

Deep Learning Approach on Information Diffusion in Heterogeneous Networks

Soheila Molaei^a, Hadi Zare^{a,*}, Hadi Veisi^a

^a*Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran.*

Abstract

There are many real-world complex systems with multi-type interacting entities that can be regarded as heterogeneous networks including human connections and biological evolutions. One of the main issues in such networks is to predict information diffusion such as shape, growth and size of social events and evolutions in the future. While there exist a variety of works on this topic mainly using a threshold-based approach, they suffer from the local viewpoint on the network and sensitivity to the threshold parameters. In this paper, information diffusion is considered through a latent representation learning of the heterogeneous networks to encode in a deep learning model. To this end, we propose a novel meta-path representation learning approach, Heterogeneous Deep Diffusion(HDD), to exploit meta-paths as main entities in networks. At first, the functional heterogeneous structures of the network are learned by a continuous latent representation through traversing meta-paths with the aim of global end-to-end viewpoint. Then, the well-known deep learning architectures are employed on our generated features to predict diffusion processes in the network. The proposed approach enables us to apply it on different information diffusion tasks such as topic diffusion and cascade prediction. We demonstrate the proposed approach on benchmark network datasets through the well-known evaluation measures. The experimental results show that our approach outperforms the earlier state-of-the-art methods.

Keywords: Heterogeneous Networks, Information Diffusion, Topic Diffusion, Cascade Prediction, Network Representation Learning, Deep Learning.

*Corresponding author

1. Introduction

Information diffusion is one of the widely studied dynamical processes on networks. Information such as news, innovations, and viruses start from a set of seed nodes and propagate throughout the network [1, 2]. Information diffusion has been investigated in a wide range of fields including health care [3, 4], complex networks [5, 6] and social networks [7, 8]. One of the most important tasks on networked systems is understanding, modeling, and predicting the rapid events and evolutions in the body of a network. This is mainly motivated by the well-known fact that discovering the structure of networks is required to predict the patterns of social events such as their shape, size, and growth known as information diffusion [9]. Many researchers have investigated various techniques and approaches to model information diffusion on homogeneous and heterogeneous networks such as [10, 11, 12, 13].

Formally, an information network denoted as a graph $G = (V, E)$ with the set V of nodes and set E of edges, is homogeneous if and only if the edges and nodes are of the same type, and heterogeneous if different types of nodes and relations on edges are involved [14]. Various studies have been conducted on homogeneous networks including semantic parsing [15], control of epidemic diseases [16, 17] and link prediction [18, 19, 20]. Recently, heterogeneous networks have been considered as an attractive research field, due to the fact of more natural assumption of heterogeneous networks on many real-world phenomena [21, 22]. Watts [23] studied the role of threshold values and network structure in information diffusion. Therefore, information diffusion, as a particular topic of interest in this regard, has been studied in [10] using different meta-paths in heterogeneous networks. In [10], MLTM-R is proposed by distinguishing the power in passing information around for different types of meta-paths. Pathsim was considered as a weight between each two nodes in this method through which predictions were conducted [24, 25].

Recently, much attention has been focused on deep learning in heterogeneous networks [26, 27]. In [28], a deep learning framework is investigated as coupled with deep learning (CDL), to address the VIS-NIR heterogeneous matching problem via the topic diffusion on networks. Tang et al. [29] proposed the LINE algorithm to learn embedding which traverses all edge types and samples one edge at a time for each edge type. Chang et al. [30]

proposed a deep architecture by embedding Heterogeneous Information Network(HIN). metapath2vec [31] was proposed as a new HIN algorithm which transcribes semantics in HINs by meta-paths. With this in mind, we point out the strengths and weaknesses of the existing approach to topic diffusion in heterogeneous networks. Our motivation is to employ a deep learning approach on information diffusion tasks such as topic diffusion and cascade prediction to alleviate the problems of the earlier ones. Most works on topic diffusion suffer from local similarity considerations and single node-based embedding [10], our latent based representational learning approach investigates a global view of the activation prediction modeling. Local similarity computations for each two nodes are time-consuming and unendurable for big networks. While there are various similarity measures, the choosing of appropriate similarity method is a challenge.

Here, we propose a novel latent representational learning approach on heterogeneous networks based on different types of meta-paths. Due to the challenges of similarity calculation, we omit this step and for representing a node we consider different graph meta-paths which in each meta-path graph, the neighbors of a node could be different. Furthermore, the proposed approach is employed on different deep neural network architectures to predict information diffusion tasks in an end-to-end framework. The experimental results demonstrate the strength of the proposed approach as compared with the earlier works. We evaluate the performance of the proposed method using real-world information graph i.e. DBLP, PubMed, ACM, APS, and Citeseer. HDD is compared with multiple strong baselines, including feature-based methods such as deepwalk [32], node2vec [33], deepcas [9], and MLTM-R [10]. HDD method significantly improves the results over these baselines.

In this paper, we have applied CNN-LSTM to generate meta-paths representation for information diffusion tasks. The CNN-LSTM architecture contains a CNN layer for feature extraction along with an LSTM to support sequential prediction. This model draws on the intuition that the sequence of features extracted from CNN can be encoded into a vector representation using LSTM architecture. CNN-LSTM can embed the the whole meta-paths in the network. Our representations are general-purpose which can be applied on heterogeneous network for information diffusion and cascade prediction tasks. To summarize, our work makes the following contributions,

- We propose a Heterogeneous Deep Diffusion (HDD) model to learn HIN node embeddings which can be retrieved for downstream tasks, such as

topic diffusion and information cascade.

- We design a feature extraction mechanism to conduct weighted aggregations of neighborhood nodes on different meta-paths for learning comprehensive embeddings.
- We conduct experiments on real-world datasets to show the superiority of our model against the prior state-of-the-art methods and give a comprehensive analysis of the learned embeddings in order to gain more insights from the datasets.

The remainder of this paper is organized as follows. In Section 2 related works are reviewed on the heterogeneous networks along with a brief motivation for the idea. Section 3 describes the proposed deep learning framework of information diffusion for heterogeneous networks. Section 4 is devoted to the experimental settings and results on the benchmark real networks. Finally, we conclude the paper in Section 5 with some suggestions for further works.

2. Literature Review

There exist a multitude of works on exploiting heterogeneous networks to uncover the structural patterns by considering the rich side of information on different nodes objects and edge attributes in these networks [34]. Sentiment classification of product reviews using heterogeneous networks of users, products, and words was addressed by Zhou et al. [35]. In this regard, Zhou et al. [35] proposed a co-ranking method which classifies the authors and documents separately based on random walks. Angelova et al. [36] presented a new classification method for mining of homogeneous information networks through their decomposition into multiple homogeneous ones. The idea of citation recommendation using heterogeneous networks was proposed by Liu et al. [37]. Information diffusion has been used in such networks. Information diffusion is mainly fallen into two categories as topic diffusion and cascade prediction which are described as follows:

2.1. Topic Diffusion

There is a variety of work on topic diffusion as a primary task to analyze the heterogeneous networks that have been employed in applied domains such as medical and public health issues. Several works considered the epidemic

modeling in heterogeneous networks like infections spreading on population systems [38], modified SIR model [39], and a survey [40]. Wang and Dai [41] addressed virus spreading in heterogeneous networks by applying an epidemic threshold on the well-known SIS model. [11] showed leveraging a heterogeneous network among people to yield more resistance against the epidemic spread of the virus. Epidemic spreading is an important issue that was considered in other networks like time-varying networks [42] and adaptive ones [43]. Nadini et al. [42] used SIR and SIS models and investigated effects of modular and temporal connectivity patterns on epidemic spreading.

Various techniques are proposed on topic diffusion in heterogeneous networks. The degree distribution is used in [44], for the modeling of information diffusion by taking the assumption of diffusion between two nodes at random times. Zhou and Liu [45] presented a social influence based clustering framework. Molaei et al. [46] predicted topic diffusion process in heterogeneous networks by considering the interactions of different meta-paths. A heterogeneous network based model was proposed for new products diffusion in two stages framework [47]. In [48], the concept of heterogeneous networks was used as an alternative definition for the infrastructure networks to explain the diffusion process. In contrast to most prior approaches to topic diffusion in multilayer networks with probabilistic graphical models, we have used deep representation learning which automatically extracts summarized features from each node in such networks.

2.2. Cascade Prediction

There are many cascade prediction methods originating from different research areas as both classification and regression problems. Recently, methods based on representation learning emerge with impressive predictive power. Graph representation learning methods have largely been based on the popular skip-gram model [49, 50] originally introduced for learning vector representations of words in the text. In particular, DeepWalk [32] used this approach to embed the nodes such that the co-occurrence frequencies of pairs in short random walks are preserved. Node2vec [33] introduced hyper-parameters to DeepWalk that tune the depth and breadth of the random walks.

Lately, some deep learning based models show good performances. These models learn to predict information cascade in an end-to-end manner. Deep-Cas [9] uses random walk to sample paths from different snapshots of the graph then uses Gated Recurrent Units (GRU) and attention mechanism

to extract features from random walk paths to predict information cascade. DeepHawkes [51] uses GRU to encode each cascade path, and employs weighted average pooling based on time decay effect to combine features from all cascade paths.

While there exist some recent works on cascade predictions with a deep learning approach [9, 51], our proposed framework differs from them due to the type of used networks, applied methods, input parameters and also more generalization capability to the different information diffusion process. We focus on the heterogeneous network and we added weights to the input with considering different meta-paths. The deep learning approach exploits the representational learning on meta-paths to generate the required features. On the other hand, we extracted features from networks differently which used more information and considered more relation in such networks.

3. Proposed Method

We present a general framework, HDD(Heterogeneous Deep Diffusion) on topic diffusion in heterogeneous networks through a deep learning approach. The flow-graph of the overall structure of the proposed approach is shown in Figure 1.

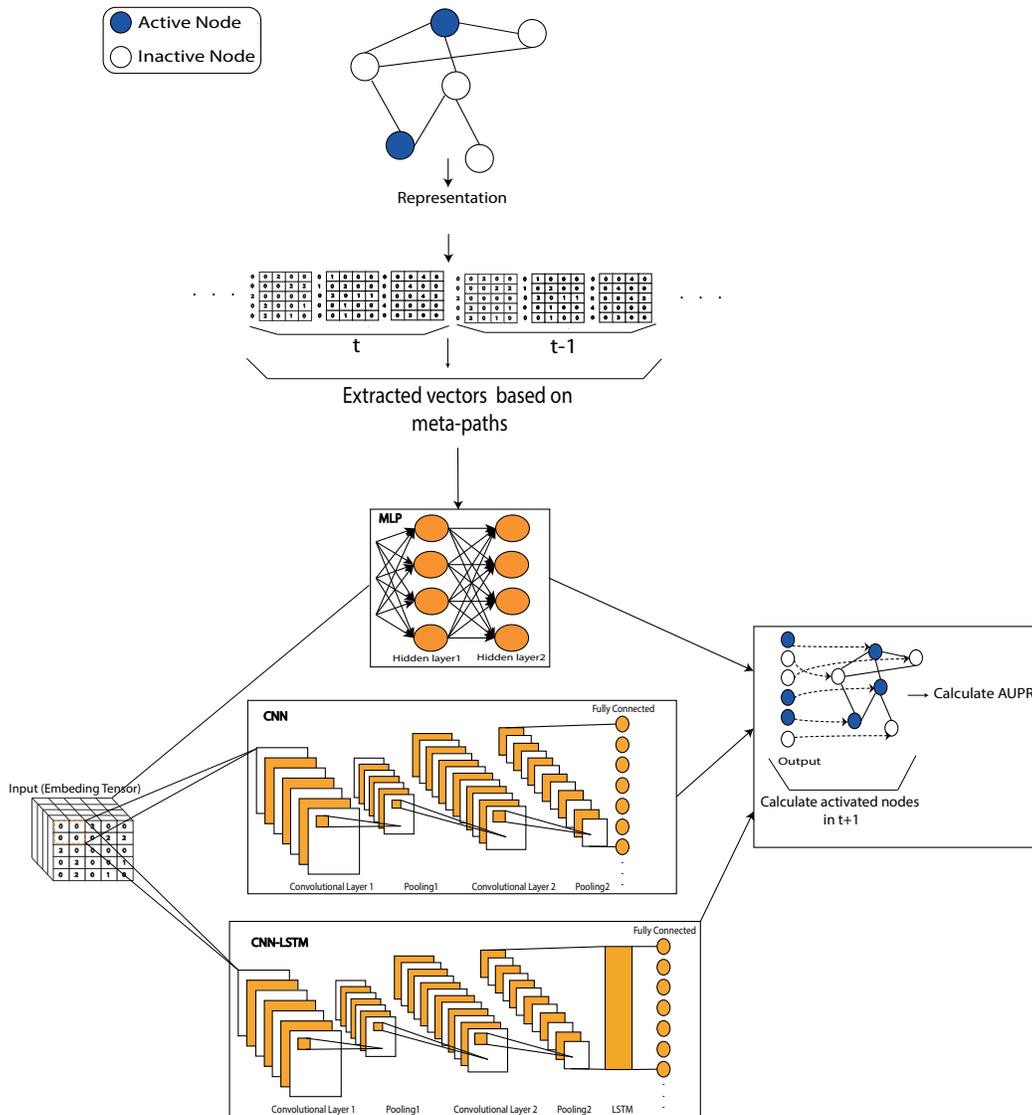


Figure 1: The overall procedure of the proposed method

Initially, the graph with distinct node types as active and inactive ones is given. In the representation stage, the continuous latent features are learned based on different meta-path definitions. Therefore, each node is represented as a continuous vector in a sequence of different meta-paths. We employ a variety of deep neural networks architectures on sequences of nodes represen-

tations to predict outgoing activation or inactivation in the body of network. In following, we describe the main information diffusion tasks in our formulation and then present our approach. The overall flowchart of the proposed method is mentioned in Figure 2.

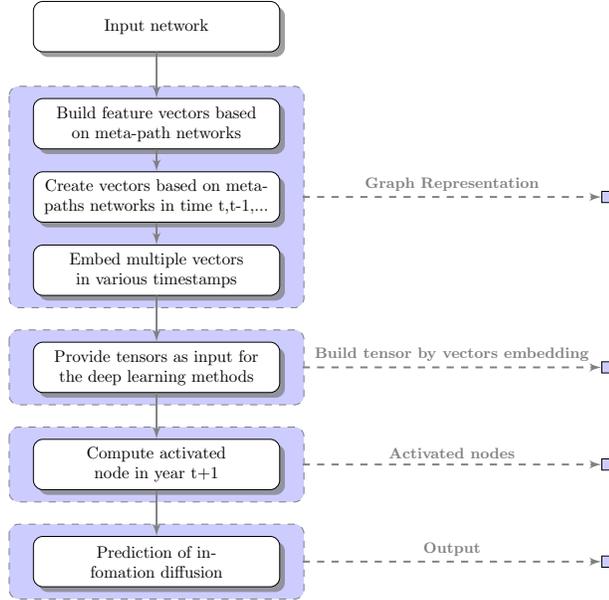


Figure 2: The overall flowchart of the proposed method

For each meta-path we create a meta-path graph and then update the embedded information of each node in these graphs within a vector. As Figure 3 shows on APA meta-path, a vector is generated for each node, in which the cells are the number of papers two nodes have co-authored on a specified topic in the year t . The last column is dedicated to output which is 1 (active node) if in $t + 1$ two nodes will co-author a paper on a specific topic, otherwise considered as 0 (inactive node). Finally, the creation of these matrices in multiple times through a variety of meta-paths yields to a tensor as the input on our selected models for downstream prediction tasks as illustrated in Figure 1.

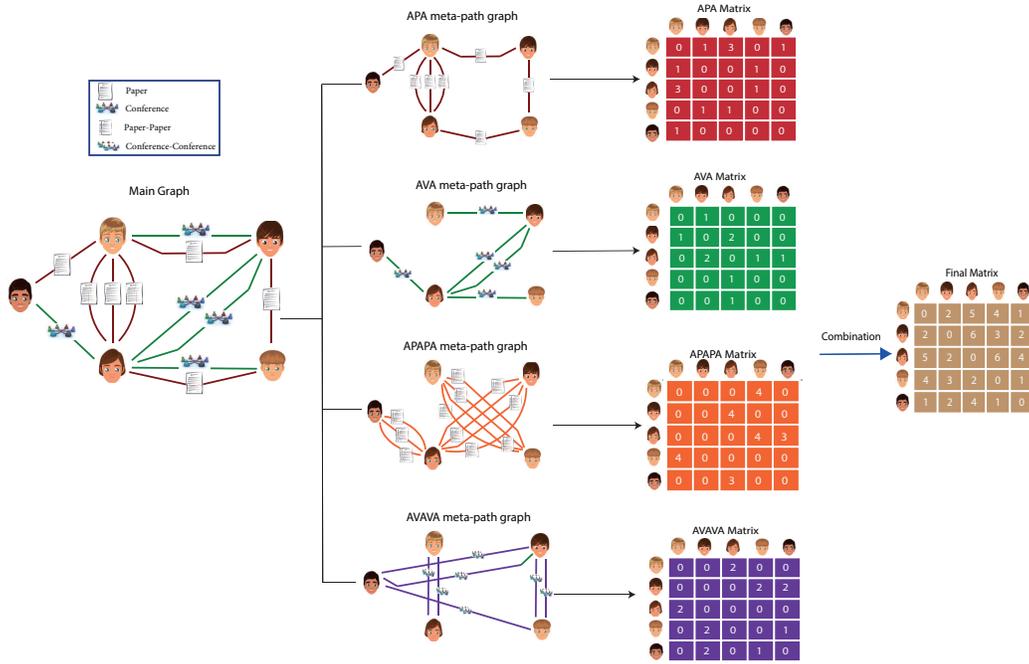


Figure 3: The procedure of a tensor input creation in a time-stamp t

3.1. Problem Statement

The two main tasks of information diffusion, *topic diffusion* and *cascade prediction* are described on heterogeneous network setting.

3.1.1. Topic Diffusion

In a general scheme of network $T_G = (A, G)$, meta-path P is defined where A and R represent the type of nodes and edges (meta-path). It is displayed as $A_1R_1A_2R_2\dots R_lA_{l+1}$.

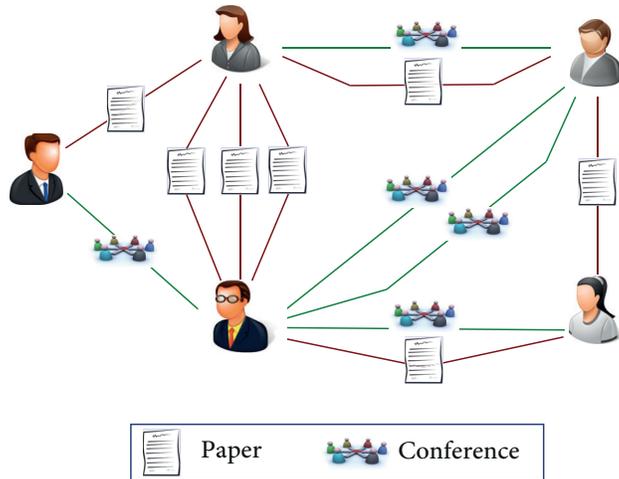


Figure 4: DBLP network

A scenario of heterogeneous collaboration network is represented in Figure 4 where an edge between two authors denotes a common publication. There may be multiple edges between two nodes for authors with multiple publications together. The APA presents the meta-path of coauthors on a paper (P) between two authors (A), and in the same way, AVA denotes authors (A) publish papers in the same conference (V).

Topic diffusion aims to solve the problem of who will write a paper on a particular topic at the time $t + 1$, when a node i wrote a paper on the same topic at time t .

Here, our aim is to construct a universal framework to employ the meta-path objects (instances) based on deep learning ideas to alleviate the learning of β_k computations in the prior works.

3.1.2. Cascade Prediction

For constructing cascades in the graph $G = (V, E)$, citation relationship between nodes i and j is required. The cascade path is defined according to the reference of its citing papers. Suppose we have N topics denoted by $N = n_i (1 < i < N)$. For each topic n_i , we use a cascade $C_i = (u_i, v_i, t_i)$ to record the topic diffusion process of n_i meaning that the author v_i cites paper of the author u_i and the time elapsed between the citation process is t_i . Figure 5 represents the cascades which are denoted by $(Paper1, t_1 = 0)$, $(Paper1, Paper2, t_2)$, $(Paper1, Paper3, t_3)$, $(Paper3, Paper4, t_4)$, $(Paper4, Paper5, t_5)$. In our framework, the aim is to

predict subsequent cascades of the specific topic at time $t+1$ from the current t -th snapshot graph.

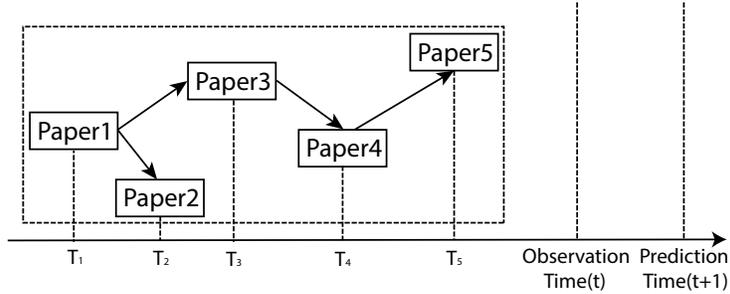


Figure 5: Cascade definition

3.2. Heterogeneous Deep Diffusion

To model topic diffusion in the heterogeneous networks, we introduce Heterogeneous Deep Diffusion(HDD) method by exploiting the meta-paths. In this part, meta-path graphs are extracted from the raw datasets to represent nodes in a heterogeneous network $G = (V, E)$. A snapshot of graph G at time t is characterized by a meta-path graph $G_t^k = (V_t^k, E_t^k)$, where V_t^k is a subset of nodes in V that have adopted meta-path k at time t . In our setting, graph snapshots are considered on different timestamps along with a variety of meta-path definitions.

3.2.1. Author Embedding

Each node is represented as a vector, $q \in R^N$ where N is the total number of authors. All users share an embedding tensor $M \in R^{N*N*T_s}$, which T_s is the number of timestamps as depicted in Figure 1.

3.2.2. Meta-path Encoding

The proposed HDD model extracts the semantic relationships among nodes in the semantic space based on different meta-paths and creates a meta-path graph for each individual meta-path; thus, the adjacency matrix in each meta-path graph is different. After extracting neighborhood features from each meta-path graph, at the final step, the extracted features are embedded together.

To represent the information flow, we use a variety of neural network architectures like LSTM, CNN, and CNN-LSTM. The main key factor here

is considering all meta-paths. All of the meta-paths author embedding tensors are merged to a unique tensor in our framework at different timestamps rather than applying ad-hoc based relationship. The meta-paths embedding in a unique tensor enriches the proposed approach from the following perspectives,

1. **Loss of information:** In the real world, two authors may have different relationships to be considered rather than single relation to prevent the loss of significant information.
2. **Time sequences:** Different snapshots of the meta-path graphs G^k are yielded to distinct author embedding cascades over time.

We encode the entire met-paths for each author by using embedding tensor as an input through an LSTM [52]. At each time step LSTM can choose to read from, write to, or reset the cell using gating mechanisms. They include four gates, generally denoted as i , o , f , and \tilde{c} , corresponding to the input, the output, the forget, and the new memory gate. In Equations (1) to (6), W_z and U_z correspond to weights of the input, x_t , and the hidden state, h_{t-1} , where z can either be the input gate, the output gate, and the forget gate or the memory gate, depending on the activation being calculated. The b_i , b_o , b_f and b_c are the biases of the input gate, the output gate, the forget gate and the memory cell, respectively. Also, the operation \odot denotes the element-wise vector product.

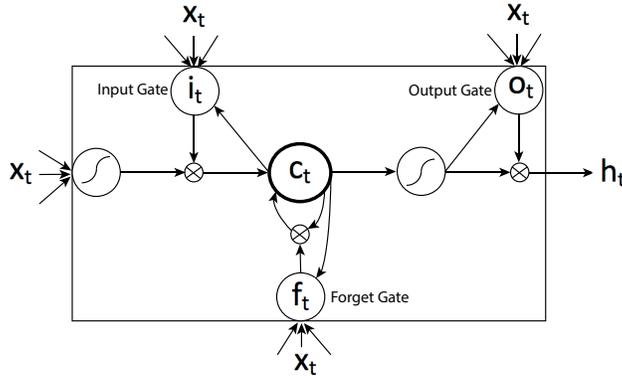


Figure 6: Long Short-Term Memory Cell

Intuitively, the gate controls the flow of information to enter and exit from the cell at each time step. In the first step, the forget gate, f_t , decides

how much of the previous state to take into account.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

Next, the input gate, i_t , decides which values to update and a new memory generation stage, \tilde{c}_t , creates a vector of new candidate values that could be added to the state.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (3)$$

Now, the new memory gate, c_t forgets the old cell state, \tilde{c}_{t-1} and gates the new memory cell, \tilde{c}_t .

$$c_t = f_t \odot \tilde{c}_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

Finally, the output gate, o_t , decides what parts of the cell state to output, then the output will be filtered by a tanh layer, to output the desired parts of it.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

We built the LSTM model from an embedding layer (of dimensionality 512), an LSTM layer (with 512 network units for each gate) with dropout regularization, and finally, a sigmoid activation function applied to the output of the LSTM.

Convolutional neural networks (CNNs)[53], are comprised of an input layer, one or more hidden layers, and an output layer. The hidden layers are a combination of convolutional layers, relu layers, pooling layers, and fully-connected layers. Convolutional layers will compute the output of neurons that are connected to only a small region in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. Relu layers will apply an element-wise activation function which zeros out negative inputs and is represented as $\max(0, x)$. Pooling layers will perform a down-sampling operation along the spatial dimensions (width and height) of the input. Fully-connected layers have neurons that are functionally similar to convolutional layers (compute dot products) but are different in that they are connected to all activations in the previous layer. The last fully-connected layer is called the output layer and it will compute

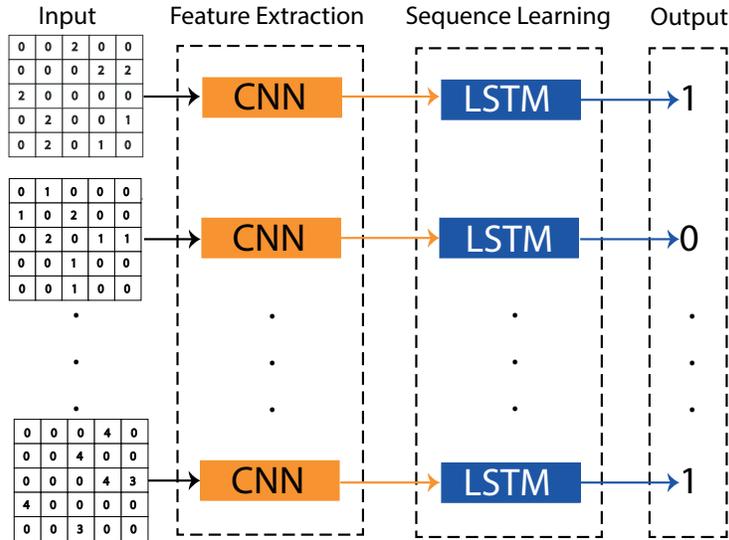


Figure 7: CNN-LSTM Architecture

the class scores. Stacking these layers will form a full CNN architecture as demonstrated in Figure 1.

CNN-LSTM is a combination of CNN for feature extraction and LSTM for summarization of the extracted features. Adopting an LSTM for aggregating the features enables the network to take the global structure into account while local features are extracted by CNN layers as represented in Figure 7. These features can be used in various heterogeneous network mining tasks, such as clustering [54], classification [55] and so on which we used for prediction. Our CNN-LSTM model uses an embedding layer (of dimensionality 1000), a one dimensional CNN layer of 5 convolutions interspersed with a 64 max pooling, an LSTM layer (with 1000 network units for each gate), and finally, a sigmoid activation function applied to the output of the LSTM.

One of the primary methods in the deep neural network is MLP. In this section this method is briefly described just for a more detailed comparison with other introduced methods [56].

The proposed method has the following properties:

1. We added weights to the input layer by considering different meta-paths in heterogeneous networks.

2. We have used deep representation learning which automatically extracts summarized features from each node in heterogeneous networks.

4. Experiments

We evaluate the proposed approach based on benchmark datasets and standard evaluation measures which are introduced in the following subsections. Then, the performance of the proposed approach is described in comparison with the earlier well-known methods. For topic diffusion and cascade prediction, initially, we need to select a topic like “data mining”. The authors who had a paper in the field of related topic considered as active. After that, we should predict the activation of other inactive nodes in the next timestamps with the proposed methods.

4.1. Datasets

The *DBLP*, *PubMed*, and *ACM* are benchmark real network datasets which are employed in our experimental studies.

- *DBLP*: This dataset is about computer science bibliography among authors in main conferences and publications [57]. Objects indicate authors in this network. Different meta-path such as APA (Author-Paper-Author), ACA (Author-Conference-Author), APAPA (Author-Paper-Author-Paper-Author), and ACACA (Author-Conference-Author-Conference-Author) are considered. Different topics are extracted from this dataset, and information diffusion about a specific topic is investigated. This dataset contains information from 1954 to 2016.
- *PubMed*: This dataset consists of medical science bibliography among authors in main conferences and publications in this domain[58]. In this network, the authors are represented by objects and meta-paths APA and APAPA are used. The dataset contains bibliographic information from 1994 to 2003.
- *ACM*: This dataset consists of the bibliographic information of publications of the Association for Computing Machinery (ACM) [57]. In this network, we used APA and APAPA meta-paths. The dataset contains information from 1959 to 2009.

Table 1: The main properties of applied datasets

Dataset	Authors	Papers
<i>DBLP</i>	215222	104940
<i>PubMed</i>	7140093	4271136
<i>ACM</i>	468114	649526
<i>APS</i>	96644	96060
<i>Citeseer</i>	–	50000

- *APS*: This dataset consists of the all American Physical Society (APS) journal articles [59]. In this network, we used APA and APAPA meta-paths. The dataset contains information from 1913 to 2016.
- *Citeseer*: This citation network extracted from the CiteSeer digital library. CiteSeer is an evolving scientific literature digital library and search engine that has focused primarily on the literature in computer and information science [60]. Nodes are publications and the directed edges denote citations. We used 50,000 nodes of this dataset.

A summary about the datasets are given in Table 1.

4.2. Evaluation Measures

For topic diffusion, we use Precision and Recall criteria to assess the performance. These measures are defined as,

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \quad (7)$$

Where True Positive (TP) is the active nodes that are correctly tagged as active by the algorithm, True Negative (TN) is the inactive nodes that are correctly tagged as inactive by the algorithm, False Positive (FP) is the active nodes that are falsely tagged as inactive by the algorithm, and False Negative (FN) is the inactive nodes that are falsely tagged as active by the algorithm.

Furthermore, AUPR(Auea under Precision Recall) curve is used by plotting Precision against Recall. The higher they are, the better the model is.

The MSE(Mean Squared Error) and AP(Average Precision) is applied to measure the accuracy of cascade prediction task. AP summarizes a precision-recall curve as the weighted mean of precisions achieved at each threshold [61].

On training and test data selection, we first consider all nodes with published papers as our particular topic of interest as active ones and the rest nodes as inactive. At time t , the training and test sets are selected as follows: *Training set*: Those within the time period from $t - 4$ to $t - 1$ and from $t - 4$ to t are considered as the training set for topic diffusion and cascade prediction separately.

Test set: Those within the time period from $t - 1$ to t are considered as the test set in topic diffusion. Additionally, the nodes tagged as active up to the time $t - 1$ are considered as the seed nodes that are activated initially in the start of the diffusion process. In the case of cascade prediction, the sequence of nodes at time t are employed as the test set.

4.3. Performance Evaluation

The proposed approach *HDD* is applied on topic diffusion and cascade prediction in comparison with the earlier well-known methods. On topic diffusion on heterogeneous networks, the performance of *HDD* is compared to *MLTM-R* which is the only related work in this category. The *Planetoid*[62] is one of the well-known embedding methods that is employed for our comparison. On cascade prediction, we employ *Node2vec* and *DeepCas* as well-known feature learning techniques that are presented in the following:

1. Planetoid[62]: It presents a learning frame-work based on graph embeddings which trains an embedding for each instance to jointly predict the class label and the neighborhood context in the graph.
2. Node2vec[33]: It learns a mapping of nodes to a low-dimensional space of features that maximizes the likelihood of preserving network neighborhoods of nodes. This method is selected as a representative work in the node embedding methods.
3. DeepCas[9]: It is initiated on paths samples from different snapshots of a graph. They used a GRU network to transform path samples into a single vector.
4. DeepWalk[32]: Inspired by the Word2Vec method [63], Perozzi et al. [32] proposed the DeepWalk that generates random paths over a graph.

It learns the new node representation by maximizing the co-occurrence probability of the neighbors in the walk.

4.3.1. ACM dataset

In this data set, the topics “Data Mining”, “Machine Learning” and “Decision Tree” are selected due to having the ground-truth about them in this dataset. Figures 8, 9 and 10 represent topic diffusion results on “Data Mining”, “Machine Learning” and “Decision Tree” topics on ACM. It can be observed that the LSTM and CNN-LSTM have a significant improvement rather than the other methods. In the dataset, LSTM increased the AUPR measure around 35% and 4% rather than MLTM-R and Node2Vec respectively and CNN-LSTM enhanced the AUPR up to 50% and 10% compared to MLTM-R and Node2Vec methods. After these methods, CNN is better than the MLTM-R and MLP which raised the AUPR by 30% in comparison with MLTM-R method.

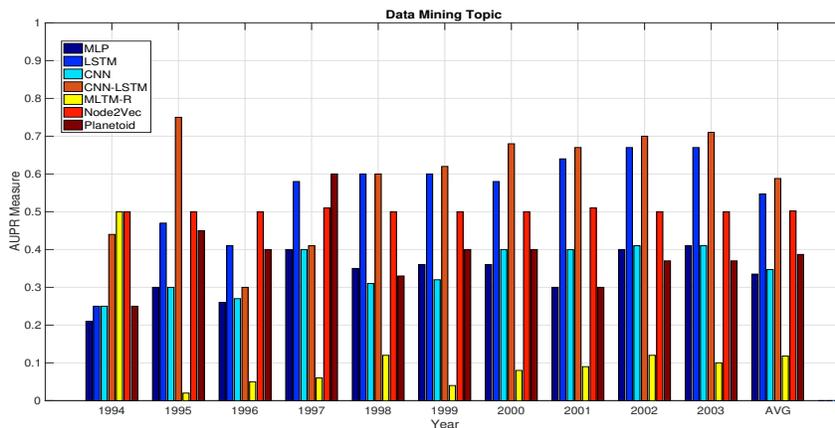


Figure 8: AUPR measure in ACM dataset for Data Mining topic (1994- 2003)

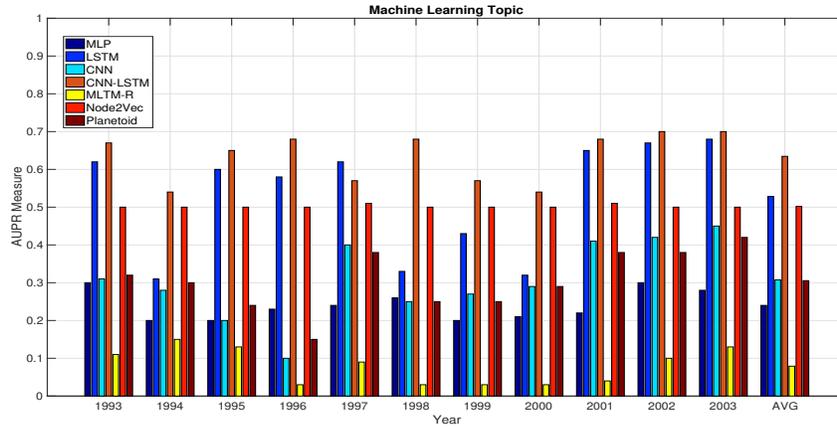


Figure 9: AUPR measure in ACM dataset for Machine Learning topic (1993- 2003)

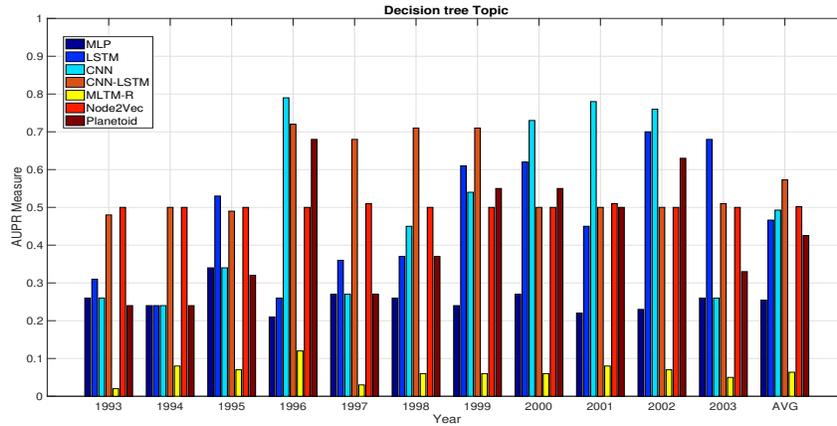


Figure 10: AUPR measure in ACM dataset for Decision tree topic (1993- 2003)

4.3.2. DBLP dataset

These topics “Data mining”, “Social Network” and “Regression” were selected in DBLP. We can infer from Figures 11, 12 and 13 that LSTM and CNN-LSTM still have better results but here, versus ACM dataset, MLTM-R method on average shows a slight growth compared with MLP. The results

comparison showed that CNN-LSTM returned high AUPR (overall AUPR about 60%) followed by LSTM and CNN.

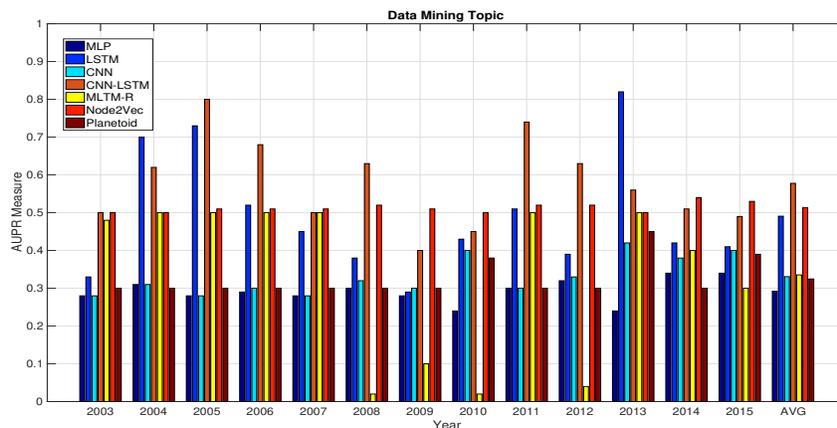


Figure 11: AUPR measure in DBLP dataset for Data Mining topic (2003- 2015)

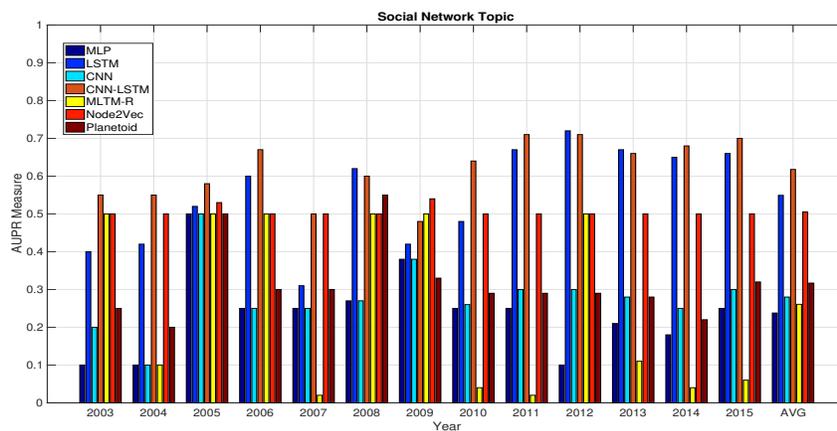


Figure 12: AUPR measure in DBLP dataset for Social Network topic (2003- 2015)

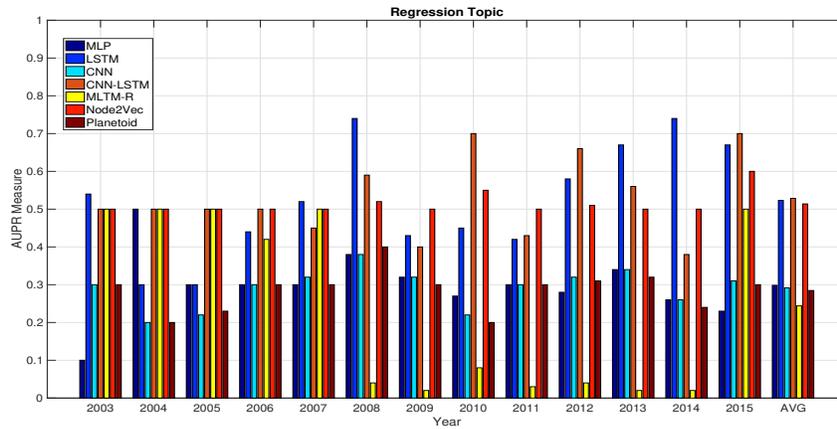


Figure 13: AUPR measure in DBLP dataset for Regression topic (2003- 2015)

4.3.3. PubMed dataset

In this dataset “*Health care*” topic was examined. As shown in the figure 14, LSTM and CNN-LSTM improved the AUPR Measure in comparison with other used methods. Here, CNN makes the results a little better against MLP due to the volume of data. According to AUPR, CNN-LSTM and LSTM outperformed MLTM-R by 25% and 18%, Node2Vec by 21% and 17% and Planetoid by 13% and 9% respectively.

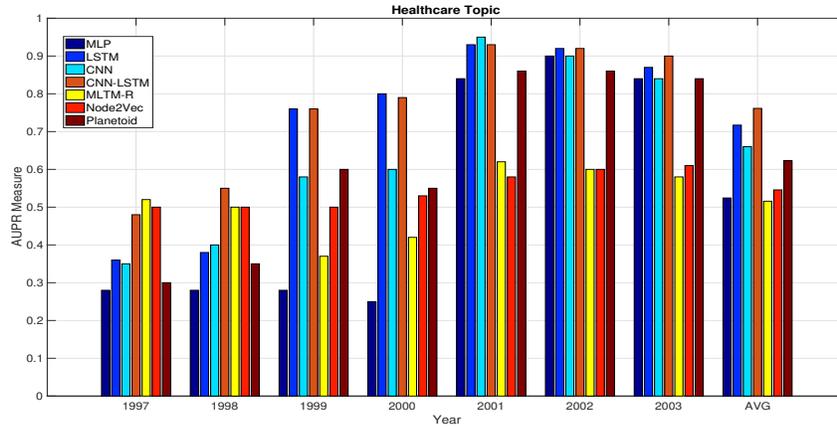


Figure 14: AUPR measure in PubMed dataset for Health care topic (1997- 2003)

4.3.4. APS Dataset

We used “*Quantum*” topic for this dataset. Figure 15 identifies further improvement (roughly 42%) of CNN-LSTM method against MLTM-R method and after this method, LSTM is highest. CNN-LSTM and LSTM also increase the AUPR against Node2Vec about 0.24% and 12% and Planetoid about 16% and 4% in order.

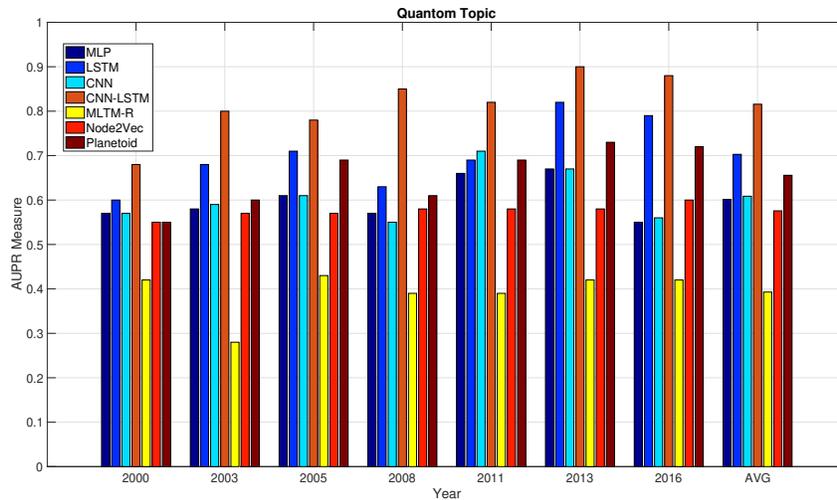
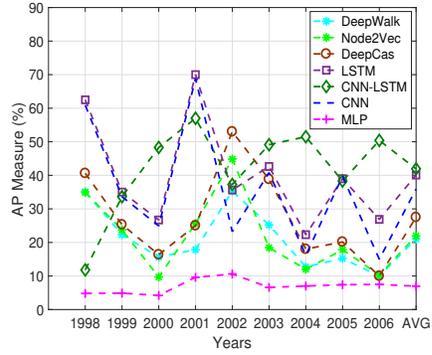
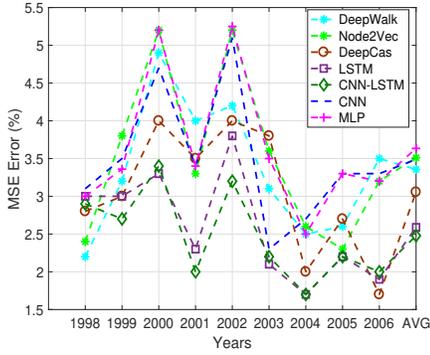


Figure 15: AUPR measure in Citeseer dataset for Quantum topic (2000- 2016)

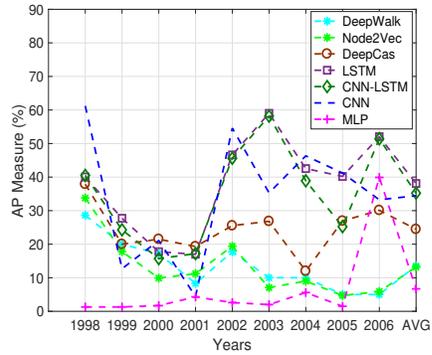
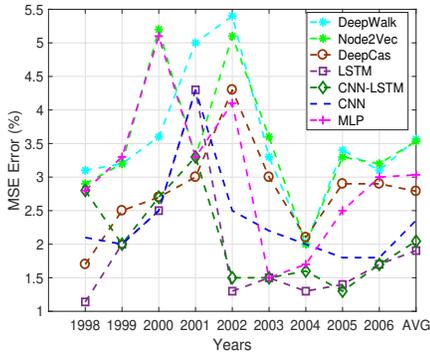
4.4. Cascade Prediction

We evaluate the prediction of cascades through the DBLP and Citeseer datasets because of existing citation relations. In DBLP dataset, we created cascades output according to PCP(paper-citation-paper) in data mining, machine learning topics and construct the meta-paths graph input based on PCP(paper-citation-paper), PVP(paper-conference-paper). As shown in Figures 16 and 17, LSTM and CNN-LSTM outperform the other methods by a significant margin.



(a) MSE error for Data Mining topic (b) AP measure for Data Mining topic

Figure 16: AP Measure and MSE error for Data Mining topic



(a) MSE error for Machine Learning topic (b) AP measure for Machine Learning topic

Figure 17: AP Measure and MSE error for Machine Learning topic

The Citeseer dataset is used in our experiments by considering PCP(paper-citation-paper) and PCPCP(paper-citation-paper-citation-paper) meta-paths. The results show the superiority of LSTM and CNN-LSTM in comparison with the other methods as represented in Figure 18.

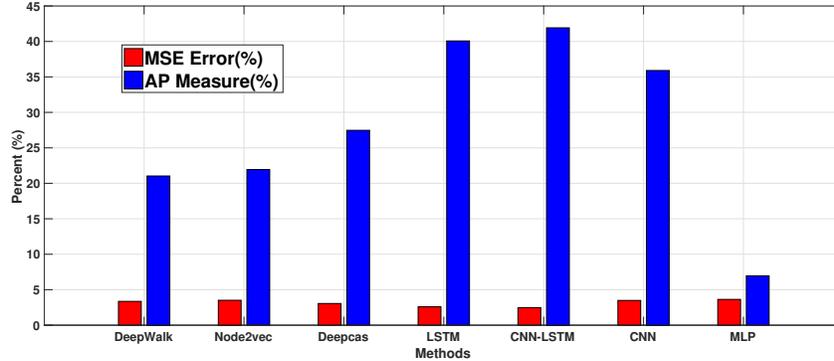


Figure 18: AP Measure and MSE error for Citeseer Dataset

The differences in results for each dataset indicate the importance of some main characteristics such as the topological structure, the context, and frequency of the specified topic in each network dataset. For example, the increasing pattern in PubMed results on topic diffusion is considerable due to the amount of topics as well as the number of nodes.

5. Conclusion

This paper studied topic diffusion and cascade prediction in heterogeneous networks. With this aim, we introduced a new end to end graph representation learning method in heterogeneous networks. The end-to-end predictor, HDD, outperformed the feature-based machine learning methods and alternative author embedding and meta-path graph embedding methods. The HDD model captures the influence of authors and cascades in different timestamps and meta-paths. Besides, employing the entire meta-paths through deep structures as an alternative to the ad-hoc based relations can significantly improve the prediction performance. In addition, the HDD model is flexible to use in other multilayer networks. We demonstrated the advantages of deep architecture methods in real-world networks such as DBLP, PubMed and ACM. Our experimental results on three real data sets verified the effectiveness and efficiency of our methods, LSTM and CNN-LSTM. The performance of the proposed method was compared with the state-of-the-art techniques. The obtained results showed that the proposed method outperforms the earlier ones. As future work, we are interested in

a combination of graph summarization and deep learning to improve the results.

References

References

- [1] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, The role of social networks in information diffusion, in: Proceedings of the 21st international conference on World Wide Web, ACM, 2012, pp. 519–528.
- [2] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, Y.-C. Zhang, Dynamics of information diffusion and its applications on complex networks, *Physics Reports* 651 (2016) 1–34.
- [3] Y. Hu, R. J. Song, M. Chen, Modeling for Information Diffusion in Online Social Networks via Hydrodynamics, *IEEE Access* 5 (2017) 128–135. doi:10.1109/ACCESS.2016.2605009.
URL <http://ieeexplore.ieee.org/document/7843668/>
- [4] P. L. Green, P. E. Plsek, Coaching and Leadership for the Diffusion of Innovation in Health Care: A Different Type of Multi-Organization Improvement Collaborative, *The Joint Commission Journal on Quality Improvement* 28 (2) (2002) 55–71. doi:10.1016/S1070-3241(02)28006-2.
URL <http://linkinghub.elsevier.com/retrieve/pii/S1070324102280062>
- [5] C. Kirst, M. Timme, D. Battaglia, Dynamic information routing in complex networks, *Nature Communications* 7 (6) (2016) 11061. doi:10.1038/ncomms11061.
URL <https://www.nature.com/articles/ncomms11061>
- [6] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, Y.-C. Zhang, Dynamics of information diffusion and its applications on complex networks, *Physics Reports* 651 (2016) 1–34. doi:10.1016/j.physrep.2016.07.002.
URL <http://dx.doi.org/10.1016/j.physrep.2016.07.002>

- [7] S. Dhamal, K. Prabuchandran, Y. Narahari, Information Diffusion in Social Networks in Two Phases, *IEEE Transactions on Network Science and Engineering* 3 (4) (2016) 197–210. doi:10.1109/TNSE.2016.2610838.
URL <http://ieeexplore.ieee.org/document/7570252/>
- [8] D. Margaris, C. Vassilakis, P. Georgiadis, Recommendation information diffusion in social networks considering user influence and semantics, *Social Network Analysis and Mining* 6 (1) (2016) 108. doi:10.1007/s13278-016-0416-z.
URL <http://link.springer.com/10.1007/s13278-016-0416-z>
- [9] C. Li, J. Ma, X. Guo, Q. Mei, Deepcas: An end-to-end predictor of information cascades, in: *Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee*, 2017, pp. 577–586.
- [10] H. Gui, Y. Sun, J. Han, G. Brova, Modeling Topic Diffusion in Multi-Relational Bibliographic Information Networks, in: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14*, ACM Press, New York, New York, USA, 2014, pp. 649–658. doi:10.1145/2661829.2662000.
URL <http://dl.acm.org/citation.cfm?doid=2661829.2662000>
- [11] H. Yang, M. Tang, T. Gross, Large epidemic thresholds emerge in heterogeneous networks of heterogeneous nodes, *Scientific Reports* 5 (1) (2015) 13122. doi:10.1038/srep13122.
URL <http://www.nature.com/articles/srep13122>
- [12] X. Wang, T. Zhao, Model for multi-messages spreading over complex networks considering the relationship between messages, *Communications in Nonlinear Science and Numerical Simulation* 48 (2017) 63–69.
- [13] F. Nian, X. Wang, Efficient immunization strategies on complex networks, *Journal of theoretical biology* 264 (1) (2010) 77–83.
- [14] Y. Sun, J. Han, Mining heterogeneous information networks: a structural analysis approach, *ACM SIGKDD Explorations Newsletter* 14 (2) (2013) 20–28. doi:10.1145/2481244.2481248.
URL <http://dl.acm.org/citation.cfm?doid=2481244.2481248>

- [15] A. Bordes, X. Glorot, J. Weston, Y. Bengio, Joint learning of words and meaning representations for open-text semantic parsing, in: *Artificial Intelligence and Statistics*, 2012, pp. 127–135.
- [16] X. Wang, T. Zhao, X. Qin, Model of epidemic control based on quarantine and message delivery, *Physica A: Statistical Mechanics and its Applications* 458 (2016) 168–178.
- [17] G. Ren, X. Wang, Epidemic spreading in time-varying community networks, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 24 (2) (2014) 023116.
- [18] A. Bordes, N. Usunier, J. Weston, O. Yakhnenko, Translating Embeddings for Modeling Multi-Relational Data, *Advances in NIPS* 26 (2013) 2787–2795. [arXiv:\protect\vrulewidth0pt\protect\href{http://arxiv.org/abs/1011.1669}](http://arxiv.org/abs/1011.1669){arXiv:1011.1669}v3, doi:10.1007/s13398-014-0173-7.2.
- [19] R. Socher, D. Chen, C. D. Manning, A. Y. Ng, Reasoning With Neural Tensor Networks for Knowledge Base Completion, *Proceedings of the Advances in Neural Information Processing Systems* 26 (NIPS 2013) (2013) 1–10[arXiv:\protect\vrulewidth0pt\protect\href{http://arxiv.org/abs/1301.3618}](http://arxiv.org/abs/1301.3618){arXiv:1301.3618}v2.
- [20] Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge Graph and Text Jointly Embedding, *Emnlp* (2014) 1591–1601.
- [21] Z. Liu, X. Wang, M. Wang, Inhomogeneity of epidemic spreading, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20 (2) (2010) 023128.
- [22] Z. Wen-Jie, W. Xing-Yuan, Inhomogeneity of epidemic spreading with entropy-based infected clusters, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23 (4) (2013) 043105.
- [23] D. J. Watts, A simple model of global cascades on random networks, *Proceedings of the National Academy of Sciences* 99 (9) (2002) 5766–5771.
- [24] J. Kuck, H. Zhuang, X. Yan, H. Cam, J. Han, Query-Based Outlier Detection in Heterogeneous Information Networks., *Advances in*

- database technology : proceedings. International Conference on Extending Database Technology 2015 (2015) 325–336. doi:10.5441/002/edbt.2015.29.
- [25] J. Shang, M. Qu, J. Liu, L. M. Kaplan, J. Han, J. Peng, Meta-path guided embedding for similarity search in large-scale heterogeneous information networks, arXiv preprint arXiv:1610.09769.
 - [26] N. Audebert, B. Le Saux, S. Lefèvre, Fusion of heterogeneous data in convolutional networks for urban semantic labeling, in: Urban Remote Sensing Event (JURSE), 2017 Joint, IEEE, 2017, pp. 1–4.
 - [27] J. Liu, M. Gong, K. Qin, P. Zhang, A deep convolutional coupling network for change detection based on heterogeneous optical and radar images, IEEE transactions on neural networks and learning systems.
 - [28] X. Wu, L. Song, R. He, T. Tan, Coupled deep learning for heterogeneous face recognition, arXiv preprint arXiv:1704.02450.
 - [29] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, Line: Large-scale information network embedding, in: Proceedings of the 24th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.
 - [30] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, T. S. Huang, Heterogeneous network embedding via deep architectures, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 119–128.
 - [31] Y. Dong, N. V. Chawla, A. Swami, metapath2vec: Scalable representation learning for heterogeneous networks, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2017, pp. 135–144.
 - [32] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 701–710.

- [33] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2016, pp. 855–864.
- [34] C. Shi, Y. Li, J. Zhang, Y. Sun, S. Y. Philip, A survey of heterogeneous information network analysis, *IEEE Transactions on Knowledge and Data Engineering* 29 (1) (2017) 17–37.
- [35] D. Zhou, S. A. Orshanskiy, H. Zha, C. L. Giles, Co-ranking Authors and Documents in a Heterogeneous Network, in: Seventh IEEE International Conference on Data Mining (ICDM 2007), IEEE, 2007, pp. 739–744. doi:10.1109/ICDM.2007.57.
URL <http://ieeexplore.ieee.org/document/4470320/>
- [36] R. Angelova, G. Kasneci, G. Weikum, Graffiti: graph-based classification in heterogeneous networks, *World Wide Web* 15 (2) (2012) 139–170. doi:10.1007/s11280-011-0126-4.
URL <http://link.springer.com/10.1007/s11280-011-0126-4>
- [37] X. Liu, Yingying Yu, C. Guo, Y. Sun, L. Gao, Full-text based context-rich heterogeneous network mining approach for citation recommendation, in: IEEE/ACM Joint Conference on Digital Libraries, IEEE, 2014, pp. 361–370. doi:10.1109/JCDL.2014.6970191.
URL <http://ieeexplore.ieee.org/document/6970191/>
- [38] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, *The European Physical Journal B* 26 (4) (2002) 521–529. doi:10.1140/epjb/e20020122.
URL <http://www.springerlink.com/index/10.1140/epjb/e20020122>
- [39] R. Yang, B.-H. Wang, J. Ren, W.-J. Bai, Z.-W. Shi, W.-X. Wang, T. Zhou, Epidemic spreading on heterogeneous networks with identical infectivity, *Physics Letters A* 364 (3-4) (2007) 189–193. doi:10.1016/j.physleta.2006.12.021.
URL <http://linkinghub.elsevier.com/retrieve/pii/S0375960106019025>
- [40] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, D. Montesi, Spreading processes in multilayer networks, *IEEE Transactions on Net-*

- work Science and Engineering 2 (2) (2015) 65–83. doi:10.1109/TNSE.2015.2425961.
- [41] L. Wang, G. Z. Dai, Global stability of virus spreading in complex heterogeneous networks, *Siam Journal on Applied Mathematics* 68 (5) (2008) 1495–1502. doi:10.1137/070694582.
- [42] M. Nadini, K. Sun, E. Ubaldi, M. Starnini, A. Rizzo, N. Perra, Epidemic spreading in modular time-varying networks, arXiv preprint arXiv:1710.01355.
- [43] G. Demirel, E. Barter, T. Gross, Dynamics of epidemic diseases on a growing adaptive network, *Scientific reports* 7 (2017) 42352.
- [44] P. Sermpezis, T. Spyropoulos, Information diffusion in heterogeneous networks: The configuration model approach, in: *Proceedings - IEEE INFOCOM*, 2013, pp. 3261–3266. doi:10.1109/INFOCOM.2013.6567148.
- [45] Y. Zhou, L. Liu, Social influence based clustering of heterogeneous information networks, in: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2013, pp. 338–346.
- [46] S. Molaei, S. Babaei, M. Salehi, M. Jalili, Information spread and topic diffusion in heterogeneous information networks, *Scientific Reports* 8 (1) (2018) 9549.
- [47] S. Li, Z. Jin, Modeling and Analysis of New Products Diffusion on Heterogeneous Networks, *Journal of Applied Mathematics* 2014 (2014) 1–12. doi:10.1155/2014/940623.
URL <http://www.hindawi.com/journals/jam/2014/940623/>
- [48] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, *Physics Reports* 544 (1) (2014) 1–122. doi:10.1016/j.physrep.2014.07.001.
- [49] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781.

- [50] W. Cheng, C. Greaves, M. Warren, From n-gram to skipgram to conc-gram, *International journal of corpus linguistics* 11 (4) (2006) 411–433.
- [51] Q. Cao, H. Shen, K. Cen, W. Ouyang, X. Cheng, Deephawkes: Bridging the gap between prediction and understanding of information cascades, in: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ACM, 2017, pp. 1149–1158.
- [52] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural computation* 9 (8) (1997) 1735–1780.
- [53] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, *The handbook of brain theory and neural networks* 3361 (10) (1995) 1995.
- [54] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, W. Gao, Multiple kernel k-means with incomplete kernels, *IEEE transactions on pattern analysis and machine intelligence*.
- [55] X. Yu, Z. Zhou, Q. Gao, D. Li, K. Ríha, Infrared image segmentation using growing immune field and clone threshold, *Infrared Physics & Technology* 88 (2018) 184–193.
- [56] F. J. Ordóñez, D. Roggen, Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition, *Sensors* 16 (1) (2016) 115.
- [57] Citation Network Dataset.
URL <https://aminer.org/billboard/citation>
- [58] Data Repository for NaN Group.
URL <http://carl.cs.indiana.edu/data/>
- [59] American Physical Society Dataset.
URL <https://journals.aps.org/datasets>
- [60] Citeseer Dataset.
URL <http://konect.uni-koblenz.de/networks/citeseer>
- [61] M. Zhu, Recall, precision and average precision, *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo 2* (2004) 30.

- [62] Z. Yang, W. W. Cohen, R. Salakhutdinov, Revisiting semi-supervised learning with graph embeddings, arXiv preprint arXiv:1603.08861.
- [63] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in neural information processing systems, 2013, pp. 3111–3119.