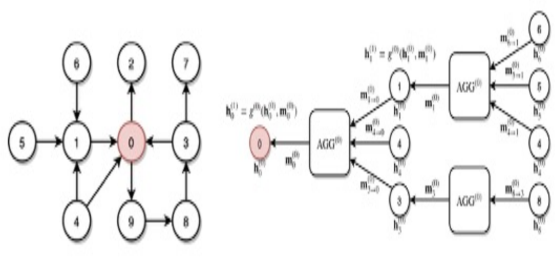
AUTHORS: Yonas Sium, Georgios Kollias, Tsuyoshi Ide, Payel Das, Naoki Abe, Aurelie Lozano, Qi Li

TITLE: DIRECTION AWARE POSITIONAL AND STRUCTURAL ENCODING FOR DIRECTED GRAPH NEURAL NETWORKS

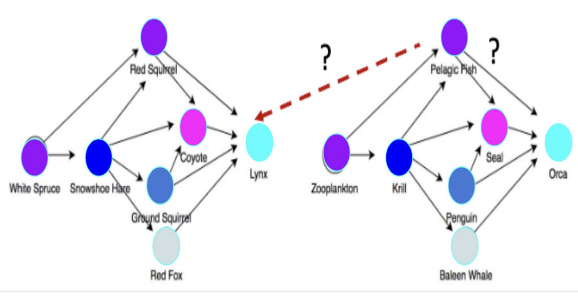
Problem Statement

Link prediction require a joint representation learning of the linked nodes



MP-GNN

- GNNs learn structural node representation via message passing between the nodes of graphs
- GNNs are designed for node classification because it learns node representations directly from k hop neighbors
- GNNs can not learn the latent link information between the linked nodes



[Srinivasan and Reibero, ICLR 2020]

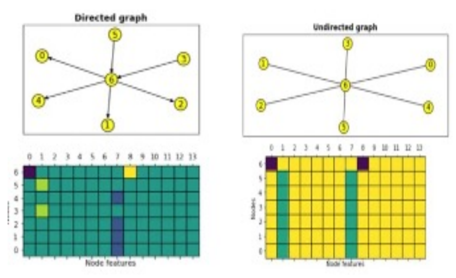
- Nodes in Identical subgraphs get same representation(GIN, Xu et al. 2019)
- Example of MP-GNN Failure for Link prediction on food web networks
- MP-GNN mis predict. The potential link from the isomorphic pray in sea and the top predator on the land

Task Formulation

Positional Encoding: Provides additional features that can help to get the structural information of a link

Citation networks

- Directed Link Prediction has many real-world applications
- For example: Recommender Systems, Citation Networks, Biomedical knowledge graph

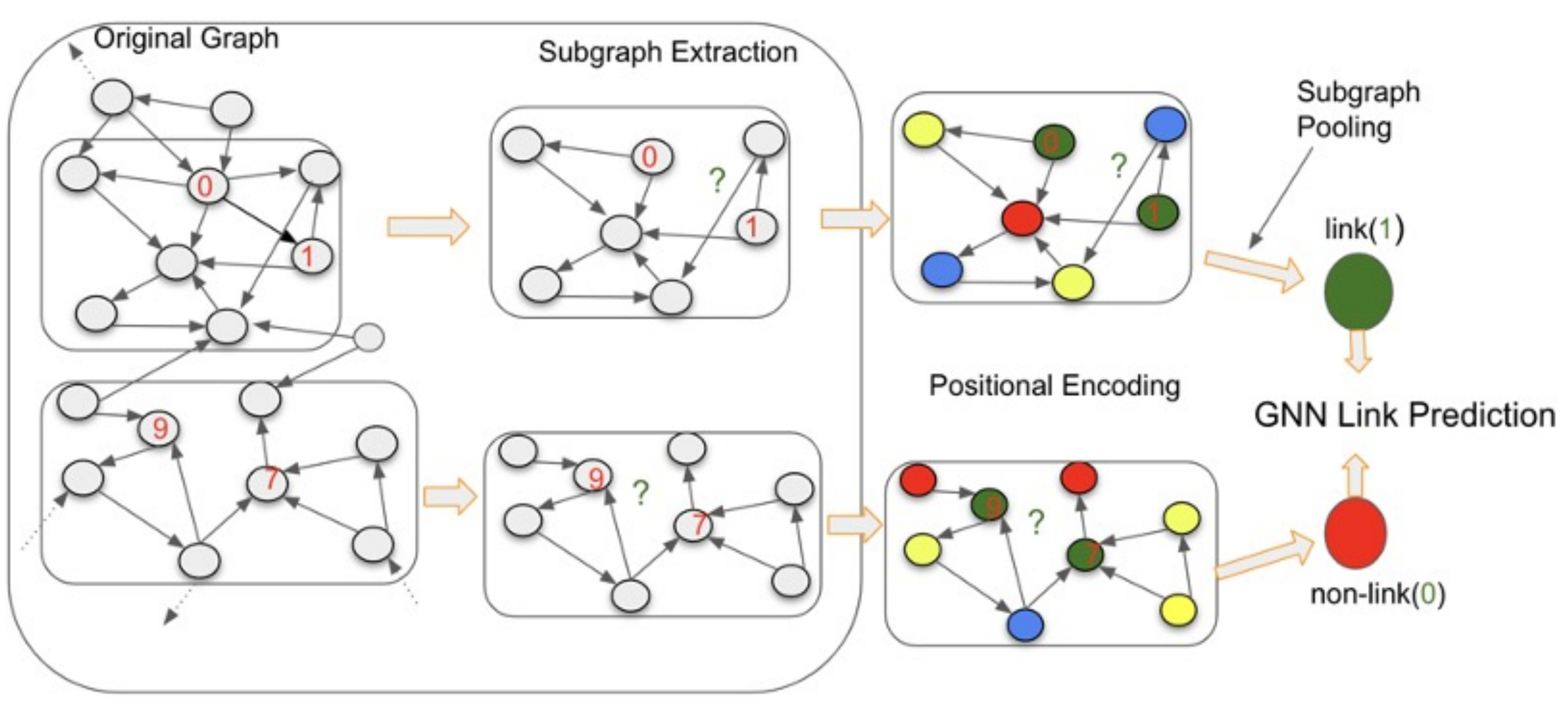


SVD based node encoding

- SVD based matrix factorizations is **valid** positional encoding and can capture **directionality** $A = U\Sigma V^T$
- For example, for the two directed and undirected subgraphs, SVD provide representation of the nodes based on the incoming and outgoing edges for the directed subgraph and based on the number of links for undirected subgraph

Proposed Solution

Direction Aware Positional Encoding



Proposed Framework

- Extracting k -hop subgraph around a directed link from $u \mapsto v$.
- we computed **Truncated SVD** and **ranking score using HITS algorithm** as a valid positional encodings
- **Truncated SVD** is an approximation of the $U_d \in \mathbb{R}^{n \times d}$ of its left singular subspace corresponding to its top singular value
- **HITS algorithm** to compute ranking score using, the authority value $\boldsymbol{a}$ and hub value $\boldsymbol{h}$
- Concatenate the computed Positional encodings of node [i] around the subgraph of directed link
- Use the computed positional encoding as initial feature during GNN training

Theory:
We prove that our positional encodings concatenated with zero-one labelling trick encodings make up a valid labelling trick. This means that a node-most expressive GNN with this labelling scheme can learn the structural representation of an ordered set of nodes, which is basically our directed edge (joint 2-node structural representation).

Experimental Results

Directed link prediction

AUC performance for Directed Link Prediction, when both truncated SVD and Rank positional encodings are used.

Model	Cornel	Texas	Wisconsin	Citeseer	CoraML
GCN(SVD + Rank)	86.16 ± 1.52	87.27 ± 2.77	82.13 ± 2.26	87.97 ± 0.57	88.15 ± 0.73
GIN(SVD + Rank)	88.01 ± 2.75	**90.72 ±2.24**	90.72 ± 1.68	89.12 ± 0.57	88.28 ± 0.25
SAGE(SVD + Rank)	88.24 ± 3.2	88.88 ± 2.72	89.13 ± 2.27	87.47 ± 1.97	87.92 ± 0.23
DGCN	82.24 ± 3.47	84.01 ± 1.67	82.89 ± 1.74	82.02 ± 0.8	82.92 ± 0.37
DiGraphIB	81.93 ± 1.65	82.72 ± 1.58	81.67 ± 1.74	84.89 ± 0.76	85.27 ± 0.62
Magnet	83.32 ± 2.71	83.01 ± 1.72	84.7 ± 1.92	86.72 ± 1.42	85.77 ± 0.42
DGCN(SVD + Rank)	89.24 ± 2.47	87.04 ± 1.92	87.21 ± 1.74	88.75 ± 0.66	90.21 ± 1.37
DiGraphIB(SVD + Rank)	87.58 ± 2.17	87.01 ± 2.87	88.11 ± 2.74	89.82 ± 0.68	89.2 ± 0.58
Magnet(SVD + Rank)	**91.98 ± 1.62**	89.98 ± 2.91	**90.82 ± 1.08**	**91.66 ± 0.81**	**93.85 ± 1.27**

AUC performance for Directed Link Prediction, when only truncated SVD positional encoding is

Model	Cornel	Texas	Wisconsin	Citeseer	CoraML
GCN(SVD)	82.03 ± 1.52	82.57 ± 1.7	77.33 ± 1.8	83.87 ± 1.47	82.87 ± 1.7
GIN(SVD)	84.25 ± 2.1	**91.35 ± 3..87**	81.32 ± 2.76	85.58 ± 0.74	87.92 ± 0.42
SAGE(SVD)	83.3 ± 2.77	82.21 ± 2.5	81.2 9 ± 1.92	82 ± 0.52	84.57 ± 0.75
DGCN(SVD)	84.86 ± 1.6	80.27 ± 1.6	85.65 ± 1.78	81.29 ± 0.87	83.47 ± 1.21
DiGraphIB(SVD)	85.86 ± 2.12	84.27 ± 2.63	87.65 ± 1.78	85.29 ± 0.87	88.87 ± 0.77
Magnet(SVD)	**90.14 ± 2.7**	89.69 ± 1.9	**91.42 ± 2.1**	**88.23 ± 1.24**	**89.08 ± 1.72**

AUC performance for Directed Link Prediction, when only Rank positional encodings is used.

Model	Cornel	Texas	Wisconsin	Citeseer	CoraML
GCN(Rank)	79.03±1.52	82.57 ± 1.7	77.33 ± 1.8	83.87 ± 1.47	82.87 ± 1.7
GIN(Rank)	**86.15 ± 1.31**	83.78 ± 2.1	82.56 ± 1.76	87.52 ± 0.42	86.43 ± 0.78
SAGE(Rank)	82.28 ± 1.97	83.79 ± 2.07	81.72 ± 1.34	87.18 ± 0.48	85.51 ± 0.28
DGCN(Rank)	82.72 ± 2.87	83.22 ± 1.73	83.72 ± 2.34	87.18 ± 0.48	83.42 ± 0.52
DiGraphIB(Rank)	83.04 ± 1.07	81.22 ± 3.34	**83.92 ± 1.34**	88.18 ± 0.28	87.23 ± 0.72
Magnet(Rank)	85.12 ± 1.89	**87.22 ± 1.47**	83.08 ± 0.87	**89.18 ± 0.48**	**90.08 ± 0.65**

Previous Work

Using DRNL distance-based encodings designed for undirected graphs

Model	Cornel	Texas	Wisconsin	Citeseer	CoraML
GCN(DRNL)	79.03	**81.27**	80.72	80.57	81.05
GIN(DRNL)	**81.07**	79.50	81.62	81.67	**82.02**
SAGE(DRNL)	79.92	75.76	**82.02**	**82.12**	81.17

Conclusion

- Adding direction aware positional encoding can help GNNs to predict directed link
- Both Truncated SVD and ranking score using HITS algorithm can make GNNs more powerful for directed link prediction
- Using distance-based encodings designed for undirected graphs (e.g. DRNL) in the directed graph settings yields much inferior performance compared to other baselines
- Using only SVD- or HITS-based positional encodings for initialization is effective; still by combining the two as we do the performance is maximized (ablation studies)