Table 3: DDI detection performance. Prec@k stands for precision@k where k equals to the number of known DDI pairs and AP is the average precision. Mean and standard deviation of 5 random splits are given for the proposed method and baselines with similarity function. Baselines without similarity learning were reported for a single test split.

Methods	Prec@k (std)	AP (std)
Proposed	**0.86 (0.005)**	**0.92 (0.003)**
DeepWalk + Similarity	0.78 (0.007)	0.86 (0.011)
node2vec + Similarity	0.79 (0.011)	0.86 (0.013)
LINE-1 + Similarity	0.72 (0.011)	0.79 (0.010)
LINE-2 + Similarity	0.74 (0.003)	0.80 (0.016)
LINE + Similarity	0.74 (0.012)	0.82 (0.016)
Deepwalk	0.56	0.54
Node2Vec	0.44	0.43
LINE-1	0.58	0.60
LINE-2	0.56	0.59
LINE	0.63	0.63

ratio. In this task, network for baseline methods includes drugs, indication, ADR nodes and pairwise links between drugs-indications, drugs-ADRs, and drug-drug DDI pairs. Another baseline would be simply creating a traditional homogeneous network for only drug nodes and the edges could represent the DDI interactions. However, this approach considers only the known edges, therefore embedding learning for drugs with unknown links is not possible for many approaches working with edge lists [9, 24]. Therefore, the baseline methods were applied to the whole network and two types of baselines were considered; (1) obtaining hyperedge similarity by learning the similarity function for the baseline embedding, (2) computing pairwise cosine similarity of drug embeddings. As it can be seen from the table, taking different types of information about drugs into account by forming hyperedges and learning embeddings to enforce high inter-hyperedge similarities gives us a better performance to detect DDIs rather than treating each node individually and considering only pairwise links between them.

We also visualize the drug nodes after embedding learning to preserve inter-hyperedge interactions by applying t-SNE on the heterogeneous hyper-network. Figure 3 demonstrates the separation between drug nodes involved in DDI pairs and unknown drugs in 2-D space. According to our observation, learning embeddings by the proposed approach for the DDI detection purpose provides a more obvious separation between the two groups of drugs with smaller intra-group distances. Among the baseline approaches, LINE with the concatenation of first and second order embeddings could also produce a separation better than other baselines. LINE-1 and 2 were not included because of their similar performances to DeepWalk and node2vec.

3.4 FAERS Dataset

The FDA Adverse Event Reporting System (FAERS) is a self-reported adverse event and associated medications database found in [5]. In this paper, the cleansed version of 2016 FAERS data [1] was used in the experiments. Each report in the FAERS dataset corresponds to one patient who had taken some drugs and suffered from several adverse drug reactions (ADRs). Thus, each report contains patient's demographic information, drugs taken and the adverse drug events encountered. Thus, the heterogeneous hyper-network constructed for FAERS data has 3 different types of nodes including patients, drugs and ADRs. Demographic information such as age group, gender, weight and type of patient's report were coded as one-hot vectors for categorical attributes and all the information was concatenated to represent patient nodes, resulting 7 attributes. Drugs and ADRs were similarly represented by sparse vectors based on their ingredient and ID information resulting 474 and 1,013 attributes, respectively. Nodes with many missing information were simply eliminated resulting total of 1,507 patients, 689 drugs and 1,006 ADR nodes. We constructed a hyper-network from FAERS dataset by forming a hyperedge for each patient. Thus, the total number of hyperedges was 1,507 with an average of 5 nodes in each. In this case, one hyperedge comprises of one patient, on average 2 drugs and 3 corresponding ADRs.

In this experiment, we observed the behavior of the patient embeddings based on age group and gender attributes. For this purpose, t-SNE was applied on the node embeddings obtained by the proposed approach and the baseline approaches and the plots of patient nodes were given in Figure 4. Here, node2vec produced a similar result as DeepWalk, and the results of LINE-1 and LINE-2 had a similar pattern as LINE. Therefore, two of the baseline approaches are given in the figure. Since the proposed approach treats different types of nodes forming one hyperedge as a whole, we could see a better separation of patients based on their demographics. Female/male patients and their age groups could be easily observed in the plot generated for HHNE. Another interesting observation is that the patient nodes locate in an order with respect to age for each gender group. These results can help to understand the dynamics between age, gender and the reported problems since the proposed method also enforces patient embeddings to be similar if there share a sufficient amount of ADR and drug nodes.

4 RELATED WORK

Embedding learning has been a widely studied topic in machine learning literature. Some of the popular early approaches focused on finding a non-linear embedding for dimensionality reduction purposes such as Local Linear Embedding (LLE) [21], Laplacian eigenmaps [2], and IsoMap [25]. Learning latent representations which can preserve the underlying structures has always drawn attention. Embedding learning is also commonly encountered in network analysis. For instance, an approach leveraging the structural information of weighted graphs to learn a representation for nodes was proposed in [3]. A network embedding method preserving local and global network structures was developed by [24] and truncated random walks were utilized by [18] to learn latent representations via an online learning algorithm. One recent study used Modularized Nonnegative Matrix factorization (M-NMF) model to incorporate the community structure into network embedding [29], while [13] presented a semi-supervised embedding approach for networks with outliers.

[5] https://open.fda.gov/data/faers/