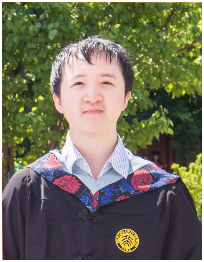
Link Prediction with Persistent Homology: An Interactive View

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Chao Chen

Peking University, IBM T. J. Watson Research Center, Stony Brook University

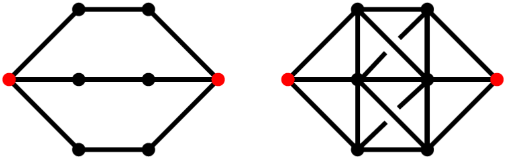
ICML | 2021

Thirty-eighth International
Conference on Machine Learning



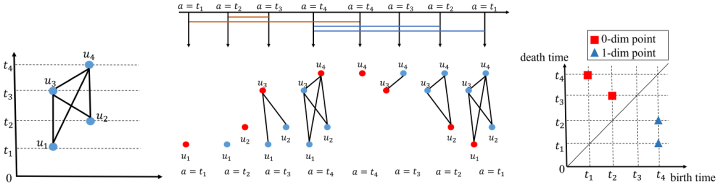
June, 2021

Introduction



- Task: link prediction
- Motivation: graph connectivity information has been used (i.e., node degree, distance to target nodes), but not enough
- We propose to use advanced topological information: Persistent Homology (PH)
  - count the number of loops
  - measure the range of distance for each loop

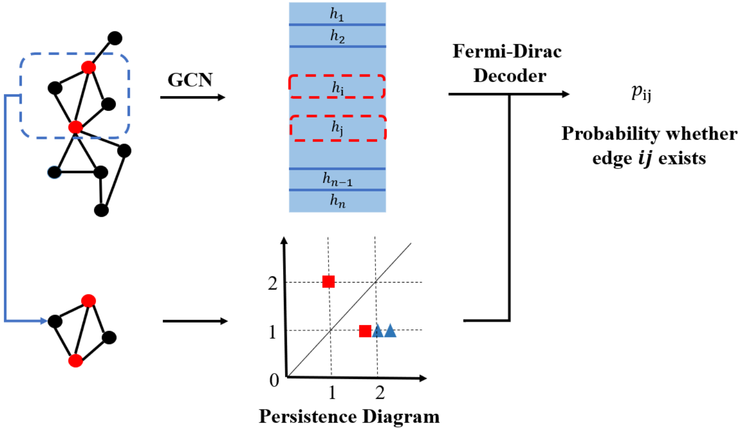
Persistent homology



(a) (b) (c)

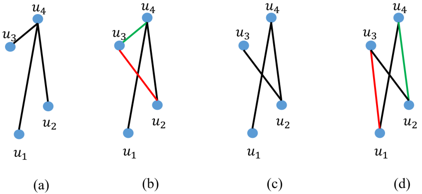
- PH: capture topological structures (connected components, loops) and encode their significance in view of the filter function
- Filter function: an observation of the elements of the graph
- Contribution:
  - Involve pairwise PH to enhance GNN for link prediction
  - Accelerate the computation of PH

TLC-GNN



- Enclosing subgraph: intersection of k-hop neighborhoods of target nodes
- Minimize the cross-entropy loss using negative sampling

A Faster Algorithm



- Maintain a rooted tree while going through the descending filtration
- Every time a new edge is added, find the corresponding persistence pair by inspecting the newly formed loop
- Update the tree and continue with the next edge
- The proposed faster algorithm achieves 1.5 to 2.5 times speedup

Table 2. Computational time (seconds per edge) evaluation.

	PUBMED	PHOTO	COMPUTERS
ALG. 1	0.0068	1.6557	4.6531
ALG. 2	0.0027	1.1176	2.7033

Experiments

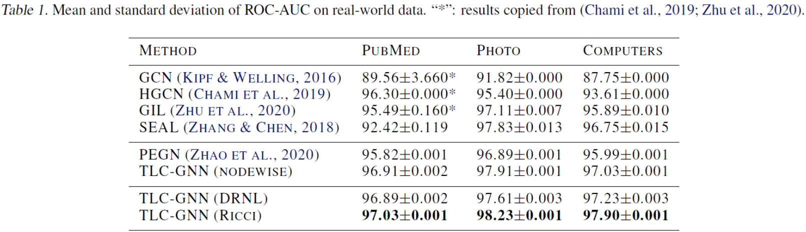


Table 1. Mean and standard deviation of ROC-AUC on real-world data. "*": results copied from (Chami et al., 2019; Zhu et al., 2020).

Method	PubMed	Photo	Computers
GCN (Kipf & Welling, 2016)	89.56±3.660*	91.82±0.000	87.75±0.000
HGCN (Chami et al., 2019)	96.30±0.000*	95.40±0.000	93.61±0.000
GIL (Zhu et al., 2020)	95.49±0.160*	97.11±0.007	95.89±0.010
SEAL (Zhang & Chen, 2018)	92.42±0.119	97.83±0.013	96.75±0.015
PEGN (Zhao et al., 2020)	95.82±0.001	96.89±0.001	95.99±0.001
TLC-GNN (nodewise)	96.91±0.002	97.91±0.001	97.03±0.001
TLC-GNN (DRNL)	96.89±0.002	97.61±0.003	97.23±0.003
TLC-GNN (Ricci)	**97.03±0.001**	**98.23±0.001**	**97.90±0.001**

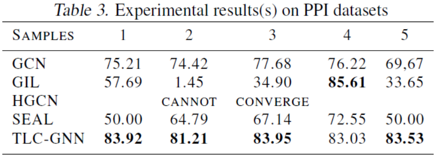


Table 3. Experimental results(s) on PPI datasets

Samples	1	2	3	4	5
GCN	75.21	74.42	77.68	76.22	69,67
GIL	57.69	1.45	34.90	**85.61**	33.65
HGCN		CANNOT	CONVERGE		
SEAL	50.00	64.79	67.14	72.55	50.00
TLC-GNN	**83.92**	**81.21**	**83.95**	83.03	**83.53**

- AUC-ROC: Achieves SOTA on nearly all the benchmarks

Thanks for listening!

Paper / code link: https://arxiv.org/abs/2102.10255