Graph Linear Convolution Pooling for Learning in Incomplete High-Dimensional Data Supplementary File

Fanghui Bi, Tiantian He, *Member, IEEE*, Yew Soon Ong, *Fellow, IEEE*, and Xin Luo, *Senior Member, IEEE*

This is the supplementary file for the paper entitled "Graph Linear Convolution Pooling for Learning in Incomplete High-Dimensional Data". The proofs, details of experimental settings, additional tables and figures are put into this file and cited.

I. PROOFS

A. Proof of Theorem 1

To prove ***Theorem* 1**, two directions of the iff conditions should be considered. If the conditions that $S_1=S_2=S$ and $\kappa\cdot\sum_{y=x,y\in X_1} g_{c_1,y}\big/\sqrt{W(y)}=\sum_{y=x,y\in X_2} g_{c_2,y}\big/\sqrt{W(y)}$ are given, for $\kappa=\sqrt{W(c_2)/W(c_1)}$ and $x\in S$. According to (6), we have:

$$\begin{cases} h(c_i,X_i)=\sum\limits_{x\in X_i}\hat{g}_{c_i x}x, \\ \hat{g}_{c_i x}=\dfrac{g_{c_i,x}}{\sqrt{W(c_i)\cdot W(x)}}. \end{cases} \tag{S1}$$

Given (S1), we derive:

$$\begin{cases} h(c_1,X_1)=\sum\limits_{x\in X_1}\hat{g}_{c_1,x}x=\sum\limits_{x\in X_1}\dfrac{g_{c_1,x}}{\sqrt{W(c_1)\cdot W(x)}}x, \\ h(c_2,X_2)=\sum\limits_{x\in X_2}\hat{g}_{c_2,x}x=\sum\limits_{x\in X_2}\dfrac{g_{c_2,x}}{\sqrt{W(c_2)\cdot W(x)}}x. \end{cases} \tag{S2}$$

With the condition that $S_1=S_2=S$, we have:

$$h(c_1,X_1)-h(c_2,X_2)=\sum_{x\in S}\left[\sum_{y=x,y\in X_1}\frac{g_{c_1,y}}{\sqrt{W(c_1)\cdot W(y)}}-\sum_{y=x,y\in X_2}\frac{g_{c_2,y}}{\sqrt{W(c_2)\cdot W(y)}}\right]\cdot x. \tag{S3}$$

Given $\kappa\cdot\sum_{y=x,y\in X_1} g_{c_1,y}\big/\sqrt{W(y)}=\sum_{y=x,y\in X_2} g_{c_2,y}\big/\sqrt{W(y)}$, for $\kappa=\sqrt{W(c_2)/W(c_1)}$, based on (S3), we directly have $h(c_1, X_1)=h(c_2, X_2)$.

If the conditions that $h(c_1, X_1)=h(c_2, X_2)$, we can prove that the conditions mentioned in ***Theorem* 1** are necessary by showing the contradictions while they are not satisfied. Given $h(c_1, X_1)=h(c_2, X_2)$, we have:

$$h(c_1,X_1)-h(c_2,X_2)=\sum_{x\in X_1}\frac{g_{c_1,x}}{\sqrt{W(c_1)\cdot W(x)}}x-\sum_{x\in X_2}\frac{g_{c_2,x}}{\sqrt{W(c_2)\cdot W(x)}}x=0. \tag{S4}$$

First, we assume $S_1\neq S_2$ for all $X_1, X_2\in\mathcal{X}$, the following equations are achieved:

$$\begin{aligned} &h(c_1,X_1)-h(c_2,X_2) \\ &=\sum_{x\in S_1\cap S_2}\left[\sum_{y=x,y\in X_1}\frac{g_{c_1,y}}{\sqrt{W(c_1)\cdot W(y)}}-\sum_{y=x,y\in X_2}\frac{g_{c_2,y}}{\sqrt{W(c_2)\cdot W(y)}}\right]\cdot x \\ &+\sum_{x\in S_1\setminus S_2}\sum_{y=x,y\in X_1}\frac{g_{c_1,y}}{\sqrt{W(c_1)\cdot W(y)}}\cdot x-\sum_{x\in S_2\setminus S_1}\sum_{y=x,y\in X_2}\frac{g_{c_2,y}}{\sqrt{W(c_2)\cdot W(y)}}\cdot x=0. \end{aligned} \tag{S5}$$

Since (S5) holds for any x, we could define a function $f(\cdot)$ as:

$$x=\begin{cases} f(x), & \text{for } x\in S_1\cap S_2; \\ f(x)-1, & \text{for } x\in S_1\setminus S_2; \\ f(x)+1, & \text{for } x\in S_2\setminus S_1. \end{cases} \tag{S6}$$

And if (S5) holds, we also infer that:

$$\begin{aligned} & h(c_1, X_1) - h(c_2, X_2) \\ &= \sum_{x \in S_1 \cap S_2} \left[\sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} - \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \right] \cdot f(x) \\ &+ \sum_{x \in S_1 \setminus S_2} \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} \cdot f(x) - \sum_{x \in S_2 \setminus S_1} \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \cdot f(x) = 0. \end{aligned} \tag{S7}$$

By substituting (S6) into (S7), we infer:

$$\begin{aligned} & h(c_1, X_1) - h(c_2, X_2) \\ &= \sum_{x \in S_1 \cap S_2} \left[\sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} - \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \right] \cdot x \\ &+ \sum_{x \in S_1 \setminus S_2} \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} \cdot (x+1) - \sum_{x \in S_2 \setminus S_1} \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \cdot (x-1) = 0. \end{aligned} \tag{S8}$$

As (S5) is equal to (S8), we have:

$$\begin{aligned} & h(c_1, X_1) - h(c_2, X_2) \\ &= \sum_{x \in S_1 \cap S_2} \left[\sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} - \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \right] \cdot x \\ &+ \sum_{x \in S_1 \setminus S_2} \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} \cdot x - \sum_{x \in S_2 \setminus S_1} \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \cdot x \\ &+ \sum_{x \in S_1 \setminus S_2} \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} - \sum_{x \in S_2 \setminus S_1} \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \\ &= \sum_{x \in S_1 \setminus S_2} \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} + \sum_{x \in S_2 \setminus S_1} \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} = 0. \end{aligned} \tag{S9}$$

Since the terms in the above summation operators are positive, i.e., (S9) cannot hold obviously. Thus, the assumption that $S_1 \neq S_2$ is false. Thus, given $h(c_1, X_1)=h(c_2, X_2)$, we have $S_1=S_2$.

Furthermore, based on $S_1=S_2=S$, we have the following inference:

$$\sum_{x \in S_1 \setminus S_2} \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} \cdot x - \sum_{x \in S_2 \setminus S_1} \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \cdot x = 0. \tag{S10}$$

Hence, according to (S5) and (S10), we have:

$$h(c_1, X_1) - h(c_2, X_2) = \sum_{x \in S} \left[\sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} - \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} \right] \cdot x = 0. \tag{S11}$$

Obviously, each term in the above summation is equal to zero:

$$\sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1) \cdot W(y)}} - \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2) \cdot W(y)}} = 0. \tag{S12}$$

Hence, we obtain:

$$\sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(c_1)} \cdot \sqrt{W(y)}} - \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(c_2)} \cdot \sqrt{W(y)}} = \frac{1}{\sqrt{W(c_1)}} \cdot \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(y)}} - \frac{1}{\sqrt{W(c_2)}} \cdot \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(y)}} = 0. \tag{S13}$$

Eq. (S13) can be rewritten as:

$$\sqrt{\frac{W(c_2)}{W(c_1)}} \cdot \sum_{y=x, y \in X_1} \frac{g_{c_1,y}}{\sqrt{W(y)}} = \sum_{y=x, y \in X_2} \frac{g_{c_2,y}}{\sqrt{W(y)}}. \tag{S14}$$

We further set $\kappa = \sqrt{W(c_2)/W(c_1)}$, we have $\kappa \cdot \sum_{y=x, y \in X_1} g_{c_1,y} \Big/ \sqrt{W(y)} = \sum_{y=x, y \in X_2} g_{c_2,y} \Big/ \sqrt{W(y)}$, for $x \in S$. Therefore, based on these inferences, ***Theorem* 1** holds.□

B. Proof of Corollary 1

Following ***Theorem* 1**, the pooling function using the strategy in (14) can be denoted as $p(c, X^{(1)}, X^{(2)},\ldots, X^{(L)})=(\alpha+\varepsilon/|X|)\cdot c+\sum_{l=\{1\sim L\},\, x\in X^{(l)}} h(c, x)$, $c, X^{(1)}, X^{(2)},\ldots, X^{(L)}\in\mathcal{X}$, where $|X^{(1)}|=|X^{(2)}|=\ldots=|X^{(L)}|=|X|$. According to ***Theorem* 1**, it is seen that when the condition that $X_1^{(l)}=\{S, \mu_1\}$, $X_2^{(l)}=\{S, \mu_2\}$, and $\kappa\cdot\sum_{y=x,y\in X_1} g_{c_1,y}\Big/\sqrt{W(y)}=\sum_{y=x,y\in X_2} g_{c_2,y}\Big/\sqrt{W(y)}$, for $\kappa=\sqrt{W(c_2)/W(c_1)}$, $x\in S$, and $l=\{1\sim L\}$ is fulfilled, $h(c_1, X_1^{(l)})-h(c_2, X_1^{(l)})=0$, for $l=\{1\sim L\}$, i.e., the aggregation function in (6) cannot distinguish different graph structures. To prove ***Corollary* 1**, it should be proved that $\mathcal{T}$ can correctly distinguish all different structures that the aggregation function in (6) fails previously. To do so, two cases require to be considered.

(1)$c_1\neq c_2$.

Given $X_1^{(l)}=\{S, \mu_1\}$, $X_2^{(l)}=\{S, \mu_2\}$, and $\kappa\cdot\sum_{y=x,y\in X_1} g_{c_1,y}\Big/\sqrt{W(y)}=\sum_{y=x,y\in X_2} g_{c_2,y}\Big/\sqrt{W(y)}$, for $\kappa=\sqrt{W(c_2)/W(c_1)}$, $x\in S$, and $l=\{1\sim L\}$, we have $h(c_1, X_1^{(l)})-h(c_2, X_2^{(l)})=0$, for $l=\{1\sim L\}$ based on ***Theorem* 1**. Hence, we have $p(c_1, X_1^{(1)}, X_1^{(2)},\ldots, X_1^{(L)})-p(c_2, X_2^{(1)}, X_2^{(2)},\ldots, X_2^{(L)})=(\alpha+\varepsilon/|X_1|)\cdot c_1-(\alpha+\varepsilon/|X_2|)\cdot c_2$. As $c_1\neq c_2$, $p(c_1, X_1^{(1)}, X_1^{(2)},\ldots, X_1^{(L)})\neq p(c_2, X_2^{(1)}, X_2^{(2)},\ldots, X_2^{(L)})$ is obvious.

(2)$c_1=c_2$.

Similarly, we have $p(c_1, X_1^{(1)}, X_1^{(2)},\ldots, X_1^{(L)})-p(c_2, X_2^{(1)}, X_2^{(2)},\ldots, X_2^{(L)})=(\alpha+\varepsilon/|X_1|)\cdot c_1-(\alpha+\varepsilon/|X_2|)\cdot c_2$. Considering the condition that $c_1=c_2$, we thereby have the following inference:

$$p\left(c_1, X_1^{(1)}, X_1^{(2)},\ldots, X_1^{(L)}\right)-p\left(c_2, X_2^{(1)}, X_2^{(2)},\ldots, X_2^{(L)}\right)=\left(\alpha+\frac{\varepsilon}{|X_1|}\right)c-\left(\alpha+\frac{\varepsilon}{|X_2|}\right)c=\varepsilon\left(\frac{1}{|X_1|}-\frac{1}{|X_2|}\right)c, \tag{S15}$$

where $|X_1|=|N(c_1)|$ and $|X_2|=|N(c_2)|$. Since $|X_1|\neq|X_2|$, $p(c_1, X_1^{(1)}, X_1^{(2)},\ldots, X_1^{(L)})\neq p(c_2, X_2^{(1)}, X_2^{(2)},\ldots, X_2^{(L)})$, meaning that the locality-enhanced holistic pooling function $\mathcal{T}$ based on (6) and (14) can successfully distinguish the graph structures that solely utilizing the aggregation function in (6) fails to distinguish previously. Thus, ***Corollary* 1** holds.□

II. General Settings

A. Details of Models for Comparison

M1. **NeuMF** [6] is a commonly-adopted neural MF baseline, which extends inner product to multilayered perception and combine them into the process of MF for achieving nonlinear node representations.

M2. **MetaMF** [16] is a federated MF-based model for LFA, which considers the privacy of nodes and learns private node representations by building a federated and meta learning framework.

M3. **GC-MC** [25] is a graph convolutional matrix completion model, which achieves a graph autoencoder to learn the first order connectivity information of nodes and proposes node dropout to improve the robustness.

M4. **NGCF** [24] is a neural graph collaborative filtering model, which performs MF with a standard graph convolutional network thereby effectively addressing the high-order connectivity in the interaction graph.

M5. **LightGCN** [27] is a light graph convolutional network-based LFA model, which removes the nonlinear activation and feature transformation to enjoy a fast learning process for accurate node representations.

M6. **LR-GCCF** [31] is a linear residual graph convolutional network-based MF model, which removes nonlinearities but reserves the feature transformation, and presents an empirical explanation of layer concatenation.

M7. **SGL-ED** [33] is a robust self-supervised graph learning-based model, which uses edge dropout to generate multiple views of a node and maximizes the agreement between different views of the same node.

M8. **DGCN-HN** [32] is a graph convolutional network-based model with hybrid normalization, which deepens the message propagation with two kinds of residual connection approaches.

M9. **LightGCN**$_{r\text{-}AdjNorm}$ [39] is a variant of lightGCN, which investigates the accuracy-novelty performance and incorporates a plugin called *r-AdjNorm* to control the normalization strength.

M10. **HMLET** [7] is a hybrid method of linear and nonlinear collaborative filtering. It is a GNN-based LFA method that leverages a gating module to choose linear or nonlinear propagation of each node.

M11. **HCCF** [28] is a self-supervised hypergraph contrastive collaborative filtering model, which constructs a hypergraph with cross-view contrastive learning to capture the local and global collaborative relations.

M12. **GTN** [29] is a graph trend filtering network-based LFA model, which notices the non-adaptive propagation and non-robustness of existing approaches and proposes a principled technique to capture the adaptive reliability.

M13. **CIGCN** [40] is a channel-independent graph convolutional network-based model, which performs transformation with diagonal parameter matrices, thus keeping independent embedding dimensions.

M14. **JMP-GCF** [41] is a GNN-based joint multi-grained popularity-aware collaborative filtering model, which constructs different granularities of popularity features and jointly learns the signals.

M15. **LightGCL** [42] is a simple graph contrastive learning-based LFA model, which uses singular value decomposition to refine the global collaborative relation and generate robust contrastive views.

M16. **GLCPN** is the proposed model in this paper.

B. Details of Training Settings

We apply the established settings to each involved model in our experiments:

(1)The initial state of trainable variables is generated randomly with Xavier method [43], and we use Adam optimizer [34] to train a target model;

(2)For objective comparison, the dimensionality of laten feature space of all involved models is fixed to 64, and considering other specific architecture of each model, we adopt the settings suggested by its original paper;

(3)We use the 70%-10%-20% train-validation-test settings on each testing case: 70% of Λ are randomly chosen as the training set Ψ for model inference, 10% as the validation set H to monitor the training process, and the remaining 20% as the testing set Ω to test the achieved model's performance, where Ψ, H, and Ω are disjoint;

(4)Each model's hyperparameters are tuned with care for its best performance: in individually-built cases Ψ and H, we tune all the hyperparameters to obtain the lowest validation error on Ψ, and then test the generalization performance on Φ with the achieved settings;

(5)We adopt ten-fold cross validation settings: the splitting and training process is performed ten times, and we record the final averaged results and standard deviations;

(6)The uniform early stopping strategies are applied: each model's training process terminates if: a) the training epoch reaches a preset threshold, i.e., 1000; or b) the representation ability keeps degrading for ten epochs.

III. Supplementary Tables

A. Results of Comparison Experiments

The results of Tables SI-SIII have been discussed in Section V.B, whose details are as follows:

- Table **SI** reports the Friedman statistical results of all involved models;
- Tables **SII-SIII** summarize the Wilcoxon signed-ranks test results.

TABLE SI
RESULTS OF THE FRIEDMAN TEST IN ESTIMATION ACCURACY (RMSE AND MAE) AND EFFICIENCY (CONVERGING TIME IN RMSE AND MAE).

No.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16
Accuracy*	14.06	12.53	9.43	14.13	3.41	9.41	9.81	6.03	3.10	6.38	11.31	8.00	7.38	12.25	7.72	**1.06**
Efficiency**	12.06	11.00	9.90	6.25	14.19	2.94	11.34	6.16	13.56	8.91	7.75	5.19	7.94	9.22	8.38	**1.22**

** High F-rank denotes low RMSE/MAE; ** High F-rank denotes low time cost to converge.*

TABLE SII
RESULTS OF THE WILCOXON SIGNED-RANKS TEST IN RMSE AND MAE CORRESPONDING TO TABLES III-IV.

Comparison	*R*+*	*R*–	*p*-value**
M16 vs M1	136	0	**2.41E-4**
M16 vs M2	136	0	**2.41E-4**
M16 vs M3	124	12	**2.05E-3**
M16 vs M4	136	0	**2.41E-4**
M16 vs M5	136	0	**2.41E-4**
M16 vs M6	136	0	**2.41E-4**
M16 vs M7	136	0	**2.41E-4**
M16 vs M8	136	0	**2.41E-4**
M16 vs M9	136	0	**2.41E-4**
M16 vs M10	136	0	**2.41E-4**
M16 vs M11	136	0	**2.41E-4**
M16 vs M12	136	0	**2.41E-4**
M16 vs M13	136	0	**2.41E-4**
M16 vs M14	136	0	**2.41E-4**
M16 vs M15	136	0	**2.41E-4**

** For M16, higher R+ values indicate higher estimation accuracy; ** With the significance level of 0.1, the accepted hypotheses are highlighted.*

TABLE SIII
RESULTS OF THE WILCOXON SIGNED-RANKS TEST ON CONVERGING TIME IN RMSE AND MAE CORRESPONDING TO TABLES V-VI.

Comparison	*R*+*	*R*–	*p*-value**
M16 vs M1	136	0	**2.41E-4**
M16 vs M2	136	0	**2.41E-4**
M16 vs M3	136	0	**2.41E-4**
M16 vs M4	120	16	**3.05E-3**
M16 vs M5	136	0	**2.41E-4**
M16 vs M6	119	17	**4.48E-3**
M16 vs M7	136	0	**2.41E-4**
M16 vs M8	136	0	**2.41E-4**
M16 vs M9	136	0	**2.41E-4**
M16 vs M10	136	0	**2.41E-4**
M16 vs M11	136	0	**2.41E-4**
M16 vs M12	127	9	**1.24E-3**
M16 vs M13	136	0	**2.41E-4**
M16 vs M14	136	0	**2.41E-4**
M16 vs M15	136	0	**2.41E-4**

** For M16, higher R+ values indicate higher computational efficiency; ** With the significance level of 0.1, the accepted hypotheses are highlighted.*

B. Results of Hyperparameter Sensitivity Test

- Table **SIV** (discussed in Section V.C) summarizes the suggested hyperparameter settings of GLCPN.

TABLE SIV
SUGGESTED HYPERPARAMETER SETTINGS OF GLCPN.

Learning Rate	L_2 Regularization Coefficient	Batch Size	L	K	α
1e-2	1e-4	2^{11}	3	64	0.1

C. Results of Ablation Studies

- Tables **SV-SVI** (discussed in Section V.D.i) summarize the training and validation errors (RMSE and MAE) of GLCPN and its several variants, including GLCPN-MF, GLCPN-A&T, GLCPN-A, and GLCPN-T;
- Tables **SVII-SVIII** (discussed in Section V.D.ii) record the RMSE, MAE, training epochs, time cost per epoch, and total time cost of GLCPN and its variants, including GLCPN-B and GLCPN-AT;
- Tables **SIX-SX** (discussed in Section V.D.iii) present the RMSE and MAE of GLCPN and its variants, including GLCPN-S, GLCPN-SS, GLCPN-M, and GLCPN-C.

TABLE SV
THE TRAINING AND VALIDATION RMSE OF GLCPN-MF, GLCPN-A&T, GLCPN-A, GLCPN-T, AND GLCPN ON D1-8.

No.		PPI Prediction				CP Prediction			
		D1	D2	D3	D4	D5	D6	D7	D8
Training RMSE	GLCPN-MF	**0.0627±3.4E-5**	0.0990±5.6E-5	**0.0701±7.6E-4**	**0.0695±5.8E-3**	**0.2180±2.2E-2**	**0.3729±3.7E-2**	**0.4610±2.3E-2**	**0.4985±4.0E-2**
	GLCPN-A&T	0.0654±9.1E-4	0.1007±1.3E-3	0.0752±1.3E-3	0.0791±1.7E-3	0.4022±1.3E-2	0.6534±4.3E-2	0.8138±1.1E-2	0.7741±6.3E-2
	GLCPN-A	0.0713±5.1E-5	0.0957±6.5E-4	0.0873±8.0E-5	0.0774±2.2E-4	0.2691±1.4E-2	0.6571±2.9E-2	0.7264±1.4E-2	0.7624±2.3E-2
	GLCPN-T	0.0692±1.5E-3	0.1022±1.8E-3	0.1009±2.5E-2	0.0931±2.6E-3	0.4432±3.0E-2	0.6819±3.1E-2	0.8024±1.0E-2	0.8054±1.3E-2
	GLCPN	0.0711±5.0E-5	**0.0951±4.9E-4**	0.0866±9.3E-5	0.0758±3.3E-4	0.2615±1.9E-2	0.5423±1.9E-2	0.5958±2.8E-2	0.6009±3.5E-2
Validation RMSE	GLCPN-MF	0.0865±6.1E-5	0.1232±1.5E-4	0.1195±2.2E-4	0.1174±2.4E-4	0.4585±8.4E-3	0.8622±3.1E-2	0.8902±5.6E-2	0.8859±5.7E-2
	GLCPN-A&T	0.0840±4.8E-4	0.1210±1.4E-4	0.1141±2.7E-4	0.1143±3.4E-4	0.4470±1.0E-2	**0.7601±2.2E-2**	0.8381±5.1E-2	0.8612±5.8E-2
	GLCPN-A	0.0835±6.1E-5	0.1195±7.7E-5	0.1125±4.2E-4	0.1118±4.6E-4	0.4393±9.5E-3	0.7640±2.6E-2	0.8466±5.0E-2	0.8676±5.8E-2
	GLCPN-T	0.0883±4.8E-4	0.1247±3.0E-4	0.1284±9.8E-3	0.1233±4.8E-4	0.4509±9.4E-3	0.7604±2.3E-2	0.8419±4.9E-2	0.8694±5.6E-2
	GLCPN	**0.0831±5.0E-5**	**0.1195±3.9E-5**	**0.1119±3.6E-4**	**0.1112±4.3E-4**	**0.4273±8.1E-3**	0.7674±2.9E-2	**0.8272±5.2E-2**	**0.8388±5.7E-2**

MF stands for matrix factorization, i.e., no graph convolution; A stands for the nonlinear activation; and T stands for the feature transformation.

TABLE SVI
THE TRAINING AND VALIDATION RMSE OF GLCPN-MF, GLCPN-A&T, GLCPN-A, GLCPN-T, AND GLCPN ON D1-8.

No.		PPI Prediction				CP Prediction			
		D1	D2	D3	D4	D5	D6	D7	D8
Training MAE	GLCPN-MF	**0.0427±1.8E-4**	0.0730±3.7E-4	**0.0535±4.7E-4**	**0.0482±7.3E-4**	0.0821±6.1E-3	**0.1266±7.0E-3**	**0.1623±7.4E-3**	**0.1715±1.3E-3**
	GLCPN-A&T	0.0445±1.1E-4	0.0732±2.1E-4	0.0543±7.1E-4	0.0556±1.6E-3	0.1553±8.3E-3	0.3173±1.0E-2	0.3235±3.1E-2	0.3146±2.6E-2
	GLCPN-A	0.0460±7.4E-5	**0.0694±6.5E-4**	0.0603±4.1E-4	0.0525±3.3E-4	**0.0748±1.2E-3**	0.3253±6.2E-3	0.3327±6.0E-3	0.3394±6.5E-3
	GLCPN-T	0.0471±1.8E-3	0.0755±1.5E-3	0.0726±1.7E-2	0.0659±1.7E-3	0.1693±8.2E-3	0.3327±1.5E-2	0.3458±2.6E-2	0.3355±2.1E-2
	GLCPN	0.0456±7.9E-5	0.0699±8.3E-4	0.0593±2.5E-4	0.0512±1.1E-4	0.0886±4.4E-3	0.2447±1.6E-2	0.1999±1.3E-2	0.2098±1.7E-2
Validation MAE	GLCPN-MF	0.0563±3.2E-5	0.0891±1.2E-4	0.0848±1.4E-4	0.0780±5.6E-3	0.1726±7.9E-4	0.4262±6.8E-3	0.3626±1.1E-2	0.3400±6.6E-3
	GLCPN-A&T	0.0543±7.2E-5	0.0867±2.3E-4	0.0802±1.8E-4	0.0786±2.0E-4	0.1846±3.9E-3	0.3799±3.6E-3	0.3594±7.6E-3	0.3467±9.0E-3
	GLCPN-A	0.0532±7.6E-5	0.0858±5.6E-5	0.0771±3.3E-4	0.0755±1.7E-4	0.1688±1.5E-3	0.3812±3.8E-3	0.3721±7.7E-3	0.3608±6.5E-3
	GLCPN-T	0.0576±6.4E-4	0.0902±1.3E-4	0.0923±8.2E-3	0.0859±3.5E-4	0.1911±1.9E-3	0.3836±4.1E-3	0.3655±6.9E-3	0.3642±9.1E-3
	GLCPN	**0.0527±6.0E-5**	**0.0857±3.0E-5**	**0.0763±2.7E-4**	**0.0749±2.2E-4**	**0.1595±1.6E-4**	**0.3686±5.3E-3**	**0.3422±9.1E-3**	**0.3286±5.5E-3**

MF stands for matrix factorization, i.e., no graph convolution; A stands for the nonlinear activation; and T stands for the feature transformation.

TABLE SVII
THE RMSE, TRAINING EPOCHS, TIME COST PER EPOCH (SEC.), AND TOTAL TIME COST (SEC.) OF GLCPN-B, GLCPN-AT, AND GLCPN ON D1-8.

No.		PPI Prediction				CP Prediction			
		D1	D2	D3	D4	D5	D6	D7	D8
RMSE	GLCPN-B	0.1135±9.5E-5	0.1392±8.6E-4	0.1142±3.8E-4	0.1125±3.4E-4	0.4325±9.0E-3	0.7734±2.7E-2	0.8376±5.2E-2	0.8502±5.8E-2
	GLCPN-AT	**0.1112±1.8E-4**	0.1370±3.3E-3	0.1266±7.7E-4	0.1235±5.9E-4	0.4328±9.4E-3	0.7782±3.0E-2	0.8481±5.1E-2	0.8630±5.6E-2
	GLCPN	0.1115±9.9E-5	**0.1363±3.0E-3**	**0.1119±3.6E-4**	**0.1112±4.3E-4**	**0.4273±8.1E-3**	**0.7674±2.9E-2**	**0.8272±5.2E-2**	**0.8388±5.7E-2**
Epochs	GLCPN-B	71±30.16	**85±12.76**	115±10.34	**100±6.26**	17±0.98	39±4.77	21±2.73	17±1.96
	GLCPN-AT	88±26.53	128±21.67	123±3.67	114±9.22	17±1.17	32±6.36	12±0.75	10±2.50
	GLCPN	**71±30.16**	86±14.12	**104±10.35**	114±19.96	**12±1.36**	**28±1.96**	**12±1.94**	**10±1.47**
Time Cost Per Epoch	GLCPN-B	35±0.08	22±0.32	28±0.05	15±0.07	0.96±0.01	0.24±0.00	1.09±0.00	2.13±0.01
	GLCPN-AT	183±0.59	109±0.21	107±0.17	55±0.75	3.54±0.05	0.97±0.01	3.56±0.06	6.85±0.20
	GLCPN	**35±0.40**	**22±0.28**	**28±0.13**	**15±0.09**	**0.96±0.01**	**0.24±0.00**	**1.09±0.01**	**2.13±0.01**
Total Time Cost	GLCPN-B	2531±1074.25	1891±296.56	3264±289.51	**1472±90.81**	16±1.10	9±1.12	22±2.92	37±4.10
	GLCPN-AT	16102±4888.56	13992±2354.94	13103±373.16	6351±553.40	59±3.90	31±0.04	42±2.85	71±16.57
	GLCPN	**2497±1060.24**	**1880±292.48**	**2966±302.48**	1681±303.08	**11±1.35**	**7±0.43**	**13±2.15**	**21±3.17**

B stands for the binary adjacency matrix; and AT stands for the self-attention mechanism.

TABLE SVIII
THE MAE, TRAINING EPOCHS, TIME COST PER EPOCH (SEC.), AND TOTAL TIME COST (SEC.) OF GLCPN-B, GLCPN-AT, AND GLCPN ON D1-8.

No.		PPI Prediction				CP Prediction			
		D1	D2	D3	D4	D5	D6	D7	D8
MAE	GLCPN-B	0.0736±5.4E-5	0.0987±5.6E-4	0.0780±2.4E-4	0.0761±1.5E-4	0.1667±4.8E-4	0.3759±5.1E-3	0.3536±8.2E-3	0.3398±6.1E-3
	GLCPN-AT	0.0738±1.5E-4	0.0991±2.3E-3	0.0892±5.0E-4	0.0854±4.6E-4	0.1637±9.5E-4	0.3782±5.6E-3	0.3651±9.0E_3	0.3508±5.7E-3
	GLCPN	**0.0721±6.7E-5**	**0.0971±1.3E-3**	**0.0763±2.7E-4**	**0.0749±2.2E-4**	**0.1595±1.6E-4**	**0.3686±5.3E-3**	**0.3422±9.1E-3**	**0.3286±5.5E-3**
Epochs	GLCPN-B	54±8.15	52±12.60	**71±13.16**	**89±12.19**	12±1.33	60±8.57	31±1.26	19±1.96
	GLCPN-AT	**42±12.58**	**37±10.79**	107±8.14	107±10.57	15±0.89	37±3.08	**12±1.10**	**9±3.20**
	GLCPN	50±12.09	56±9.68	75±12.27	146±15.77	**11±0.75**	**27±6.31**	20±3.01	12±2.04
Time Cost Per Epoch	GLCPN-B	35±0.09	22±0.33	29±0.09	15±0.08	0.96±0.01	0.24±0.00	1.09±0.00	2.13±0.01
	GLCPN-AT	183±0.61	109±0.56	107±0.10	56±0.76	3.55±0.05	0.97±0.01	3.56±0.07	6.85±0.18
	GLCPN	**35±0.40**	**22±0.32**	**29±0.19**	**15±0.08**	**0.96±0.01**	**0.24±0.00**	**1.09±0.01**	**2.13±0.01**
Total Time Cost	GLCPN-B	1923±287.08	**1172±295.15**	**2018±369.88**	**1312±179.24**	12±1.19	15±2.21	34±1.28	40±4.12
	GLCPN-AT	7626±2289.23	4053±1191.31	11493±874.54	5932±611.35	53±2.80	36±0.03	43±3.22	65±22.74
	GLCPN	**1744±412.10**	1242±198.53	2140±340.74	2137±221.26	**10±0.75**	**6±1.47**	**22±3.36**	**25±4.41**

B stands for the binary adjacency matrix; and AT stands for the self-attention mechanism.

TABLE SIX
THE RMSE OF GLCPN-S, GLCPN-SS, GLCPN-M, GLCPN-C, AND GLCPN ON D1-8.

No.	PPI Prediction				CP Prediction			
	D1	D2	D3	D4	D5	D6	D7	D8
GLCPN-S	0.1331±9.1E-5	0.1660±2.4E-4	0.1677±8.5E-4	0.1593±6.8E-4	0.4418±7.4E-3	0.7816±2.6E-2	0.8445±5.0E-2	0.8593±5.8E-2
GLCPN-SS	0.0866±3.4E-5	0.1220±2.5E-4	0.1168±3.0E-4	0.1161±5.8E-4	0.4443±9.2E-3	0.7780±2.6E-2	0.8453±4.8E-2	0.8593±5.4E-2
GLCPN-M	0.0845±6.9E-5	0.1216±9.7E-5	0.1128±2.3E-4	0.1139±4.8E-4	0.4297±7.7E-3	0.7690±2.9E-2	0.8276±5.1E-2	0.8396±5.7E-2
GLCPN-C	0.0864±8.6E-5	0.1230±9.5E-5	0.1200±2.9E-4	0.1191±4.9E-4	0.4378±8.8E-3	0.7912±2.9E-2	0.8392±5.2E-2	0.8492±5.8E-2
GLCPN	**0.0831±5.0E-5**	**0.1195±3.9E-5**	**0.1119±3.6E-4**	**0.1112±4.3E-4**	**0.4273±8.1E-3**	**0.7674±2.9E-2**	**0.8272±5.2E-2**	**0.8388±5.7E-2**

S stands for that a model only outputs the single final layer; SS stands for that a model only outputs the final layer but with self-loop message propagation; M stands for that a model outputs the mean of all the layers; and C stands for that a model concatenates the feature transformation.

TABLE SX
THE MAE OF GLCPN-S, GLCPN-SS, GLCPN-M, GLCPN-C, AND GLCPN ON D1-8.

No.	PPI Prediction				CP Prediction			
	D1	D2	D3	D4	D5	D6	D7	D8
GLCPN-S	0.0909±3.1E-4	0.1187±7.3E-4	0.1152±1.9E-3	0.1123±5.0E-4	0.1812±1.1E-3	0.3905±4.8E-3	0.3705±8.5E-3	0.3594±5.5E-3
GLCPN-SS	0.0565±7.2E-5	0.0872±7.8E-5	0.0821±1.8E-4	0.0801±2.4E-4	0.1777±1.7E-3	0.3844±7.2E-3	0.3621±5.6E-3	0.3513±1.2E-2
GLCPN-M	0.0548±6.3E-5	0.0875±7.7E-5	0.0794±1.7E-4	0.0782±2.0E-4	0.1647±6.7E-4	0.3694±6.0E-3	0.3454±8.9E-3	0.3330±5.3E-3
GLCPN-C	0.0562±8.0E-5	0.0889±8.6E-5	0.0853±1.8E-4	0.0826±2.6E-4	0.1671±1.1E-3	0.3820±5.4E-3	0.3497±8.8E-3	0.3358±5.4E-3
GLCPN	**0.0527±6.0E-5**	**0.0857±3.0E-5**	**0.0763±2.7E-4**	**0.0749±2.2E-4**	**0.1595±1.6E-4**	**0.3686±5.3E-3**	**0.3422±9.1E-3**	**0.3286±5.5E-3**

S stands for that a model only outputs the single final layer; SS stands for that a model only outputs the final layer but with self-loop message propagation; M stands for that a model outputs the mean of all the layers; and C stands for that a model concatenates the feature transformation.

IV. SUPPLEMENTARY TABLES

The hyperparameter sensitivity test results have been drawn in Figs. S1-6, which are discussed in Section V.C. Their details are described as follows:

- Figs. **S1-2** plot the errors and epochs of GLCPN as L varies;
- Figs. **S3-4** plot the errors and epochs of GLCPN as K varies;
- Figs. **S5-6** plot the errors and epochs of GLCPN as α varies.

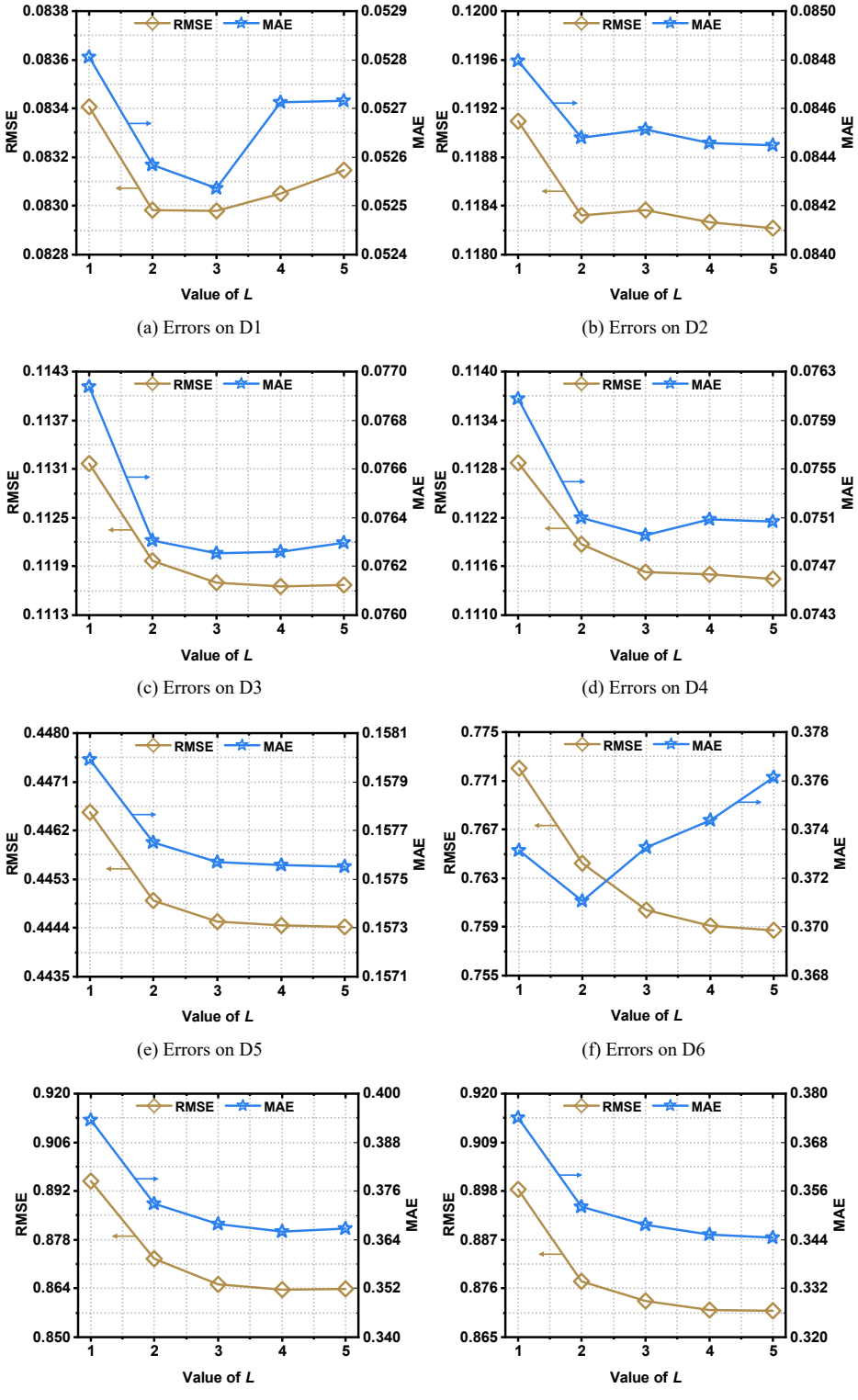
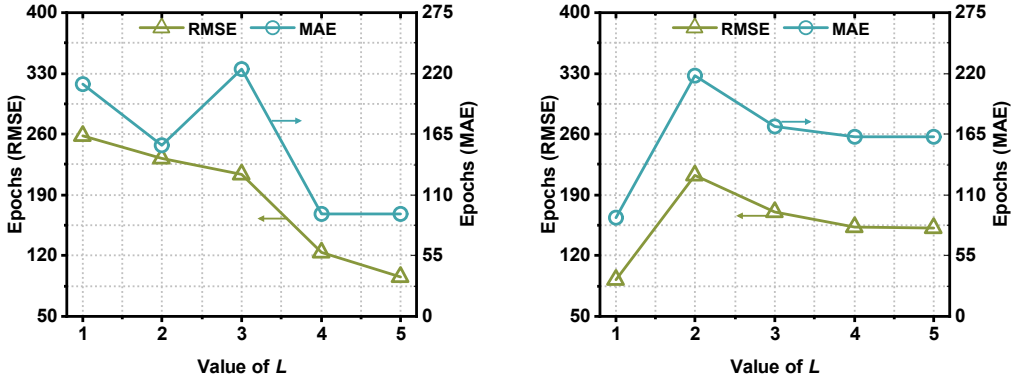
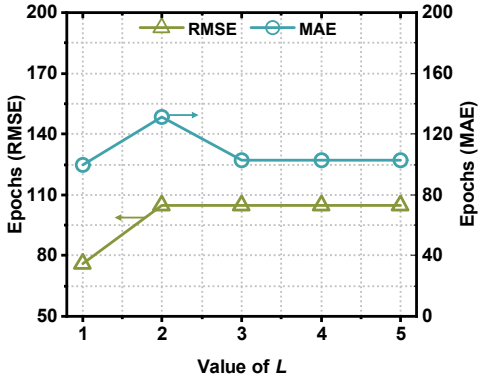


Fig. S1. Errors of GLCPN as L varies while other hyperparameters are being fixed on D1-8.

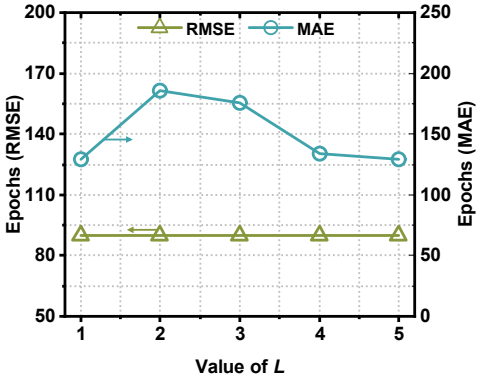


(a) Epochs on D1

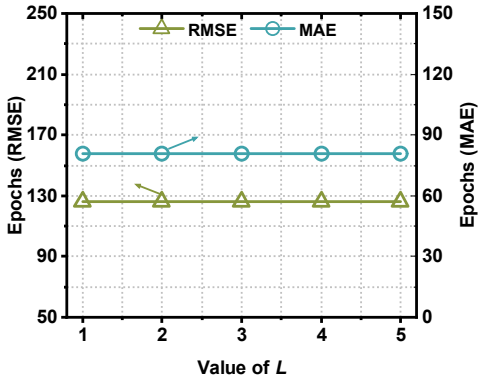
(b) Epochs on D2



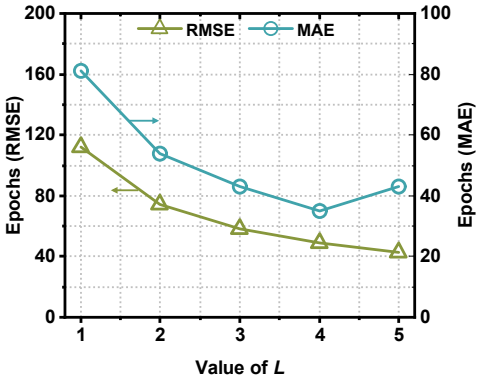
(c) Epochs on D3



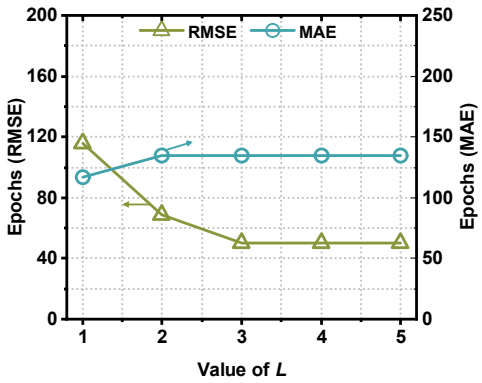
(d) Epochs on D4



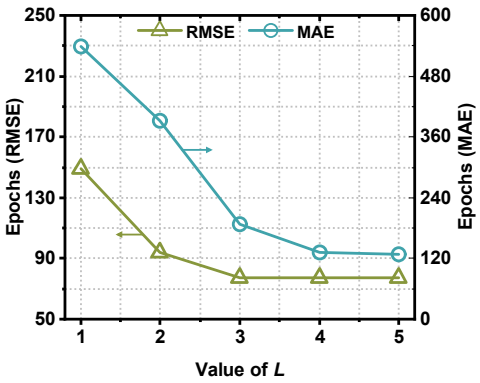
(e) Epochs on D5



(f) Epochs on D6

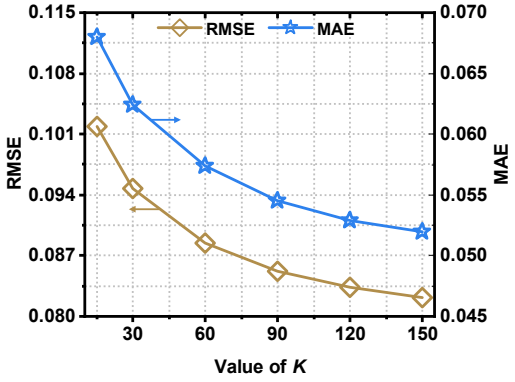


(g) Epochs on D7

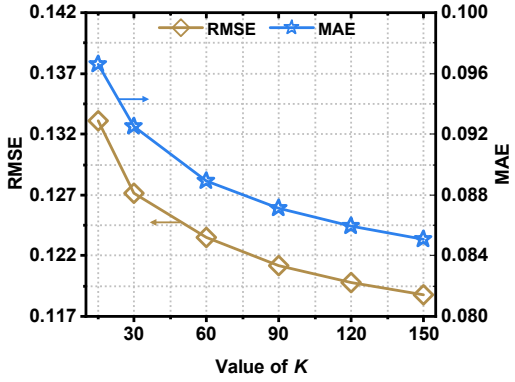


(h) Epochs on D8

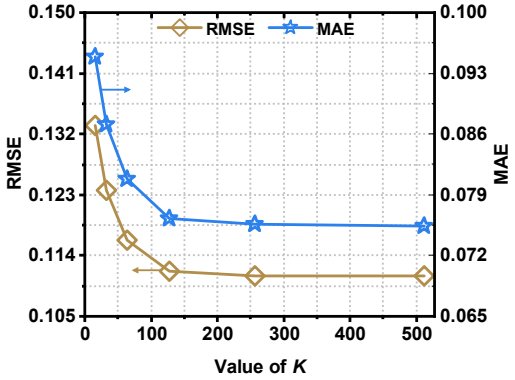
Fig. S2. Training epochs of GLCPN as L varies while other hyperparameters are being fixed on D1-8.



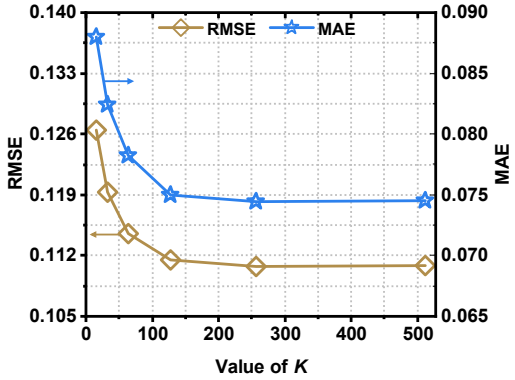
(a) Errors on D1



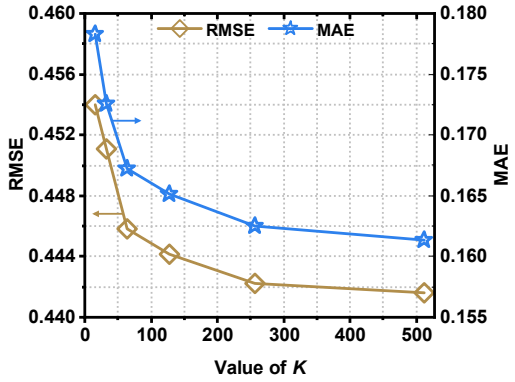
(b) Errors on D2



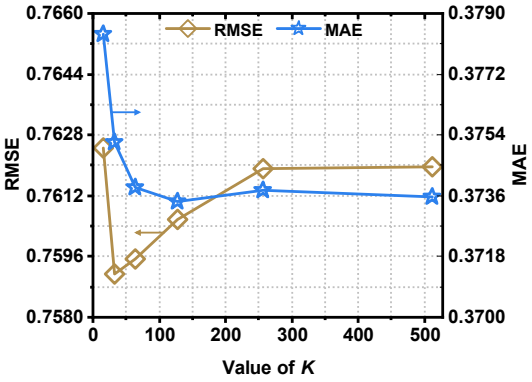
(c) Errors on D3



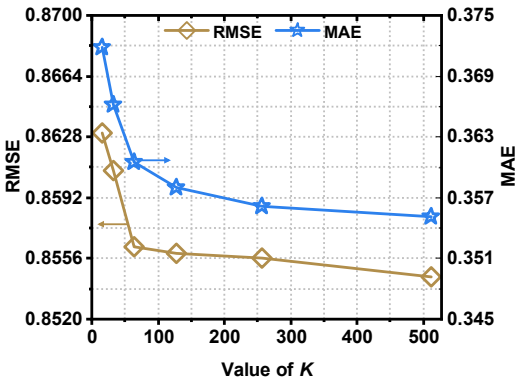
(d) Errors on D4



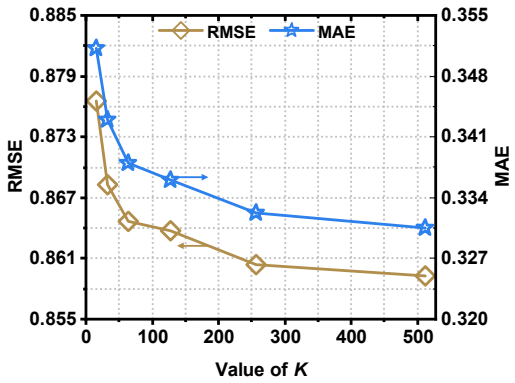
(e) Errors on D5



(f) Errors on D6

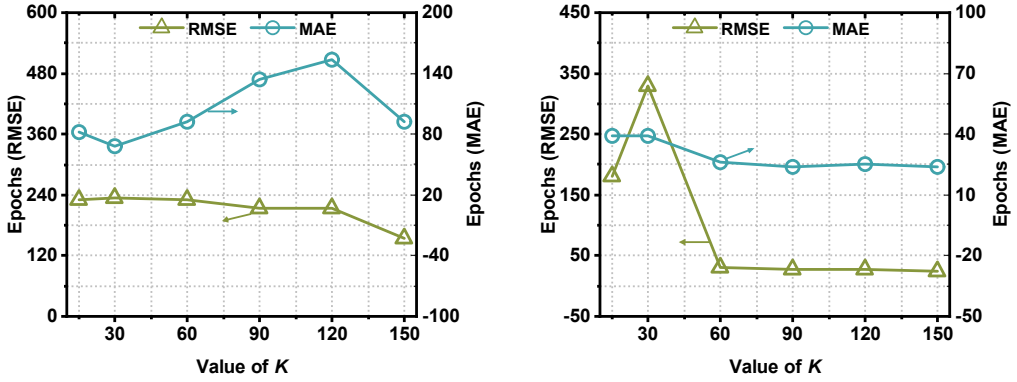


(g) Errors on D7



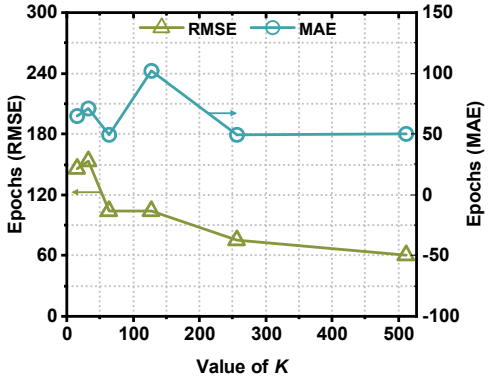
(h) Errors on D8

Fig. S3. Errors of GLCPN as *K* varies while other hyperparameters are being fixed on D1-8.

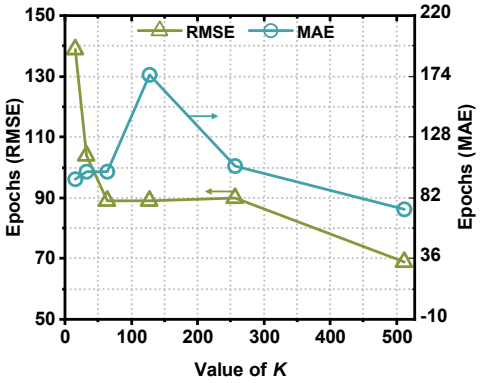


(a) Epochs on D1

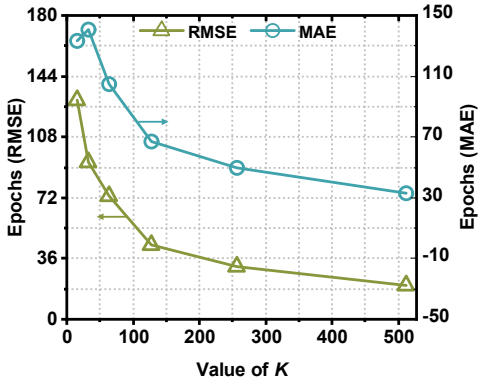
(b) Epochs on D2



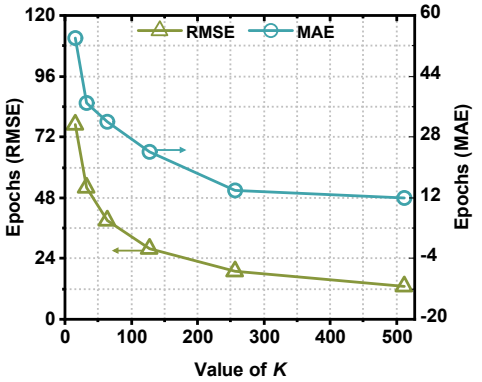
(c) Epochs on D3



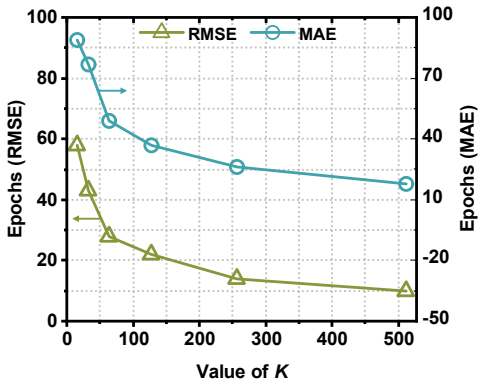
(d) Epochs on D4



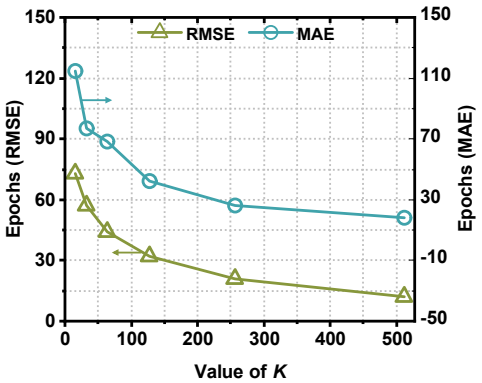
(e) Epochs on D5



(f) Epochs on D6

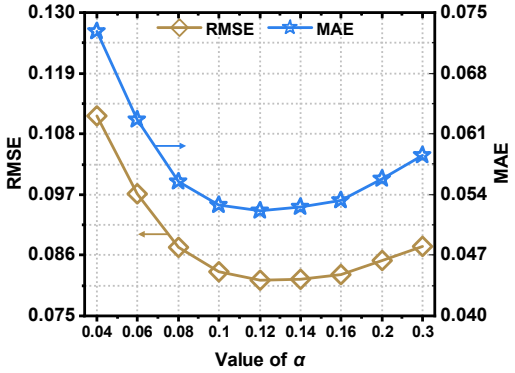


(g) Epochs on D7

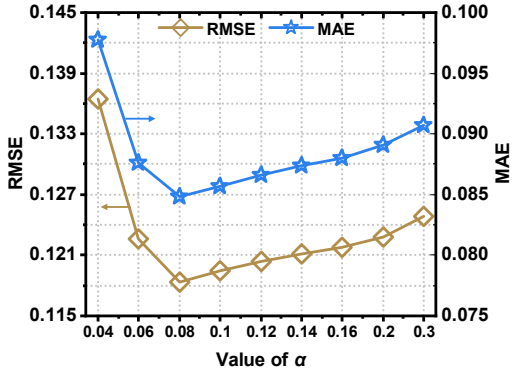


(h) Epochs on D8

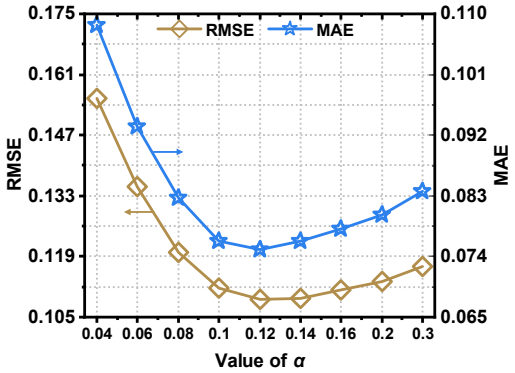
Fig. S4. Training epochs of GLCPN as *K* varies while other hyperparameters are being fixed on D1-8.



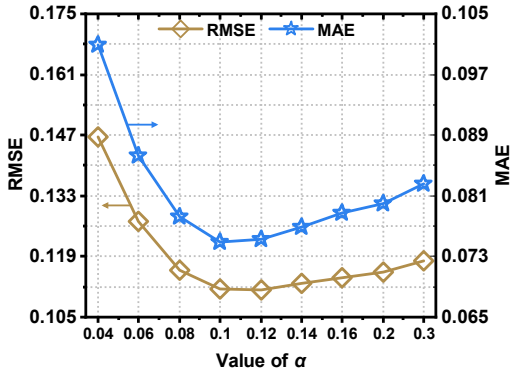
(a) Errors on D1



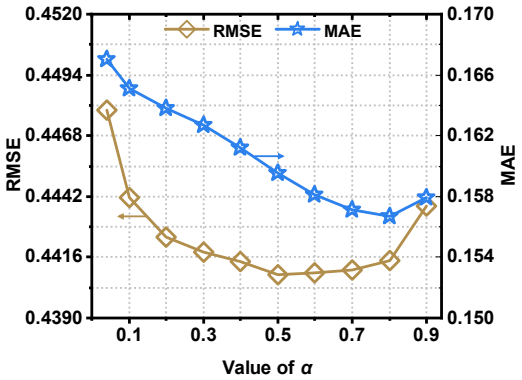
(b) Errors on D2



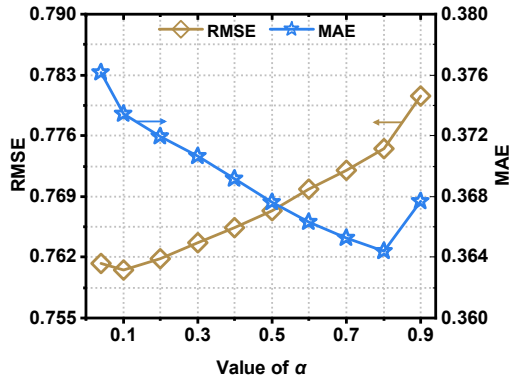
(c) Errors on D3



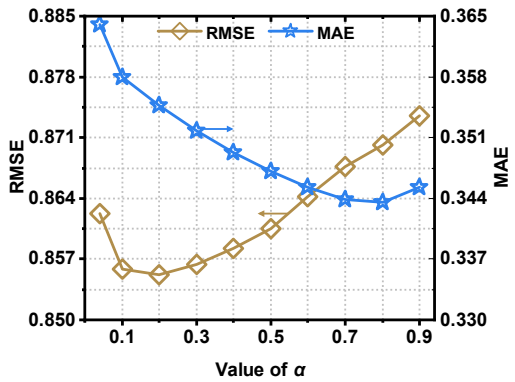
(d) Errors on D4



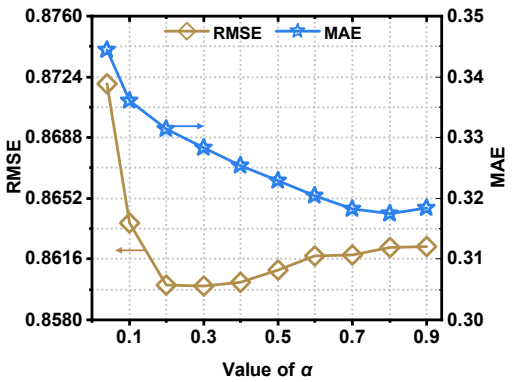
(e) Errors on D5



(f) Errors on D6

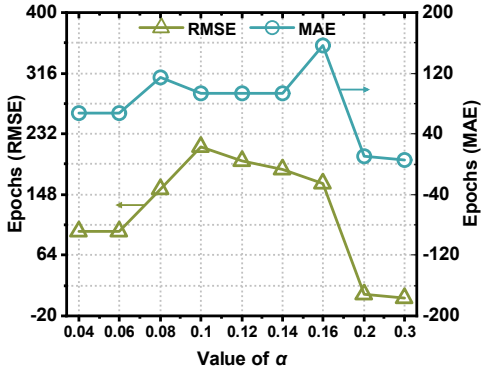


(g) Errors on D7

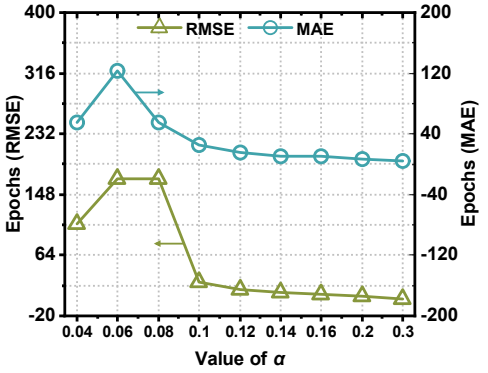


(h) Errors on D8

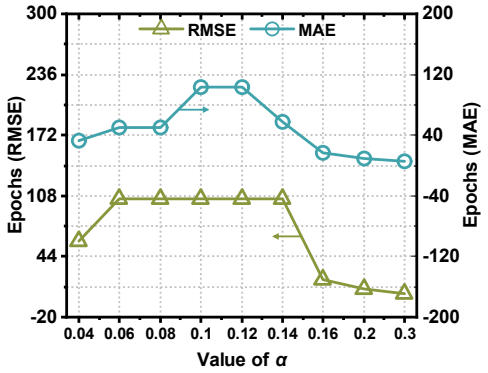
Fig. S5. Errors of GLCPN as α varies while other hyperparameters are being fixed on D1-8.



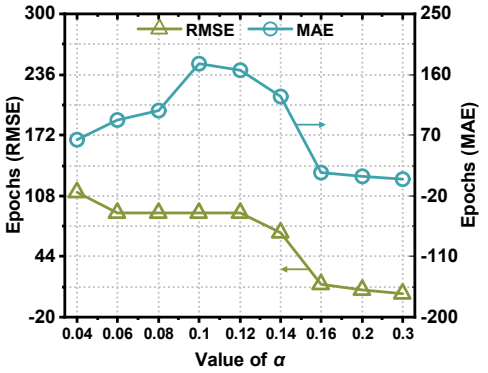
(a) Epochs on D1



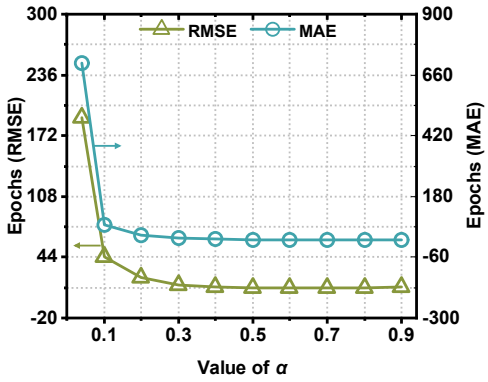
(b) Epochs on D2



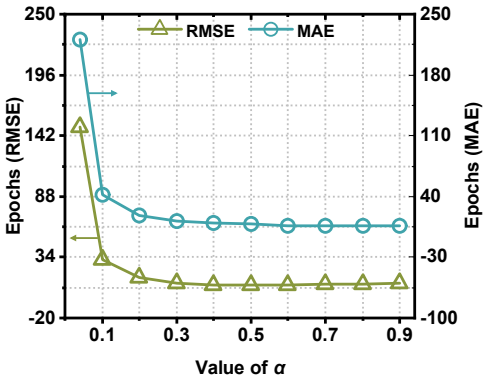
(c) Epochs on D3



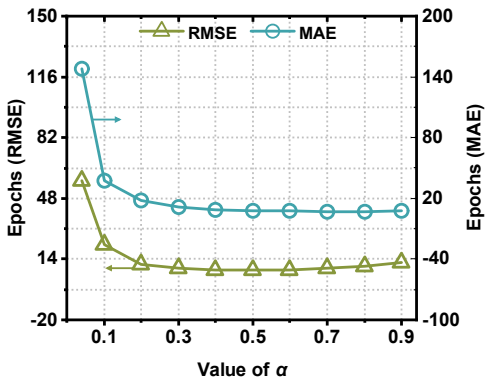
(d) Epochs on D4



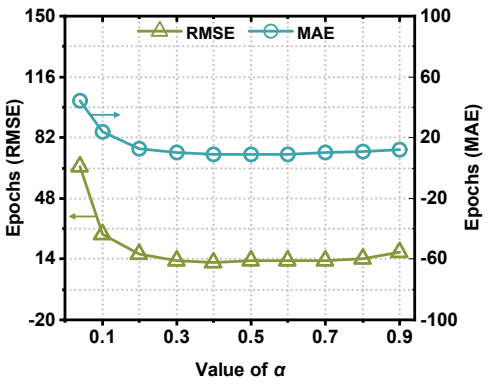
(e) Epochs on D5



(f) Epochs on D6



(g) Epochs on D7



(h) Epochs on D8

Fig. S6. Training epochs of GLCPN as α varies while other hyperparameters are being fixed on D1-8.