations of nodes, DeepWalk reduces the computational complexity of subsequent influence maximization algorithms.

2. Enhanced node representation: DeepWalk provides rich and meaningful illustrations for nodes in the graph, enabling influence maximization algorithms to capture the influence-spreading characteristics of different nodes better. These representations can capture both local and global structural information, helping identify influential nodes more accurately.

3. Increased influence spread: DeepWalk-based influence maximization methods have shown improved performance in maximizing influence spread in social networks. By leveraging the learned node embeddings, these methods can identify seed nodes likely to impact the diffusion process significantly.

4. Robustness to node label noise: DeepWalk-based techniques can be more resilient to noise in node labels or attributes. Since DeepWalk learns node representations solely based on the graph structure, it can still capture meaningful information even when node labels are incomplete or noisy. This robustness can improve the performance of influence maximization algorithms in real-world scenarios.

5. Generalizability across domains: DeepWalk-based techniques can be applied to various types of graphs beyond social networks, such as citation networks, recommendation systems, and biological networks. This versatility makes DeepWalk a valuable tool for influence maximization across different domains.

6. When for different datasets like YouTube, Facebook, Reddit, Instagram, Gab, and Twitter, the proposed framework is compared with the existing algorithms to find out the seeds for maximum influence maximization and the accuracy, recall, F1-score, and precision are high as compared to other existing algorithms.

7. The Multi neighbor seed selection approach is used to overcome problems during seed selection due to a lack of discriminatory power, in which the neighbor degree value (NDV) is calculated to estimate the influence strength of the seeds. Furthermore, the user

interaction was ignored in the selection of Judgment leaders, which the Judgment leader addresses pick weighting grade. As a result, the proposed framework effectively addresses the issues in viral marketing and maximizes its influence. The proposed framework is useful for trust management using opinion leaders in ubiquitous services.

8. Evolutionary algorithms were studied to find opinion leader specifically Ant Colony. It is important to identify OL for an opinion leadership intervention to establish whether or not the dissemination of innovation succeeds. Few analyses have provided a comprehensive framework for researchers to determine opinion leaders. The work aimed to present a synthesis of the methodologies for identifying OL using SN analysis techniques based on this need. This research combined communication network data gathering methods with social network analysis to determine opinion leaders.

7.3 Scope of the Future Research

The proposed methodology DeepWalk Based Influence Maximization, Effectual Seed Pick Framework, Multi Neighbour Seed selection for Influence Maximization, and Evolutionary Algorithms Based Seed Pick for Influence Maximization are just starting points in the use of machine learning and social network analysis for the Influence Maximization problem by using Seeds in the network, the scope of research is vast. Some of the future research areas are:

1. Use of proposed Methodology in other application areas like terrorism, politics, agriculture, and outlier detection in the networks.
2. More context based algorithms can be developed depending on the business problem for which Influence Maximization is required.
3. Scalability Issues can be covered for the Influence Maximization problem in Social Networks as the dataset used can be vast.
4. Computational time for calculating seeds in the network can also be considered for future research work.
5. Other diffusion models other that the Cascade diffusion model and Linear Threshold Model can also be considered for further research.

References

A. Bozorgi, H. H. (2016). INCIM: A community-based algorithm for influence maximization problem under the linear threshold model. *Information Processing and Management , 52 no.6*, 1188–1199.

A. Goyal, W. L. (2011). CELF++: Optimizing the greedy algorithm for influence maximization in social networks. *Proc. 20th International Conference Companion World Wide Web, WWW*, (pp. 47–48).

A. Lancichinetti, J. S. (2010). Characterizing the community structure of complex networks. *PLoS One , 5* (8), 1–8.

A. Landherr, B. F. (2010). A Critical Review of Centrality Measures in Social Networks. *Business Information System Engineering , 6*, 371–385.

Á. MacDermott, M. M. (2022). Using Deep Learning To Detect Social Media 'trolls. *Forensic Science Int. Digit Investing* .

A.Kanan, A. A. (2022). Machine Learning Approaches For Developing Land Cover Mapping. *Applied Bionics Biomech* .

Agarwal, S. &. (2020). Effective influence estimation in twitter using temporal, profile, structural and interaction characteristics. Information Processing & Management. *Effective influence estimation in twitter using temporal, profile, structural and interaction characteristics. Information Processing & Management* , 57(6).

Aghadam, S. M. (2021). Opinion Leaders Selection with Grey Wolf Optimizer Algorithm on. *13* (2).

April, C. P. (2009). Corrected 24 april 2009. *Science* , 892-896.

Aslay, C. B.-Y. (2014). Online topic-aware influence maximization queries. *In Advances in Database Technology - EDBT 2014: 17th International Conference on Extending Database Technology, Proceedings* (pp. 295–306). OpenProceedings.org, University of Konstanz, University Library.

B. Cai, L. Z. (2019). Community detection method based on node density, degree centrality, and k-means clustering in complex network. *Entropy , 21* (12).

B. Liu, G. C. (2014). Influence spreading path and its application to the time constrained social influence maximization problem and beyond. *IEEE Transaction Knowledge Data Engineering , 26* (8), 1904–1917.

B. M. Tabak, M. T. (2014). Directed clustering coefficient as a measure of systemic risk in complex banking networks. *Physica A: Statistical Mechanics and Its Applications , 394*, 211–216.

Banati, H. &. (2013). Evolutionary Approach for k-Max Influence Problem. 163–172.

Banerjee, S. J. (2020). A survey on influence maximization in a social network. . *Knowledge and Information Systems* , 3417–3455.

Berger, J. (2012). What Makes Online Content Viral? *SAGE Journals* .

Bhagat, S. G. (2012). Maximizing product adoption in social networks. *In WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining* , (pp. 603–612).

Bharadwaj, P. K. (2021). *Pattern Recognition and Machine Learning.* EAI/Springer Innovations in Communication and Computing.

Bodendorf, F. &. (2010). Detecting opinion leaders and trends in online communities. *In 4th International Conference on Digital Society, ICDS 2010, Includes CYBERLAWS 2010: The 1st International Conference on Technical and Legal Aspects of the e-Society* .

Bonchi, F. C. (2011). Social network analysis and mining for business applications. *ACM Transactions on Intelligent Systems and Technology , 3* (2).

Borgatti, S. P. (2009). Network analysis in the social sciences. *Science* .

Borgatti, S. P. (2016). Analyzing social networks. *Global Journal of Human-Social Science: A Arts & Humanities - Psychology* .

Brown, J. O. (2007). Online Communities : Conceptualizing the Online Social Network . *Journal of Interactive Marketing* , 2–20.

Bu, Z. L. (2020). Graph K-means Based on Leader Identification, Dynamic Game, and Opinion Dynamics. *IEEE Transactions on Knowledge and Data Engineering , 32* (7), 1348–1361.

Bucur, D. &. (2016). Influence maximization in social networks with genetic algorithms. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9597*, pp. 379–392. Springer Verlag.

C. Borgs, M. B. (2014). Maximizing social influence in nearly optimal time. *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, (pp. 946–957).

C. C. Huang, W. Y. (2020). Identification of opinion leaders and followers—a case study of green energy and low carbons. *Applied Science , 10* (23), 1–16.

C. S. Sanghee Oh, J. O. (2008). The use of information sources by internet users in answering questions. *Semantic Scholar* .

C. Wang, W. C. (2012). Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, *vol. 25, no. 3*, pp. 545–576.

Cao, J. W. (2022). Compactness Preserving Community Computation Via a Network Generative Process. *IEEE Transactions on Emerging Topics in Computational Intelligence , 6* (5), 1044–1056.

Carley, W. W. (2015). Measuring temporal patterns in dynamic social networks. *ACM Transactions on Knowledge Discovery from Data , 10.*

Celentano, D. D. (2010). Social Networks and Health: Models, Methods, and Applications: By Thomas W. Valente. *American Journal of Epidemiology* , 488–488.

Centola, D. (2010). The spread of behavior in an online social network experiment. . *Science* , 1194–1197.

Chen W., W. C. (2010). Scalable Influence Maximization for Prevalent Viral. *16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1029–1038). ACM.

Chen, T. S. (2020). Modeling Public Opinion Polarization in Group Behavior by Integrating SIRS-Based Information Diffusion Process. *Complexity* .

Chen, W. L. (2016). Real-time topic-aware influence maximization using preprocessing. *Computational Social Networks* .

Chen, W. W. (2010). Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks . *Microsoft Research Technical Report. Presentations* .

Chen, W. W. (2009). Efficient influence maximization in social networks. *In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , (pp. 199–207).

Chen, W. Y. (2010). Scalable influence maximization in social networks under the linear threshold model. *In Proceedings - IEEE International Conference on Data Mining, ICDM* , (pp. 88–97).

Cheng, J. J. (2013). An epidemic model of rumor diffusion in online social networks. *European Physical Journal B* .

Chien, I. L. (2018). Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. *In International Conference on Artificial Intelligence and Statistics, AISTATS 2018* (pp. 871–879). PMLR.

Creswell, J. W. (2018). *Designing and Conductiong Mixed MethosResearch.* SAGE Publication, Inc.

Crossley, N. B. (2017). Social Network Analysis for Ego-Nets. *SAGE Publications Ltd.*

D. Posner, E. (2015). *Machine Learning in Python.* Indianapolis: John Wiley & sons, Inc.

Dey, A. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies* , 1174–1179.

Dietz, L. S. (2007). Unsupervised prediction of citation influences. *ACM International Conference Proceeding Series*, (pp. 233-240).

Domingos, P. (2012). A Few Useful Things to Know About Machine Learning. *Communications of the ACM* .

Domingos, P. R. (2001). Mining the network value of customers. *In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 57–66). Association for Computing Machinery (ACM).

E. Bakshy, W. A. (2011). Everyone's an influencer: Quantifying influence on twitter. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM*, (pp. 65-74).

E. Cohen, D. D. (2014). Sketch-based influence maximization and computation: Scaling up with guarantees. *CIKM 2014 - Proceedings of the 2014 ACM International Conference on Information and Knowledge Management*, (pp. 629–638).

Estrada, E. &.-V. (2005). Spectral measures of bipartivity in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics , 72* (4).

Fan, J. Q. (2018). Octopus: An online topic-aware influence analysis system for social networks. *In Proceedings - IEEE 34th International Conference on Data Engineering, ICDE 2018*, (pp. 1573–1576).

Fortunato, S. (2010). Community detection in graphs. *Physics Reports , 486* (3-5), 75–174.

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry , 40* (1).

G. Cordasco, L. G. (2018). Evangelism in social networks: Algorithms and complexity. *Networks , 71* (4), 346–357.

Galhotra, S. A. (2015). ASIM: A scalable algorithm for influence maximization under the independent cascade model. *In WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web* (pp. 35–36). Association for Computing Machinery, Inc.

Galhotra, S. A. (2016). Holistic influence maximization: Combining scalability and efficiency with opinion-aware models. *In Proceedings of the ACM SIGMOD International Conference on Management of Data . 26-June-2016*, pp. 743–758. Association for Computing Machinery.

Gawali, E. A. (2022). Analysis of Machine Learning Techniques for Sentinel-2A Satellite Images. *Journal of Electrical and Computer Engineering.*

Gazan, R. (2010). Microcollaborations in a social Q&A community. *Information Processing and Management , 46* (6), 693–702.

Géron, A. (2019). *Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow.* O'Reilly Media.

Ghosh, M. G. (2021). Aspect-Based Unsupervised Negative Sentiment Analysis. In *In Lecture Notes on Data Engineering and Communications Technologies* (Vol. 57, pp. 335–344). Springer Science and Business Media Deutschland GmbH.

Goldenberg, J. L. (2001). Using Complex Systems Analysis to Advance Marketing Theory Development. *Academy of Marketing Science Review , 2001* (3), 1–18.

Gong, M. Y. (2016). Influence maximization in social networks based on discrete particle swarm optimization. *Information Sciences* , 367–368,600–614.

Goyal, A. B. (2008). Discovering leaders from community actions. . *In International Conference on Information and Knowledge Management, Proceedings*, (pp. 499–508).

Goyal, A. B. (2010). Learning influence probabilities in social networks. *In WSDM 2010 - Proceedings of the 3rd ACM International Conference on Web Search and Data Mining* , (pp. 241–250).

Goyal, A. L. (2011). CELF++: Optimizing the greedy algorithm for influence maximization in social networks. *In Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011* , (pp. 47–48).

Granovetter, M. (1978). *Threshold Models of Collective Behavior1 BACKGROUND AND DESCRIPTION OF THE MODELS.* Source: American Journal of Sociology .

Guo Zhen, Z. Z. (2014). A two-level topic model towards knowledge discovery from citation networks. *IEEE Transactions on Knowledge and Data Engineering* , 780-794.

Guo, J. Z. (2013). Personalized influence maximization on social networks. *In International Conference on Information and Knowledge Management,Proceedings* , (pp. 199–208).

Guo, L. Z. (2017). Influence maximization in trajectory databases (Extended abstract). *In Proceedings - International Conference on Data Engineering. 0*, pp. 27–28. IEEE Computer Society.

H. T. Nguyen, M. T. (2017). A Billion-Scale Approximation Algorithm for Maximizing Benefit in Viral Marketing. *IEEE/ACM Transactions on Networking , 25* (4), 2419–2429.

H. T. Nguyen, M. T. Stop-and-Stare: Optimal sampling algorithms for viral marketing in billion-scale networks. *Proc. ACM SIGMOD Int. Conf. Manag. Data*, *26-June-20*, pp. 695–710.

H. T. Nguyen, M. T. (2016). Stop-and-Stare: Optimal sampling algorithms for viral marketing in billion-scale networks. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, *26-June-20*, pp. 695–710.

He., W. Z. (2013). Identifying and evaluating the internet opinion leader community based on k-clique clustering. *Acedemia* .

HosseiniBamakanaIldarNurgalievabQiangQ, S. M. (2019). Opinion leader detection: A methodological review. *Expert Systems with Applications* .

Huang, L. F. (2010). Artificial intelligence. *2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010* (pp. 575-578). IEEE.

Ienco, D. B. (2010). The meme ranking problem: Maximizing microblogging virality. *In Proceedings - IEEE International Conference on Data Mining, ICDM* , (pp. (pp. 328–335)).

J. Kim, S. K. (2013). Scalable and parallelizable processing of influence maximization for large-scale social networks? *Proceedings - International Conference on Data Engineering*, (pp. 266–277).

J. Leskovec, A. K. (2007). Cost-effective outbreak detection in networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 420–429).

J. Leskovec, A. K. (2007). Cost-effective outbreak detection in networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 420–429).

J. Leskovec, D. C. (2010). Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research* , 985–1042.

J. Weng, J. W. (2010). Institutional Knowledge at Singapore Management University Twitterrank : Finding topic-sensitive influential Twitterers TwitterRank : Finding Topic-sensitive Influential Twitterers. *WSDM '10 Proc. third ACM International. Conference.* ACM.

J. Xie, B. K. (2011). SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. *Proceedings - IEEE International Conference on Data Mining, ICDM*, *September 2011*, pp. 344–349.

J., G. B. (2001). Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters* , 211-223.

J.Tang. (2009). Social Influence Analysis In Large Scale Networks. *International Conference Knowledge Discovery Data Mining* (pp. 807-815). ACM SIGKDD.

Jaradat, S. B. (2018). *Community Structure Detection Using Firefly Algorithm.*

Jiang, L. G. (2013). BBS opinion leader mining based on an improved PageRank algorithm using MapReduce. *In Proceedings - 2013 Chinese Automation Congress, CAC 2013* , 392–396.

Jiang, Q. S. (2011). Simulated annealing based influence maximization in social networks. *In Proceedings of the National Conference on Artificial Intelligence* , *1*, pp. 127–132.

Jin, J. (2015). Why users contribute knowledge to online communities. *ACM diigital library* .

Jing, L. &. (2014). Identification of microblog opinion leader based on user feature and interaction network. *In Proceedings - 11th Web Information System and Application Conference, WISA 2014* , (pp. 125–130).

Johnson, C. A. (2007). Social capital and the search for information: Examining the role of social capital in information seeking behavior in Mongolia. *Journal of the American Society for Information Science and Technology , 58* (6), 883–894.

Jonard, R. C. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control* , 1557–1575.

K. Jung, W. H. (2012). IRIE: Scalable and robust influence maximization in social networks. *Proceedings - IEEE International Conference on Data Mining, ICDM*, (pp. 918–923).

K. Rahimkhani, A. A. (2015). A fast algorithm for finding most influential people based on the linear threshold model. *Expert Systems with Applications*, *42*, pp. 1353–1361.

Katarya, L. J. (2019). Discover opinion leader in online social network using firefly algorithm. *Expert Systems with Applications , 122*, 1-15.

Katherine, S. w. (2012). *Social Network Analysis.* Cambridge University Press.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika , 18* (1), 39–43.

Kempe David, K. J. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 137-146). ACM.

Kempe, D. K. (2015). Maximizing the spread of influence through a social network. *Theory of Computing* , 105–147.

Kim, A. J. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research* , 1480–1486.

Kim, H. B. (2014). Finding influential neighbors to maximize information diffusion in twitter. *In WWW 2014 Companion - Proceedings of the 23rd International Conference on World Wide Web (pp. 701–706)*, (pp. 701–706).

Kim, J. K. (2013). Scalable and parallelizable processing of influence maximization for large-scale social networks? *In Proceedings - International Conference on Data Engineering*, (pp. 266–277).

Kimura, M. &. (2006). Tractable models for information diffusion in social networks. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, *4213 LNAI*, p. 259.

L. Jain, R. K. (2019). Opinion leader detection using whale optimization algorithm in online social network. *Expert Systems with Applications , 142*.

L. M. Norz, V. D. (2022). Measuring Social Presence In Online Based Learning:An Exploratory Path Analysis Using Log Data and Social Network Analysis. *Internet High. Education* .

Lam, H. W. (2009). Finding influential ebay buyers for viral marketing - A conceptual model of buyerrank. *In Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, (pp. 778–785).

Langville, A. N. (2011). Google's PageRank and beyond: The science of search engine rankings. *Princeton University Press* , 1–224.

Lee, W. K. (2012). LCT-IC: Continuously activated and Time-restricted independent cascade model for viral marketing. *In Proceedings - IEEE International Conference on Data Mining, ICDM* , (pp. 960–965).

Leskovec, J. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web* .

Leskovec, J. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web* .

Leskovec, J. K. (2007). Celf Algorithm - In order to monitor the pipeline network epidemic situation, the sensor location selection is judged by response time, response possibility, and influencing population factors.

Li, J. X. (2012). Training opinion leaders in microblog: A game theory approach. *In Proceedings - 2nd International Conference on Cloud and Green Computing and 2nd International Conference on Social Computing and Its Applications, CGC/SCA*, (pp. 754–759).

Li, J. Y. (2016). Robust influence blocking maximization in social networks. *Jisuanji Yanjiu Yu Fazhan/Computer Research and Development , 5* (3), 601–610.

Li, W. R. Choosing leaders for multi-objective PSO algorithms using differential evolution. In *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artificial Intelligence Lecture Notes Bioinformatics)* (Vol. 5361 LNAI, pp. 249–258).

Li, Y. T. (2017). Discovering your selling points: Personalized Social influential tags exploration. *In Proceedings of the ACM SIGMOD International Conference on Management of Data . Part F127746*, pp. 619–634. Association for Computing Machinery.

Li, Y. Z. (2015). Real-time targeted influence maximization for online advertisements. In Proceedings of the VLDB Endowment Association for Computing Machinery.

Liu, B. C. (2012). Time constrained influence maximization in social networks. *In Proceedings - IEEE International Conference on Data Mining, ICDM*, (pp. 439–448).

Liu, B. J. (2017). PennState. *Identifying Predict. desire to Help Soc. Question answering* .

Liu, B. (2011). *Web Data Mining.* Springer Berlin Heidelberg.

Liu, N. M. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. *Expert Syst. Appl. , 41* (4 PART 1), 1357–1368.

Liu, N. M. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. *Expert System with Application , 40* (4), 1357–1368.

Liu, N. M. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. *Expert System with Application , 41* (4), 1357–1368.

M. Gomez-Rodriguez, D. B. (2011). Uncovering the temporal dynamics of diffusion networks. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, (pp. 561–568).

M. L. G. Zhenya Tang, L. C. (2018). How to keep brand fan page followers? The lens of person-environment fit theory. *Emarland* .

Mao, J. (2014). Social media for learning: A mixed methods study on high school students' technology affordances and perspectives. *Computers in Human Behavior* , 213–223.

Martínez-Torres, M. R.-M. (2018). Identifying the features of reputable users in eWOM communities by using Particle Swarm Optimization. *Technological Forecasting and Social Change , 133*, 220–228.

McAdam, D. (1982). Political Process and the Development of Black Insurgency, 1930-1970. *University of Chicago Press* , 36-59.

McAlexander, J. H. (2002). Building brand community. *Journal of Marketing* , 38–54.

Mccarty, C. K. (2001). Comparing two methods for estimating network size . *ABI/INFORM Global Human Organization; Springer* , 28.

Meißner, N. (2014). Opinion leaders as intermediaries in audience building for independent films in the Internet age. *SAGE Journals* .

Meyffret, S. (2014, October). RED : a Rich Epinions Dataset for Recommender Systems To cite this version : RED : a Rich Epinions Dataset for Recommender Systems .

Mikolov, T. C. (2013). Efficient estimation of word representations in vector space . *In 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings. International Conference on Learning.*

Mikolov, T. C. (2013). Efficient estimation of word representations in vector space. *In 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings. International Conference on Learning Representations, ICLR.*

Mirtabatabaei, A. &. (2012). Opinion dynamics in heterogeneous networks: Convergence conjectures and theorems. *SIAM Journal on Control and Optimization , 50* (5), 2763–2785.

MNaik, S. A. (2013). Maximizing influence of viral marketing via evolutionary user selection. *In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013* (pp. 1435–1436). Association for Computing Machinery.

Murphy, K. P. (2012). *Machine Learning A Probabilistic Perspective.* London,England: The MIT Press Cambridge, Massachusetts London,.

N. T. H. Gayraud, E. P. (2015). Diffusion maximization in evolving social networks. *COSN 2015 - Proceedings of the 2015 ACM Conference on Online Social Networks*, (pp. 125–135).

Nekovee, M. M. (2007). Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and Its Applications* , 457–470.

Nguyen, H. T. (2016). Cost-aware Targeted Viral Marketing in billion-scale networks. *In Proceedings - IEEE INFOCOM (Vol. 2016-July). Institute of Electrical and Electronics Engineers Inc.*

P. Parau, C. L. (2017). Opinion Leader Detection. *Sentiment Analysis in Social Networks* .

P., D. (2012). A Few Useful Things to Know About Machine Learning. *Communications of the ACM* , 79–88.

Perozzi, B. &. (2014). DeeWalk. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , 701–710.

Qiang, X. H. (2021). Criterial Based Opinion Leader's Selection for Decision-Making Using Ant Colony Optimization. *Scientific Programming* .

Qiu, J. W. (2016). A survey of machine learning for big data processing. *Eurasip Journal on Advances in Signal Processing. Springer International Publishing* .

Ramakrishnan, J. M. (2020). Opinion Mining using Machine Learning Approaches: A Critical Study. *In 2020 International Conference on Computing and Information Technology, ICCIT.* nstitute of Electrical and Electronics Engineers Inc.

Razghandi, A. &. (2020). Improved Leach algorithm for detecting opinion leaders in social networks.

Reingen, P. H. (1994). Structural Holes: The Social Structure of Competition. *Journal of Marketing* .

Richardson Matthew, D. P. (2002). Mining knowledge-sharing sites for viral marketing. *SIGKDD 02,Association for Computing Machinery* .

Richert, W. &. (2015). *Building Machine Learning Systems with Python.*

Rui, X. Y. (2020). A neighbour scale fixed approach for influence maximization in social networks. *Computing , 102* (3), 427–449.

S, W. H. (2021). Big Data Analysis and Perturbation using Data Mining Algorithm. *Journal of Soft Computing Paradigm , 3* (1), 19–28.

S. Cheng, H. S. (2013). StaticGreedy: Solving the scalability-accuracy dilemma in influence maximization. *International Conference on Information and Knowledge Management, Proceedings*, (pp. 509–518).

S. Peng, G. W. (2017). Social Influence Analysis in Social Networking Big Data: Opportunities and Challenges. *IEEE Network* , 11-17.

S. S. Lokesh Jain, R. K. (2020). Recognition of opinion leaders coalitions in online social network using game theory. *Knowledge-Based Systems* .

Saraswathi, K. R. (2020). Multi ant colony optimization for opinion classification. *International Journal of Scientific and Technology Research , 9* (3), 4554–4560.

Sarmento, H. M. (2014). Match analysis in football: a systematic review. *Journal of Sports Sciences* , 1831–1843.

Schölkop, M. G.-R. (2012). Influence maximization in continuous time diffusion networks. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, *1*, pp. 313–320.

Schölkopf, M. G.-R. (2012). Influence maximization in continuous time diffusion networks . *Proc. 29th Int. Conf. Mach. Learn. ICML*, (pp. 313–320).

Scott, J. W. (1996). Social Network Analysis: Methods and Applications . *The British Journal of Sociology* .

Scott, J. (2012). *What is Social Network Analysis?* Bloomsbury Academic.

SergeyBrin, L. (2000). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks* .

Shang, J. Z. (2017). CoFIM: A community-based framework for influence maximization on large-scale networks. *Knowledge-Based Systems , 117*, 88–100.

Shuo Chen, J. F.-l. (2015). Online Topic-Aware Influence Maximization. *Proceedings of the VLDB Endowment , 8* (6), 666–677.

Snijders, T. A. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks* , 44-60.

Song, X. T. (2006). Personalized recommendation driven by information flow. *In Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 509–516). Association for Computing Machinery (ACM).

Sonia, K. S. (2023). DWIM: Influence Maximization Using Deep Learning. *Intelligent Automation & Soft Computing, IASC , vol.35* (no.1).

Sonia, K. S. (2023). Effectual Seed Pick Framework focusing on Maximizing Influence in Social Networks. *Wireless Communications and Mobile Computing , Volume 2023*, 19.

Squillero, G. B. (2016). Applications of evolutionary computation: 19th European conference, evoapplications 2016 Porto, Portugal, march 30 – april 1, 2016 proceedings, part I. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics , 9597.*

Sugawara, E. &. (2014). Properties of AdeABC and AdeIJK efflux systems of Acinetobacter baumannii compared with those of the AcrAB-TolC system of Escherichia coli. *Antimicrobial Agents and Chemotherapy , 58* (12), 7250–7257.

T. Guan, L. W. (2018). Knowledge contribution behavior in online Q&A communities: An empirical investigation. *Computers in Human Behavior , 81*, 137–147.

T. Veeramakali, A. N. (2022). Preserving The Privacy Of Healthcare Data Over Social Networks Using Machine Learning. *Computational Intelligent Neuroscience* .

Tang, X. &. (2012). Ranking user influence in healthcare social media. *ACM Transactions on Intelligent Systems and Technology* .

Trusov, M. B. (2010). Determining influential users in internet social networks. *Journal of Marketing Research , 47* (4), 643–658.

Tsai, C. W. (2016). A Genetic NewGreedy Algorithm for Influence Maximization in Social Network. *In Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, (pp. 2549–2554).

Tsukanova, S. A. (2019). On the Algorithms of Identifying Opinion Leaders in Social Networks. *Procedia Computer Science , 162* (Itqm 2019), 778–785.

Van Der Merwe, R. &. (2009). Finding and utilizing opinion leaders: Social networks and the power of relationships. South African Journal of Business Management. *South African Journal of Business Management , 40* (3), 65–76.

W.Chen. (2011). Influence maximization in social networks when negative opinions may emerge and propagate. *Proceedings of the 11th SIAM International Conference on Data Mining, SDM 2011*, (pp. 379–390).

Walker, S. A. (2012). Identifying influential and susceptible members of social networks. *Science* , 337–341.

Wang, C. C. (2012). Scalable influence maximization for independent cascade model in large-scale social networks. *In Data Mining and Knowledge Discovery , 25*, 545–576.

Wang, J. &. (2008). Particle swarm optimization with a leader and followers. *Progress in Natural Science , 18* (11), 1437–1443.

Wang, X. D. (2018). Targeted influence minimization in social networks. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) , 10939 LNAI*, pp. 689–700.

Wang, X. Z. (2016). Distance-aware influence maximization in geo-social network. *In 2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016* (pp. 1–12). Institute of Electrical and Electronics Engineers Inc.

Wasserman, S. &. (1994). Social network analysis: Methods and applications.

Watts, D. J. (2007). Inf luentials, networks, and public opinion formation. *Journal of Consumer Research , 34* (4), 441–458.

Weskida, M. &. (2019). Finding influentials in social networks using evolutionary algorithm. *Journal of Computational Science , 31*, 77–85.

Won, R. (2018). *Deep-learning boost.* Nature Photonics. Nature Publishing Group.

X. Deng, Y. L. (2013). Parallel Micro Blog Crawler Construction for Effective Opinion Leader Approximation. *AASRI Procedia , 5*, 170–176.

X. He, G. S. (2012). Influence blocking maximization in social networks under the competitive linear threshold model. *Proceedings of the 12th SIAM International Conference on Data Mining, SDM 2012*, (pp. 463–474).

X. Li, X. C. (2018). Community-based seeds selection algorithm for location aware influence maximization. *Neurocomputing , 275*, 1601–1613.

X. Qiang, Z. H. (2021). Criterial Based Opinion Leader's Select ion for Decision-Making Using Ant Colony Opt imization. *Science Program* .

Xuanhao Chen, Y. Z. (2021). Efficient Similarity-aware Influence Maximization in Geo-social Network. *IEEE Transactions on Knowledge and Data Engineering , 34* (10), 1-14.

Y. C. Chen, W. Y. (2014). CIM: Community-based influence maximization in social networks. *ACM Transactions on Intelligent Systems and Technology , 5* (2).

Y. Chen, T. D. (2021). What concerns consumers about hypertension? A comparison between the online health community and the q&a forum. *nternational Journal of Computational Intelligence Systems , 14* (1), 734–743.

Y. Li, D. Z. (2015). Real-time targeted influence maximization for online advertisements. *Proceedings of the VLDB Endowment*, (pp. 1070-1081).

Y. Tang, X. X. (2014). Influence maximization: Near-optimal time complexity meets practical efficiency. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, (pp. 75–86).

Y. Xiong, Z. C. (2018). Accumulation mechanism of opinion leaders' social interaction ties in virtual communities: Empirical evidence from China. *Computers in Human Behavior , 82*, 81-93.

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research* , 452–473.

Zelen, K. S. (1989). Rethinking centrality: Methods and examples. *Social Networks , 11* (1), 1–37.

Zhao, Y. K. (2018). Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. *Information Sciences* , 131–147.

Zhao, Y. K. (2018). Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. *Information Sciences , 426*, 131–147.

Zhou, C. L. Scalable graph embedding for asymmetric proximity. *In 31st AAAI Conference on Artificial Intelligence, AAAI 2017* (pp. 2942–2948). AAAI press.

Zhou, T. C. (2015). Location-based influence maximization in social networks. *In International Conference on Information and Knowledge Management, Proceedings*, *19-23-Oct-2015*, pp. 1211–1220.