Journal of Ambient Intelligence and Humanized Computing (2023) 14:4157–4174
https://doi.org/10.1007/s12652-022-04507-3

ORIGINAL RESEARCH

Enhanced link prediction using sentiment attribute and community detection

Debadatta Naik[1] · **Dharavath Ramesh**[1] · **Naveen Babu Gorojanam**[1]

Received: 27 October 2020 / Accepted: 1 September 2022 / Published online: 26 December 2022

Abstract
In social network analysis, link prediction is an important area where the researchers can find the missing links and the future links possible among the users. Often, link prediction is made by analyzing the social linkage of the users in the given networks, i.e., the Topological structure of the networks. However, this approach leads to inconsistencies when researchers want to emphasize topics on which users have mainly engaged their selves in discussions. Mainly, this approach predicts future links based on available network structures without considering the topics on which the users are participating. This can be enhanced by incorporating the sentiment attributes and the community structure of the users in the network. In this paper, we propose an algorithm that incorporates the sentiment attribute of users and community structures along with the topological features. To evaluate the same, we have crawled the tweets of various countries concerning COVID-19 from Twitter. Experimental results show that users exhibiting the same emotion and belonging to the same community will influence other users to connect, thereby improving the performance of the link prediction.

Keywords Link prediction · Sentiment analysis · SVM · Community detection

1 Introduction

Social Network is one of the most popular communication tools used for communication and sharing opinions through the Internet. Some of the most popular Micro-blogging sites on the internet are Twitter, Tencent Weibo, Whatsapp, Facebook, where millions of users interact with each other worldwide. The users often share information in the form of insights, knowledge, emotions, opinions, and experiences using the effective and open nature of these micro-blogging sites. Posting a short message called Tweet is the main activity in the micro-blogging sites (Wu et al. 2012). The users of each site follow each other to form a network structure, i.e., the social network. Each user can view the tweets that are posted by the users he or she follows or vice versa. This implies that the underlying network structure can experience rapid changes over a while, i.e., change in nodes and links (Valverde-Rebaza and de Andrade Lopes 2012). This occurs due to the change in the following and follower structures.

Link prediction has many applications such as; Friend recommendation, Movie recommendation, Identifying missing or hidden criminals in a terrorist network, Author collaboration, Knowledge graph completion, Metabolic network reconstruction, etc. Conventional link prediction is made by studying the structure of the network. Various features like similarity-based features, edge-based features, node-based features, etc., are studied to predict the missing links in the networks (Lü and Zhou 2011). Some more algorithms like shortest path finding algorithms have been used to predict the most probable links based on the shortest paths formed in the future. All these methods focused only on the topological structure of the network but ignored the underlying semantic features of the network. As a result, maximum links were predicted between the users from different topics of discussion. Whereas, very less number of users from the same topics were get linked. This problem occurred due to the high emphasis given to topological structures only. Since the link prediction is being made on a social network,

✉ Dharavath Ramesh
drramesh@iitism.ac.in

Debadatta Naik
deba.uce03@gmail.com

Naveen Babu Gorojanam
naveenbabugorojanam@gmail.com

[1] Department of Computer Science and Engineering, Indian Institute of Technology (ISM), Dhanbad 826004, India

studying the behavior and future links among the users is an important task.

Since incorporating semantic values of the users in the network is an important topic for link prediction, various methods have been developed which used the sentiment scores of users to predict the results. The sentiment values are calculated for the unconnected node pair. The pair which produces the highest score is marked to have the highest probability for a link in the future. But the disadvantage of such prediction is that it does not incorporate the network structure. For example, a user is residing in *America* and another user is residing in *Singapore*. The user in *America* might like *Burgers* and the user in *Singapore* might like *pizza*. Since the sentiment values of both the users matched, predicting a link between them is incorrect. In the network structure, the users might not have any mutual friends.

So, the sentiment values must be matched between the users coming under the same topic while predicting the links based on sentiment similarity. Also, the users belonging to the same community are known to exhibit many similar properties. Therefore, a link between users of the same community is much higher than users belonging to different communities. For example there are three users named *Sam*, *Smith*, and *Jane*. *Sam* knows both *Smith* and *Jane* from work i.e., from same community. Consider another person *Denver* knows *Smith* from work and *Jane* from gym i.e., different communities. The probability that *Smith* and *Jane* will be connected is higher by *Sam*, who knows them from same community rather than *Denver* who knows them from different communities. Therefore considering the community information has higher importance in link prediction.

In this work, we consider three major topological properties of the social network. Along with the topological features, we also include the sentiment values of the users in the network and the community information of the users in the network. The sentiment values are analyzed based on polarity and similarity to having a clear knowledge of the sentiment values of the users. The community information has been extracted based on the topics used by the users in discussion rather than the conventional community detection algorithms. This is because the conventional community detection algorithms perform community detection based on the topological structure of the network. Whereas, in this work, we want to maximize the semantic probability using community information. To achieve that, we consider the topics and distributed the users into various topic-based communities. We combined all the above values and produced an optimized link prediction. The detailed architecture of link prediction using sentiment analysis is shown in Fig. 1. The major contributions of this paper is summarized as follows:

- In this work, an algorithm that incorporates heterogeneous information such as; sentiment attributes, topological features, and community structures are considered to predict the optimized links in the network structures.
- Tweet data related to COVID-19 topics and other hot topics are crawled from various countries.
- The most popular classifier called Support Vector Machine (SVM) predicts the presence or absence of links between the users inside a community.
- The proposed algorithm is compared with the J48 algorithm, and it is observed that the proposed algorithm outperforms in terms of Precision, Recall, and F-Measure metrics.

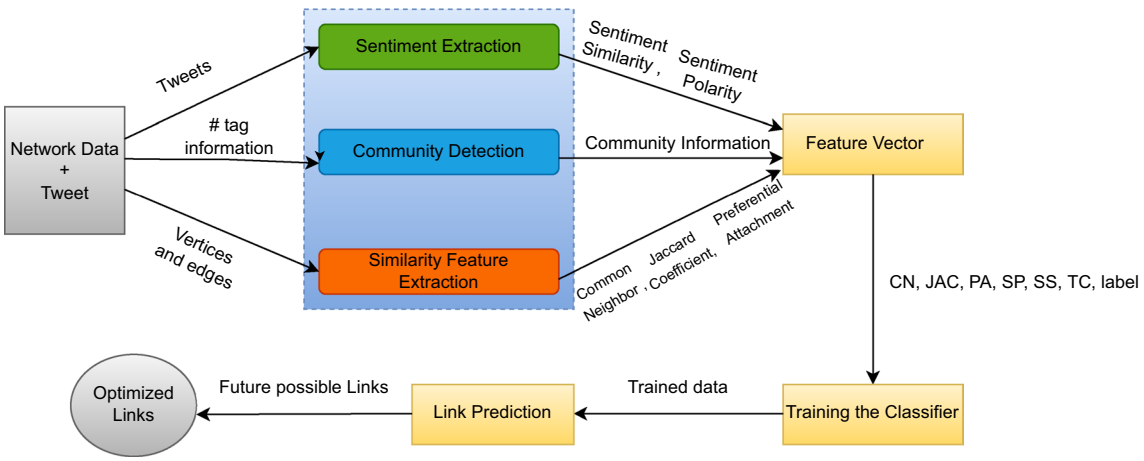

Fig. 1 Architecture of multi-feature Link Prediction

The rest of the paper is structured as follows. Section 2 contains the works of various works contributed earlier in the field of link prediction employing various techniques like sentiment detection, automata based enhancing etc., Section 3 contains the proposed methodology for enhancing the Link prediction using sentiment detection and community detection. Section 4 contains the implementation of the proposed method on a sample dataset. Section 5 depicts the experimental results of the proposed approach. The dataset used for experiment contains the tweets related to COVID-19 from various countries. The threats to validity are mentioned in Section 6. Section 7 contains the conclusions and scope of future work.

2 Related work

The study of link prediction in social networks has drawn a lot of attention in the recent years by the community of data mining. Link prediction is the process of predicting the missing or future links among the nodes of the network (Lü and Zhou 2011; Liben-Nowell and Kleinberg 2007). The research on link prediction can help in understanding the evolution of the complex networks. Practically, the Link prediction is used for suggesting new friends as a part of improving the user experience, detecting the crime members, and many more. In recent developments of the social networks, link prediction is even used to suggest new post/tweet recommendations based on previous engagements.

To improve the results produced by link prediction, many algorithms have been proposed over some time. Out of these algorithms, the similarity-based algorithm is the simplest one. Considering the local structure of the network, there are some feature-based tools such as Common Neighbours, Jaccard Coefficient, Adamic Adar, Resource Allocation, and Preferential Attachment (Valverde-Rebaza and de Andrade Lopes 2012; Yu and Wang 2014). Additionally, the semantic similarity of the users in the network is considered to improve the performance of the prediction.

Studies show that the link prediction that was generated only using the topological features of the network or the semantic information of the users has ignored the sentiment information of the users. As a result, the accuracy of link prediction is not good. But in social sciences, it has been proved that considering the emotion or the sentiment plays an important role in social life (Hu et al. 2013). It has been pointed out that emotion or sentiment plays significant role in the interpersonal communication (Tang 1996). In the recent times, author Thelwall found out that emotion homophily is important in social networks (Thelwall 2010). In the paper (Tan et al. 2011), Tan et al., found out that probability of the users connected to each other share the same sentiment than random users in the network. Lu and Zhou proposed that the network structure and the sentiment network are integrated to perform link prediction (Wang et al. 2018). Though this approach considered the sentiment values, the community structure has not been considered. Instead, the social network has been considered for link prediction. Using community information for link prediction has shown to improve the accuracy of the predicted links. Authors Soundarajan and Hopcroft showed how community information is used along with the similarity-based link prediction results (Soundarajan and Hopcroft 2012). The results show that the quality of links predicted has improved a lot. But the drawback of this classification is not able to consider the emotions or the sentiment values of the users in the network. In the paper (Naik et al. 2020), authors Naik et al. shown that how the users belonging to the same community are influenced under sentiment attributes. In this work, Ekman's Emotional model is employed to detect the emotion of individual users. The main drawback associated with their work is the classification of emotion into six different types, which creates less probability of using this sentiment model for link prediction. Authors Sa et al. proposed a methodology, where supervised link prediction has been done (De Sá and Prudêncio 2011). In their work, the community information is first trained through supervised learning and used to improve the link prediction. The weighted network has been considered to improve the quality of link prediction, but the sentiment values of the users have not been considered. Wang et al. used a hyperbolic mapping where the community information of the users is mapped with the network structure for link prediction (Wang et al. 2016). In the paper (Bastami et al. 2019), a gravitation-based link prediction has been proposed by the Bastami et al., where the community information is trained to an unsupervised learning approach and the same has been used for link prediction. The problem with both of these classifier methods is not considering the sentiment features of the users present in the network. Mallek et al. proposed the link prediction two-phase techniques using the uncertain framework of belief function theory (Mallek et al. 2019). In the first phase, a graph-based model is used to handle the uncertainty in links, followed by the second phase, where a novel method is used to discover the new links using the belief function tool. The generalization of Bayesian theory is taken into consideration in the belief function tool, which allows making decisions by combining pieces of information (neighborhood, common group information) from different sources and enables the quantification, management, and representation of evidence. The new connections are discovered using these neighborhood and common group information. The method outperforms the traditional approaches. However, the method considers only the local and group information. A new link prediction algorithm proposed by Tuan et al. is based on similarity and semi-supervised fuzzy clustering (Tuan et al. 2019). It

determines the membership degree and other measures of NS (Fuzzy neutrosophic set) by introducing an extension of measures in NS. The strength of connections between pair of nodes is calculated using NS. The fuzzy method is very simple, easy to implement, and straightforward. However, in this method, only topological features are considered. A learn function that maps the subgraph patterns to link existence is addressed by the authors Zhang and Chen to identify the future connections between the nodes (Zhang and Chen 2018). This is an automatic heuristic learning approach in which decaying theory is incorporated to learn suitable heuristics from a network. Heuristics are gathered and approximated from the local subgraphs in the first phase. Following this, GNN is applied to learn heuristics from the subgraphs. The proposed method shows consistent and unprecedented performance in diverse problems. The time required to learn the graph structure is the main disadvantage of this method. Authors Jalili et al. proposed a meta-path based method to predict the future connections between pair of nodes in the multiplex networks (Jalili et al. 2017)

Multiplex networks are a sort of multi layer network in which nodes of the same type are connected by various types of links. To predict a missing inter or intralayer link, effective algorithms which are based on connection information from the target layer and other layers are required. The authors considered the two social networks (Twitter and Foursquare) and studied the impact of a variety of nodal features on the Foursquare link prediction problem. To perform this task, meta-path-based features are considered, and high accuracy is achieved. However, it is a very difficult task to obtain the multiplex network for all the users. A scalable matrix completion-based method is proposed by author Zhou et al. to predict the future links in large networks (Zhou and Kwan 2018). The algorithm produces the link prediction results by executing three different steps iteratively. The matrix value obtained after executing the last iteration is the desired output of the link prediction problems. Experiments were conducted on two different datasets and found better performance than state-of-the-art algorithms. The main disadvantage of this method is that it suffers from a sparsity problem. Inspired by Newton's gravitational law, authors Wahid-Ul-Ashraf et al. proposed a link prediction method in which degree centrality and the shortest distance between two nodes is considered as mass and distance respectively (Wahid-Ul-Ashraf et al. 2017). The proposed algorithm is compared with seven other methods, and it outperforms the other methods in terms of performance. The main disadvantage of this method is that it considers the only local structure of the node. To detect the future links between pairs of nodes in the networks, Samad et al. (Samad et al. 2019) proposed a method in which the SAM similarity score is computed considering each node separately. It is observed that higher degree nodes are more likely to connect with lower degree nodes. SAM similarity between any two nodes (u, v) is the average of SAM similarity between u to v and SAM similarity between v to u. The method outperforms the other traditional algorithms. However, it considers only the topological information of the networks.Authors Wang et al. proposed a novel, flexible method called SHINE to predict the unobserved sentiment links using multiple deep auto encoders (Wang et al. 2018). First, the entity-level sentiment extraction method is used to obtain the labeled heterogeneous sentiment information (profile knowledge, social relation, and sentiment relation). Then, the SHINE framework is used to extract users' latent representations and the sign of unobserved sentiment linkages. More accuracy is achieved due to consideration of sentiments, social relationships, and user profiles. However, the complexity is very high due to the deep learning approach. A comparative study of earlier methods is depicted in Table 1.

Author Marinho et al. proposed a method to find the authorship attribution task using the motifs concept (Marinho et al. 2016). The network is formed by representing words as a node and syntactic relation between co-occurrence words as an edge. Classification features are obtained by extracting absolute frequencies of all the thirteen motifs with three nodes from co-occurrence words. Then the ability of motifs is quantified by applying a machine learning classifier (SVM). To capture the short texts produced in neuro-psychological evaluation, author Santos et al. proposed a model that transforms the transcripts into complex networks and enhances them with word embedding (CNE) (Santos et al. 2017). While performing word embedding, the word is treated as a node, and links are generated between any pair of words if the cosine similarity of that pair exceeds the threshold value. Mild Cognitive Impairment (MCI) is automatically identified in the transcript network using well-known classifiers. Experimental result shows that CNE achieved the highest accuracy. Based on the author's collaborative patterns and topological properties of collaborative networks, Amancio et al. developed a hybrid strategy (Amancio et al. 2015). Their experimental results are promising. However the main limitation is that some information are not always found in DBLP. A new link prediction method is proposed by the author Coskun et al. using fast LRC-Katz algorithm (Coşkun et al. 2021). The LRC-Katz algorithm is based on indexing and low-rank correlation. The locality of the network is exploited in this method. This method outperformed the other link prediction methods based on Vanilla and truncated Katz. Amancio et al. proposed a complex network-based hybrid method to classify the texts (Amancio et al. 2012). The structure and semantic comparison of texts are involved in Katz similarity, which is highly correlated with NIST measurements. Classification of text is significantly improved by this hybrid method.

Table 1 Comparative analysis of related studies

Authors	Methodology used	SN	DN	Topological features		SF	CI	Advantage	Disadvantage
				Global	Local				
(Jalili et al. 2017)	Machine learning	Y	N	N	Y	N	N	Performance increases due to the inclusion of cross layer information	It is always not possible to get the cross layer information of all the users
(Wahid-Ul-Ashraf et al. 2017)	Similarity based on Newton's law	Y	N	N	Y	N	N	Similarity measure can accomodate global structure of the network	Time complexity is high and topological informations are only considered
(Zhang and Chen 2018)	gamma-Decaying theory and GNN	Y	N	N	Y	N	Y	It outperforms other heuristic similarity measures due to inclusion of learning heuristic paradigm	Considers only topological features and complexity is high.
(Wang et al. 2018)	Deep learning	Y	N	N	Y	Y	N	Flexible, Scalable and free from cold start problem	High complexity and high time consumption for large graphs
(Zhou and Kwan 2018)	Matrix completion technique	Y	N	N	Y	N	N	The method handle large scale network	It may suffers from network sparsity problem
(Mallek et al. 2019)	Dempster shafer theory	Y	N	N	Y	N	N	It can be used for uncertain network	Considers only the topological information
(Samad et al. 2019)	Similarity based	Y	N	N	Y	N	N	It creates more chances to lower degree nodes to link with higher degree nodes	Consider only the topological information
(Tuan et al. 2019)	Fuzzy and neutrosophic measures	Y	Y	N	Y	N	N	It handles the quick change in the network structure and uncertainty in the structure	Consider only topologcal structure
2020	Proposed method	Y	N	N	Y	Y	Y	It uses the heterogeneous information	Computational Complexity is less

SN – Static network; DN – Dynamic network; SF – Sentiment feature; CI – Community information; Y– Yes; N– No

The community detection algorithms like Louvain Algorithm (De Meo et al. 2011), Girvan Newman algorithm (Bolla 2011) etc., use the topological features for the extraction of community information. Using these algorithms in further improvement of link prediction is not advisable because the topological information is already extracted through similarity based features separately. Any improvement that needs to be done on the similarity based features needs to be done with semantic information. So, in this paper, we propose to use the Topic based communities (Zhao et al. 2012) for enhancing the results of link prediction. To overcome the drawbacks of previous models where either community information or sentiment values have been used for link prediction, we propose to use both these as features in our method. We tend to obtain an optimal result of link prediction. To achieve the goal of optimal link prediction, we employ multiple optimization techniques to optimize our multi-feature vector and produce an optimal result.

3 Proposed work

The social network is represented as a graph $G(V, E)$, where V represents the number of nodes and E represents the number of edges between the nodes. The complete graph is denoted by U, which contains all the $|V|(|V|-1)/2$ possible edges or links among the nodes of the network. The main goal of link prediction is to find out whether the non existent links i.e., $U - E$ are true or not (Valverde-Rebaza and de Andrade Lopes 2012). The complete procedure of the proposed approach is depicted in Algorithm 1.

Algorithm 1: Enhanced Link prediction with Community information and Sentiment Attributes

Input: The Social network G and Tweet Data of each user in the Network.
Output: Link Prediction

1. Calculate the Similarity Features for each pair of unconnected nodes.
2. Determine the sentiment values of each node in the Network.
3. Calculate the Sentiment Polarity and Sentiment Similarity for each pair of unconnected nodes.
4. Determine the Topic based communities based on the network information.
5. Construct the Feature vector as {CN, JAC, PA, SP, SS, TC, *label*}.
6. Train the classifier with existing links in Network and predict the future possible links;
7. Optimize the Feature vectors with true labels to do the Link Prediction.

3.1 Similarity based features

Consider a node a. Then the structural definition of the node $a \epsilon V$ in a graph is its neighbourhood $\Gamma(a) = \{b|(a,b)\epsilon(b,a)\epsilon E\}$. This definition is used to obtain various similarity based features as shown below.

3.1.1 Common Neighbours(CN)

If there are two nodes in a network a and b, they are more likely to get linked in future if they have more common neighbours. The common neighbors are determined by the number of nodes matching between a and b. In terms of social network, it can be defined as more the number of mutual friends, they are more likely to connect. The same has been shown in the Eq. 1.

$$S_{ab}^{CN} = |\Lambda_{ab}| = |\Gamma(a) \cap \Gamma(b)| \tag{1}$$

3.1.2 Jaccard Coefficient(JAC):

This is a measure to check whether the two nodes have a significant number of common nodes regarding their total neighbors. This was defined by Jaccard a hundred years ago. For the undirected network, it is defined as in the Eq. 2.

$$S_{ab}^{JAC} = \frac{|\Gamma(a) \cap \Gamma(b)|}{|\Gamma(a) \cup \Gamma(b)|} \tag{2}$$

3.1.3 Preferential Attachment(PA):

If there are two nodes in a network a and b, then the probability that a new link will be formed between a and b is proportional to $|\Gamma(a)| \times |\Gamma(b)|$. Based on this measure, the similarity index between two users in a network can be defined as Eq. 3.

$$S_{ab}^{PA} = |\Gamma(a)| \times |\Gamma(b)| \tag{3}$$

3.2 Sentiment extraction

Sentiment analysis is one of the prominent technologies to aid the users in the huge amount of users generated data in social network (Tan et al. 2011). Unlike the traditional data, the data that has been collected from micro-blogging sites are noisy. The data are often short, incomplete, and contain unnecessary data which are not related to the topic. In recent years, a lot of effort has been made in the field of sentiment analysis (Liu 2012).

Emotional words have been used to compute the sentiment value of a sentence. In this work, we have considered sentence as a tweet by the user. We have used SentiwordNet 3.0 which contains 117,659 emotional words, and their corresponding sentiment values (Baccianella et al. 2010). Apart from using the emotional words, we have manually determined the negation words which change the meaning of the sentence. When the negation word is used, the sentiment value of a sentence or tweet becomes negative.

In addition to the above, the exclamation mark (!) is identified. It conveys a negative score to the tweet. If a tweet is modified by a negation word or an exclamation mark, then the sentence or tweet carries a negative sentiment score.

3.3 Sentiment polarity

In this step, the sentiment values of each user are calculated. Then the users with sentiment polarity are made into groups. For example; consider there are two users *Bob* and *Alice*. Let S_{Bob} and S_{Alice} represent the sentiment values of *Bob* and *Alice* respectively. *Bob* and *Alice* can fall into the same group if any of the following case is true .

Case 1. *PositiveGroup*, if $S_{Bob} > 0$ and $S_{Alice} > 0$
Case 2. *NegativeGroup*, if $S_{Bob} < 0$ and $S_{Alice} < 0$

Case 3. *NeutralGroup*, if $S_{Bob} = 0$ and $S_{Alice} = 0$

3.3.1 Sentiment similarity

The sentiment similarity can be defined as the similarity distance between the users in terms of sentimental value. Here, we calculate the absolute difference of the sentiment scores between the two users that belong to either a positive sentiment group or negative sentiment group. The users belonging to the neutral sentiment group have zero sentiment value, due to which the sentiment similarity will also be zero for their case. The two users should have sentiment polarity for this. If the users exhibit sentiment polarity, i.e., exhibit the same sentiment values, then there is a higher chance for an edge to be formed between them. If the users do not have polarity in terms of sentiment, then we assign a score of -100 for sentiment similarity, which is impossible to achieve.

$$\begin{aligned} &\textit{Sentiment Similarity}(Bob, Alice) \\ &= \begin{cases} |S_{Bob} - S_{Alice}|, & if\ SP_{Bob} = SP_{Alice} \\ -100, & Otherwise \end{cases} \end{aligned} \tag{4}$$

Since the sentiment grouping is made earlier, the process of calculating sentiment similarity can be done simultaneously for Positive sentiment group and negative sentiment group. Since the neutral sentiment similarity is always zero, we do not consider the users in neutral polarity for this step.

3.4 Topic based community detection

Community detection is one of the important feature for the analysis of the users belonging to a social network. Studying the community structure of a network helps in understanding the interesting properties shared by the users. The identified communities are often used to provide collaborative recommendations (Yang et al. 2007), knowledge sharing (Zeng et al. 2008), information spreading (McCallum et al. 2005), and other applications. But we do not employ traditional community detection algorithms like Girvan Newman or Louvain etc., because these algorithms detect communities based on the topological properties only. Communities are detected either by studying the linkage of the nodes in the network or the modularity of the nodes in the network. Since the topological features have already been extracted, we proposed to use the Topical based communities as presented in (Zhao et al. 2012).

The topics for our communities are the HashTags, which are most popular in the tweet dataset. Based on the most similar topics among the tweets of the users, the hot topics are determined. Based on these topics, users are formed into topic-based communities.

3.5 Feature vector

Sentiment polarity, sentiment similarity, and topical communities are calculated after calculating the similarity based measures. Then we move into the next step called construction of feature vector. The feature vector will be constructed for each pair of nodes which are not connected in the network. The feature vector for each pair of nodes can be represented as {CN, JAC, PA, SP, SS, TC,*label*}. Since the sentiment similarity between the users in Neutral sentiment community is zero, we mark $SS = 0$ in feature vector for the links predicted in neutral sentiment community. The *label* here represents the whether a link is present or not. The *label* value will be 1 if the link is present between a pair of nodes and 0 if there is no link between the pair of nodes. The feature vector features are mentioned below:

- CN = Common Neighbours
- JAC = Jaccard Coefficient
- PA = Preferential Attachment
- SP = Sentiment Polarity
- SS = Sentiment Similarity
- TC = Topic based Community
- label = Predicted link(1 = True , -1 = False)

3.6 Link prediction

The classifier is trained with the feature vector data where the *label* is true (i.e., having value 1). Based on the trained data, Link prediction is done for unconnected pairs of nodes. The prediction is done on the all the feature vectors representing the future links possible. The final result is a set of feature vectors with *label* values as -1 or 1. The feature vectors whose *label* value as 1 are the links that are predicted as true, and the feature vectors with *label* as -1 are the links that predicted as false.

3.6.1 Support Vector Machine

The classifier predicts the presence of a link or an absence of a link in the network. To do the same, SVM, i.e., Support Vector Machine tool, has been used as the classifier. The training data for the SVM is the links that are already present in the network. We represent the links that are present in the network with *label* = +1. All the other missing links will be represented by *label* = -1.

3.7 Optimization

Since the result is displayed in the form of a Feature vector, the result needs to be optimized. The feature vector contains multiple features. Some features need to be maximized while some need to be minimized. For example, the Common Neighbours

needs to be maximized but the feature Sentiment Similarity needs to be minimized. In such cases, the result should be optimized to get the optimized link prediction results. In this paper, we employed the below-mentioned optimization algorithms to find the best link prediction result.

3.7.1 Weighted Arithmetic Mean

The weighted mean or the weighted arithmetic mean is just like the average, but the contribution of each weight could be altered according to the requirements. If all the weights of the features are considered the same, then this is the same as the average.

Consider the vector with the values $\{a_1, a_2, ..., a_n\}$ and a corresponding non negative weights $\{ w_1, w_2, ..., w_n\}$. Then the result can be determined by Eq. 5.

$$\bar{a} = \frac{\sum_{i=1}^{n} w_i a_i}{\sum_{i=1}^{n} w_i} \tag{5}$$

Equation 5 can be further expanded as,

$$\bar{a} = \frac{w_1 a_1 + w_2 a_2 + ... + w_n a_n}{w_1 + w_2 + ... + w_n} \tag{6}$$

3.7.2 TOPSIS

The technique used for preference by similarity to obtain the ideal solution is simply known as TOPSIS. It is used for multi-criteria decision making. It works by comparing a set of alternatives by identifying the weights for each criterion and then assigning the normalized scores for each criterion. The best solution is found out by calculating the geometric distance between the alternatives and the ideal alternative. TOPSIS works on the assumption that the criteria or the features are monotonically increasing or decreasing. Normalization is done on the results because the features are often in incongruous dimensions.

3.8 Complexity of the proposed method

Analysis of computational complexity is important in social networks. In our work, we have computed topological information of users, Sentiment values of the sentence delivered by each user using SentiwordNet 3.0 tool, and Community information of each user using Hashtag information of the tweet. Topological information are computed using UNION and INTERSECTION operations of set theory. The computational complexity of local information link prediction algorithms is low and it is $O(n.k^2)$, where n is the number of users and k is the average degree of all the nodes. The time complexity of our method mainly depends upon the classification task and Optimization task. For classification task we have used the SVM classifier. Since we have used LIBSVM package, its time complexity is $O(n^2)$. TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)algorithm is used to predict the future possible links. This algorithm complexity is in the part of attribute normalization and weighting which results in $O(mn)$, where m represents the number of features.

3.9 Performance evaluation

To evaluate the link predictions produced by our method, some metrics are used. To use the metrics, consider the result of link prediction as either predicting a link can be formed or a link cannot be formed. Consider all the links that can be formed to be a positive group, and the not-linking be a negative group. Our prediction method can predict a link in a positive group or negative group. If the prediction is made as a positive group, then the prediction can be correctly known as correct-positive (CP); otherwise, it is known as wrong-positive (WP). Likewise, if the prediction is made as a negative group, then the prediction can be correct known as correct-negative (CN); otherwise, it is known as wrong-negative (WN). Based on these values, the metrics known as Precision and Recall can be defined as shown in Eqs. 7 and 8.

$$Precision = \frac{|CP|}{|CP| + |WP|} \tag{7}$$

$$Recall = \frac{|CP|}{|CP| + |WN|} \tag{8}$$

Based on the above metrics 7 and 8, another metric F-measure is defined. F-measure compares the predictors numerically. It is a harmonic mean of precision and recall. It is defined as shown in Eq. 9 .

$$F - measure = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{9}$$

4 Implementation

A social network can be represented as a graph structure as shown in Fig. 2. The tweets are mentioned at the nodes of each user. Here we noticed that the node E does not have any connection, i.e., linkage with the other users in the network. So, the node E is not considered in further proceedings.

Figure 3 represents the network after the sentiment values have been calculated for each node. The Red color signifies a Negative sentiment value, Green signifies a Positive sentiment value, and Grey color signifies a neutral sentiment value, i.e., 0 sentiment value. The sentiment similarity

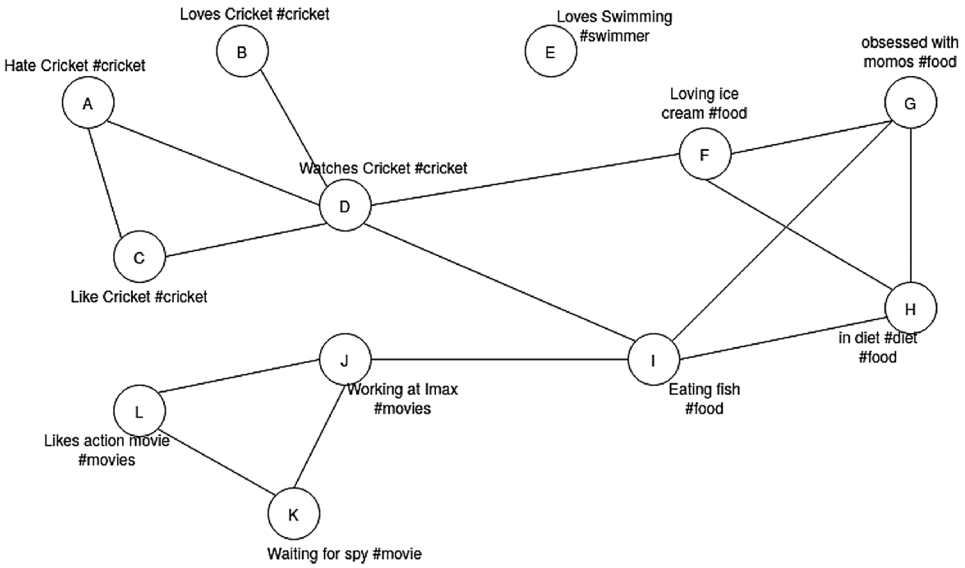

Fig. 2 Sample representation of the Data

Fig. 3 Sentiment analysis on Nodes of the Network

values will be calculated only for the nodes that are not connected to each other.

Figure 4 shows how the nodes in the network are made into communities based on the topics. These topics are the #'s, i.e., HashTags that are mentioned in their tweets. Based on the participation of the users in the relevant topic, they are included in their respective communities. From Fig. 4, it is observed that in our sample network, three topic-based communities are determined, i.e., Cricket community, Food community, and movie community. One more community is determined (Swimmer), but it is not included for further processing because it belongs to a singleton node.

In the Fig. 5, the future possible links are mentioned with dotted lines. The Feature vectors are constructed for all these possible links. But the challenging part is determining which link among these has the highest probability. To do the same, the result should be optimized with respect to multiple features.

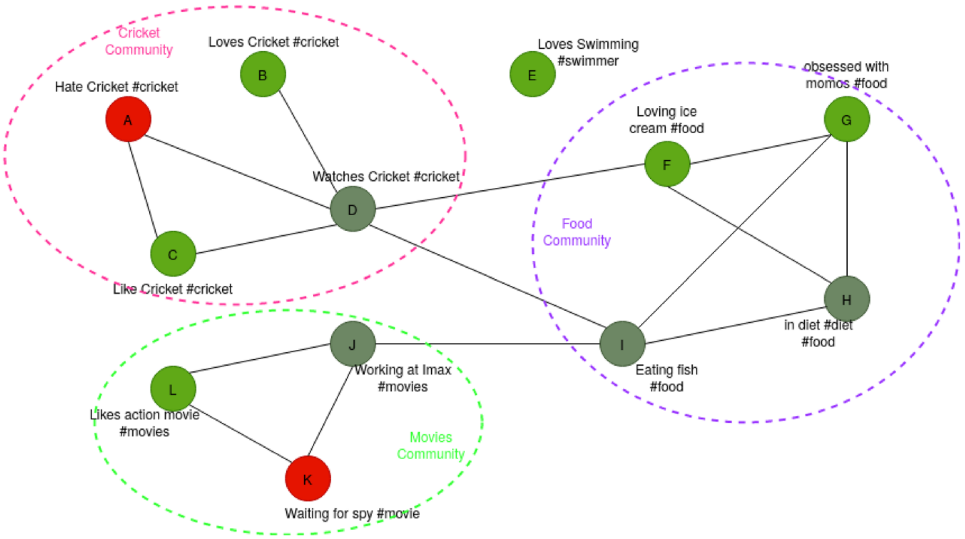

Fig. 4 Topic based Community detection

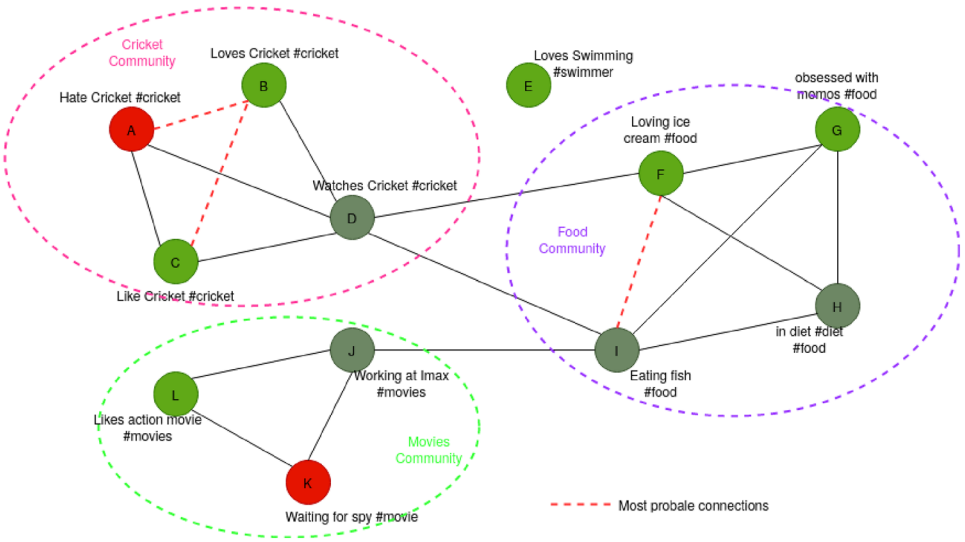

Fig. 5 Links possible in the Network

Table 2 Data set

Country	Tweets	Users	Users with link	Links
India	13241	7802	368	912
China	20988	6905	207	302
France	12392	4103	186	794
USA	18592	6388	212	545
UK	12487	6929	394	1198
Iran	11232	5394	258	198
Switzerland	17116	2999	72	145
Germany	15238	9172	215	440
Spain	19923	9288	240	252
Italy	22103	8614	297	398

5 Experimental results

In this work, we have considered the tweets that are related to COVID-19. For the same, the tweets from various countries that are affected by the COVID-19 are considered for the purpose of link prediction.

5.1 Dataset description

We have extracted tweets from the twitter application. These are the real time data of the users that have tweeted about COVID-19 from the top 11 countries. The data has been extracted during the duration from 01/04/2020 to 05/04/2020. The details about the data that have been extracted are mentioned in Table 2.

5.2 Dataset pre-processing

There are various pre-processing steps carried out before using the tweet data for the experiment. To make the data suitable for experimental setup, we perform the following processing steps.

5.2.1 Removing nodes with no linkage

The nodes that do not have a connection with any of the other nodes in the network are excluded from the dataset. This step is necessary because if a node does not connect with at least one node in the network, the topological values cannot be calculated for that node. The similarity-based values will be 0, which makes the node an outlier for the current dataset.

5.2.2 Excluding tweets based on language

In this work, we have used SentiwordNet 3.0 (Baccianella et al. 2010), using which sentiment analysis can be performed only on tweets that are in english language. The tweets that are written in any other languages are excluded from the dataset. If a proper WordNet can be defined, then the tweets in other languages can also be used. Also the tweets that are invalid and unable to extract the sentiment information are excluded from the final dataset.

5.2.3 Excluding tweets with no meaning

The tweets that convey no meaning or hard to analyze the sentiment values are excluded for the experiment. Tweets that contain *URL* to some websites or the tweets that have only smileys or special characters are also excluded. This is done to consider the tweets that are relevant to the topic.

5.3 Similarity feature extraction

In this work, similarity based features are calculated for unconnected pairs of nodes that are present in the network. The similarity based features calculated for unconnected pairs are Common Neighbours as mentioned in Eq. 1, Jaccard coefficient as mentioned in Eq. 2 and Preferential attachment as mentioned in Eq. 3. The features thus obtained are based on the topological properties of the network structure. These similarity-based feature values are further used in the construction of the Feature vector.

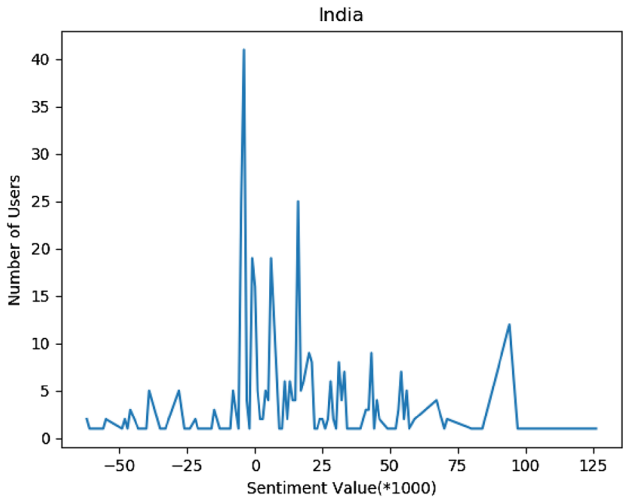

Fig. 6 Sentiment values in India

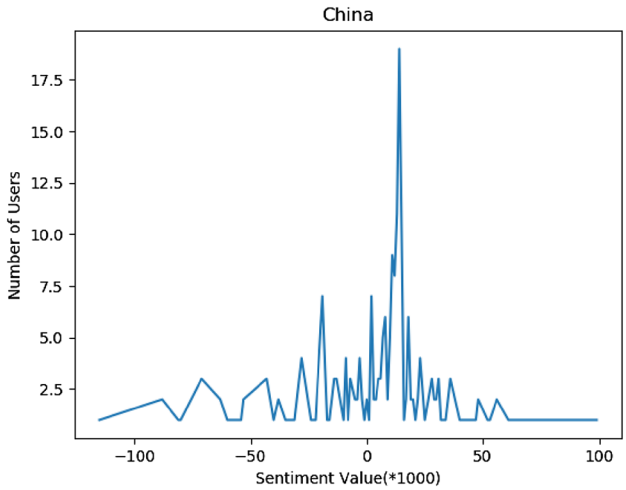

Fig. 7 Sentiment values in China

5.4 Sentiment analysis and grouping

In this step, the sentiment values of each user are calculated. Then the users are grouped based on the sentiment polarity. As per the sentiment polarity values, three different groups are formed. These are the Positive sentiment group, Negative sentiment group, and Neutral sentiment group. **sentiment similarity** and **sentiment polarity** are calculated for the users falling under the positive group and negative group. Since the users in the Neutral group do not have any sentiment values, we calculate only **sentiment polarity** for them. The sentiment grouping of users concerning each country has been shown in Table 3. Figures 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15 depicts the sentiment values of each country against their number of users.

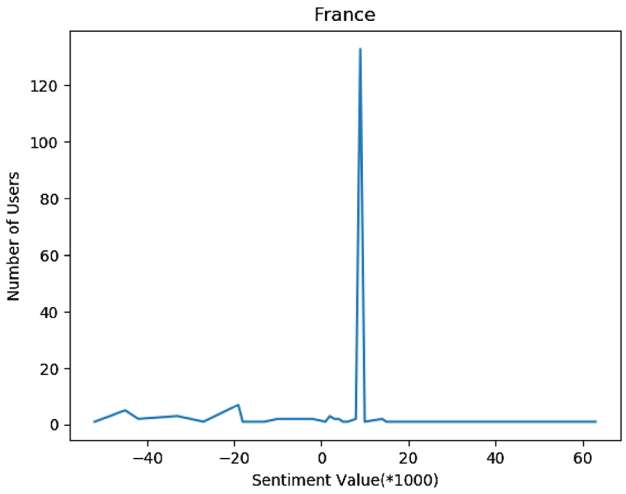

Fig. 8 Sentiment values in France

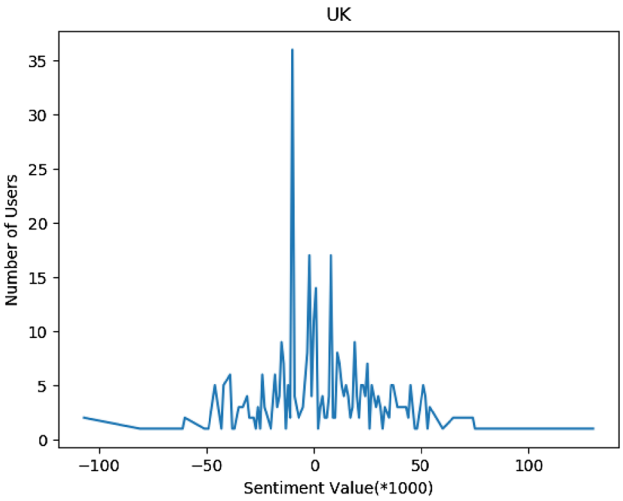

Fig. 10 Sentiment values in UK

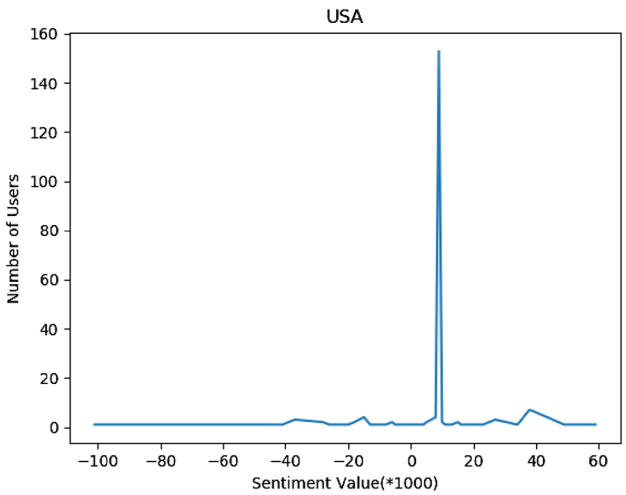

Fig. 9 Sentiment values in USA

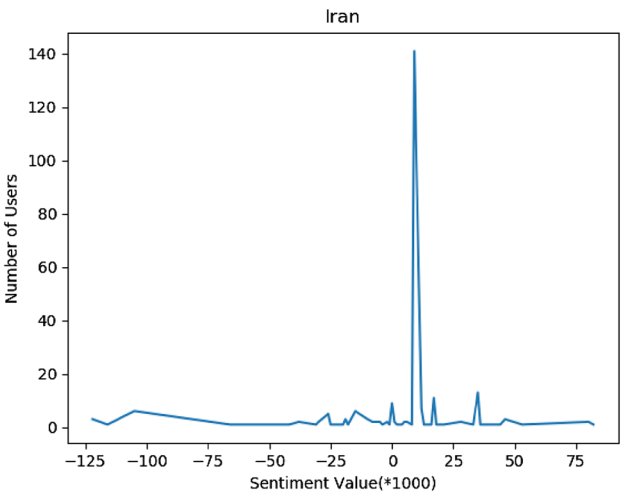

Fig. 11 Sentiment values in Iran

From Fig. 6, it is observed that more variation in the number of users occurs for the positive sentiment value. The number of users is maximum when the number of users attained for sentiment values is negative. In Fig. 7, China's number of users versus sentiment value is shown. More number of users have shown positive sentiment values. In China, the highest value of the number of users is achieved for the positive centrality value. In Figs. 8 and 9, it is observed that the change in the number of users against the sentiment values is very similar in nature. Both the countries achieved the peak value of the number of users for the positive sentiment value between 0 to 20. However, fewer users have shown negative opinions in their tweets in both countries. In Fig. 10, the difference between the positive opinion users and negative opinion users is much less compared to other countries. In the country Iran, as shown in Fig. 11, many users are showing their positive opinion in their tweets. This country's maximum positive sentiment value is in the approximate range of 75 to 90. The highest number of users showed the sentiment similarity value in the range of 0 to 25. However, the minimum negative sentiment value achieved in this country is approximately -125. From Fig. 12, it is observed that sentiment values of the users from Switzerland country lie in between the approximate range of -200 to 50. More variation in the number of users exists when the sentiment value exceeds the sentiment value -50. In Fig. 13, the maximum variation in the number of users occurs in the positive sentiment value between 0 to 50. The number of users are maximum when the sentiment value is approximately 50.

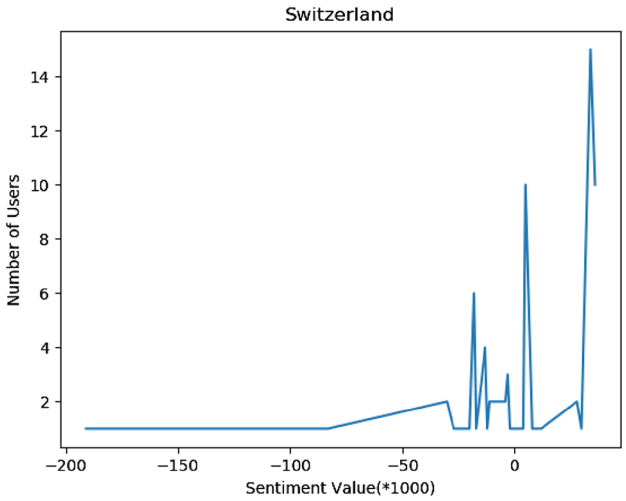

Fig. 12 Sentiment values in Switzerland

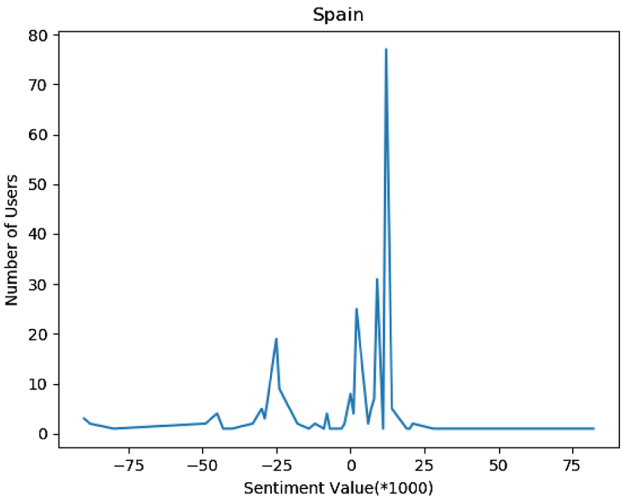

Fig. 14 Sentiment values in Spain

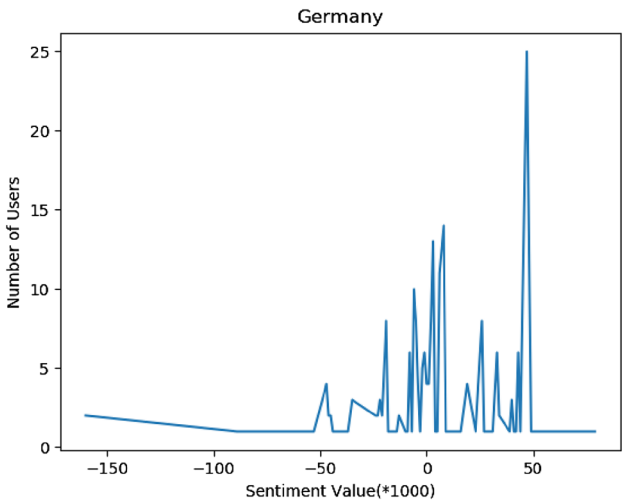

Fig. 13 Sentiment values in Germany

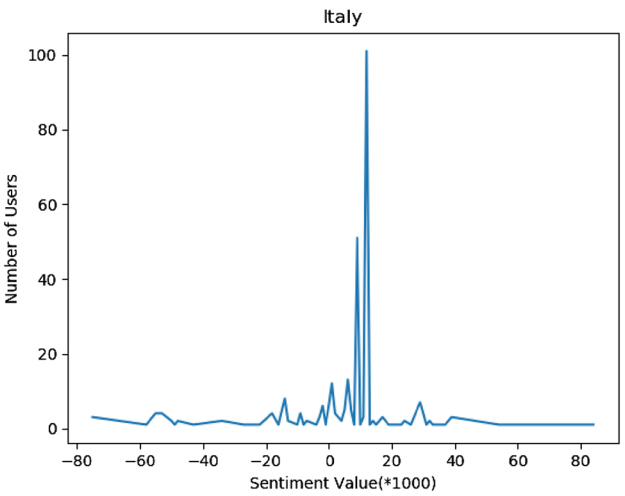

Fig. 15 Sentiment values in Italy

Figure 14 presents the sentiment value graph of the users of Spain's country. Many users show a positive sentiment value in the approximated range of 0 to 25. From Fig. 15, it is observed that the sentiment values of different users of the country Italy lie in the approximate range of -80 to 100. More change in the number of users occurs for the positive sentiment value. More number of users showed a positive sentiment value in their tweets. From these figures, it is inferred that, in the country of India, a large difference occurs between the number of users showing the positive sentiment value and negative sentiment value. Among all the countries, Switzerland country has achieved the minimum sentiment value. In comparison, the maximum sentiment value is achieved in the country UK.

Table 3 Sentiment grouping

Country	Positive group	Negative group	Neutral group
India	321	43	4
China	135	71	1
France	153	32	1
USA	187	24	1
UK	216	177	1
Iran	207	48	3
Switzerland	44	27	1
Germany	126	88	1
Spain	164	73	3
Italy	231	65	1

Table 4 Topic based communities part-1

Country	Topic based communities	Number of users
India	COVID19	172
	Breaking	68
	Tabhleegijamaat	46
	Coronavirus	40
	Indiafightscorona	20
	India	11
	COVID19pandemic	9
	Stayathomesavelives	5
	Stpiindia	3
	Coronavirusoutbreak	3
China	COVID19	127
	China	27
	Coronavirus	23
	Cuba	7
	Eeuu	6
	Estevirusloparamosunidos	4
	Quedateencasa	4
	fReeleahsharibu	3
	Renosnuggets	3
	Ccp	3
France	COVID19	104
	Coronavirus	24
	France	19
	Confinement	11
	Confinementjour17	11
	Masques	6
	Macron	4
	Coronavirusfrance	3
	Turquie	2
	COVID19france	2
USA	COVID19	119
	USA	37
	Coronavirus	27
	Trump	5
	Flattenthecurve	4
	Trumpownseverydeath	4
	Turquie	3
	China	5
	COVID-19	3
	Manosfueradevenezuela	5

Table 5 Topic based communities part-2

Country	Topic based communities	Number of users
UK	COVID19	292
	Coronavirus	54
	UK	11
	Stayhomesavelives	6
	Lockdown	6
	COVID19pandemic	6
	COVID19uk	5
	Proudmalaysian	5
	NHS	5
	Cop26	4
Iran	COVID19	164
	Iran	45
	Coronavirus	31
	Trump	5
	Turkey	4
	COVID-19	5
	USA	4
Switzerland	COVID19	40
	Switzerland	8
	Coronavirus	7
	Stayathomeandstaysafe	5
	Sarscov2	4
	Versusvirus	2
	Italy	2
	Spain	2
	Iran	1
	Police	1
Germany	COVID19	143
	Coronavirus	27
	Germany	23
	Italy	5
	Spain	3
	France	3
	Europe	4
	Stayhome	3
	Flatteningthecurve	2
	Yesywecan	2

5.5 Topic based community detection

In this step, the frequent topics or hot topics concerning each country are determined. This is done by analyzing the tweets of the users for each country network. Then the users are grouped based on which topic community they are categorized into. Tables 4, 5 and 6 show the topic based communities that have been determined for each country.

5.6 Link prediction

In this step, feature vectors are calculated that contain the similarity-based features, sentiment values (Sentiment Polarity and Sentiment Similarity), and the community

Table 6 Topic based communities part-3

Country	Topic based communities	Number of users
Spain	COVID19	102
	Coronavirus	65
	Spain	36
	Italy	11
	Coronavirusoutbreak	7
	Strongtogether	6
	Newzroom405	4
	Coronavirussouthafrica	3
	Nato	3
	COVID19pandemic	3
Italy	COVID19	120
	Coronavirus	54
	Italy	54
	Spain	17
	Cop26	14
	Strongtogether	11
	Coronavirusoutbreak	8
	China	8
	Nato	6
	COVID-19	5

information of the possible links in the network. The feature vector is represented as $\{CN, JAC, PA, SP, SS, TC, label\}$. The feature vectors are constructed for each pair of unconnected nodes in the network w.r.t. each country data. Then the classifier is trained with the links existing in the network, i.e., with the feature vectors having *label* values as 1. With this training, the link prediction is made on the network.

5.7 Optimization of link prediction

Since the result contains several features obtained from the topological features, sentiment values, and community information, the result needs to be optimized to predict the best possible links. To perform the same, we optimize the result using various approaches like Weighted Sum and TOPSIS. The optimization is done to display the feature vectors in descending order of priority.

5.8 Performance of link prediction

To evaluate the performance of the proposed algorithm, we have calculated the values of metrics defined in Eqs. 7, 8, and 9. We have compared the result of our proposed algorithm against the J48 for Supervised learning and the GNN-based Weisfeiler-Lehman Neural Machine (Zhang and Chen 2017) method. The J48 algorithm is the Weka (Zhou et al. 2009) adaption of the C4.5 Algorithm. The Feature Vector

Table 7 Performance of link prediction

Country	Method	Precision	Recall	F-measure
India	Base	0.89992	0.93642	0.91780
	Enhanced	0.90909	0.95238	0.93023
	WLNM	0.90102	0.94032	0.92025
China	Base	0.81976	0.92679	0.86999
	Enhanced	0.84333	0.94339	0.89055
	WLNM	0.82681	0.94001	0.87978
France	Base	0.89276	0.98176	0.93514
	Enhanced	0.90285	0.98954	0.94420
	WLNM	0.89854	0.98457	0.93958
USA	Base	0.86956	0.83334	0.85106
	Enhanced	0.88695	0.89047	0.88870
	WLNM	0.87562	0.85102	0.86314
UK	Base	0.90909	0.85684	0.88219
	Enhanced	0.92592	0.90900	0.91738
	WLNM	0.91462	0.88372	0.89890
Iran	Base	0.81967	0.85642	0.83764
	Enhanced	0.79365	0.89909	0.84308
	WLNM	0.83750	0.88125	0.85881
Switzerland	Base	0.80638	0.93642	0.86654
	Enhanced	0.81645	0.94646	0.94866
	WLNM	0.80986	0.93875	0.86955
Germany	Base	0.86656	0.93546	0.89969
	Enhanced	0.87719	0.98564	0.92825
	WLNM	0.86967	0.95632	0.91093
Spain	Base	0.88495	0.93642	0.90995
	Enhanced	0.89285	0.94733	0.91928
	WLNM	0.90842	0.96410	0.93543
Italy	Base	0.89285	0.91175	0.90220
	Enhanced	0.90901	0.92681	0.91786
	WLNM	0.89842	0.92132	0.90972

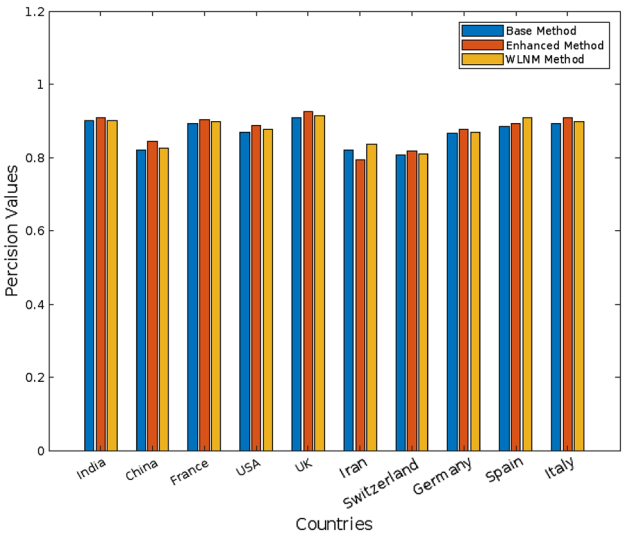

Fig. 16 Precision

for this algorithm will be $\{CN, JAC, PA, AA, RA, label\}$ and let us denote it as **Base Method** for representation purposes. WLNM method is GNN based link prediction algorithm, which learns only the topological features of the given graph. These features are further used for prediction purposes. Basically, WLNM has three important steps; extracting enclosed sub-graphs, labeling and encoding the nodes, and training the neural network. In the first step, the enclosed sub-graph of the targeted link is extracted. Following this, a novel Weisfeiler-Lehman (WL) is used to label the nodes as per their topological role in the sub-graph. Finally, labeled sub-graphs are trained to learn the various heuristics present in the local patterns. Out of the predicted links, part of the data has been used as training data and the rest of the data as testing data. The metrics for evaluation are calculated by testing the correctness of the training data. The results of Precision, Recall, and F-measure is shown in Table 7. In Fig. 16, the Precision value obtained using the Base method, Enhanced method, and WLNM method is demonstrated for all the countries. Using the proposed (Enhanced) method, the UK achieved the highest value of precision among all other countries. In contrast, Iran has the lowest value of precision using the enhanced method. In sparse graphs like; Iran and Spain, the WLNM method shows good precision as compared to the other two methods.

The enhanced method achieves a high precision value as compared to the other two methods (Base method, WLNM method) in well-connected graphs except for the sparse graphs. The node's explicit information such as; topic-based community, sentiment similarity, sentiment polarity value, and the vital reason behind these outcomes. These extra information plays an important role to influence the performance metrics in prediction tasks. However, the enhanced method unable to capture all the hidden topological information in sparse graphs. WLNM shows high performance in terms of Precision, Recall, and F-measure than the Base method in all the graphs. Using the Base method, country UK and Switzerland have the maximum and minimum precision values, respectively.

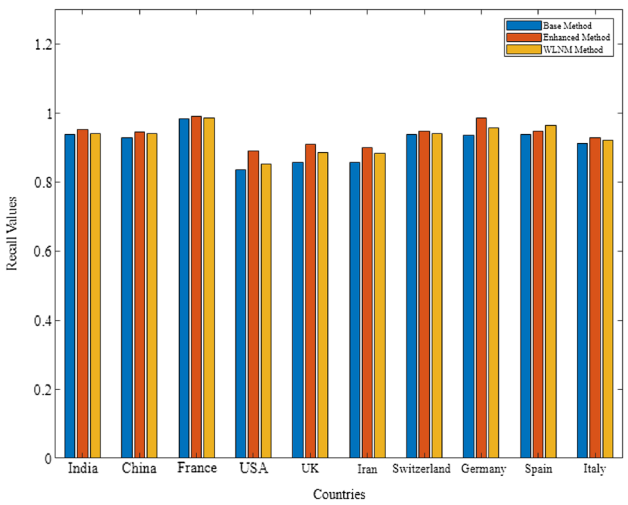

Fig. 17 Recall

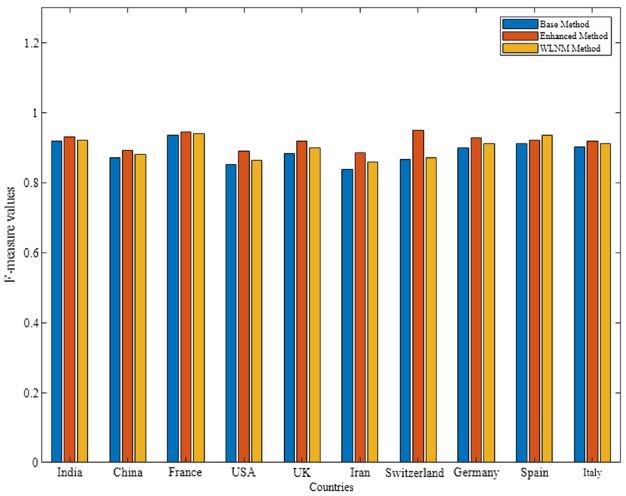

Fig. 18 F-Measure

Using the WLNM method, the country UK has a high precision value, and the country Switzerland has a low precision value. In Fig. 17, the Recall values of all the countries are shown. The country France and USA show the highest and lowest recall value respectively, in all the methods. Our proposed method outperforms the Base method in all the well-connected graphs. However, due to sparsity in the graph, WLNM shows high performance than the other two methods in terms of Recall value. Because WLNM automatically learns the meaningful topological features from the extracted sub-graph and does depend on the common neighborhood information. In Fig. 18, F- Measure values for all the countries are shown, and it is observed that our method has better performance than the other two methods in all most all the graphs. Overall, due to the presence of extra node information, our proposed method (Enhanced method) shows good performance than the other two state-of-the-art methods. In the sparse graph WLNM method shows good performance than our proposed method. However, the computational cost of the WLNM method has very as compared to the other two methods. This is because WLNM learns the meaningful pattern (heuristics) from the extracted sub-graph during the training period.

6 Threats to validity

The tweets that convey the meaning in sarcasm are not able to be distinguished with SentiNet 3.0. This is because the tweets conveying the message in sarcasm mean the same as a positive tweet. But the underlying meaning of the sarcasm

tweet conveys a negative sentiment. Efforts have been made to make use of smileys in the tweet to find if the tweet conveys sarcasm. But the cases where it is not possible to distinguish, the tweet will be considered in a positive sentiment, which leads to a slight impact on the results. Some users have privacy settings enabled on their profile. When the account is private, the follower's information cannot be extracted. Thus, the link present with this user will not be identified, further leading to less connectivity in the user network obtained.

7 Conclusion and future work

In this work, we have proposed a novel method to incorporate sentiment values of the users and the community information of the network for the link prediction. Topic based community information has been used, due to which the trending topical communities are incorporated to enhance the results of link prediction. It has been found that, the more the connections in the network, the higher the quality of the links predicted. This is because the more information we have i.e., more links in the network, we can train classifier with more data and thereby increasing the quality of link prediction.

In the future, we propose to use a metric known as community distance between the nodes of the network. Using this metric, we aim to find the distance between the user's community and use this information to improve the link prediction. Also, folding the graph can construct a densely connected social network. In addition to Twitter data, we will further collect the data from other social networks, such as; Facebook, Instagram, LinkedIn, etc., using the corresponding API and incorporate the heterogeneous GNN model on these data to enhance the proposed method.

Instead of predicting a link as +1 or -1, we want to determine the strength of each positive prediction and negative prediction. Based on this prediction score, the predicted links can be presented in a sorted manner as an end result.

Funding Not applicable

Data availability The data sets analysed during the current study are available with the authors. The same will be shared once the manuscript is accepted/published.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

Amancio DR, Oliveira ON Jr, Costa LdF (2012) Structure-semantics interplay in complex networks and its effects on the predictability of similarity in texts. Physica A: Stat Mech Appl 391(18):4406–4419

Amancio DR, da F Costa L et al (2015) Topological-collaborative approach for disambiguating authors' names in collaborative networks. Scientometrics 102(1):465–485

Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Lrec, vol 10, pp 2200–2204

Bastami E, Mahabadi A, Taghizadeh E (2019) A gravitation-based link prediction approach in social networks. Swarm Evol Comput 44:176–186. https://doi.org/10.1016/j.swevo.2018.03.001

Bolla M (2011) Penalized versions of the newman-girvan modularity and their relation to normalized cuts and k-means clustering. Phys Rev E 84(1):016108. https://doi.org/10.1103/PhysRevE.84.016108

Coşkun M, Baggag A, Koyutürk M (2021) Fast computation of katz index for efficient processing of link prediction queries. Data Min Knowl Discov 35(4):1342–1368

De Meo P, Ferrara E, Fiumara G, Provetti A (2011) Generalized louvain method for community detection in large networks. In: 2011 11th international conference on intelligent systems design and applications. IEEE, pp 88–93. https://doi.org/10.1109/ISDA.2011.6121636

De Sá HR, Prudêncio RB (2011) Supervised link prediction in weighted networks. In: The 2011 international joint conference on neural networks. IEEE, pp 2281–2288. https://doi.org/10.1109/IJCNN.2011.6033513

Hu X, Tang L, Tang J, Liu H (2013) Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the sixth ACM international conference on Web search and data mining, pp 537–546. https://doi.org/10.1145/2433396.2433465

Jalili M, Orouskhani Y, Asgari M, Alipourfard N, Perc M (2017) Link prediction in multiplex online social networks. R Soc Open Sci 4(2):160863. https://doi.org/10.1098/rsos.160863

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031. https://doi.org/10.1002/asi.20591

Liu B (2012) Sentiment analysis and opinion mining. Synth Lectures Hum Lang Technol 5(1):1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Lü L, Zhou T (2011) Link prediction in complex networks: a survey. Physica A: Stat Mech Appl 390(6):1150–1170. https://doi.org/10.1016/j.physa.2010.11.027

Mallek S, Boukhris I, Elouedi Z, Lefèvre E (2019) Evidential link prediction in social networks based on structural and social information. J Comput Sci 30:98–107. https://doi.org/10.1016/j.jocs.2018.11.009

Marinho VQ, Hirst G, Amancio DR (2016) Authorship attribution via network motifs identification. In: 2016 5th Brazilian conference on intelligent systems (BRACIS). IEEE, pp 355–360

McCallum A, Corrada-Emmanuel A, Wang X (2005) Topic and role discovery in social networks. https://scholarworks.umass.edu/cs_faculty_pubs/3

Naik D, Gorojanam NB, Ramesh D (2020) Community based emotional behaviour using ekman's emotional scale. In: International conference on innovations for community services. Springer, pp 63–82. https://doi.org/10.1007/978-3-030-37484-6_4

Samad A, Qadir M, Nawaz I (2019) Sam: a similarity measure for link prediction in social network. In: 2019 13th international conference on mathematics, actuarial science, computer science

and statistics (MACS). IEEE, pp 1–9. https://doi.org/10.1109/MACS48846.2019.9024762
Santos DLB, Corrêa Jr EA, Oliveira Jr ON, Amancio DR, Mansur LL, Aluísio SM (2017) Enriching complex networks with word embeddings for detecting mild cognitive impairment from speech transcripts. Preprint at arXiv:1704.08088
Soundarajan S, Hopcroft J (2012) Using community information to improve the precision of link prediction methods. In: Proceedings of the 21st international conference on world wide web, pp 607–608. https://doi.org/10.1145/2187980.2188150
Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P (2011) User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 1397–1405. https://doi.org/10.1145/2020408.2020614
Tang E (1996) A cultural framework of "chinese learn english'': a critical review of and reflections on. Compar Educ Rev 41(1):3–26
Thelwall M (2010) Emotion homophily in social network site messages. First Monday. https://doi.org/10.5210/fm.v15i4.2897
Tuan TM, Chuan PM, Ali M, Ngan TT, Mittal M et al (2019) Fuzzy and neutrosophic modeling for link prediction in social networks. Evol Syst 10(4):629–634. https://doi.org/10.1007/s12530-018-9251-y
Valverde-Rebaza J, de Andrade Lopes A (2012) Structural link prediction using community information on twitter. In: 2012 fourth international conference on computational aspects of social networks (CASoN). IEEE, pp 132–137. https://doi.org/10.1109/CASoN.2012.6412391
Wahid-Ul-Ashraf A, Budka M, Musial-Gabrys K (2017) Newton's gravitational law for link prediction in social networks. In: International conference on complex networks and their applications. Springer, pp 93–104. https://doi.org/10.1007/978-3-319-72150-7_8
Wang H, Zhang F, Hou M, Xie X, Guo M, Liu Q (2018) Shine: signed heterogeneous information network embedding for sentiment link prediction. In: Proceedings of the eleventh ACM international conference on web search and data mining, pp 592–600. https://doi.org/10.1145/3159652.3159666
Wang Z, Wu Y, Li Q, Jin F, Xiong W (2016) Link prediction based on hyperbolic mapping with community structure for complex networks. Physica A: Stat Mech Appl 450:609–623. https://doi.org/10.1016/j.physa.2016.01.010
Wu H, Sorathia V, Prasanna VK(2012) Predict whom one will follow: followee recommendation in microblogs. In: 2012 international conference on social informatics. IEEE, pp 260–264. https://doi.org/10.1109/SocialInformatics.2012.74
Yang B, Cheung W, Liu J (2007) Community mining from signed social networks. IEEE Trans Knowl Data Eng 19(10):1333–1348. https://doi.org/10.1109/TKDE.2007.1061
Yu Y, Wang X (2014) Link prediction in directed network and its application in microblog. Math Probl Eng. https://doi.org/10.1155/2014/509282
Zeng J, Zhang S, Wu C (2008) A framework for www user activity analysis based on user interest. Knowl-Based Syst 21(8):905–910. https://doi.org/10.1016/j.knosys.2008.03.049
Zhang M, Chen Y (2017) Weisfeiler-lehman neural machine for link prediction. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp 575–583
Zhang M, Chen Y (2018) Link prediction based on graph neural networks. In: Advances in neural information processing systems, pp 5165–5175
Zhao Z, Feng S, Wang Q, Huang JZ, Williams GJ, Fan J (2012) Topic oriented community detection through social objects and link analysis in social networks. Knowl-Based Syst 26:164–173. https://doi.org/10.1016/j.knosys.2011.07.017
Zhou J, Kwan C (2018) Missing link prediction in social networks. In: International symposium on neural networks. Springer, pp 346–354. https://doi.org/10.1007/978-3-319-92537-0_40
Zhou T, Lü L, Zhang Y-C (2009) Predicting missing links via local information. Eur Phys J B 71(4):623–630. https://doi.org/10.1140/epjb/e2009-00335-8