Friend Recommender System to Influence Friends on Social Networks Based on B-Mine Method

Tingting Feng, Wenya Jin*, Wei Li
School of Journalism and Communication, Hebei Institute of Communications, Shijiazhuang 050000, China

***Abstract*—Social networks are linked by one or more particular kinds of connections, including web links, friends, family, and the sharing of ideas and money. Graph theory is used to investigate social relationships in social network analysis. The individuals within the networks are the vertices, and the connections among them are the edges. Between vertices, there can be a wide variety of edges. Due to the rise in Internet usage, online shopping, and social media usage in recent years, recommender systems have become more and more popular. Numerous websites have been successful in putting this recommender system into place. This thesis introduced an approach that uses the B-mine method to explore common patterns and enhance the accuracy of identifying influential nodes in social networks. In this method, two user similarity criteria—coverage and confidence—were used simultaneously to improve the recommender system. The behavior of previous users is analyzed, and recommendations are made to the current user based on friends' behavior and similarity, as well as on their interactions and preferences across different groups. According to the simulation results, the suggested approach performs satisfactorily, with accuracy and sensitivity of 89% and 76%, respectively.**

Keywords—Influential nodes; recommender; social networks and B-mine

I. Introduction

A number of websites have been effective in adopting recommender systems, and their use has grown in recent years as a result of consumers using the Internet more frequently, making more purchases online, and interacting on social media [1]. For this system, a number of strategies have been proposed, including collaborative filtering, content-based filtering, knowledge-based filtering, and others. Nevertheless, these strategies face several obstacles, including scalability issues and cold start problems [2]. The cold start issue arises when a newly registered user is not properly assessed, making it impossible to make ideas or recommendations for them [3]. Thus, to address this issue in the majority of circumstances, the approach numerous techniques have been employed for clustering, which might offer suggestions to novice users, including hierarchical clustering and K_means clustering [4]. However, a drawback of some of these clustering techniques is that their accuracy tends to decline with increasing data sizes.

A social network is a framework for social interaction made up of individual or group nodes. Social networks are linked by one or more particular kinds of connections, including web links, friends, family, and the sharing of ideas and money [5]. Graph theory is used to investigate social relationships in social network analysis. The individuals within the networks are the vertices, and the connections among them are the edges [6]. There are several kinds of edges that can connect vertices [7]. The findings of numerous studies demonstrate that social network analysis may be applied at many different levels, both personally and socially, to find communities, develop social interactions, discover relationship kinds, examine the graph to find patterns, and other aspects of goal-achieving. Social networks are essential for corporate growth and success because they give organizations a means of information gathering, competition avoidance, and even price and policy-setting cooperation [8].

The fact that the algorithms used to conduct this exploration frequently yield a massive collection of alternating patterns as an answer is one of the main issues in the subject of alternating pattern exploration [9]. The narrower threshold they select, the more obvious this problem becomes. The primary cause of this problem is the alternating nature of all subgroups inside an item set [10]. This means that the number of intermittent item sets (i.e., the number of subsets of the large item set) that are available in the transaction database can rise exponentially in the presence of a large intermittent item set in the transaction database [11].

One of the biggest problems with social networks is locating pertinent knowledge and information among a huge number of members; this may be a very time-consuming and even irritating activity [12]. Analyzing social networks to find the people who engage in the most interactions and talk with one another is one method to keep an eye on this problem. Because the recommendations and opinions of those who are used through influencer friends, it is necessary to find influencer friends in order to present each user with the most relevant and desirable options from a vast array of information and products [13]. He developed a relationship with them or gained their trust in order to receive more precise and useful advice [14].

Finding relevant information and knowledge among a large amount of work information has become difficult and even frustrating due to the rapid pace at which new information, advertisements, products, etc., are produced in the virtual environment, particularly in social networks [15]. By finding friends (users) with the help of influential friends, recommendations and opinions of people who share interests can be used, so it seems essential that each user receives the most relevant and interesting information from a wide range of products conditions and their unique characteristics [16]. Since friends and the ideas and thoughts shared by effective friends can be trusted, more accurate and useful information can be obtained without wasting time. That's why you need to find influential friends, interact with them and categorize them. With this, it is possible to reduce the time and cost of extracting useful content by interacting with influential friends and choosing appropriate parameters such as popularity, identity and power [17].

The challenge of examining intermittent substructures can be approached from two basic perspectives: the first is based on Apriori, while the second is based on pattern growth. Using an Apriori-based technique, they first search the given set of structures for alternating small substructures. [18]. From then on, each step creates a new substructure connected by a node to another substructure. Only the nodes identified as alternating nodes in the first stage are used to add nodes to an alternating substructure. The process of generating a new substructure involves scanning the set of structures to ascertain if the new substructure is periodic or non-periodic [19]. By adding edges to a periodic substructure in every conceivable location, algorithms using a pattern expansion approach enlarge it and produce larger alternating substructures [20]. A possible issue arising from edge expansion is the possibility of repeatedly creating and examining a graph. One way to find knowledge in data mining is to investigate alternating patterns. Since this approach is intended to handle discrete data, any data used in continuous data must first be quantified. The final results that are used in this research to locate important friends from the suggested B_mine algorithm may be affected by the loss of some data and the addition of fictitious data to the data space caused by this effort. Finding influential friends is possible with the use of the suggested algorithm.

In this study, it is attempted to determine which users have the potential to be more influential on the network by taking into account the social network's graph structure. For this reason, the organizations in the network that are subgraphs of the main network are first identified using the valid algorithms available in this field. Following this, users are scored using a proposed numerical criterion based on two characteristics of the influence and connection of a user within the organization and between organizations. The final step involves identifying and reporting the users with the most influence across the network.

However, a large portion of research on social networks focuses on static or cross-sectional networks. However, this analysis looks at user interactions across several periods. The index developed in this study is adaptable to many social networks and gives each parameter a weight that indicates its relative relevance in relation to other characteristics within each social network. This technique involves reviewing users' connections in the network at various intervals, strengthening or weakening the ties between network nodes, and altering the network's structure in response to user activity. In order for the algorithm to be able to coordinate with the intricacies of the actual world, it attempts to resolve or lessen the problems that are now present in social networks. This method's innovation is that, in the first instance, an individual is chosen from a list of people who have become friends because they share interests. Next, the system applies the rules based on the selected individual, and it should attempt to extract the selected individual's friends from the interest groups. Identifies the current reason why the chosen individual is buddies with other people. Selecting a threshold for the HI-Counter to pick more influential individuals is the innovative part. The writers' contributions to this study are as follows.

- Improving the precision of recognizing significant friends.
- A more difficult time making meaningful friends.

This is how the rest of the article is structured. The backdrop of the research and its fundamental ideas are presented in the Section II. The proposed approach is presented in Section III and its application and evaluation in Section IV. In Section V, a summary of the findings is presented and outlines future projects.

II. Related Works

The cold problem in recommender systems has been addressed by introducing a hybrid method [21] to boost correlation. With this approach, the movement data set is used to gather data based on the user's demographics (gender, age, and nation). The idea behind demographic filtering is that individuals who share similar traits should be given equal importance. Consequently, the person correlation strategy is used for neighbourhood building in this recommender system, which combines the population filtering approach and the joint filtering approach. Neural networks have also been used to anticipate new ranks by integrating the outcomes of the joint filtering strategy and the demographic filtering method. This system has been evaluated using the correlation evaluation criterion, and it has been demonstrated that combining these techniques improves the proposer's accuracy. Additionally, the system can be expanded using a variety of techniques, including knowledge- and content-based approaches.

The combination of the clustering method and the weighted similarity assessment using the evolutionary algorithm is another hybrid way to build a recommender system based on joint filtering. As a result, data clusters are first formed, and the values from these clusters are then analyzed using a genetic algorithm. The method's assessment criterion is weighted similarity, and the objective is to identify similarities between values. It is desirable to cluster and derive similarity criteria. The fig below [22] depicts the architecture's basic layout. There is no requirement for a hybrid model to implement this suggested approach, which can be employed in any cooperative filtering system. Gupta and Patli present a hierarchical clustering approach [23], wherein clusters are generated from user data via the application of a hierarchical clustering algorithm. The rating of a specific item is predicted by a voting method. Therefore, electronic business applications are a better fit for this recommender system. There are two stages to the suggested hierarchical algorithm. An initial graph is produced in the first phase, which its division follows into several partitions and their grouping to form clusters in the second phase. Furthermore, two parameters are used to generate the clusters: the relative correlation parameter (RI) and the relative closedness parameter (RC). The definition of these parameters is as follows:

One of the most widely used hard clustering methods is the K-means clustering approach. The way this algorithm operates is by defining k centers at random for every cluster. The following stage connects every piece of data in the input data set to the closest center. The first phase is completed when there is no data to verify. Subsequently, the masses acquired from the preceding stage are used to recalculate the new centers. The data from each set and the closest center found are connected in the following phase. This loop is repeated until it is observed that K moves positions at each step until there are no more moves, at which point the algorithm terminates [24].

In study [25], a neural network and the K-Means clustering method are used to create a system for user behavior prediction. Each page in the session is given a weight based on how long it takes to see and how often it is viewed in order to display the impact of each page. In this system, user survey patterns are extracted using the K-Means clustering algorithm, and the best cluster is then determined online using a neural network.

A fuzzy technique known as HU-FCF is presented in study [26] as a solution to the cold start problem. It finds related users by using population data and the hard clustering method. Using this method, demographic data is collected, and a fuzzy similarity matrix between the new user and others is built by fuzzy computation. In the opposite way, provided scores are extracted from the data, and a matrix is constructed based on these scores. It generates the best recommendations for the customer but is slow due to the combination of these two matrices. As can be seen in Table I, an overview of previous works and their used methods is given.

TABLE I. OVERVIEW OF THE METHODS AND TECHNIQUES OF THE WORK DONE

Ref	Assessment Area	Interaction Information	Technique used
[15]	Repetitive pattern exploration	Not support	Algorithm for concatenation of paths without common edges
[19]	Repeated pattern search	Not support	Affectionate tree
[21]	Repetitive pattern exploration	Frequency	FSG algorithm
[24]	Repeated pattern search	Not support	FP_growth algorithm
[25]	Repetitive pattern exploration	Frequency	AGM algorithm
[27]	Repetitive pattern exploration	Frequency	Based on Apriori associative rules
[28]	No interactive exploration techniques	Not support	An independent colony model
[29]	No interactive exploration techniques	Not support	Through the number of messages exchanged
[30]	Repetitive pattern exploration	Not support	Algorithm for frequent weight exploration of item sets using IF-tree

III. PROPOSED METHOD

The creation of knowledge discovery algorithms has become the focus of research due to the abundance of data and the dearth of fresh, practical, and understandable knowledge. One data mining method for knowledge discovery is the examination of alternating patterns. In order to apply this approach to continuous data, which requires quantification of the data, it is intended to handle discrete data. The final results that are used in this research to locate important friends from the suggested B_mine algorithm may be affected by the loss of some data and the addition of fictitious data to the data space caused by this effort. Listed here are the reasons why the proposed method is suitable for solving the problems presented in the text:

Using the B-mine method: This method is suitable for analyzing continuous and discrete data, which is possible in social networks, due to its ability to extract intermittent patterns and discover repetitive patterns.

Combining two user similarity measures: Using coverage and confidence measures simultaneously to improve the accuracy of the recommender system can increase efficiency.

Analysis of past user behavior: By analyzing the behavior of users in the past, the system is able to provide the best recommendations for current users and use the patterns and preferences of their friends.

Use of social networks: by considering the structure of social networks and the changes that occur in them, the proposed method can provide the best recommendations and increase its accuracy.

These reasons show that the proposed method is suitable for solving these problems, considering the unique characteristics of social networks and the problems raised.

The suggested method's block diagram is displayed in Fig. 1.

A. Repeating Pattern

Finding fresh, legitimate, helpful, and intelligible patterns in the data is a scientific method for gaining knowledge from the database. The most crucial step in this process is data exploration, which pulls patterns out of the database using specialized algorithms. The process of retrieving knowledge and information from databases that may be hidden and potentially helpful is known as data mining. The increasing volume and diversity of data in today's world have made this problem crucial [27]. A collection of methods known as data mining enables one to go beyond standard data processing and aids in the discovery of information hidden in the volume of data. An important area of research in data mining is finding and extracting valuable information from a set of dependency relationships. With the growing use of large data banks and transaction warehouses, many researchers have recently focused on developing practical methods for extracting these relationships. Dependency rules are the implicit relationships that exist between data values in a database. Discovery or investigation of dependence laws is the process of locating dependency laws. Identifying the collection of repeated things is the most crucial step in investigating dependency rules.

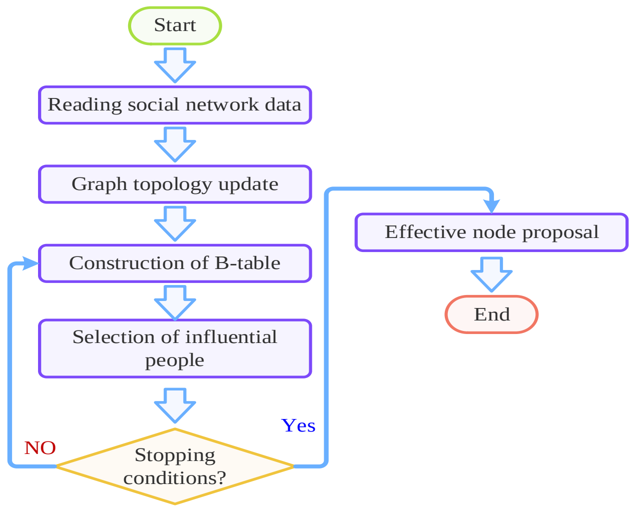

Fig. 1. Flowchart of the proposed method.

Let us consider the set I = {a, b, c, d, e}, which consists of five members. Let us consider a set D consisting of database records. Each record T within this set can be defined as a subset of elements, denoted as $T \subseteq I$. Every record is assigned a distinct numerical identifier known as TID. Recurring patterns are characterized by their frequent occurrence within a given data set [28], [30]. In the context of a store's purchase history database, milk and bread can be regarded as instances of repeating patterns due to their frequent co-buy occurrence. The set of elements that consists of k elements is denoted as a k-itemset. An illustration of an item can be seen in the set {computer, monitor}. The frequency of a collection of elements corresponds to the count of records that encompass said set of elements. A set of elements is considered to be repeated when the count of occurrences of that set of elements is equal to or exceeds the product of the minimal support threshold and the total number of records in the dataset D. A collection of elements that possesses the requisite number of transactions to accomplish a certain objective is referred to as a set of recurring elements.

B. *B-Mine Proposed Algorithm*

The technique of intermittent item set exploration is employed for the purpose of investigating information within transaction databases. In this particular approach, a B-table is utilized to depict the association between transactions, wherein the values of zero and one are present. A value of zero signifies the absence of a relationship between the items within a transaction. The non-existence of a value in a table signifies the absence of a relationship between the items in a transaction. Conversely, if the value of one is present in the table, it indicates the presence of a relationship between the items. In the context of this study, the value of one serves as a measure of efficient communication among friends.

1) *Frequent pattern extraction*: It consists of two modes of a specific friend and selecting a group of friends, which are explained in detail below:

a) The first mode is to select a specific friend: A person is chosen from the list of individuals who have bonded via shared interests. It then searches for friends in interest groups who are related to the selected person. Finally, the system determines the causal relationships between the chosen individual and other friends based on the chosen individual. There are three stages to this system.

The first step is to create the table that the system needs, which is made from all of the friends' databases. First, a user is chosen by the system, which then generates a table pertaining to the friends that are desired. In actuality, all users that have the aforementioned buddy are chosen. The system will filter out those who don't have the desired friend in their member groups by checking those individuals. The threshold limit is used to pick friends more accurately; it is thought to be equivalent to half the number of persons who have the desired friend. Phase two: During this stage, the group of buddies who are wanted is acquired [31]. Friends whose cumulative frequency exceeds the threshold are chosen based on how frequently each person is contacted overall. Any set that was fewer than the threshold (the total number of choices made by both friends at the same time) is eliminated. The sets of two members are formed from the friends from the previous stage. Each pair consists of the individual's chosen friend and one additional friend. Up until the largest set whose frequency is equal to or more than the threshold is produced, the previous step keeps creating larger sets of friends [34].

The latter stage: The collected collections are examined in this phase based on two criteria: the coverage relationship and the confidence relationship. The best guidelines are extracted from any collection of friendships that are confirmed by the two relations mentioned above. The next example, which displays the list of individuals who have become friends through common interests in Table II, serves to explain the procedure that has been discussed. Every member of list L_i shares a common interest. For instance, members of L_1, which includes Anna, Beato, Davy, Eva, and Fabio, are interested in social computing, whereas members of L_2, which includes Beato, Carlos, Eva, and Fabio, are interested in another area, like data collecting. In the following table, there are several transactions in it. This table displays the identities of the friends of each of the n=6 individuals in groups with various interests.

TABLE II. TABLE OF FRIENDS IN GROUPS WITH DIFFERENT INTERESTS

List of people (groups with different interests)	Friends list
L1	Ana,Beto,Davi,Eva,Faboi
L2	Beto,Carlos,Eva,Faboi
L3	Ana,Carlos,Davi,Faboi
L4	Beto,Carlos,Davi,Eva
L5	Ana,Davi,Faboi
L6	Ana,Beto,Carlos,Davi,Eva

TABLE III. RELATIONSHIP TABLE OF PEOPLE

		Friends list					
		A	***B***	***C***	***D***	***E***	***F***
list of people (groups)	L1	1	1	0	1	1	1
	L2	0	1	1	0	1	1
	L3	1	0	1	1	0	1
	L4	0	1	1	1	1	0
	L5	1	0	0	1	0	1
	L6	1	1	1	1	1	0

TABLE IV. PEOPLE WHO HAVE FRIENDS A

		Friends list					
		A	***B***	***C***	***D***	***E***	***F***
list of people (groups)	L1	1	1	0	1	1	1
	L3	1	0	1	1	0	1
	L4	1	0	0	1	0	1
	L6	1	1	1	1	1	0

The system requires the creation of a table, which is made from the database of all friends. An effective friend relationship indicator is shown in Table III, and a value of zero denotes that there is no relationship between the elements in a transaction.

For example, Ana is selected from the target group. Four people are friends with Anna(A), and the threshold limit is set to 2 (half the number of people who are friends with A). Only people who have friend A are examined (Table IV).

Initially, the frequency of each buddy is computed.

{A}=4, {B}=2, {C}=2, {D}=4, {E}=2, {F}=3

Due to the fact that the frequency of all friends exceeds the predetermined threshold, no friends are eliminated, resulting in the formation of set L1.

L1= {A, B, C, D, E, F}

The two-member set is formed with the element L1.

{A,B}=2	{B,C}=1	{C,D}=2	{D,E}=2	{E,F}=1
{A,C}=2	{B,D}=2	{C,E}=1	{D,F}=3	
{A,D}=4	{B,E}=2	{C,F}=1		
{A,E}=2	{B,F}=1			
{A,F}=3				

Any two-person groupings with a frequency below the specified threshold are eliminated.

L2= {{A, B}, {A, C}, {A, D}, {A, E}, {A, F}, {B, D}, {B, E}, {C, D}, {D, E}, {D, F}}

The L3 set is formed by utilizing the L2 set, which consists of three members.

{A,B,C} =1	{A,C,D} =2	{A,D,E} =2	{A,E,F} =1	{B,D,E} =2
{A,B,D} =2	{A,C,E} =1	{A,D,F} =3		{D,E,F} =1
{A,B,E} =2	{A,C,F} =1			
{A,B,F} =1				

Any group of three that has a frequency lower than the specified threshold is eliminated.

L3= {{A, B, D} {A, B, E} {A, C, D} {A, D, E} {A, D, F} {B, D, E}}

The L4 set is derived from the L3 set.

{A,B,D,E}= 2	{A,B,D,F}= 1	{A,D,E,F}= 1	{A,B,D,C}= 1

Any group consisting of four elements that have a frequency lower than the specified threshold is eliminated.

L4= {{A, B, D, E}}

The cumulative total of L2, L3, L4, and L1 is as follows:

Answer={A,B}, {A,C}, {A,D},{A,E}, {A,F},{B,D}, {B,E}, {C,D},{D,E},{D,F}, {A,B,D} {A,B,E} {A,C,D} {A,D,E} {A,D,F} {B,D,E},{A,B,D,E} }

In the subsequent stage, it is important to compute the amount of confidence and coverage for the ultimate set. In this computation, it is necessary to evaluate the validity and comprehensiveness of the rule A→X, ensuring that X represents the components within the solution set. In this particular instance, the level of certainty is 70%, as indicated in Table V.

a) The second mode is to select a group of friends: In this scenario, the individual does not designate a specific friend as the target but rather specifies a preferred group of friends, such as a sports group. The multi-friend system then generates suggestions for the individual based on the provided information. In the event that an individual selects a friend who has not been picked by any other individual, the system will propose the utilization of the second way of self-friendship. Hence, the operational procedure is outlined as follows: a) The individual selects the social circle. The system generates a table that is associated with the specified group. c) Generates a comprehensive set of rules based on the specified threshold. d) The individual is supplied with suggestions for laws that are deemed appropriate in terms of their extent of coverage and dependability.

TABLE V. ASSURANCE AND COVERAGE OF RULES

Collection	Law	confidence	cover	Condition
{A,B}	{A→B}	0.5	0.33	weak
{A,C}	{A→C}	0.5	0.33	weak
{A,D}	{A→D}	1	0.66	Strong
{A,E}	{A→E}	0.5	0.33	weak
{A,F}	{A→F}	0.75	0.5	Strong
{A,B,D}	{A→B,D}	0.5	0.33	weak
{A,B,E}	{A→B,E}	0.5	0.33	weak
{A,D,E}	{A→D,E}	0.5	0.33	weak
{A,C,D}	{A→C,D}	0.5	0.33	weak
{A,D,F}	{A→D,F}	0.75	0.66	Strong
{A,B,D,E}	{A→B,D,E}	0.5	0.33	weak

IV. Discussion and Evaluation

The MATLAB programming language was utilized for conducting simulations, with multiple stages being taken into account for evaluating each criterion. The dataset size progressively increased at each stage, and numerous simulations were executed to obtain the results for each stage. The average of these results was then considered as the final output. The proposed method has been successfully used in a portable system with the following specifications: a CPU working at a frequency of 2.53 GHz, a physical memory (RAM) capacity of 8 GB, the Windows 8 operating system, and the MATLAB software implementation tool. In order to conduct a comparative analysis, the present study has employed the methodologies employed by other researchers, as referenced in Section II, namely [18], [23].

A. Evaluation Criteria

The simulation results are compared using the following standards.

Operation time: How long does it take the system to process the request and return the desired outcome?

Number of dependence rules: Determine how many of the dataset's dependency rules are helpful and which are not.

Coverage definition: Given an item set x and a set I comprising all items, they say that the coverage of x in the database equals ℓ if and only if the item set x's number of occurrences in the database equals ℓ.

$$\text{Support(x)}=\ell \text{Support } (x \to y) = \frac{\text{support } (x \cup y)}{M} \quad (1)$$

Degree of confidence: Set I include all items, and sets of items x and y are assumed. The degree of certainty of the rule x=>y is equal to: x∩y=ø

$$\text{Conf}(x \to y) = \frac{\text{support } (x \cup y)}{\text{Support } (x)} \quad (2)$$

Accuracy criterion: The relationship that follows is used to calculate this criterion. In this sense, the number of data that are appropriately identified as true positive (TP) and the number of data that are incorrectly identified as false positive (FP) are related.

$$\text{Precision } = \frac{Tp}{TP+FP} \quad (3)$$

Call criterion: The number of data accurately identified as true positives (TP) and the number of data incorrectly identified as false negatives (FN) are the two variables in the relationship used for the calculation.

$$\text{Recall} = \frac{Tp}{TP+FN} \quad (4)$$

F1-Measure: Recall and precision are measured, and this relationship is used to calculate the measure:

$$F1 = \frac{2*RECALL*PRECISION}{RECALL+PRECISION} \quad (5)$$

B. Methods Compared

The publications [32], [33], which are discussed in the second section, have been utilized to compare the suggested methodology.

a) Execution Time: The length of time the program ran in this circumstance and the results extracted were assessed. The algorithm's quality increases with a reduced execution time. The simulation execution time is displayed in Fig. 2.

b) Support Criterion: The ten rules with the greatest Support values in each program output step were averaged to compute this measure, which is shown in Fig. 3.

The analysis of the simulation outcomes indicates that the proposed approach exhibits superior performance in extracting frequent rules based on the support criterion.

a) Confidence Criterion: In order to assess this criterion, the dataset grows in size at each execution cycle, and as Fig. 4 illustrates, 10 extracted rules are averaged at each stage.

The suggested approach performs better and extracts rules with a greater validity based on the results for the aforementioned criterion.

b) Recall Criterion: The recall criterion is a crucial factor in assessing the extracted rules. In order to compute this criterion, Fig. 5 displays the average values' output following each step.

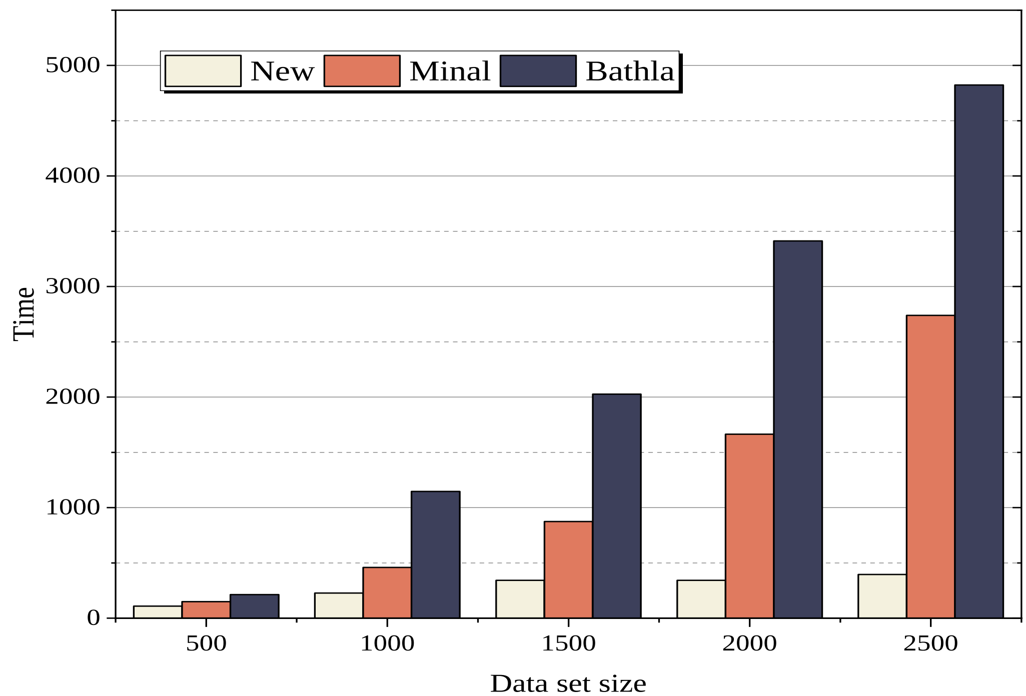

Fig. 2. Simulation duration.

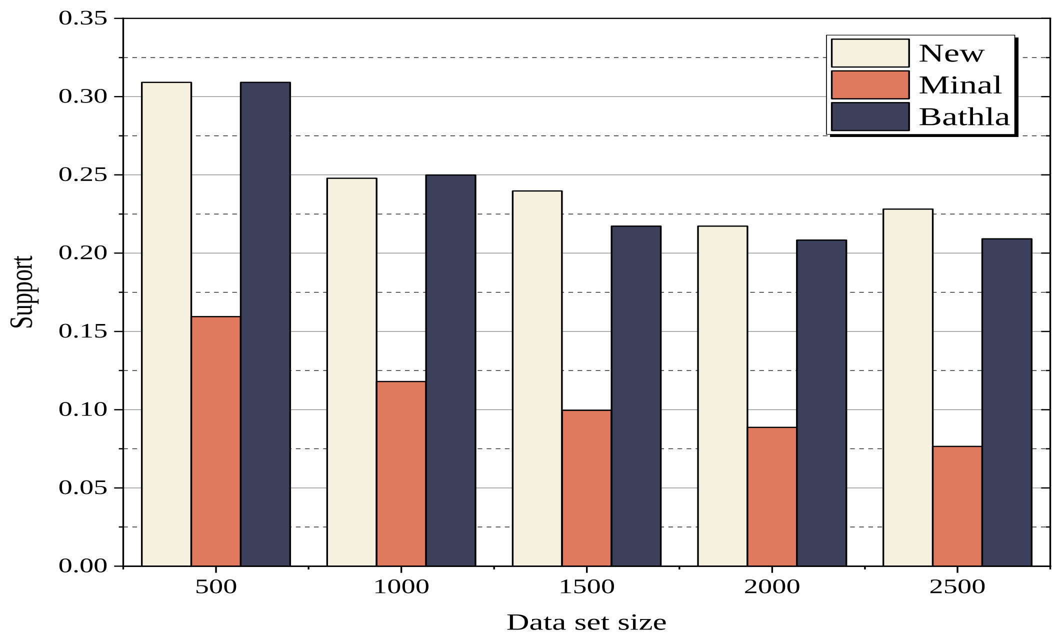

Fig. 3. Comparison of support criterion.

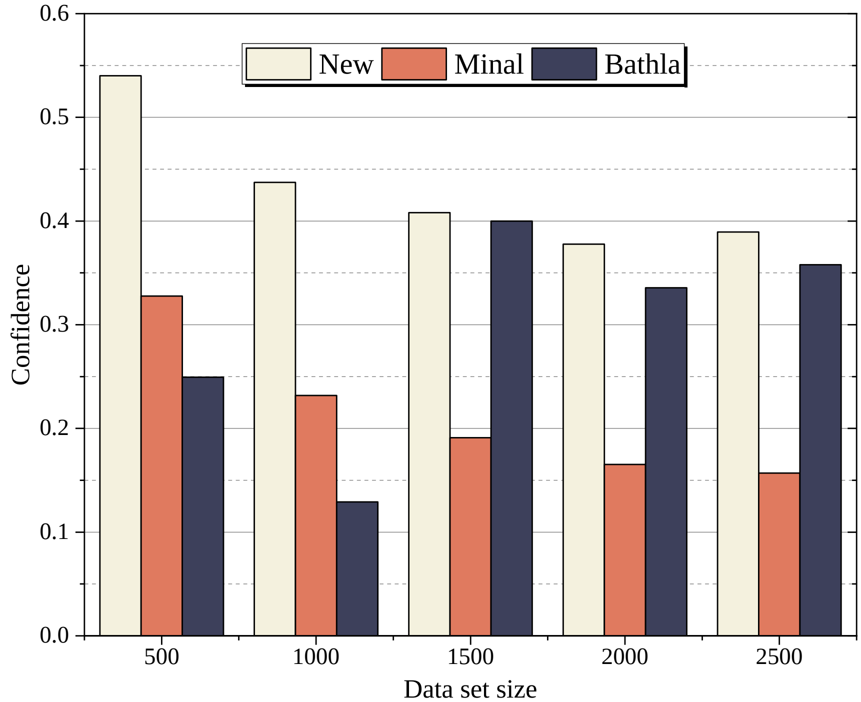

Fig. 4. Comparison of Confidence criteria.

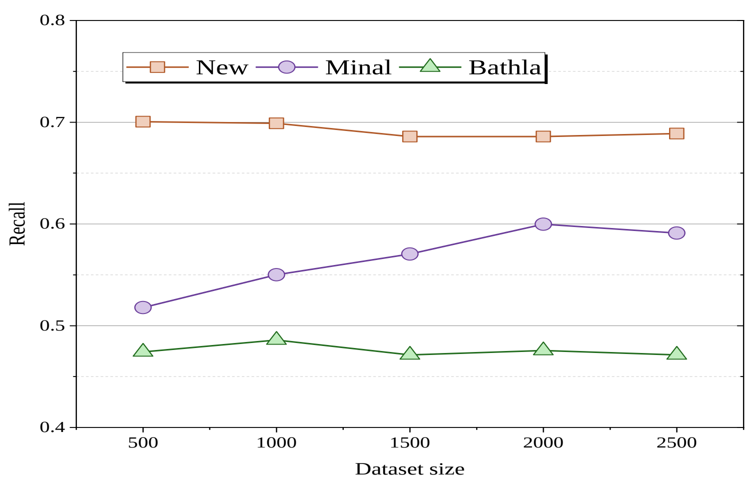

Fig. 5. Comparison of recall criteria.

c) Precision Measure: Fig. 6 displays the outcome of the comparison of the aforementioned criteria. The average values of the aforementioned criteria are determined at each phase; the larger the value, the more trustworthy the derived law is.

The result shows that the proposed method has performed better than the previous two methods for the above criterion.

a) F1-Measure: This criterion is calculated according to the two criteria, recall and precision, according to the following relationship: The result of this criterion for simulation and comparison can be seen in Fig. 7.

$$\mathrm{F-MESURE} = \frac{2*\mathrm{RECALL}*\mathrm{PRECISION}}{\mathrm{RECALL}+\mathrm{PRECISION}} \quad (6)$$

The results obtained from the simulation shown in the graph show that the proposed method has a 12% improvement.

2) First experiment (Dolphin social network): The goal of this experiment was to determine which social network within dolphin had the greatest influence. To that end, numerical figs in the form of tables and graphs were presented based on the three criteria of sensitivity, accuracy, and F1 score. In this study, time intervals ranging from one to five months are taken into consideration, with a penetration range of k between three and five. Table VI and Fig. 8 present the results of the sensitivity criterion calculation for the proposed method under various incursion scenarios and time periods. It is evident that the suggested method's sensitivity rises as penetration does, and at k=5, the genetic optimization algorithm's sensitivity reaches 100%.

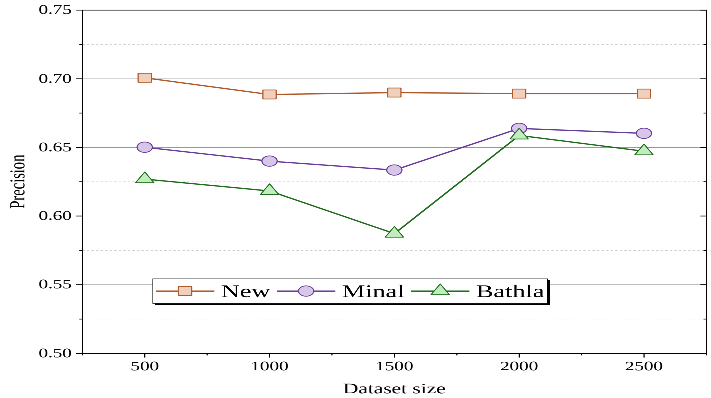

Fig. 6. Comparison of precision criteria.

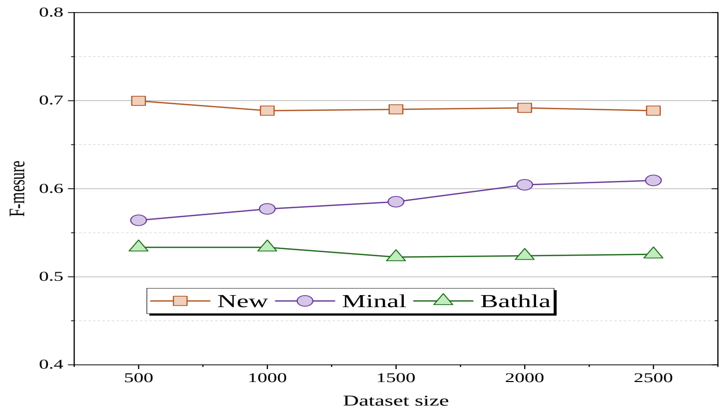

Fig. 7. F1-Measure comparison.

TABLE VI. SENSITIVITY OF THE PROPOSED METHOD AT DIFFERENT TIMES

		k		
		3	***4***	***5***
t	1	31.87	16.22	3.3
	2	26.71	11.14	3.3
	3	28.88	16.24	1.1
	4	35.70	8.10	2.2
	5	31.89	17.24	1.1

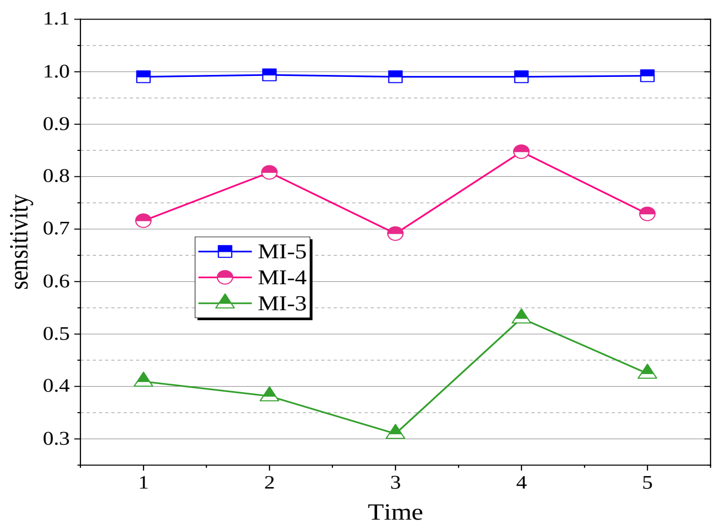

Fig. 8. The sensitivity of the proposed method in finding the largest penetration.

Table VII and Fig. 9 demonstrate the suggested method's validity throughout varying time periods (months). It is evident that when penetration rises, the accuracy of the suggested approach rises as well, meeting the sensitivity condition in the process. At k=5, the genetic optimization algorithm achieves a sensitivity of 100%.

In addition to these two factors, each algorithm's capacity to identify influence inside a social network is assessed using the evaluation of the suggested algorithm with varying k in terms of F1 value. The F1 value for the suggested approach of detecting penetration is displayed in Fig. 10. It is evident that the suggested genetic optimization approach can locate nodes in the dolphin social network with bigger K values in particular to have an effective influence.

TABLE VII. ACCURACY OF THE PROPOSED METHOD AT DIFFERENT TIMES

		k		
		3	***4***	***5***
t	1	31.47	16.16	3.3
	2	26.41	11.11	3.3
	3	28.46	16.17	1.1
	4	35.46	8.10	2.2
	5	31.44	18.17	1.1

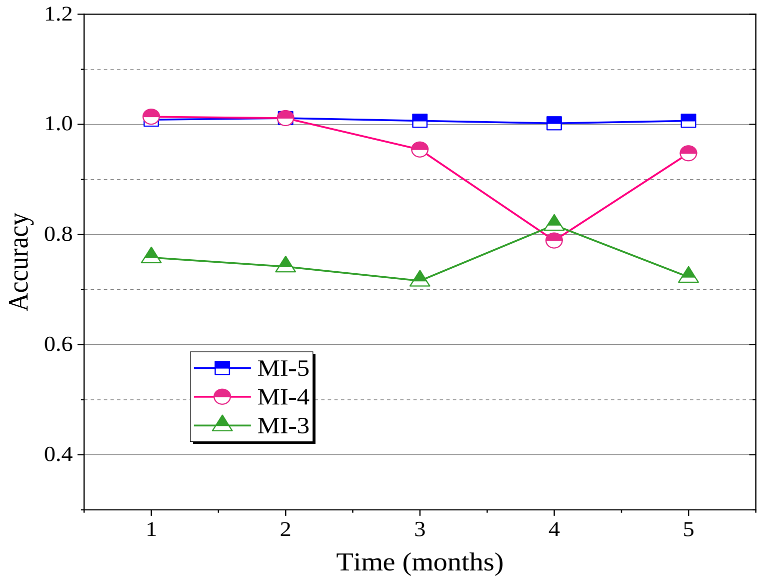

Fig. 9. The accuracy of the proposed method in finding the largest penetration.

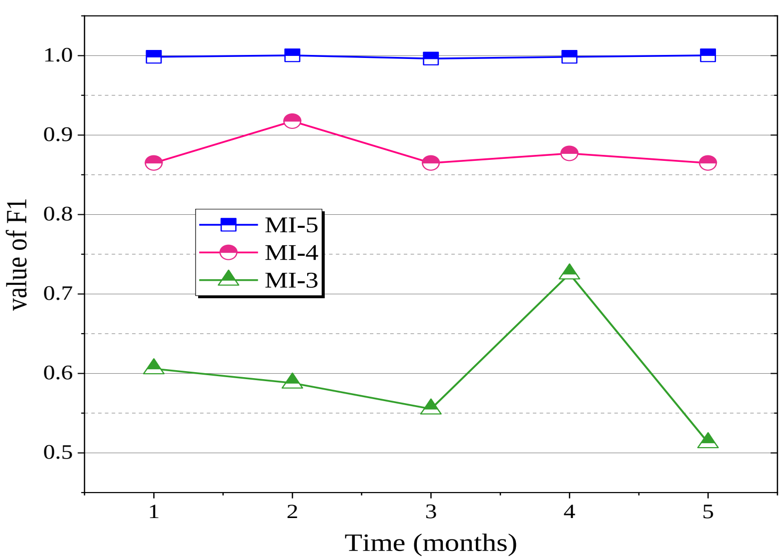

Fig. 10. F1 value of the proposed method in finding penetration.

3) Second test (standard graphs (DIMACS): The outcomes of the suggested strategy are displayed on 17 common graphs in Table VIII. A graph's vertex count can range from 64 to 1024, whereas its edge count can range from 1200 to 518656.

The table indicates that the suggested algorithm for Johnson, Hammer, Keller, and C-fat graphs consistently locates the largest influence. Additionally, improved outcomes for Sanr graphs have been attained. Of the 17 graphs that were examined, 15 had the most impact. In the majority of graphs, the average penetration size yields the best results. The suggested strategy performs better in assessing and forecasting effective friends, and it can be applied in appropriate situations, according to simulation and comparison results. Fig. 11 illustrates the suggested method's level of improvement in relation to each comparative criterion.

TABLE VIII. RESULTS OF THE PROPOSED METHOD ON DIFFERENT GRAPHS

Row	Graph name	number of vertices	number of edges	stock count	method output
1	c-fat200-1	200	1534	12	12
2	c-fat200-2	200	3235	24	24
3	c-fat200-5	200	8473	58	58
4	c-fat500-1	500	4453	14	14
5	c-fat500-2	500	9139	26	26
6	c-fat500-5	500	23191	64	64
7	Johnson8-4-4	70	1200	14	14
8	Johnson16-2-4	120	5460	8	8
9	Johnson32-2-4	496	107880	16	16
10	Keller4	171	9435	11	11
11	Keller5	776	225990	27	27
12	hamming6-2	64	1824	32	32
13	hamming8-2	256	31616	128	128
14	hamming10-2	1024	518656	512	512
15	Sanr200-0.7-1	200	13868	18	17
16	Sanr400-0.5	400	39984	13	13
17	San1000	1000	250500	15	14

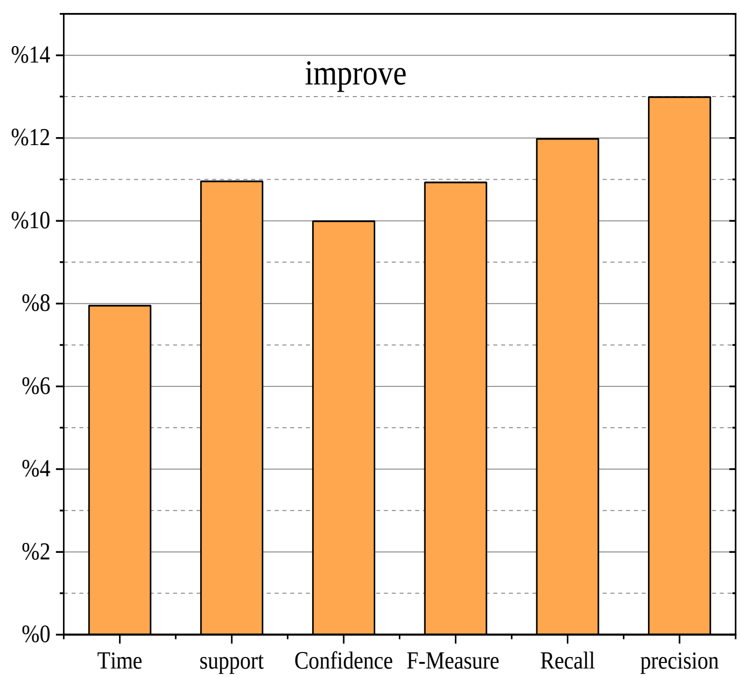

Fig. 11. The improvement rate of the proposed method.

V. Conclusions

Societal network users may differ in their behavior or preferences because of their differing views on societal problems (religion, politics, etc.) or personal interests (family, friends, shopping, health, etc.). Data mining techniques are one of the most effective methods for obtaining behavioral patterns. One artificial intelligence technology that has been created to examine massive amounts of data and find significant patterns and rules is data mining. Through the application of artificial intelligence and statistical approaches, data mining techniques are able to extract user behavioral patterns and gather a wealth of information about them. Many marketing decisions can be supported by data mining. The identification of recurring items and associative rules are the two most crucial data mining tasks. You can identify the relationships and dependencies among the data in a database by applying associative rules. These days, recommender systems are widely used in both industrial markets and educational settings. These days, when the Internet is growing at an exponential rate and the amount of data is huge, there is a need for systems that can suggest the most relevant content to users and others who share common interests. Therefore, systems that perform this function are referred to as recommender systems. In order to find the most relevant items—such as data, information, and friends recommender systems employ a variety of algorithms and specialized techniques. They then propose the buddy who most closely matches the user's preferences. This article employs a social computing approach that encompasses social behavior on media, modeling and analysis techniques, social network categorization, and the identification of influential friends.

In addition, a new method is introduced to combine data mining techniques to extract influential friends from a database of historical user behavior. It is evaluated and recommended to the present user based on the behavior and resemblance of friends in relation to the interactions and interests of individuals in various groupings. The results of the simulation indicate that the suggested approach performs acceptably.

One of the most crucial knowledge extractions from data that can result in significant time and cost savings is the extraction of association rules of dependence. In the course of this work, by presenting a method that allows, in addition to extracting dependency rules, to predict dependency rules in the future. This means that, based on the next set of data, the rules associated with it are most accurately predicted by the recorded data and the extracted rules. The limitations of the existing method for the possible problem in this research are:

Sensitivity to the amount of data: this method may have limitations against the large amount of data that exists in social networks. When the data becomes very large, the performance and efficiency of the method may decrease.

Data reliability: Due to the dependence of this method on the input data and its accuracy, if the data contains noise or incorrect information, the accuracy and efficiency of the method may be affected.

Computational complexity: Using complex algorithms to analyze data and extract patterns may increase computing time and increase computing costs.

Dependence on diagnostic criteria: This method relies on criteria to detect patterns and recommendations that may lead to deviations or errors in the final recommendations.

These limitations show that when dealing with large data, the accuracy of the analysis and recommendations provided by the method may decrease and these limitations should be carefully faced.

The following are, in brief, the future objectives: a) Obtaining frequently used rules from users based on their profiles. b) Offer a profit forecasting model. c) Offering a model for analyzing user behavior after making a purchase.

Acknowledgments

This work was supported by 2022 Sports Science and Technology Research Project of Hebei Provincial Sports Bureau,“Research on Sports Communication Innovation under the Background of Beijing Tianjin Hebei Coordinated Development”(No.2023CY09), and the First Graduate Education and Teaching Reform Research Project in Hebei Province,“Research on Graduate Teaching Scenarios from the Dimensions of Space and Information”(No.YJG2023110).

References

[1] M. Scholz, V. Dorner, G. Schryen, and A. Benlian, “A configuration-based recommender system for supporting e-commerce decisions,” Eur J Oper Res, vol. 259, no. 1, pp. 205–215, 2017.

[2] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: a survey,” Decis Support Syst, vol. 74, pp. 12–32, 2015.

[3] A. L. V. Pereira and E. R. Hruschka, “Simultaneous co-clustering and learning to address the cold start problem in recommender systems,” Knowl Based Syst, vol. 82, pp. 11–19, 2015.

[4] Esmaeili, N., & Bamdad Soofi, J. (2022). Expounding the knowledge conversion processes within the occupational safety and health management system (OSH-MS) using concept mapping. International Journal of Occupational Safety and Ergonomics, 28(2), 1000-1015.

[5] A. Gupta, H. Shivhare, and S. Sharma, “Recommender system using fuzzy c-means clustering and genetic algorithm based weighted similarity measure,” in 2015 International Conference on Computer, Communication and Control (IC4), IEEE, 2015, pp. 1–8.

[6] U. Gupta and N. Patil, “Recommender system based on hierarchical clustering algorithm chameleon,” in 2015 IEEE international advance computing conference (IACC), IEEE, 2015, pp. 1006–1010.

[7] Mokhlesi Ghanevati, D., Khorami, E., Boukani, B., & Trik, M. (2020). Improve replica placement in content distribution networks with hybrid technique. Journal of Advances in Computer Research, 11(1), 87-99.

[8] Hosseini, A. (2023). Predictive machine learning model for the future trend of energy consumption in fully electricity homes considering occupancy status of the building (Doctoral dissertation, Wichita State University).

[9] Khezri, E., Zeinali, E., & Sargolzaey, H. (2023). SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks. Plos one, 18(4), e0282031.

[10] M. Aivazoglou et al., “A fine-grained social network recommender system,” Soc Netw Anal Min, vol. 10, pp. 1–18, 2020.

[11] O. Stitini, S. Kaloun, and O. Bencharef, “Towards the detection of fake news on social networks contributing to the improvement of trust and transparency in recommendation systems: trends and challenges,” Information, vol. 13, no. 3, p. 128, 2022.

[12] Khezri, E., Yahya, R. O., Hassanzadeh, H., Mohaidat, M., Ahmadi, S., & Trik, M. (2024). DLJSF: Data-Locality Aware Job Scheduling IoT tasks in fog-cloud computing environments. Results in Engineering, 21, 101780.

[13] Khezri, E., & Zeinali, E. (2021). A review on highway routing protocols in vehicular ad hoc networks. SN Computer Science, 2(2), 71.

[14] R. Chen, Q. Hua, Y.-S. Chang, B. Wang, L. Zhang, and X. Kong, "A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks," IEEE Access, vol. 6, pp. 64301–64320, 2018.

[15] A. Bin Suhaim and J. Berri, "Context-aware recommender systems for social networks: review, challenges and opportunities," IEEE Access, vol. 9, pp. 57440–57463, 2021.

[16] Khosravi, M., Trik, M., & Ansari, A. (2024). Diagnosis and classification of disturbances in the power distribution network by phasor measurement unit based on fuzzy intelligent system. The Journal of Engineering, 2024(1), e12322.

[17] Z. Wang, Z. Jin, Z. Yang, W. Zhao, and M. Trik, "Increasing efficiency for routing in internet of things using binary gray wolf optimization and fuzzy logic," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 9, p. 101732, 2023.

[18] Wang, G., Wu, J., & Trik, M. (2023). A novel approach to reduce video traffic based on understanding user demand and D2D communication in 5G networks. IETE Journal of Research, 1-17.

[19] Liao, Y., Tang, Z., Gao, K., & Trik, M. (2024). Optimization of resources in intelligent electronic health systems based on Internet of Things to predict heart diseases via artificial neural network. Heliyon.

[20] Li, Y., Wang, H., & Trik, M. (2024). Design and simulation of a new current mirror circuit with low power consumption and high performance and output impedance. Analog Integrated Circuits and Signal Processing, 1-13.

[21] M. Samiei, A. Hassani, S. Sarspy, I. E. Komari, M. Trik, and F. Hassanpour, "Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare," J Cancer Res Clin Oncol, pp. 1–15, 2023.

[22] J. Sun, Y. Zhang, and M. Trik, "PBPHS: a profile-based predictive handover strategy for 5G networks," Cybern Syst, pp. 1–22, 2022.

[23] Saidabad, M. Y., Hassanzadeh, H., Ebrahimi, S. H. S., Khezri, E., Rahimi, M. R., & Trik, M. (2024). An efficient approach for multi-label classification based on Advanced Kernel-Based Learning System. Intelligent Systems with Applications, 21, 200332.

[24] M. Trik, A. M. N. G. Molk, F. Ghasemi, and P. Pouryeganeh, "A hybrid selection strategy based on traffic analysis for improving performance in networks on chip," J Sens, vol. 2022, 2022.

[25] Xiao, L., Cao, Y., Gai, Y., Khezri, E., Liu, J., & Yang, M. (2023). Recognizing sports activities from video frames using deformable convolution and adaptive multiscale features. Journal of Cloud Computing, 12(1), 167.

[26] Cao, C., Wang, J., Kwok, D., Cui, F., Zhang, Z., Zhao, D., ... & Zou, Q. (2022). webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic acids research, 50(D1), D1123-D1130.

[27] Zhang, H., Zou, Q., Ju, Y., Song, C., & Chen, D. (2022). Distance-based support vector machine to predict DNA N6-methyladenine modification. Current Bioinformatics, 17(5), 473-482.

[28] Ding, X., Yao, R., & Khezri, E. (2023). An efficient algorithm for optimal route node sensing in smart tourism Urban traffic based on priority constraints. Wireless Networks, 1-18.

[29] Fakhri, P. S., Asghari, O., Sarspy, S., Marand, M. B., Moshaver, P., & Trik, M. (2023). A fuzzy decision-making system for video tracking with multiple objects in non-stationary conditions. Heliyon, 9(11).

[30] M. Trik, H. Akhavan, A. M. Bidgoli, A. M. N. G. Molk, H. Vashani, and S. P. Mozaffari, "A new adaptive selection strategy for reducing latency in networks on chip," Integration, vol. 89, pp. 9–24, 2023.

[31] Zhu, J., Hu, C., Khezri, E., & Ghazali, M. M. M. (2024). Edge intelligence-assisted animation design with large models: a survey. Journal of Cloud Computing, 13(1), 48.

[32] Zhang, L., Hu, S., Trik, M., Liang, S., & Li, D. (2024). M2M communication performance for a noisy channel based on latency-aware source-based LTE network measurements. Alexandria Engineering Journal, 99, 47-63.

[33] Asghari, A., Zoraghchian, A. A., & Trik, M. (2014). Presentation of an algorithm configuration for network-on-chip architecture with reconfiguration ability. International Journal of Electronics Communication and Computer Engineering (IJECCE), 5(5), 124-136.

[34] Y.-L. Lee, T. Zhou, K. Yang, Y. Du, and L. Pan, "Personalized recommender systems based on social relationships and historical behaviors," Appl Math Comput, vol. 437, p. 127549, 2023.