






NODDLE: Node2vec Based Deep Learning Model for Link Prediction

Kazi Zainab Khanam , Aditya Singhal  , and Vijay Mago 

Lakehead University, Thunder Bay, ON P7B 5E1, Canada
{kkhanam, asinghal, vmago}@lakeheadu.ca

Abstract. Computing the probability of an edge's existence in a graph network is known as link prediction. While traditional methods calculate the similarity between two given nodes in a static network, recent research has focused on evaluating networks that evolve dynamically. Although deep learning techniques and network representation learning algorithms, such as node2vec, show remarkable improvements in prediction accuracy, the Stochastic Gradient Descent (SGD) method of node2vec tends to fall into a mediocre local optimum value due to a shortage of prior network information, resulting in failure to capture the global structure of the network. To tackle this problem, we propose NODDLE (integration of NDe2vec and Deep Learning mEthod), a deep learning model which incorporates the features extracted by node2vec and feeds them into a four layer hidden neural network. NODDLE takes advantage of adaptive learning optimizers such as Adam, Adamax, Adadelta, and Adagrad to improve the performance of link prediction. Experimental results show that this method yields better results than the traditional methods on various social network datasets.

Keywords: Graph learning · Social networks · Link prediction · Web information systems

1 Introduction

Link prediction is a fundamental problem of network analysis, mainly because of its importance in social network applications such as designing recommendation systems for social media platforms and e-commerce websites, identification of credit card fraud, and even locating terrorist groups based on their criminal activities [3, 10, 12, 32, 37, 44]. The field of bioinformatics often uses link prediction for predicting protein-protein interactions containing important information about biomolecular behavior. Such interactions can reveal answers about diseases and cures [4], and therefore, predicting such upcoming links is a crucial component of graph mining.

This research is funded by NSERC Discovery Grant (RGPIN-2017-05377), held by Dr. Vijay Mago.

The main objective of the link prediction problem is to predict the unseen edges that will emerge in a graph. Based upon the *snapshot assumption*, when a snapshot of a graph $G(t)$ at time t is given, link prediction is used to compute which new upcoming links will emerge in the future graph $G(t')$ within the time period $[t, t']$, where $t' = t + n$ (n is the sequence of snapshots) [34]. Link prediction is implemented on real-world network graphs, which are often too massive and *dynamic*, as such graphs are evolving at an extremely high speed. In addition, link prediction uses proximity-based measures, such as the Jaccard coefficient, Resource Allocation, and Adamic Adar metric, to measure the probability of the upcoming links in the network [8]. The features extracted are based on the local nodal properties as these functions use the information available only from the local proximity of the nodes. Although these metrics are used widely in multiple applications because of their simplicity and interpretability, the problem arises when social network graphs become large with numerous users. As a result, predicting future links with these measures becomes a very challenging task. Most importantly, hidden and meaningful knowledge lies between the nodes and edges of networks [4], and analyzing these graphs is extremely difficult when large-scale network data comprises billions of nodes and edges [42].

Traditional approaches calculated link prediction statically by using only a single network snapshot to predict future links. However, the prediction task is a time-dependent problem, where a network evolves over time [38]. Hence, the dynamic network concept was initiated in which the structure of the network is captured in multiple snapshots over a span of time [50]. Dynamic-link prediction is considered more valuable and challenging than static link prediction. The evolution of the network structure offers much more information that adds a whole new dimension in network analysis and helps achieve a better link prediction performance [39]. The problem arises when the number of edges and nodes increases at a faster rate as it becomes very challenging to extract or infer any reasoning and information from the whole network [19]. Dimensional reduction techniques have been used to solve this issue, which transform the nodes of a graph into lower-dimensional latent representations [6]. These representations can be used as features for executing tasks in graph mining, such as clustering and link prediction [45]. Similarly, network representation learning algorithms such as node2vec have also been used to tackle this issue. Node2vec conducts high order proximity by escalating the probability of finding successive neighboring nodes within a fixed length of random walk [19]. This method can efficiently find the equilibrium position between breadth-first search (BFS) and depth-first search (DFS) graphs by developing random biased walks. As a result, it can succeed in embedding rich quality data, enabling node2vec to preserve the structural balance of the node communities.

Although node2vec has successfully achieved high link prediction performance, it still has many shortcomings [14]. Firstly, it follows a local approach that takes short random walks to get exposed to only the local neighborhood of nodes [13] and hence ignores the global relationship of nodes that might have

longer distances. Due to this, the learned representation may be unable to comprehend the essential global structure of the model. Secondly, node2vec uses the Stochastic Gradient Descent (SGD) method to resolve non-convex optimization problems, where the non-convex constraints may have various regions and many locally optimal points within each region [7, 23]. The algorithm repeatedly gets updated when SGD is used to optimize the objective function. This causes the optimal points to oscillate frequently and possibly causes them to get stuck in a local minima. Due to the complexity of the growing networks, recent research has focused on applying deep learning techniques to evaluate the complex relationships that exist in graphs and visualize the hidden patterns [43]. To tackle these problems we propose NODDLE (integration of **NO**de2vec and **D**eep Learning **m**ethod), a deep learning model that combines the features extracted by node2vec algorithm and feeds them into four layers of hidden neural network. It optimizes the performance by using different types of optimizers, which include Adaptive Moment Estimation (Adam), Adamax, An Adaptive Learning Rate Method (Adadelta), and Adaptive Gradient Algorithm (Adagrad). We have compared our approach with the benchmark methods that include Adamic Adar, Jaccard coefficient, and Preferential Attachment [17, 27, 47].

The rest of the paper is organized as follows. Section 2 presents background on the previous studies conducted on link prediction of social networks with heuristic-based, machine learning, and deep learning approaches. Section 3 introduces our proposed approach in detail, including our data preparation method and the method for combining node2vec with the deep learning model. Later, in Sect. 4, we validate our approach on real-world social network data and analyze the results. Finally, Sect. 5 concludes the paper.

2 Related Work

2.1 Heuristic Similarity Metrics

Liben-Nowell and Kleinberg proposed a link prediction problem for social networks using multiple heuristic functions [32]. They found that topological features can be used to predict a future edge between two nodes that showed high “similarity” or “proximity” between the target nodes. Furthermore, their findings conveyed that the heuristics such as Adamic/Adar and Katz centrality measure notable correlation with the predicted future links [2, 24].

Many research works emphasized enhancing the performance of heuristic functions by increasing the neighbor-based attributes to second, third, or higher adjacency degrees. For instance, Yao et al. presented an improved common neighbors heuristic algorithm that includes nodes with a distance of two hops and used time-decay for recent snapshots to have a greater weight [46]. Kaya et al. used progressive events to calculate the possibility of future links in a time-weighted

fashion [25], and Deylami and Asadpour proposed a community detection algorithm to identify high activity clusters [15]. Similarity metrics have also been used to detect social and cognitive radio network events for common link prediction problems [21, 22, 49].

2.2 Machine Learning and Deep Learning

Both supervised and unsupervised techniques have been employed to predict links in the network. Unsupervised methods comprise developing the heuristic approaches to determine the score for the likelihood of each upcoming link [41]. Similarity metrics are most commonly used to measure the intensity of the relationship between the nodes. Topological features of the nodes such as common neighbors and graph distances are used to measure the strength of the interaction between the nodes [5]. Conversely, supervised methods involve treating the link prediction problem as a binary classification task in which the edges and non-edges of a network model are employed for training a classifier [29].

Compared to heuristic-based approaches, machine learning techniques have proven better at link prediction tasks as these models have received higher accuracy. Yet, the major problem with them is representing the graphical features, since it is impossible to use the large-scale graphs as input into the machine learning models. As a result, researchers have attempted to extract features. For example, Hasan et al. have extracted multiple graph features and implemented the features with various machine learning algorithms such as Decision trees, Naive Bayes, and k-Nearest Neighbors [5]. Similarly, Bechettara et al. have implemented topological-based features of bipartite graphs with decision trees [9], and Doppa et al. proposed a supervised feature vector-based approach with k-means classifier for link prediction [16]. Even though machine learning techniques have been shown to achieve better prediction accuracy, these methods rely highly on features developed by human intelligence. Thus, engineering such features is extremely tedious and slow. As a result, most state-of-art link prediction techniques utilize deep neural networks for their exceptional learning ability.

A deep neural network model is defined as a group of models in machine learning consisting of multiple connected layers. The layers generate output-yielding nodes where the parameters of the neural network layers are tuned in continuous iterations to reduce the error between the final output and the original value [11]. Li et al. have explored a neural network structure as a conditional temporal Restricted Boltzmann Machine (ctRBM), which expands on the architecture of an RBM to integrate the temporal elements of a dynamic changing network [31]. Furthermore, Zhang et al. suggested the neural network model as a means of feature representation by using the term Social Pattern and External Attribute Knowledge (SPEAK); these features are used as input in deep neural network models [48]. Ozcan A has proposed a link prediction algorithm that extracts multi-variable features from heterogeneous networks and is based

upon non-linear autoregressive neural networks [36]. This method was tested on various datasets and has outperformed the existing algorithms that focus on only single variable features. Zhang et al. proposed a framework that uses graph neural networks to learn general graph features for link prediction [48]. Graph neural networks are defined as a message-passing algorithm, in which the message represents the features extracted from each node in a graph, and their effects on the edges and nodes are learned by neural networks [35]. Their framework has also shown promising results in the online social networking Stanford Facebook dataset [48]. Therefore, state-of-the-art research has mainly focused on learning multiple features from graphs at an extensive level as such features contain hidden and meaningful insights into link probability. With the rise of complex growing networks, deep learning techniques have produced highly accurate results. Besides, deep learning can model the complex relationships hidden in the network data and can reveal unseen patterns hidden beneath the billions of nodes and edges [38].

Further research is being conducted to improve link prediction performance by applying both supervised, unsupervised, as semi-supervised approaches [33, 52]. Semi-supervised learning combines a small proportion of labeled data with a large pile of unlabeled data during the training process. As mentioned earlier, a semi-supervised approach such as node2vec has outperformed existing supervised approaches since it can maintain the community structure and embed better quality information [19]. In addition, neural networks are also currently being used to enhance link prediction performance. These novel methods have proven to be highly effective [30]. Such methods can produce promising link prediction results in large complex networks. Even so, a primary disadvantage to such approaches is that the training and prediction process is highly time-consuming.

3 Proposed Approach

This section explains the strategy of solving the problem with the deep learning method. Algorithms 1 and 2 provide insights into how we have aggregated the connected and unconnected pairs from the network used to build the training dataset. The overall steps for preparing a graph with connected and unconnected pairs from the raw network graph are explained in Algorithm 3. Then, we present the node2vec model for extracting the features from the training network dataset. Finally, we show how we have developed the deep neural network model with improved optimizers for executing AUC scores for the link prediction of the network. Figure 1 provides the overview of our proposed approach.

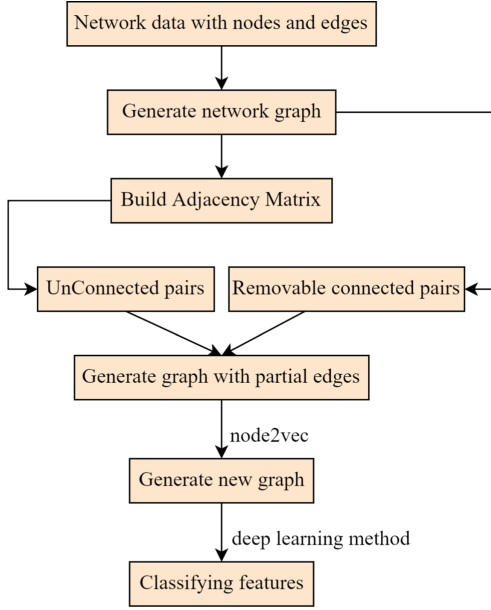


Fig. 1. Overall structure of the proposed approach

3.1 Problem Statement

A sequence of snapshots in time from t to $t + n$ is defined as a dynamic network in which the set of edges in each snapshot depicts the links present at time t . The link prediction problem is that given snapshots from t to $t + n$, return the score for the possibilities of edges at time $t + n$. Figure 2 and Fig. 3 show a dynamic network with two snapshots. Given the information at time t , we would like to predict the likelihood of link prediction at time $t + n$.

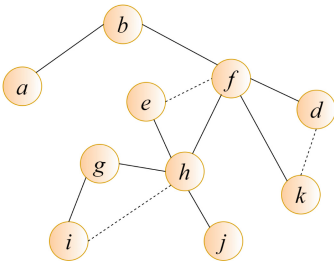


Fig. 2. At time t , the dotted lines represents the future links

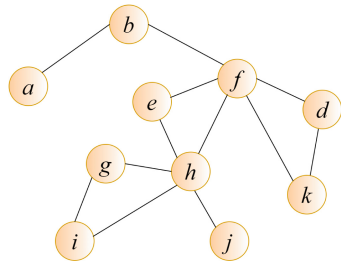


Fig. 3. At time $t+n$, the new links have formed

3.2 Data Preparation

In real-time scenarios, network data is extremely large and highly imbalanced as it contains a higher number of unconnected nodes than connected nodes. Therefore, it is always challenging for the model to learn features from connected nodes since the connected node pairs are often much fewer than the unconnected nodes. Hence, we provide a way for preparing the data computationally economically to extract the unconnected and connected node pairs from a large number of imbalanced network data. Below we discuss the process of sampling the positive (connected) and the negative (unconnected) node pairs.

Aggregation of Unconnected Samples. To find the negative sample that depicts the unconnected nodes, we build an adjacency matrix with the aid of the *networkx* library. The connected and unconnected nodes are represented as rows and columns for each node. As the values in the matrix are the same for above and below the diagonal of the adjacency matrix, we only focus on finding the positions of the unconnected nodes from above the diagonal to make the approach computationally efficient. Algorithm 1 shows the complete steps for finding the unconnected nodes. After experimentation with other configurations, we use *networkx* library to find the shortest path between the unconnected nodes and only select the ones within distance 3. The unconnected node pairs were labeled as ‘0’, as the unconnected node pairs represent the negative links.

Aggregation of Connected Samples. Some of the edges from the graph will be randomly removed and labelled ‘1’ since, these edges connect the nodes and show the presence of links. Thus, when training the model, it will predict such potential links at time $t + n$. However, it is essential to ensure that the graph’s nodes do not become completely isolated when dropping the edges since taking such a step can misrepresent data, and the model will be trained poorly. Thus, when removing an edge, we ensure this does not lead to splitting the graph, and the number of connected nodes is ≥ 1 . If the removed edge satisfies both of the conditions, only then the edge is dropped, and the process is repeated for the next pairs of nodes. Algorithm 2 shows the steps for accumulating the positive samples.

3.3 NODDLE (Integration of NODe2vec and Deep Learning mMethod)

Given that node2vec is a local approach that is limited to the structure around the node, it uses a short random walk to find the local neighborhood of nodes. Such attention to local structure implicitly overlooks the long-distance relationship in the whole network, and the representation may not reveal the important global structure model [19]. We propose NODDLE, a deep learning model in which features extracted from the node2vec are fed into a 4 layer hidden neural

Algorithm 1: Finding all Unconnected pairs

Input: Adjacency Matrix as $AdjM$, Graph $G = (N, E)$
Output: All unconnected pairs as UCP

```

1  $UCP \leftarrow \phi$  // empty dataframe for all unconnected pairs
2 for each row in  $AdjM$  do
3   for each column in  $AdjM$  do
4     if row.index  $\neq$  column.index then
5       if FindShortestLength ( $G$ , row, column)  $\leq 3$  then
6         /* using networkx for shortest length function */
7         if  $AdjM$  [row, column] == 0 then
8           Append  $N$ .row,  $N$ .column to  $UCP$ 
9
10  Return  $UCP$ 

```

Algorithm 2: Finding all Connected pairs

Input: Graph $G = (N, E)$
Output: Connected Pairs as CP

```

1  $CP \leftarrow \phi$  // Empty set of connected pairs
2 for each  $e$  in  $E$  do
3    $G' \leftarrow RemoveEdge(G)$  // remove edge from a node pair and generate
4   a new graph as  $G'$ 
5   if the new nodes are not completely isolated then
6      $CP \leftarrow Append(G'.node, G'.edge)$ 
7
8 Return  $CP$ 

```

network. The prediction performance is enhanced by applying various optimizers, including Adaptive Moment Estimation (Adam), Adamax, An Adaptive Learning Rate Method (Adadelta), and Adaptive Gradient Algorithm (Adagrad), respectively. We illustrate a brief description of node2vec, its limitations, and explain the need for this deep learning model.

Node2vec. Node2vec algorithm is a feature extraction method used to generate vector representations of nodes on a graph. It is mainly a local approach that uses random walk to search for the local neighborhood of nodes. The algorithm uses direct encoding and a product-based decoder. Therefore, node2vec embedding is defined as such:

$$DE(s_i, s_j) \cong \frac{e^{z_i^T z_j}}{\sum_{v_k \in V} e^{z_i^T z_k}} \approx (P, R(v_j|v_i)) \tag{1}$$

In this Eq. (1), $DE(s_i, s_j)$ represents the decoded product based proximity value, the probability of visiting to node target node v_j from the source node v_i with fixed length of random walk R , denoted by $(P, R(v_j|v_i))$. $(P, R(v_j|v_i))$ can be calculated for both random and undirected graphs. Cross entropy loss for node2vec is calculated by the following formula:

Algorithm 3: Data Preparation for generating Model

```

Input: Node as  $N$  and Edge as  $E$ 
Output: New Graph as  $G'$ 
1  $G \leftarrow \phi$  // Graph
2  $AdjM \leftarrow \phi$  // Adjacency Matrix
3  $UCP \leftarrow \phi$  // Empty set of unconnected pairs
4  $CP \leftarrow \phi$  // Empty set of connected pairs
5 for each  $n, e$  in  $N, E$  do
6    $G \leftarrow (n, e)$  // Creating network Graph with networkx library
7    $AdjM \leftarrow AdjacencyMatrix(G)$  /* Create Adjacency Matrix from nodes
   and edges with networkx */
8    $UCP \leftarrow UnconnectedPairs(AdjM, G)$  // Algorithm 1
9    $CP \leftarrow ConnectedPairs(G)$  // Algorithm 2
10   $G' \leftarrow CreateNewGraph(UCP, CP)$ 
11 Return  $G'$ 

```

$$Loss = \sum_{(v_i, v_j) \in Deno} -\log(DE(s_i, s_j)) \quad (2)$$

The training set is generated by collecting random walks from a source node v_i in which the N pairs of v_i for each node are collected from the probabilistic distribution of $(v_i, v_j) \sim (P, R(v_j|v_i))$. However, it is extremely expensive to calculate the cross entropy loss because of the high computational costs for evaluating $O(|Deno||V|)$, as $O(|V|)$ has a high time complexity when computing the denominator $Deno$ of (1). As a result, node2vec uses various optimization and approximation methods for computing the cross entropy loss. It uses the ‘‘Negative sampling’’ approximation method to evaluate (2). The node2vec takes a random set of negative samples for approximately calculating the normalization factor instead of letting the entire set of vertices be normalized [19]. Additionally, it applies two hyper parameters p and q . The probability of going back to a previous node after visiting a new node is controlled by p . The hyper parameter q controls the possibility of exploring the graph’s new nodes. When these hyper parameters are employed, node2vec can interpolate between the walks much more smoothly, and the approach becomes similar to BFS and DFS. Grover et al. also demonstrated that when the two hyper parameters are well-adjusted, it enables node2vec to preserve the structural balance between the nodes [19]. However, node2vec still has its drawbacks. It uses SGD method for solving the non-convex optimization problem [18, 28]. The algorithm constantly updates when SGD is used as the objective function, which causes the optimal points to oscillate frequently, leading the optimal points to dismount into the local minimum range.

Besides, SGD keeps the learning rate constant when the parameters are updated. As a result, SGD cannot adapt the learning rate and adjust it for carrying out greater updates on lower frequency features [40]. Hence, Adam, Adamax, and Adadelta optimizers have been introduced to resolve this issue. These optimizers can incorporate different learning rates with different parameters.

Compared to SGD, these optimizers are more compatible for large network datasets in high dimensional spaces and, most importantly, non-convex optimization objective functions. Furthermore, deep learning techniques are also applied to study the complex relationships with the growing networks. Hence, we are focusing on improving the performance of link prediction by fusing node2vec with a deep learning model, in which the model is supported with improved optimizers.

Algorithms 4 and 5, show the steps of the node2vec algorithm. The algorithm first learns the representations of the nodes by generating a random walk with a length of l , which starts from each of the nodes. When the step is taken in each of the walks, sampling is conducted with the transitional probability of θ_{vx} . The transitional probability θ_{vx} of the second order Markov chain is at first calculated so that node sampling can be computed efficiently by using the alias method in $O(1)$ time. In the final phase, the transitional probability preprocessing is conducted sequentially, and optimization of SGD is used.

Algorithm 4: Node2Vec Algorithm

Input: Graph $G' = (N, E)$, dimension dim , Walks per node r , Walk Length l , Context size h , Return p , In-out q
Output: final Stochastic gradient descent function as f

- 1 $\theta = \text{PreprocessModifiedWeights}(G, p, q)$ $G' = (V, E, \theta)$ Initialize *walks* to empty
- 2 **for** *iter* 1 to r **do**
- 3 **for** all nodes $u \in V$ **do**
- 4 $walk = \text{node2vecWalk}(G', u, l)$
- 5 Append *walk* to *walks*
- 6 $f = \text{Stochastic gradient descent}(h, dim, walks)$
- 7 **Return** f

Algorithm 5: node2vecWalk Algorithm

Input: Graph $G' = (N, E)$, Start node u , Length l , Walk Length l , Context size h , Return p , In-out q
Output: *walk*

- 1 **node2vecWalk**(Graph $G' = (N, E)$, Start node u , Length l , Walk Length l , Context size h , Return p , In-out q) Initialize *walk* to $[u]$
- 2 $G' = (V, E, \theta)$
- 3 Initialize *walks* to empty
- 4 **for** *walk* from 1 to l **do**
- 5 $curr = walk[-1]$
- 6 $V_{current} = \text{GetNeighbours}(Current, G')$
- 7 $s = \text{AliasSample}(V_{current}, \theta)$
- 8 Append s to *walk*
- 9 **Return** *walk*

Deep Learning Model. In the final step, we build a deep learning network in which the features extracted from node2vec are fed into a four layer hidden neural network. As mentioned earlier, the SGD optimization function of node2vec has limited capabilities to adapt to different learning rates. As a result, for boosting up the performance of the link prediction task, we built the deep learning model with adaptive learning rate optimizers: Adam, Adamax, Adadelata, and Adagrad, respectively. This approach is treated as a supervised classification problem, where the network aims to yield a single value representing the probability for a given edge. Thus, we end the deep learning model with a sigmoid activation function to score between 0 and 1.

Optimizers. Below we have discussed the different types of optimizers that were used with the model.

- Adagrad: Adaptive gradient, or AdaGrad, divides the learning rate by the square root of v , which is mainly the cumulative sum of current and past squared gradients up to time t [17]. Moreover, the gradient component is unchanged just like in SGD. The Adagrad is defined as such:

$$w_{t+1} = w_t - \frac{\rho}{\sqrt{v_t + \epsilon}} \cdot \frac{\partial L}{\partial w_t} \quad (3)$$

In Eq. (3), w_t is the current weight at time step t that needs to be updated, ρ represents the learning rate and $\frac{\partial L}{\partial w_t}$ denotes the gradient descent to update the weight at w_t and ϵ is a constant value.

- Adadelata: Adadelata is a much more powerful extension of Adagrad that emphasizes the learning rate component [47]. The optimizer is based on updating gradient using the sliding window technique instead of aggregating all the previous gradients. In Adadelata, the difference between the current and updated weights is denoted by ‘delta’. Furthermore, the learning rate parameter is replaced by T , the exponential moving average of squared deltas and is defined in 4).

$$w_{t+1} = w_t - \frac{\sqrt{T_{t-1} + \epsilon}}{\sqrt{v_t + \epsilon}} \cdot \frac{\partial L}{\partial w_t} \quad (4)$$

- Adam: Adaptive moment estimation, or Adam, focuses on the gradient component by using \hat{s} , which estimates the exponential average of the moving gradients [27]. In addition, the learning rate component is calculated by dividing the learning rate ρ by the square root of v , which is the exponential moving average of squared gradients. The equation is defined as below:

$$w_{t+1} = w_t - \frac{\sqrt{T_{t-1} + \epsilon}}{\sqrt{v_t + \epsilon}} \cdot \hat{s}_t \quad (5)$$

- Adamax: AdaMax is a variation of the Adam optimizer, which uses infinity norms [27]. The infinity norm is used to calculate the absolute values of the v components in a vector space (‘max’), and \hat{s} refers to the estimated value

of the exponential average of moving gradients, and v is the exponential moving average of previous p -norm of gradients, that is approximately the max function as defined below:

$$w_{t+1} = w_t - \frac{\rho}{v_t} \cdot \hat{s}_t \quad (6)$$

4 Experiment

This section will evaluate the proposed model on real-world data network datasets and examine how it is more effective than the existing benchmark methods, including Adamic Adar, Preferential Attachment, and Jaccard Coefficient.

4.1 Datasets

We evaluate our model on Facebook¹ and Twitter² datasets that consists of nodes and edges. Table 1 shows the overview of the five network datasets. The first four of these datasets were collected from the SNAP website. Also, with the aid of Twitter API we have extended a Twitter dataset of around 7,000 users who have followed Twitter medical accounts [26]. The extended dataset contains the follower and following IDs of the users working in the medical profession. Public biographical contents of the users were used for finding the occupation of the users.³

Table 1. Details of the datasets

Dataset	Number of nodes	Number of edges
Twitter	81,306	1,768,149
Facebook1	4,039	88,234
Facebook2	1,046	27,794
Facebook3	546	5,360
Occupation	6,754	470,168

4.2 Experimental Results and Discussions

Our proposed model was implemented in Python 2.8.6, and the experiment was conducted on HPC (High Configuration GPU enabled PC)⁴. In our model, we have used four layer fully connected deep neural network with 1024 ReLU neurons in each of the hidden layers. Then, we developed our model using Adagrad,

¹ <https://snap.stanford.edu/data/egonets-Facebook.html>.

² <https://snap.stanford.edu/data/egonets-Twitter.html>.

³ https://github.com/ZainabKazi22/occupation_twitter.

⁴ https://github.com/ZainabKazi22/link_prediction_with_gbm.

Adadelta, Adam, and Adamax optimizers to improve the performance of the link prediction task.

We calculate the Area Under ROC Curve (AUC) scores to evaluate the performance of our approach of combining node2vec and deep learning model with each of the optimizers, respectively. The AUC score is defined in (7):

$$AUC = \frac{D_0 - n_0(n_0 + 1)/2}{n_0n_1} \quad (7)$$

In Eq. (7), n_0 and n_1 denotes the number of positive and negative class links, respectively and $D_0 = \sum r_i$, where r_i represents the rank of the index i in the positive class link in terms of similarity index. Also, $AUC \in [0, 1]$, in which the higher the value of AUC , the higher the link prediction accuracy of the algorithm. We have compared the performance of our approach with the traditional link prediction benchmark methods: Adamic Adar (AA), Jaccard Co-efficient (JC), and Preferential Attachment (PA).

In AA, the association between two neighboring nodes with a smaller degree may occur more than a node with a higher degree [51]. For instance, two celebrity fans are likely not to know each other. Yet, if two users follow someone who has fewer fans, then those two users will have a higher chance to have similar interests or tend to be in the same social circle. JC believes that the probability of the presence of links is proportional to the number of two nodes' neighbors [20]. If any two Twitter users tend to have similar interests, they have a higher chance of having some type of connections with each other. Research has shown that the rate of an edge to be connected to a node is proportional to the degree of the node [1]. Thus, PA states that the chances of a new edge to be connected to a node are related to the degree of the node. Two popular celebrities will have a higher chance to know each other since they have a higher degree compared

Table 2. AUC Scores of the Link Prediction Algorithms

Link Prediction Algorithm	AUC Score				
	Twitter	Occupation	Facebook1	Facebook2	Facebook3
Node2vec	0.895	0.876	0.938	0.873	0.861
Node2vec + DL (Adam)	0.902	0.931	0.934	0.862	0.855
Node2vec + DL (Adamax)	0.916	0.945	0.941	0.879	0.882
Node2vec + DL (Adagrad)	0.911	0.911	0.932	0.845	0.851
Node2vec + DL (Adadelta)	0.924	0.932	0.908	0.871	0.863
Adamic Adar	0.897	0.711	0.898	0.878	0.734
Jaccard Co-efficient	0.897	0.748	0.901	0.856	0.699
Preferential Attachment	0.891	0.803	0.835	0.801	0.76

to two ordinary persons. The equations of the following algorithms are stated in (8), (9), and (10):

$$s_{AA} = \sum_{x \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log k_x} \quad (8)$$

$$s_{JC} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (9)$$

$$s_{PA} = k_x \cdot k_y \quad (10)$$

Table 2 shows the AUC scores obtained from the link prediction algorithms. Overall, node2vec and the node2vec optimized algorithm (Node2Vec+DL) have performed better than the traditional benchmark methods. This might be because node2vec algorithms can learn high-level features from the network data [19]. Moreover, as high-end robust computational engines like GPU are readily available, it is possible to execute the deep learning models. Whereas predicting future links from large network data is challenging for the existing benchmark methods. Node2vec with Adamax optimizer has received the highest AUC score in Occupation and Facebook1 and Facebook2 datasets among the deep learning models. The node2vec with Adadelta optimizer has performed best in Twitter and Facebook3 datasets. The model with Adamax optimizer has performed better than the rest of the optimizers across three datasets, proving that the Adamax optimizer modified over Adam optimizer performs better than the Adam optimizer. Similarly, the model with the Adadelta optimizer has performed better for Twitter and Facebook3 datasets than the Adagrad optimizer. This has demonstrated that the Adadelta optimizer, an improved version of Adagrad optimizer, has achieved a better performance score than the Adagrad optimizer. Thus, from the results in Table 2, we can see that optimizers of the DL model have increased the performance of the node2vec algorithm. The model proposed in this paper has acquired higher AUC scores than the existing benchmark and node2vec method. Also, the AUC scores of the node2vec with improved optimizers of the DL model are highest across all the datasets.

5 Conclusion

In this paper, we explored the drawbacks of the node2vec algorithm when boosting up non-convex functions. In other words, the likelihood of falling into a local minimum due to lack of network knowledge and SGD optimizer’s incapacibilities to execute adaptive adjustment of the learning rate. Hence, such a scenario makes it extremely difficult for node2vec to process sparse social networks. As a result, we proposed NODDLE, a deep learning model where we have merged the features aggregated by the node2vec algorithm and used them as inputs into a multi-layer neural network optimizing its performance by using different types of improved optimizers such as Adam, Adamax, Adadelta, and Adagrad. Compared to the various baselines, the results of experiments on real-world social networks

proved that our approach enhances the prediction accuracy and is much more effective and efficient.

Acknowledgements. The authors thank DaTALab members & Lakehead University's HPC (High Configuration GPU enabled PC) for executing the models, and Punardeep Sikkha, Arunim Garg, Bart, and Abhijit Rao for proofreading and reviewing the manuscript.

References

1. Abbasi, A., Hossain, L., Leydesdorff, L.: Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *J. Informet.* **6**(3), 403–412 (2012)
2. Adamic, L.A., Adar, E.: Friends and neighbors on the web. *Soc. Netw.* **25**(3), 211–230 (2003)
3. Aiello, L.M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., Menczer, F.: Friendship prediction and homophily in social media. *ACM Trans. Web (TWEB)* **6**(2), 1–33 (2012)
4. Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9**(Sep), 1981–2014 (2008)
5. Al Hasan, M., Chaoji, V., Salem, S., Zaki, M.: Link prediction using supervised learning. In: *SDM06: Workshop on Link Analysis, Counter-terrorism and Security*, vol. 30, pp. 798–805 (2006)
6. Al-Rfou, R., Perozzi, B., Zelle, D.: DDGK: learning graph representations for deep divergence graph kernels. In: *The World Wide Web Conference*, pp. 37–48 (2019)
7. Amari, S.I.: Backpropagation and stochastic gradient descent method. *Neurocomputing* **5**(4–5), 185–196 (1993)
8. Ayoub, J., Lotfi, D., El Marraki, M., Hammouch, A.: Accurate link prediction method based on path length between a pair of unlinked nodes and their degree. *Soc. Netw. Anal. Min.* **10**(1), 1–13 (2020). <https://doi.org/10.1007/s13278-019-0618-2>
9. Benchettara, N., Kanawati, R., Rouveirol, C.: Supervised machine learning applied to link prediction in bipartite social networks. In: *2010 International Conference on Advances in Social Networks Analysis and Mining*, pp. 326–330. IEEE (2010)
10. Bressan, M., Chierichetti, F., Kumar, R., Leucci, S., Panconesi, A.: Counting graphlets: space vs time. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 557–566 (2017)
11. Canziani, A., Paszke, A., Culurciello, E.: An analysis of deep neural network models for practical applications. arXiv preprint [arXiv:1605.07678](https://arxiv.org/abs/1605.07678) (2016)
12. Chen, H., Perozzi, B., Hu, Y., Skiena, S.: HARP: hierarchical representation learning for networks. arXiv preprint [arXiv:1706.07845](https://arxiv.org/abs/1706.07845) (2017)
13. Chen, H., Perozzi, B., Hu, Y., Skiena, S.: HARP: hierarchical representation learning for networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32 (2018)
14. Chen, J., et al.: N2VSCDNNR: a local recommender system based on node2vec and rich information network. *IEEE Trans. Comput. Soc. Syst.* **6**(3), 456–466 (2019)
15. Deylami, H.A., Asadpour, M.: Link prediction in social networks using hierarchical community detection. In: *2015 7th Conference on Information and Knowledge Technology (IKT)*, pp. 1–5. IEEE (2015)

16. Doppa, J.R., Yu, J., Tadepalli, P., Getoor, L.: Learning algorithms for link prediction based on chance constraints. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010. LNCS (LNAI), vol. 6321, pp. 344–360. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15880-3_28
17. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **12**(7) (2011)
18. Goldberg, Y., Levy, O.: Word2vec explained: deriving Mikolov et al.’s negative-sampling word-embedding method. arXiv preprint [arXiv:1402.3722](https://arxiv.org/abs/1402.3722) (2014)
19. Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864 (2016)
20. Gupta, A.K., Sardana, N.: Significance of clustering coefficient over jaccard index. In: 2015 Eighth International Conference on Contemporary Computing (IC3), pp. 463–466. IEEE (2015)
21. Hu, W., Wang, H., Peng, C., Liang, H., Du, B.: Retracted: an event detection method for social networks based on link prediction (2017)
22. Hu, W., Wang, H., Qiu, Z., Nie, C., Yan, L., Du, B.: An event detection method for social networks based on hybrid link prediction and quantum swarm intelligent. *World Wide Web* **20**(4), 775–795 (2017)
23. Jain, P., Kar, P.: Non-convex optimization for machine learning. arXiv preprint [arXiv:1712.07897](https://arxiv.org/abs/1712.07897) (2017)
24. Katz, L.: A new status index derived from sociometric analysis. *Psychometrika* **18**(1), 39–43 (1953)
25. Kaya, M., Jawed, M., Bütün, E., Alhajj, R.: Unsupervised link prediction based on time frames in weighted-directed citation networks. In: Missaoui, R., Abdessalem, T., Latapy, M. (eds.) Trends in Social Network Analysis. LNSN, pp. 189–205. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53420-6_8
26. Khanam, K.Z., Srivastava, G., Mago, V.: Identifying health related occupations of twitter users through word embedding and deep neural networks. In: Proceedings of The 19th Asia Pacific Bioinformatics Conference Accepted. In press (2021)
27. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
28. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: International Conference on Machine Learning, pp. 1188–1196 (2014)
29. Li, J.C., Zhao, D.L., Ge, B.F., Yang, K.W., Chen, Y.W.: A link prediction method for heterogeneous networks based on BP neural network. *Phys. A: Stat. Mech. Appl.* **495**, 1–17 (2018)
30. Li, T., Zhang, J., Philip, S.Y., Zhang, Y., Yan, Y.: Deep dynamic network embedding for link prediction. *IEEE Access* **6**, 29219–29230 (2018)
31. Li, X., Du, N., Li, H., Li, K., Gao, J., Zhang, A.: A deep learning approach to link prediction in dynamic networks. In: Proceedings of the 2014 SIAM International Conference on Data Mining, pp. 289–297. SIAM (2014)
32. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. *J. Am. Soc. Inform. Sci. Technol.* **58**(7), 1019–1031 (2007)
33. Liu, D., Li, Q., Ru, Y., Zhang, J.: The network representation learning algorithm based on semi-supervised random walk. *IEEE Access* **8**, 222956–222965 (2020)
34. Liu, H., Kou, H., Yan, C., Qi, L.: Link prediction in paper citation network to construct paper correlation graph. *EURASIP J. Wirel. Commun. Netw.* **2019**(1), 1–12 (2019). <https://doi.org/10.1186/s13638-019-1561-7>

35. Maron, H., Ben-Hamu, H., Serviansky, H., Lipman, Y.: Provably powerful graph networks. In: *Advances in Neural Information Processing Systems*, pp. 2156–2167 (2019)
36. Ozcan, A., Oguducu, S.G.: Link prediction in evolving heterogeneous networks using the NARX neural networks. *Knowl. Inf. Syst.* **55**(2), 333–360 (2018)
37. Pezeshkpour, P., Tian, Y., Singh, S.: Investigating robustness and interpretability of link prediction via adversarial modifications. arXiv preprint [arXiv:1905.00563](https://arxiv.org/abs/1905.00563) (2019)
38. Rahman, M., Hasan, M.A.: Link prediction in dynamic networks using graphlet. In: Frasconi, P., Landwehr, N., Manco, G., Vreeken, J. (eds.) *ECML PKDD 2016. LNCS (LNAI)*, vol. 9851, pp. 394–409. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46128-1_25
39. Rahman, M., Saha, T.K., Hasan, M.A., Xu, K.S., Reddy, C.K.: Dylink2vec: effective feature representation for link prediction in dynamic networks. arXiv preprint [arXiv:1804.05755](https://arxiv.org/abs/1804.05755) (2018)
40. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint [arXiv:1609.04747](https://arxiv.org/abs/1609.04747) (2016)
41. Tabakhi, S., Moradi, P., Akhlaghian, F.: An unsupervised feature selection algorithm based on ant colony optimization. *Eng. Appl. Artif. Intell.* **32**, 112–123 (2014)
42. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: LINE: large-scale information network embedding. In: *Proceedings of the 24th International Conference on World Wide Web*, pp. 1067–1077 (2015)
43. Wang, H., Shi, X., Yeung, D.Y.: Relational deep learning: a deep latent variable model for link prediction. In: *AAAI*, pp. 2688–2694 (2017)
44. Wang, P., Xu, B., Wu, Y., Zhou, X.: Link prediction in social networks: the state-of-the-art. *Sci. China Inf. Sci.* **58**(1), 1–38 (2015)
45. Xu, L., Wei, X., Cao, J., Yu, P.S.: Interaction content aware network embedding via co-embedding of nodes and edges. In: Phung, D., Tseng, V.S., Webb, G.I., Ho, B., Ganji, M., Rashidi, L. (eds.) *PAKDD 2018. LNCS (LNAI)*, vol. 10938, pp. 183–195. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93037-4_15
46. Yao, L., Wang, L., Pan, L., Yao, K.: Link prediction based on common-neighbors for dynamic social network. *Procedia Comput. Sci.* **83**, 82–89 (2016)
47. Zeiler, M.D.: ADADELTA: an adaptive learning rate method. arXiv preprint [arXiv:1212.5701](https://arxiv.org/abs/1212.5701) (2012)
48. Zhang, C., Zhang, H., Yuan, D., Zhang, M.: Deep learning based link prediction with social pattern and external attribute knowledge in bibliographic networks. In: *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 815–821. IEEE (2016)
49. Zhang, L., Zhuo, F., Bai, C., Xu, H.: Analytical model for predictable contact in intermittently connected cognitive radio ad hoc networks. *Int. J. Distrib. Sens. Netw.* **12**(7), 1550147716659426 (2016)
50. Zhang, Z., Wen, J., Sun, L., Deng, Q., Su, S., Yao, P.: Efficient incremental dynamic link prediction algorithms in social network. *Knowl.-Based Syst.* **132**, 226–235 (2017)
51. Zhou, T., Lü, L., Zhang, Y.C.: Predicting missing links via local information. *Eur. Phys. J. B* **71**(4), 623–630 (2009)
52. Zhu, J., Zheng, Z., Yang, M., Fung, G.P.C., Tang, Y.: A semi-supervised model for knowledge graph embedding. *Data Min. Knowl. Disc.* **34**(1), 1–20 (2020)