International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

# On Enhancement of Supervised Link Prediction in Social Networks using Topological Features and Node2Vec

Mohamed Badiy[1], Fatima Amounas[2], Moha Hajar[3]
[1]PHD Student, RO.AL&I Group, Faculty of Sciences and Technics, Moulay Ismail University, Errachidia, Morocco
[2]RO.AL&I Group, Computer Sciences Department, Faculty of Sciences and Technics, Moulay Ismail University, Errachidia, Morocco
[3]RO.AL&I Group, Mathematical Department, Faculty of Sciences and Technics, Moulay Ismail University, Errachidia, Morocco

***Abstract*—*Social network analysis has attracted much attention in today's world. Link prediction is major issue of Social network analysis. Large systems that illustrate connectivity between millions of social units are known as social networks. One of the key research areas in social network analysis is the study of their patterning. It involves the link prediction problem, which is the prediction of future relationships between unlinked node. There are different features in a social network that can be used for link prediction. The feature extraction step is one of the most crucial phases in the model for predicting the no existing links. Machine learning techniques have recently been discovered to be a good contender for solving the link prediction problem in social network analysis. In this paper, we attempt to propose the supervised link prediction approach with improved structural features. Our contribution considers a set of topological features and Node2Vec for training the supervised machine learning classifiers. The standard evaluation metrics such as accuracy, AUC, precision and F-measures have been used for validation purpose. The experimental results indicate that the combination of topological features and node2Vec generates the best performances.*

***Keywords*—*Social network Analysis, Link prediction, Social network, supervised learning, topological features, Node2Vec.***

## I. INTRODUCTION

In the last decade, Social Network Analysis (SNA) is a broad field of research dealing with techniques and strategies for the study and analysis of social network. Social network is a popular way to model the interaction among the people in a group or community. It can be represented by graph where node represents a user and edges represent relationship among various users. Due to the enormous growth and dynamics of social networks, social network analysis (SNA) has attracted significant attention in recent years. Social network analysis is in depth analysis of social networks. SNA allows us to analyze and determine the relations between the nodes (people, groups, organizations, etc) and many other connected entities which provide some knowledge and information. In particular, link prediction is the current trend to analyze the social network, which exploits existing networks information, like the characteristics of the nodes and edges to predict the potential links to be formed in the future. The problem of link prediction in graphs is one of the most interesting challenges. Link prediction is a process of predicting future connections among entities in the graph based on existing connections. Applications of link prediction include discovering interactions between proteins in bioinformatics [1], creation of the recommendation system for electronic commerce [2,3], prediction of transportation networks [4], forecasting the behavior of terrorism campaigns and social bots [5], and so on. Due to wide range of application in different areas, link prediction problem has attracted more attention for the researchers. Many link prediction algorithms have been proposed to solve this problem. There are various successful approaches on link prediction analysis in social networks. For example, the authors in [6] proposed the taxonomy of link prediction in social networks. They extensively studied the art of link prediction approaches. Figure 1 illustrates the different approaches of link prediction in social networks. According to the link prediction taxonomy, the approaches can be classified under four main categories: similarity, probabilistic, algorithmic and hybrid. The first one called "Similarity approaches" is based on the assumption that a node pair with higher similarity score is assumed to establish a link in the future. These approaches can be investigated under three main categories: local, quasi-local, and global approaches [7,8]. Next, the Probabilistic approaches are based on statistical probability model that fits with the network structure. The model, specified with parameters, computes a mathematical statistic to produce a probability value for each pair of nodes. Several probabilistic models have been proposed in the literature to predict missing links in the networks [9,10]. Algorithmic approaches are another high performance approach that is widely explored among researchers in the link prediction literature [11-15]. The algorithmic approaches often use extra information from the network and consider other factors that affect link formation. The Hybrid approaches are a combination of two approaches to predict future links. For instance, when a similarity approach is combined with an algorithmic approach or vice versa [16,17].

In the recent years, most of the existing works have used the supervised learning algorithm based on topological features of the network [18,19]. It has been observed from the literature that topological features are not enough to capture the link information in all the networks. Therefore, predicting the links is still challenging. So, it is necessary to obtain a prediction approach with high performance. In the present paper, we propose an approach of supervised link prediction.



Mohamed Badiy, Fatima Amounas, and Moha Hajar, "On Enhancement of Supervised Link Prediction in Social Networks using Topological Features and Node2Vec," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 1, pp. 54-60, 2022.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

The motivation behind the proposed approach is to effectively utilize the hybridization of topological information of the network and Node2Vec for enhancing the link prediction accuracy.

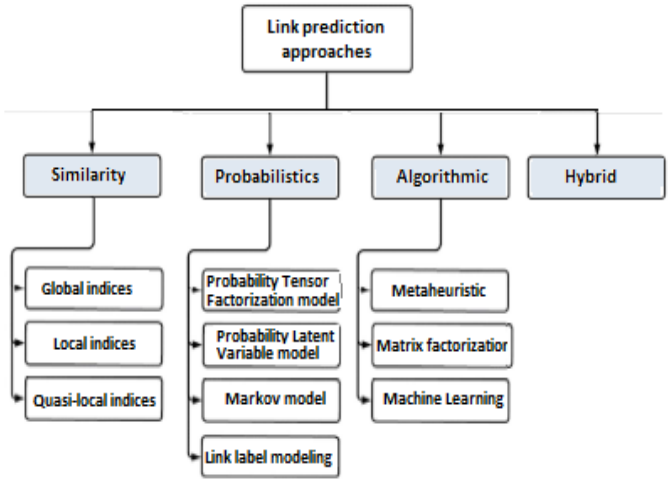


Fig. 1. Link prediction approaches

The remainder of this paper is organized as follows: Section 2 summarizes the previous works related to link prediction in social networks. Section 3 investigates the basic theory of link prediction approaches. Section 4 describes the proposed approach and presents the implementation part of the work. The experimental results and the analysis part have been also discussed in this section. The last section presents conclusion and future work.

## II. Related Works

Link prediction problem is found to be one of the broad areas for researchers in the recent years. Due to its importance in social network analysis, several techniques are adopted by many authors to solve the link prediction problem, but still there is a scope to improve the previous approaches. In the field of link prediction, the first one introduced by Liben-Nowell and al. [20], where the authors observed that the topological features in greatly effective for predicting future link in the network. Subsequently, Deylami and al. [21] extended the survey by considering community information for improve the performance of similarity-based link prediction methods. However, the result shows that the information of communities often increases the efficiency and accuracy of link prediction. Furthermore, the authors in [22] proposed dimensionality reduction to a global based approach for link prediction in co-authorship network DBLP. Next, the authors in [23] proposed an efficient hybrid approach using two similarity-based link prediction methods, common neighbors and preferential attachment to improve the performance of link prediction. Recently, machine learning is one of the emerging solutions to solve link prediction problem in social network analysis. In this context, many researchers have used supervised machine learning methods for link prediction in social network. Pecli and al. [24] applied three automatic variables selection strategies on the supervised learning approach in link prediction, the result shows that their feature selection method improves the efficiency of classifiers. Fu and al. [25] have used weighted network and some learning models like GBDT, Random Forest (RF) and Support Vector Machine (SVM). Butun and Kaya [26] proposed a supervised link prediction approach to improve link prediction accuracy of Triad Closeness in directed complex networks. Rezaeipanah and al. [27] proposed an efficient solution to the problem of link prediction in multiplex social network by considering two layer social networks. Furthermore, Kumari and et al. [28] implemented many supervised link prediction algorithms to predict the possibilities of establishing the links in future. They have utilized both local and global approaches as the features vectors for machine learning model. Malhotra Deepanshu and Rinkaj Goyal [29] proposed a novel technique to predict the future links in single-layer and multiplex networks. They constructed a set of topological features and utilize them for the supervised learning approaches. The multiplex network makes use of data from many layers to increase the algorithm's performance.

## III. Background Information

In this section, we recall briefly the notion of link prediction problem and we introduce some of the useful methods for the prediction of nonexistence links, which are likely to appear in the future. Similarity-based link prediction metrics can be classified into two types: Node-based similarity measure and Path-based similarity measure.

### A. *Link prediction problem*

Link prediction problem can be defined as a binary classification problem, where non-existing links are classified as Predicted (P) or Not Predicted (NP). Considering a given social network G = (V, E) where V denotes set of users and E denotes set of relationships among the users. The objective is to find out the label of non-existing links $E_{xy}$. It can be mathematically formulated as below:

$$\begin{cases} \mathrm{L} = \mathrm{F}(v_x, v_y)\ \forall (x \neq y) \\ E_{xy} \notin E(G) \end{cases}$$

Where F is a function used to predict the label through various link predictors, L is the binary class label (either P or NP), $E_{xy}$ is the non-existence edges between $v_x$ and $v_x$. More precisely, if there is an snapshot of a social network at time t, the purpose of link prediction is to predict the edges that are likely to be established in the period between t and t+1 (t < t+1), as shown below:

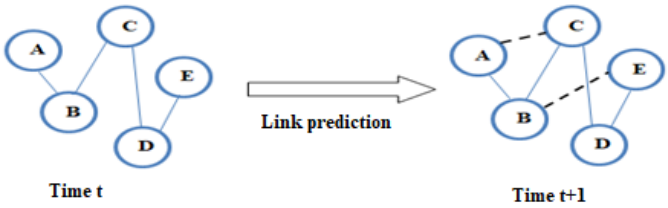


Fig. 2. Example of link prediction



Mohamed Badiy, Fatima Amounas, and Moha Hajar, "On Enhancement of Supervised Link Prediction in Social Networks using Topological Features and Node2Vec," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 1, pp. 54-60, 2022.

### B. Similarity Based Methods

Similarity-based methods are the simplest one in link prediction. The basic idea behind these approaches is that the similarity score is measured by considering either local or global information in the network. According to the characteristics of these methods, they can be divided into two categories: local similarity-based methods and Global similarity-based methods. Here, we present some notations that we use in the following definitions: let N(x) and N(y) are neighbors set of x and y, and let |N(x)| and |N(y)| are number of neighbors of node x and y.

#### 1) Local similarity-based metrics

Local similarity-based metrics are based on the assumption that, if node pairs have common neighbor structures, they will probably form a link in the future. Some of the important metrics are identified below:

- *Common Neighbors* (CN): The common neighbor between a pair of vertices is the length of the set of vertices that have a connection to both the given vertices. It can be calculated using the equation below:

$$CN(x,y)= |\mathbf{N(x)} \cap \mathbf{N(y)}|$$

- *Jaccard coefficient* (JC): the Jaccard index normalizes the size of common neighbors by considering the total of numbers of shared and non-shared neighbors [30]. It can be defined as follows:

$$JC\ (x, y) = \frac{|\mathbf{N(x)} \cap \mathbf{N(y)}|}{|\mathbf{N(x)} \cup \mathbf{N(y)}|}$$

- *Preferential attachment* (PA): the preferential attachment index calculates the similarity score for each pair of nodes using their degree information. It can be computed as below:

$$PA\ (x, y) = |\mathbf{N(x)}|.|\mathbf{N(y)}|$$

- *Adamic Adar* (AA): Initially, this method was proposed by Adamic and Adar for calculation of the similarity between two web pages [31]. It calculates the common neighbors by assigning more weight to the neighbor with fewer connections. It can be mathematically represented as follows:

$$AA\ (x, y) = \sum_{z \in |\mathbf{N(x)} \cap \mathbf{N(y)}|} \frac{1}{Log(|\mathbf{N(z)}|)}$$

- *Resource allocation Index* (RA): this metric proposed by zhou and al. [32], is very similar to adamic and adar index. The Resource Allocation uses the degree magnitude instead of logarithmic value. It can be calculated as follows:

$$RA\ (x, y) = \sum_{z \in |\mathbf{N(x)} \cap \mathbf{N(y)}|} \frac{1}{|\mathbf{N(z)}|}$$

#### 2) Global similarity-based metrics

The Global similarity-based metrics called "path-based methods" use the whole topology of the network to compute the similarity between node pairs. Some of the important metrics are defined below:

- *Shortest Path*: this measure determined the similarity of two nodes x and y by takes into account the shortest path between x and y, reflecting the expectation that nodes with low degrees of separation are likely to interact.
- *Katz measure*: the katz index can be considered as a variant of the shortest path metric [33]. It counts all paths between two nodes and discounts the longer path exponentially, and can be calculated as:

$$\text{Score } (\mathbf{x}, \mathbf{y}) = \sum_{l=1}^{\infty} \beta^l . \left| paths_{x,y}^{l} \right|$$

where Score(x,y) is the link predicted score using Karz measure, β is the small constant value, and $paths^{l}_{x;y}$ is the set of all paths lengths from x to y.

- SimRank: the SimRank index computes the similarity starting from the hypothesis that two nodes are similar if they are connected to similar nodes [34]. It is defined as follows:

$$\text{Score } (x, y) = \begin{cases} 1, & si\ x = y \\ \gamma . \frac{\sum_{a \in N(x)} \sum_{b \in N(y)} \text{Score}(x,y)}{|\,N(x)|.|\,N(y)|} & \text{Otherwise} \end{cases}$$

where Score (x,y) is the similarity score between x and y using SimRank, is the real number lies between 0 to 1.

### C. Supervised Machine Learning Methods

Machine supervised learning is one of the main paradigms of link prediction, which is based on three different approaches: the topological structure of the network, the semantic similarity among the properties of the vertices, and the description of the network behavior by means of probabilistic models. In this section, we briefly review the supervised machine learning techniques that we use in our experiments.

- *Random Forest* (RF): is one of the most effective machine learning algorithms for classification introduced by Breiman [35]. Random Forest is composed of multiple independent decision trees where every tree depends on the random vector distributed among all trees in forest.
- *Support vector machine* (SVM): is a supervised machine learning algorithm that can be used for classification learning tasks. This algorithm divides data into classes by determining the optimum hyperplane that separates all data points in one class from those in the other.
- *K-Nearest Neighbor* (KNN): is one of the most used learning algorithms proposed by Fix and Hodges [36]. The basic assumption of this algorithm is that similar things are closer to each other. The distance between the data points is calculated mathematically Euclidean measure or any other measures. The parameter k depends on the chosen number of neighbors.
- *Bagging*: is also called as Boostrap Aggregation is one of the most famous and successful learning methods. The main idea of bagging is to create the various classifiers and then combine these classifiers. This method selects some classifier algorithms for training data sets process.
- *Light gradient boosting machine* (Lightgbm) : is one of the most successful machine learning algorithms based on Gradient Boosting Decision Tree (GBDT) proposed by Microsoft in 2017 [37]. It has the advantages of improved faster training efficiency, lower memory usage, higher accuracy, and large-scale data processing. This method is based on two techniques, namely Gradient-Based oneside sampling (GOSS) and Exclusive Feature Bundling (EFB).



Mohamed Badiy, Fatima Amounas, and Moha Hajar, "On Enhancement of Supervised Link Prediction in Social Networks using Topological Features and Node2Vec," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 1, pp. 54-60, 2022.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

- *Multilayer Perceptron* (MLP): is an artificial neural network, which consists of three or more layers an input and an output layer with one or more hidden layers. MLP uses the supervised training method called back propagation training. It adjusts the weight values that are calculated from input-output mappings and minimize the error between the correct output value and the target value.

## IV. Main Results

### A. Proposed Approach

The main intention of this paper is to present a supervised approach with some features vectors to solve link prediction problem in social network. More precisely, we evaluate the effectiveness of supervised learning methods using topological features and Node2Vec for link prediction in social network. Our goal is to see how effective these features are in terms of predicting the future links in a social network. Now, we discuss the proposed approach in greater detail to explain its working. Here, we use six well-known supervised machine learning methods. These methods include Multilayer Perceptron (MLP), Support vector machine (SVM), K-Nearest Neighbor (KNN), Light gradient boosting machine (Lightgbm), Random Forest (RF) and Bagging (Boostrap Aggregation). The proposed architecture of the link prediction process using the supervised learning algorithms is shown in Figure 3.

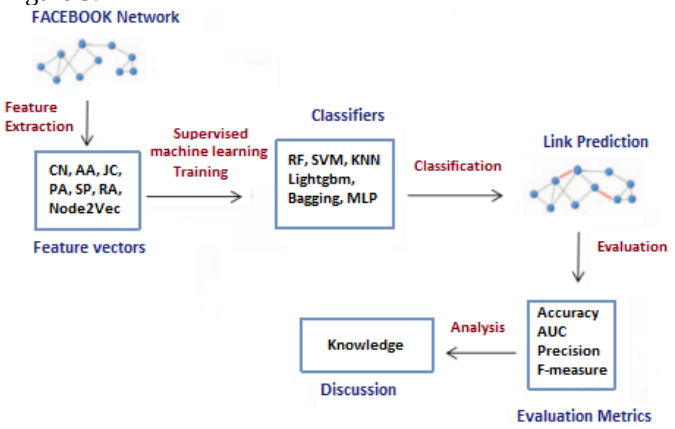


Fig. 3. Architecture of the proposed approach

In our case, the various steps involved in the link prediction process are given as follows:
*Step 1*: extract the features vectors from the Facebook dataset.
*Step 2*: represents the training of the choosing classifiers.
*Step 3*: is the classification process itself in which the new links are predicted.
*Step 4*: presents the evaluation stage where different metrics are used to quantify the efficiency of the classifiers.
*Step 5*: consists of the analysis the obtained results.

### B. Implementation

In this paper, we evaluate the effectiveness of topological features and Node2Vec for link prediction in social networks. The considered topological features are CN, AA, JC, PA, SP and RA. Another effective feature for link prediction is Node2Vec algorithm. Here, we attempt to extract the mentioned features from the Facebook dataset and construct the training dataset and test data. For evaluation purposes, we adopted the different evaluation measures like accuracy, AUC, Precision and F-measure.

#### 1) Data set considered

We used the Facebook dataset in our experiments to evaluate the performance of our models. This data was provided by Stanford Network Analysis Project (SNAP), it is available for download at [38]. Facebook is an undirected and unweighted social network that represents user friendship. In this network, each user is represented by node and friendship is represented by associated link between two nodes. The basic parameters of Facebook network include the number of nodes, number of edges and average clustering coefficient. Table I shows the statistics of the dataset considered.

TABLE I. Data set considered

| Dataset | Number of nodes | Number of edges | Average clustering coefficient |
|---|---|---|---|
| Facebook | 4039 | 88234 | 0.6055 |

The Figure 4 shows the visualization of Facebook dataset using NetworkX and MatplotLib. The Graphical representation shows the link relationships for all nodes in the network.

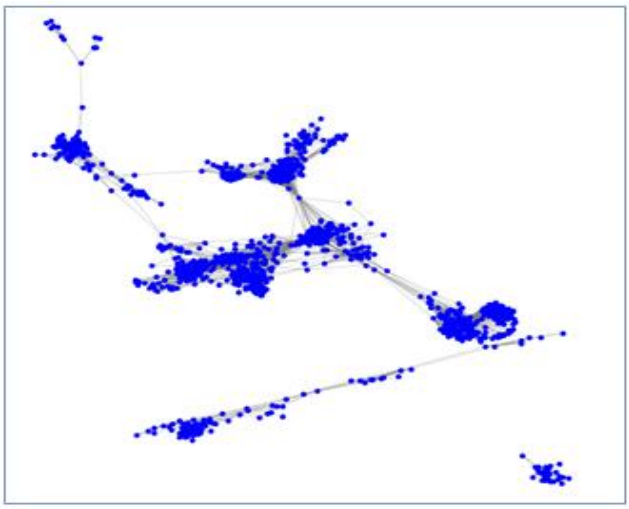

Fig. 4. Network diagram of Facebook dataset

#### 2) Feature Selection

In the link prediction problem, a set of features have be defined in order to be used in building the classification model. It very important to choose the right features for any classifier. In our work, we have used two types of features, topological features and node embedding features. Here, we considered six topological features including Common Neighbors (CN), Jaccard coefficient (JC), Preferential attachment (PA), Adamic Adar (AA), Resource allocation Index (RA) and Shortest Path (SP) as mentioned in section 3. We also use Node2Vec as node embedding features, which represent an embedding method that transforms networks into numerical representations. Node2Vec uses graph embedding approach and generates features by simulating biased random



Mohamed Badiy, Fatima Amounas, and Moha Hajar, "On Enhancement of Supervised Link Prediction in Social Networks using Topological Features and Node2Vec," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 1, pp. 54-60, 2022.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

walks in a vector representation, and storing each node in a d-dimensional space according to their features [39].

*3) Performance evaluation Metrics*

In this section, we present the performance measures that we use for the evaluation of our models.

- *Accuracy*: is defined as the ratio of number of links that are correctly classified to the total number of the instances in a given dataset.
- *AUC* (Area Under Curve): this can be interpreted as the probability that a randomly chosen future edge in the test graph is given a higher score than a randomly chosen nonexistent link in the test graph. Thus, we pick a missing link and a non-existent link and compare their scores repeatedly.
- *Precision*: is the ratio of TP to the summation of TP and FP.
- *F-measure*: is considered to be the harmonic mean of the precision and recall. Here the Recall represents the proportion of True Positives that were correctly identified.

Mathematically, the previous measures are illustrated in Table II. Some of the notations related to evaluation metrics are discussed as below:

- TP: number of links that are labeled as predicted and also actually predicted by the link predictor.
- FP : number of links that labeled as not to be predicted but are being predicted by the link predictor.
- TN : number of links that are labeled as not to be predicted and are actually not predicted by the link predictor.
- FN : number of links that are labeled as to be predicted and are not predicted by the link predictor.
- n: the number of comparisons.
- n': the number of times the future edge has a higher score.
- n": the number of times they have the same score.

TABLE II. Evaluation Metrics

| Evaluation metric | Mathematical representation |
|---|---|
| Accuracy | Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ |
| AUC | AUC = $\frac{n' + 0.5\, n''}{n}$ |
| Precision | Precision = $\frac{TP}{TP+FP}$ |
| F-measure | F-measure = $\frac{2TP}{2TP+FP+FN}$ |

## V. Experimental Results and Discussion

In this section, we explain how we performed our experiments. We also analyze the results obtained from practical implementation. All the experiment has been performed on a i5 processor having 3.2GHz clock rate. All algorithms have been tested in "google colab" environment.
After preparing our data and extracting our favorite features, we applied the supervised learning algorithms to predict the links in the Facebook dataset. As mentioned earlier, we measured the performance of our models using different evaluation metrics. Table III shows the performance of different supervised learning models on the training dataset.

TABLE III. Performance of the models on the Training data set

| Model | Accuracy | AUC | Precision | F-measure |
|---|---|---|---|---|
| MLP | 0.993 | 0.994 | 0.985 | 0.990 |
| SVM | 0.948 | 0.932 | 0.955 | 0.919 |
| KNN | 0.949 | 0.947 | 0.910 | 0.925 |
| Lightgbm | 0.909 | 0.872 | 0.963 | 0.849 |
| RF | 0.992 | 0.985 | 0.998 | 0.997 |
| Bagging | 0.988 | 0.983 | 0.999 | 0.982 |

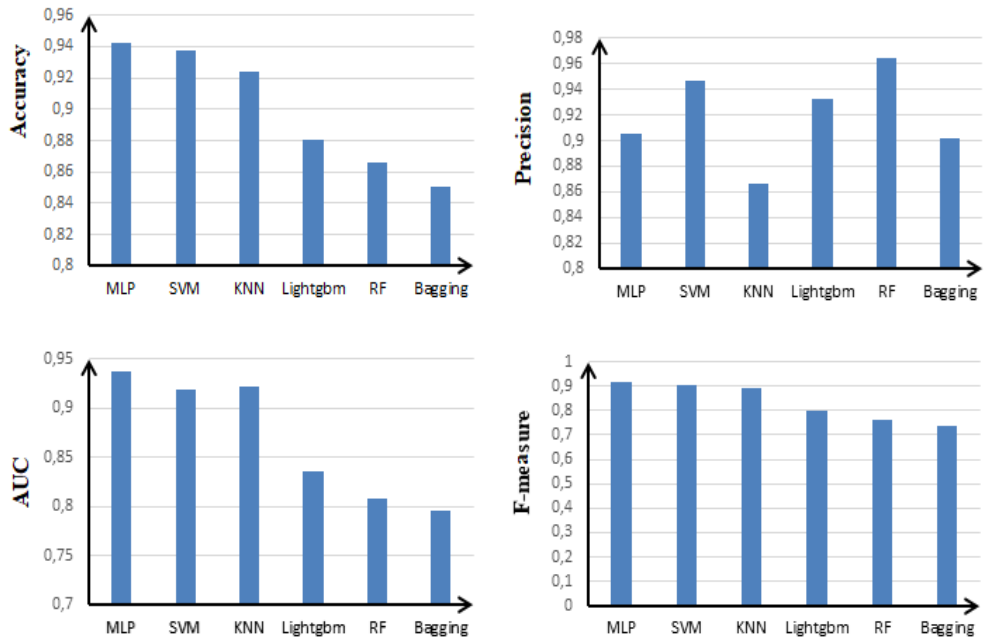


Fig. 5. Performance analysis of different link prediction models.



Mohamed Badiy, Fatima Amounas, and Moha Hajar, "On Enhancement of Supervised Link Prediction in Social Networks using Topological Features and Node2Vec," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 1, pp. 54-60, 2022.

From the above results, it observed that all the models, except the Lightgbm model have a performance greater than 90%. Our results show that the MLP model achieved 99% in all metrics, and Lightgbm model has the lowest AUC and F-measure values (0.872 and 0.849 respectively). Next, we investigated the performance of different models on our test data. The obtained results are summarized in the Table IV. Graphical representation of the above described table is shown in Figure 5.

TABLE IV. Performance of the models on the Testing dataset

| Model | Accuracy | AUC | Precision | F-measure |
|---|---|---|---|---|
| MLP | 0.942 | 0.937 | 0.905 | 0.915 |
| SVM | 0.937 | 0.918 | 0.946 | 0.902 |
| KNN | 0.923 | 0.922 | 0.866 | 0.890 |
| Lightgbm | 0.881 | 0.835 | 0.932 | 0.797 |
| RF | 0.866 | 0.807 | 0.964 | 0.759 |
| Bagging | 0.852 | 0.795 | 0.902 | 0.738 |

From the results obtained on the test data set, it can be observed that MLP and SVM algorithms provide better scores for all evaluation metrics as compared with other supervised machine learning algorithms. KNN achieves the best accuracy and AUC values around 92%. Although the other models (Lightgbm, RF and Bagging) achieve the best precision value (0.932, 0.964 and 0.902 respectively). In general, it is observed that all the classifiers achieve greater than 80% accuracy, AUC and precision scores. This shows that the features extracted a wide range of data for the machine learning classifier to perform a superior performance.

Finally, it can be noted from the analysis of link prediction results mentioned above that the hybridization of the topological features and Node2Vec helps to achieve the better performances in link prediction.

## VI. Conclusion

Link prediction is growing as a promising research field in the social network analysis domain. It has applications in other fields such as ecommerce, information retrieval, bioinformatics and recommendation systems. There are different features that can be used for the link prediction. Most published link prediction methods cannot take into account the effect of features on link prediction in social networks. In this paper, we suggested an approach of supervised link prediction using the topological features and Node2Vec as features vectors in social networks. Some of similarity-based methods and Node2Vec are considered combinely as features in the link prediction of the future link through several machine learning models. We evaluated the effectiveness of these features for supervised link prediction process in the Facebook dataset. Our experiments on supervised link prediction revealed satisfactory results and shows better performance as compared to the existing link prediction methods [18,19]. Based on our experiments, we observed that the combination of topological features and Node2Vec is a very effective feature for supervised link prediction. Finally, we like to point out that the use of these features will provide better performance in this regard. As a future work, the proposed approach can be extended for link prediction in big social networks. Furthermore, the future work should be focused to analyze other types of social network features and to examine their impact on link prediction.

## References

[1] Nasiri, Elahe, and al. “A novel link prediction algorithm for protein protein interaction networks by attributed graph embedding.” Computers in Biology and Medicine, 137, 104772, 2021.

[2] Folino, F. and Pizzuti, C., “Link prediction approaches for disease networks”, In International Conference on Information Technology in Bio-and Medical Informatics. Springer, Berlin, Heidelberg, pp. 99-108, 2012.

[3] Esslimani, I., Brun, A. and Boyer, A., “Densifying a behavioral recommender system by social networks link prediction methods”, Social Network Anal. Mining, vol 1, Issue 3, pp.159-172, 2011.

[4] Ma, Y.; Liang, X., Huang, J., Cheng, G. “Intercity transportation construction based on link prediction”, In Proceedings of the 2017 IEEE 29th International Conference on Tools with Artificial Intelligence, pp. 1135-1138, 2017.

[5] Heidari, Maryam, and James H. Jones. “Using BERT to extract topic independent sentiment features for social media bot detection.”, 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, IEEE, pp. 0542-0547, 2020.

[6] Nur Nasuha Daud and al., “Applications of Link Prediction in Social Networks: A Review”, Journal of Network and Computer Applications, Vol.166, 102716, 2020.

[7] Hou, Lei, and Kecheng Liu. “Common neighbour structure and similarity intensity in complex networks.” Physics Letters A 381.39 pp. 3377-3383, 2017.

[8] Muniz, Carlos Pedro, Ronaldo Goldschmidt, and Ricardo Choren, “Combining contextual, temporal and topological information for unsupervised link prediction in social networks.” Knowledge-Based Systems, vol. 156, pp. 129-137, 2018.

[9] Javari, Amin, et al. “Statistical link label modeling for sign prediction: Smoothing sparsity by joining local and global information.” 2017 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 1039-1044, 2017.

[10] Das, Sima, and Sajal K. Das. “A probabilistic link prediction model in time-varying social networks.” 2017 IEEE International Conference on Communications (ICC). IEEE, pp. 1-6, 2017.

[11] Srilatha, P., and R. Manjula. “Structural similarity based link prediction in social networks using firefly algorithm.” 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE, pp. 560-564, 2017.

[12] Li, Taisong, et al. “Restricted Boltzmann machine-based approaches for link prediction in dynamic networks.” IEEE Access, vol. 6, pp. 29940-29951, 2018.

[13] Shao, Hao, Lunwen Wang, and Yufan Ji. “Link prediction algorithms for social networks based on machine learning and HARP.”, IEEE Access, vol. 7, pp. 122722-122729, 2019.

[14] Ghorbanzadeh, H., Sheikhahmadi, A., Jalili, M., and Sulaimany, S. “A hybrid method of link prediction in directed graphs.”, Expert Systems with Applications, vol. 165, 113896, 2021.

[15] RAEINI, Mohammad G. “Link Prediction Using Supervised Machine Learning based on Aggregated and Topological Features”, arXiv preprint arXiv:2006.16327, 2020.

[16] Wang, Zhiqiang, Jiye Liang, and Ru Li.”A fusion probability matrix factorization framework for link prediction.” Knowledge-Based Systems, vol. 159, pp. 72-85, 2018.

[17] Srilatha, P., and R. Manjula. “Structural similarity based link prediction in social networks using firefly algorithm.” 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon). IEEE, pp. 560-564, 2017.

[18] Sa, Hially Rodrigues de and Ricardo B. C. Prudencio, “Supervised link prediction in weighted networks”, The 2011 International Joint Conference on Neural Networks, pp.2281-2288, 2011.

[19] Pujari M, Kanawati R. “Link prediction in complex networks by supervised rank aggregation”. In: Proceedings of IEEE 24th

International Conference on Tools with Artificial Intelligence, Athens, Greece, pp.782-789, 2012.
[20] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," Journal of the American society for information science and technology, vol. 58, no. 7, pp. 1019–1031, 2007.
[21] H. A. Deylami and M. Asadpour, "Link prediction in social networks using hierarchical community detection", 7th Conference on Information and Knowledge Technology (IKT), pp. 1-5, doi: 10.1109/IKT.2015.7288742, 2015.
[22] Coskun, M. and Koyuturk, M. "Link Prediction in Large Networks by Comparing the Global View of Nodes in the Network", In 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp.485-492, 2015.
[23] Zeng, S. "Link prediction based on local information considering preferential attachment", Physica A: Statistical Mechanics and Its Applications, 443, 537-542, 2016.
[24] Pecli, A., Cavalcanti, M. C. and Goldschmidt, R. "Automatic feature selection for supervised learning in link prediction applications: a comparative study", Knowledge and Information Systems, vol. 56, No.1, pp.85-121, 2018.
[25] Fu, Chenbo and al. "Link Weight Prediction Using Supervised Learning Methods and Its Application to Yelp Layered Network", IEEE Transactions on Knowledge and Data Engineering, vol. 30, No. 8, pp.1507-1518, 2018.
[26] Butun E, Kaya M. "A pattern based supervised link prediction in directed complex networks", Physica A: Statistical Mechanics and its Applications, vol.525, pp.1136-1145, 2019.
[27] Rezaeipanah, A., Ahmadi, G. and Matoori, S. S. "A classification approach to link prediction in multiplex online ego-social networks", Social Network Analysis and Mining, vol. 10, No 1, pp.1-16, 2020.
[28] Kumari, A., Behera, R. K., Sahoo, K. S., Nayyar A., Kumar Luhach A. and Prakash Sahoo, S. "Supervised link prediction using structured based feature extraction in social network". Concurrency and Computation: Practice and Experience, e5839, 2020.
[29] Malhotra Deepanshu, and Rinkaj Goyal, "Supervised-learning link prediction in single layer and multiplex networks", Machine Learning with Applications, vol. 6, 100086, 2021.
[30] Cukierski, William, Benjamin Hamner, and Bo Yang. "Graph-based features for supervised link prediction", The 2011 International Joint Conference on Neural Networks. IEEE, pp. 1237-1244, 2011.
[31] Adamic, Lada A., and Eytan Adar. "Friends and neighbors on the web", Social networks vol. 25, Issue 3, pp. 211-230, 2003.
[32] Zhou, T., L¨u, L. and Zhang, YC. "Predicting missing links via local information" Eur. Phys. J. B 71, pp. 623-630, 2009.
[33] H. Tong, C. Faloutsos, J.-Y. Pan, "Fast random walk with restart and its applications", in: Proceedings of the Sixth International Conference on Data Mining, in: ICDM '06, IEEE Computer Society, Washington, DC, USA, pp. 613-622, 2006.
[34] Jeh, Glen, and Jennifer Widom. "Simrank: a measure of structural context similarity", Proceedings of the eighth ACM SIGKDD, International conference on Knowledge discovery and data mining, pp. 538-543, 2002.
[35] Breiman, L. "Random Forests", Machine Learning, vol. 45, pp. 5-32, 2001.
[36] Fix E, Hodges J L. "Discriminatory analysis-nonparametric discrimination: Consistency properties" International Statistical Review, vol. 57, No. 3, pp. 238-247, 1989.
[37] Meng, Qi et al. "A Communication-Efficient Parallel Algorithm for Decision Tree", NIPS, 2016.
[38] "Facebook wall posts network dataset - KONECT," Apr. 2017.
[39] Grover, Aditya and Jure Leskovec, "Node2vec: Scalable feature learning for networks", Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016.