Graph Neural Network-Incorporated Linear Latent Feature Analysis Supplementary File

Authors

This is the supplementary file for the paper entitled "Graph Neural Network-Incorporated Linear Latent Feature Analysis". The proofs, additional tables and figures are put into this file and cited by the paper.

1. PROOFS

1.1 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 Eq. (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 Eq. (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:

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

Since Eq. (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 Eq. (S5) holds, we also infer that:

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

By substituting Eq. (S6) into Eq. (S7), we have:

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

As Eq. (S5) is equal to Eq. (S8), we infer:

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

Since the terms in the above summation operators are positive, i.e., Eq. (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 Eqs. (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.□

1.2 Proof of Corollary 1

Following ***Theorem* 1**, we can denote the pooling function using the strategy in Eq. (14) 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 Eq. (6) cannot distinguish different graph structures. To prove ***Corollary* 1**, we need to prove that $\mathcal{T}$ can correctly distinguish all different structures that the aggregation function in Eq. (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 Eqs. (6) and (14) can successfully distinguish the graph structures that solely utilizing the aggregation function in Eq. (6) fails to distinguish previously. Thus, ***Corollary* 1** holds.□

2. Supplementary Tables

2.1 Results of Ablation Studies (RQ1)

- Tables **S1-2** (discussed in Section 5.2.1) summarize the training and validation errors (RMSE and MAE) of GL²FA and its several variants, including GL²FA-MF, GL²FA-A&T, GL²FA-A, and GL²FA-T;
- Tables **S3-4** (discussed in Section 5.2.2) record the RMSE, MAE, training epochs, time cost per epoch, and total time cost of GL²FA and its variants, including GL²FA-B and GL²FA-AT;
- Tables **S5-6** (discussed in Section 5.2.3) present the RMSE and MAE of GL²FA and its variants, including GL²FA-S, GL²FA-SS, GL²FA-M, and GL²FA-C.

TABLE S1
THE TRAINING AND VALIDATION RMSE OF GL²FA-MF, GL²FA-A&T, GL²FA-A, GL²FA-T, AND GL²FA ON D1-8.

No.		D1	D2	D3	D4	D5	D6	D7	D8
Training RMSE	**GL²FA-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**
	GL²FA-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
	GL²FA-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
	GL²FA-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
	GL²FA	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	**GL²FA-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
	GL²FA-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
	GL²FA-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
	GL²FA-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
	GL²FA	**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 S2
THE TRAINING AND VALIDATION RMSE OF GL²FA-MF, GL²FA-A&T, GL²FA-A, GL²FA-T, AND GL²FA ON D1-8.

No.		D1	D2	D3	D4	D5	D6	D7	D8
Training MAE	**GL²FA-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**
	GL²FA-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
	GL²FA-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
	GL²FA-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
	GL²FA	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	**GL²FA-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
	GL²FA-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
	GL²FA-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
	GL²FA-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
	GL²FA	**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 S3
THE RMSE, TRAINING EPOCHS, TIME COST PER EPOCH (SEC.), AND TOTAL TIME COST (SEC.) OF GL²FA-B, GL²FA-AT, AND GL²FA ON D1-8.

No.		D1	D2	D3	D4	D5	D6	D7	D8
RMSE	**GL²FA-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
	GL²FA-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
	GL²FA	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	**GL²FA-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
	GL²FA-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
	GL²FA	**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	**GL²FA-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
	GL²FA-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
	GL²FA	**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	**GL²FA-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
	GL²FA-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
	GL²FA	**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 S4
THE MAE, TRAINING EPOCHS, TIME COST PER EPOCH (SEC.), AND TOTAL TIME COST (SEC.) OF GL^{2}FA-B, GL^{2}FA-AT, AND GL^{2}FA ON D1-8.

No.		D1	D2	D3	D4	D5	D6	D7	D8
MAE	**GL^{2}FA-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
	GL^{2}FA-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
	GL^{2}FA	**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	**GL^{2}FA-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
	GL^{2}FA-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**
	GL^{2}FA	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	**GL^{2}FA-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
	GL^{2}FA-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
	GL^{2}FA	**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	**GL^{2}FA-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
	GL^{2}FA-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
	GL^{2}FA	**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 S5
THE RMSE OF GL^{2}FA-S, GL^{2}FA-SS, GL^{2}FA-M, GL^{2}FA-C, AND GL^{2}FA ON D1-8.

No.	D1	D2	D3	D4	D5	D6	D7	D8
GL^{2}FA-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
GL^{2}FA-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
GL^{2}FA-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
GL^{2}FA-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
GL^{2}FA	**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 S6
THE MAE OF GL^{2}FA-S, GL^{2}FA-SS, GL^{2}FA-M, GL^{2}FA-C, AND GL^{2}FA ON D1-8.

No.	D1	D2	D3	D4	D5	D6	D7	D8
GL^{2}FA-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
GL^{2}FA-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
GL^{2}FA-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
GL^{2}FA-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
GL^{2}FA	**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.

2.2 Results of Hyperparameter Sensitivity Test (RQ2)

- Table **S7** (discussed in Section 5.3) summarizes the suggested hyperparameter settings of GL^{2}FA.

TABLE S7
SUGGESTED HYPERPARAMETER SETTINGS OF GL^{2}FA.

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

2.3 Results of Comparison Experiments (RQ3)

- Tables **S8-9** (discussed in Section 5.4(1)) record the RMSE and MAE of GL^{2}FA;
- Tables **S10-11** (discussed in Section 5.4(2)) record the time cost in RMSE and MAE of GL^{2}FA;
- Table **S12** (discussed in Section 5.4) reports the Friedman statistical results of all involved models;
- Tables **S13-14** (discussed in Section 5.4) summarize the Wilcoxon signed-ranks test results.

TABLE S8
THE RMSE, WIN/LOSS COUNTS, AND FRIEDMAN TEST RESULTS OF M1-16 ON ALL TESTING CASES.

No.	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
M1	0.1064±4.6E-5	0.1375±7.6E-5	0.1880±9.1E-2	0.1352±3.1E-4	0.5058±9.1E-2	0.9384±9.0E-3	1.0164±9.4E-3	0.9979±3.1E-3	8/0
M2	0.1117±1.4E-3	0.1393±2.9E-3	0.1438±2.4E-3	0.1413±5.8E-4	0.5129±4.0E-3	0.8310±3.0E-2	0.8650±1.5E-3	0.8492±2.5E-3	8/0
M3	0.0944±1.6E-4	0.1284±9.2E-5	0.1280±5.4E-4	0.1276±2.4E-4	0.4741±1.4E-2	**0.7931±2.7E-2**	0.8385±8.2E-3	0.8241±2.6E-2	7/1
M4	0.0979±8.0E-5	0.1337±3.7E-4	0.1370±1.4E-3	0.1417±2.2E-3	0.6277±1.1E-2	1.0414±2.0E-2	1.0547±1.2E-2	1.0315±1.8E-2	8/0
M5	0.0894±6.2E-5	0.1244±1.5E-4	0.1187±3.3E-4	0.1183±3.6E-4	0.4673±1.6E-2	0.8182±3.1E-2	0.8348±9.1E-3	0.8169±1.8E-2	8/0
M6	0.0916±6.9E-5	0.1266±3.3E-5	0.1265±2.5E-4	0.1241±2.9E-4	0.4933±1.6E-2	0.8725±2.7E-2	0.8685±1.7E-2	0.8441±1.7E-2	8/0
M7	0.1089±1.1E-4	0.1594±2.3E-4	0.1495±3.9E-4	0.1437±4.2E-4	0.4636±1.7E-2	0.8479±2.3E-2	0.8461±9.3E-3	0.8210±1.7E-2	8/0
M8	0.0905±5.9E-5	0.1246±1.0E-4	0.1232±4.1E-4	0.1222±2.7E-4	0.4757±1.7E-2	0.8430±3.2E-2	0.8482±1.0E-2	0.8229±1.9E-2	8/0
M9	0.0894±5.7E-5	0.1242±7.5E-5	0.1185±3.4E-4	0.1179±3.2E-4	0.4698±1.7E-2	0.8199±3.3E-2	0.8361±9.2E-3	0.8186±1.8E-2	8/0
M10	0.0900±4.5E-5	0.1252±1.8E-4	0.1202±2.8E-4	0.1200±2.9E-4	0.4764±1.7E-2	0.8517±3.3E-2	0.8562±1.1E-2	0.8286±1.7E-2	8/0
M11	0.0992±1.0E-3	0.1335±4.9E-4	0.1299±2.1E-4	0.1329±5.4E-4	0.4972±1.6E-2	0.8551±3.5E-2	0.8695±1.1E-2	0.8498±1.9E-2	8/0
M12	0.0995±3.1E-5	0.1313±1.4E-4	0.1271±3.5E-4	0.1267±3.4E-4	0.4744±1.7E-2	0.8261±3.2E-2	0.8383±9.6E-3	0.8178±1.8E-2	8/0
M13	0.0918±1.6E-4	0.1271±9.3E-5	0.1247±2.4E-4	0.1220±2.7E-4	0.4756±1.8E-2	0.8712±3.4E-2	0.8692±1.3E-2	0.8356±1.8E-2	8/0
M14	0.1330±1.1E-3	0.1611±1.7E-3	0.1682±1.3E-3	0.1636±1.7E-3	0.4823±1.7E-2	0.8255±3.2E-2	0.8455±9.1E-3	0.8242±1.7E-2	8/0
M15	0.0929±2.7E-5	0.1284±5.2E-5	0.1324±4.6E-4	0.1284±2.6E-4	0.4751±1.8E-2	0.8343±3.3E-2	0.8451±1.0E-2	0.8235±1.8E-2	8/0
M16	**0.0876±6.3E-5**	**0.1229±9.7E-5**	**0.1161±3.7E-4**	**0.1164±2.8E-4**	**0.4527±1.7E-2**	0.8159±3.0E-2	**0.8281±8.7E-3**	**0.8015±1.8E-2**	—

TABLE S9
THE MAE, WIN/LOSS COUNTS, AND FRIEDMAN TEST RESULTS OF M1-16 ON ALL TESTING CASES.

No.	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
M1	0.0710±5.2E-5	0.0998±5.8E-5	0.1429±8.7E-2	0.0949±2.3E-4	0.1997±4.0E-2	0.4856±4.0E-3	0.4557±5.1E-3	0.4564±4.4E-3	8/0
M2	0.0698±1.1E-3	0.0992±2.0E-3	0.0918±2.2E-3	0.0917±5.4E-4	0.2367±6.8E-3	0.4471±1.3E-2	0.4151±4.7E-3	0.4018±3.2E-3	8/0
M3	0.0644±7.5E-4	0.0950±6.1E-4	0.0954±1.6E-3	0.0907±4.4E-4	0.2231±3.9E-3	0.4536±2.0E-2	0.4291±5.2E-3	0.4057±1.2E-2	8/0
M4	0.0672±1.7E-4	0.0987±3.3E-4	0.1007±1.3E-3	0.1003±1.9E-3	0.3002±6.7E-3	0.7225±1.0E-2	0.6819±1.2E-2	0.6546±1.1E-2	8/0
M5	0.0584±5.2E-5	0.0898±8.9E-5	0.0839±1.9E-4	0.0818±1.7E-4	0.1813±3.6E-3	0.3894±7.4E-3	0.3629±4.1E-3	0.3518±2.2E-3	8/0
M6	0.0600±5.2E-5	0.0916±3.7E-5	0.0905±2.8E-4	0.0870±2.8E-4	0.2138±3.3E-3	0.4471±8.9E-3	0.4002±8.9E-3	0.3858±2.6E-3	8/0
M7	0.0726±6.7E-5	0.1181±1.1E-4	0.1088±3.0E-4	0.1029±2.7E-4	0.1724±3.5E-3	0.4020±6.8E-3	0.3610±3.9E-3	0.3437±2.1E-3	8/0
M8	0.0594±3.9E-5	0.0899±4.2E-5	0.0876±2.1E-4	0.0851±1.4E-4	0.1856±3.6E-3	0.3984±8.4E-3	0.3612±4.0E-3	0.3489±1.7E-3	8/0
M9	0.0583±4.6E-5	0.0896±4.6E-5	0.0837±2.1E-4	0.0813±1.3E-4	0.1815±3.7E-3	0.3870±7.8E-3	0.3612±4.1E-3	0.3508±1.7E-3	8/0
M10	0.0592±7.9E-5	0.0904±8.4E-5	0.0849±1.8E-4	0.0831±1.4E-4	0.1833±3.9E-3	0.4083±9.2E-3	0.3646±4.6E-3	0.3488±2.0E-3	8/0
M11	0.0642±7.6E-4	0.0931±4.7E-4	0.0940±2.4E-4	0.0939±3.2E-4	0.2148±2.8E-3	0.4287±9.5E-3	0.3859±4.5E-3	0.3681±3.7E-3	8/0
M12	0.0681±2.9E-5	0.0951±7.6E-5	0.0906±2.0E-4	0.0895±1.6E-4	0.1921±3.5E-3	0.3942±8.2E-3	0.3647±4.1E-3	0.3540±1.8E-3	8/0
M13	0.0603±1.2E-4	0.0922±7.3E-5	0.0890±8.7E-5	0.0849±1.5E-4	0.1758±3.7E-3	0.4100±9.5E-3	0.3632±3.8E-3	0.3454±2.5E-3	8/0
M14	0.0903±8.8E-4	0.1121±4.5E-4	0.1175±2.1E-3	0.1208±2.4E-3	0.2052±3.9E-3	0.4036±7.5E-3	0.3762±3.9E-3	0.3640±1.7E-3	8/0
M15	0.0613±1.3E-5	0.0932±7.1E-5	0.0957±2.8E-4	0.0907±1.8E-4	0.1845±3.9E-3	0.3933±8.2E-3	0.3609±4.1E-3	0.3508±2.0E-3	8/0
M16	**0.0569±3.6E-5**	**0.0882±8.3E-5**	**0.0810±3.2E-4**	**0.0800±9.7E-5**	**0.1628±3.7E-3**	**0.3709±8.9E-3**	**0.3340±3.5E-3**	**0.3200±2.4E-3**	—

TABLE S10
THE TRAINING TIME COST IN RMSE (SEC.), WIN/LOSS COUNTS, AND FRIEDMAN TEST RESULTS OF M1-16 ON ALL TESTING CASES.

No.	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
M1	23639±3281.13	15262±2313.04	3784±1328.82	2847±888.49	3124±1319.35	705±42.48	2485±516.66	3041±843.18	8/0
M2	86717±8762.90	100478±34208.66	16636±5717.27	21886±775.19	19±0.28	9±1.63	36±0.28	61±13.49	8/0
M3	111765±25822.30	53694±8577.76	7264±2205.15	3598±2076.02	24±3.35	17±2.33	51±4.53	77±12.26	8/0
M4	27007±2315.49	15677±1372.02	2394±232.15	1801±168.76	180±97.91	7±1.26	14±3.72	21±5.98	8/0
M5	86409±4060.43	47689±2030.77	6504±91.71	3641±79.94	291±18.47	39±2.60	140±11.96	270±46.18	8/0
M6	23776±382.51	12475±346.35	**812±12.78**	522±12.12	56±35.34	16±7.04	24±5.96	36±1.19	7/1
M7	80054±9638.93	29526±807.53	4109±152.29	2698±41.52	110±2.65	18±1.68	206±269.24	133±18.17	8/0
M8	27716±332.19	16827±11.01	1909±36.29	1024±17.55	60±2.82	15±0.64	45±2.21	77±9.62	8/0
M9	80932±622.02	47631±210.16	6386±68.83	3507±47.87	299±24.09	38±1.41	126±9.59	213±62.21	8/0
M10	40161±3369.92	15008±148.85	2706±74.28	1274±3.51	140±13.83	29±2.40	69±6.52	105±13.40	8/0
M11	31986±9000.12	13559±1278.28	2148±458.26	1307±119.87	98±3.94	36±2.25	69±4.48	92±6.29	8/0
M12	23748±1717.06	15310±697.19	5418±537.70	**218±315.74**	44±3.44	8±0.64	25±2.68	44±8.21	7/1
M13	29599±322.72	14960±598.34	1793±48.35	962±26.48	91±3.36	43±1.99	91±3.42	134±3.75	8/0
M14	64241±20949.85	37383±16126.03	6290±2004.81	4118±758.49	123±8.32	16±1.57	96±3.01	230±29.74	8/0
M15	41094±218.09	22592±278.36	3390±42.10	1900±27.05	65±2.76	28±1.30	54±2.22	75±4.14	8/0
M16	**18053±805.84**	**10682±1625.76**	1243±328.76	337±50.30	**10±0.66**	**5±0.36**	**14±0.32**	**19±1.52**	—

TABLE S11
THE TRAINING TIME COST IN MAE (SEC.), WIN/LOSS COUNTS, AND FRIEDMAN TEST RESULTS OF M1-16 ON ALL TESTING CASES.

No.	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
M1	23721±3496.87	15111±2262.50	3546±1292.37	2993±970.69	3291±1367.57	717±32.25	2477±515.63	3059±836.57	8/0
M2	86846±8947.68	85208±37758.77	17226±4821.14	21905±790.42	571±80.59	11±2.98	57±19.28	75±13.74	8/0
M3	54936±13745.18	17854±5042.50	2543±242.30	2248±876.67	32±5.97	23±3.85	64±5.28	97±14.80	8/0
M4	28132±3352.69	14422±1371.31	2242±159.11	1834±123.43	335±41.57	8±1.95	19±4.85	29±6.76	8/0
M5	78178±3396.68	43853±1915.68	5763±122.28	3500±78.80	293±15.35	47±5.34	206±7.42	435±30.94	8/0
M6	20161±818.08	10973±459.04	**712±41.90**	491±20.89	30±4.12	19±9.76	19±3.09	27±3.61	7/1
M7	102976±10275.42	23993±212.93	2487±84.92	2433±38.51	95±6.41	21±1.64	198±209.61	182±7.46	8/0
M8	25686±760.97	15297±485.69	1681±32.21	973±22.12	64±2.17	13±0.71	39±3.85	78±6.25	8/0
M9	74130±997.08	43757±289.41	5758±92.65	3394±34.17	320±11.69	53±3.67	213±13.72	417±20.05	8/0
M10	36989±2171.40	13019±360.84	2285±71.03	1200±42.06	152±16.69	27±1.92	74±9.68	141±17.84	8/0
M11	25744±5430.47	20815±5353.89	1801±294.71	1317±137.64	64±4.11	25±0.66	44±2.60	56±2.80	8/0
M12	21000±1241.13	13453±556.08	4944±899.33	2100±256.52	49±2.66	6±0.59	23±2.97	47±7.44	8/0
M13	27687±845.58	12942±614.46	1549±56.24	846±23.25	79±1.85	37±1.22	82±3.54	130±1.91	8/0
M14	24117±7551.16	12243±3910.36	1103±625.83	791±335.47	77±16.68	20±4.30	100±17.31	204±35.47	8/0
M15	40326±463.23	20638±204.73	2982±23.92	1731±23.45	59±1.03	21±0.68	39±3.49	65±3.94	8/0
M16	**14200±1727.31**	**8362±1621.56**	728±136.53	**440±133.24**	**6±0.01**	**3±0.18**	**9±0.40**	**11±0.60**	—

TABLE S12
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 S13

RESULTS OF THE WILCOXON SIGNED-RANKS TEST IN RMSE AND MAE CORRESPONDING TO TABLES S8, S9, AND S12.

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 S14

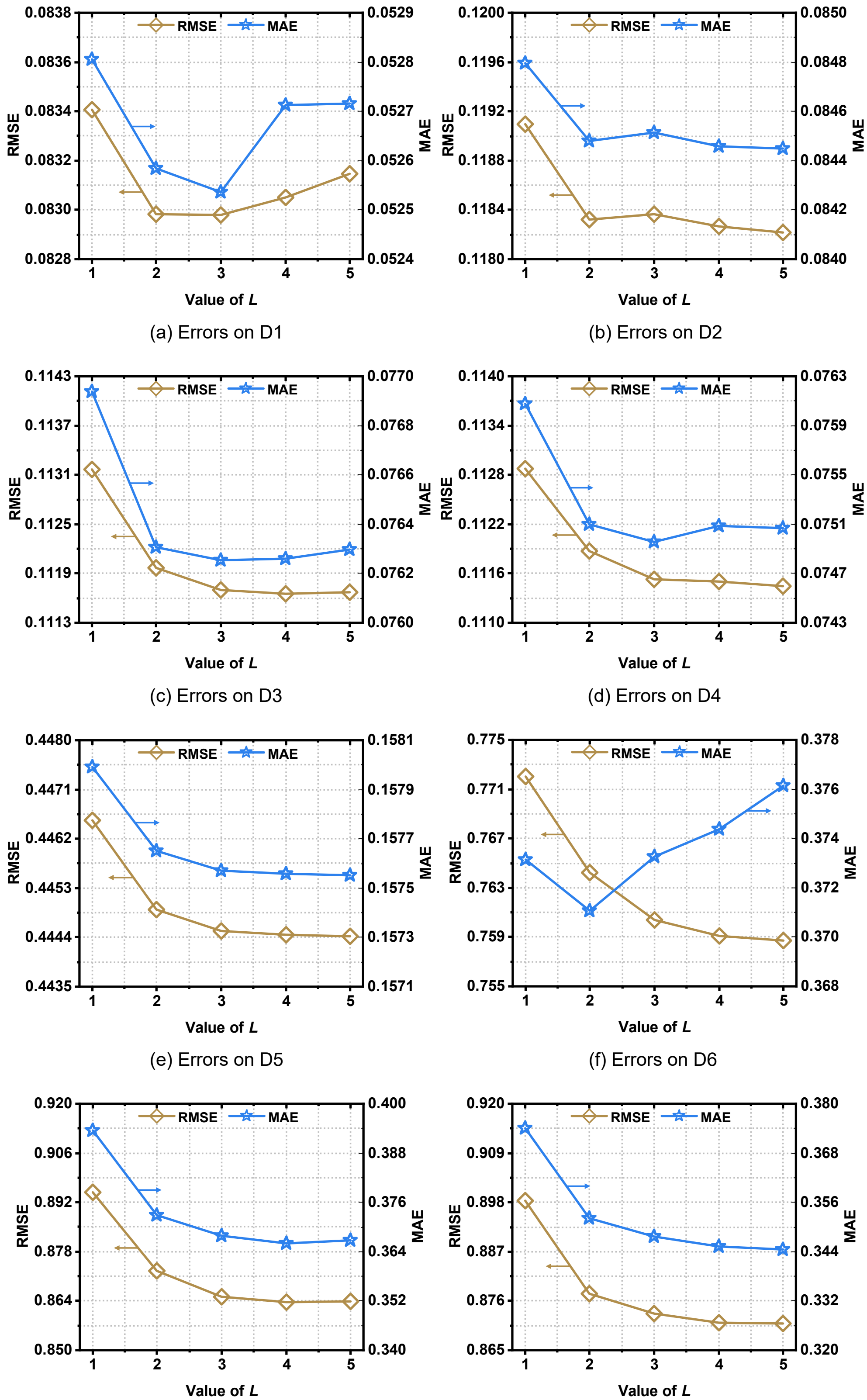
RESULTS OF THE WILCOXON SIGNED-RANKS TEST ON CONVERGING TIME IN RMSE AND MAE CORRESPONDING TO TABLES S10, S11, AND S12.

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-5**
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.*

3. Supplementary Figures

- Figs. **S1-2** (discussed in Section 5.3) plot the errors and epochs of GL²FA as L varies;
- Figs. **S3-4** (discussed in Section 5.3) plot the errors and epochs of GL²FA as K varies;
- Figs. **S5-6** (discussed in Section 5.3) plot the errors and epochs of GL²FA as α varies.



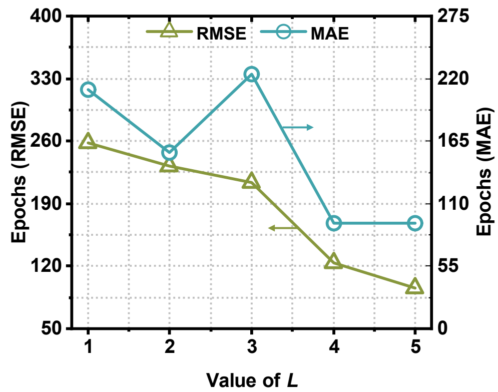
(a) Errors on D1 (b) Errors on D2

(c) Errors on D3 (d) Errors on D4

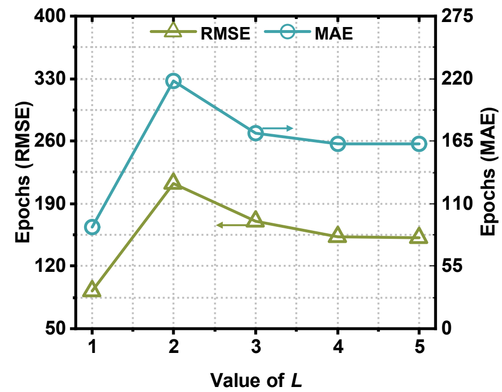
(e) Errors on D5 (f) Errors on D6

(g) Errors on D7 (h) Errors on D8

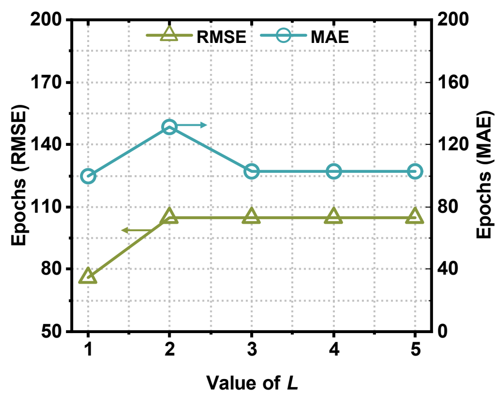
Fig. S1. Errors of GL^{2}FA as L varies while other hyperparameters are being fixed on D1-8.



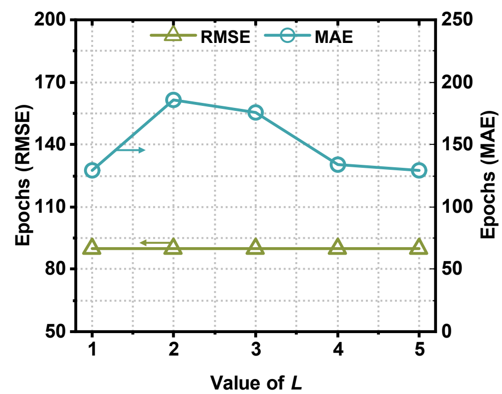
(a) Epochs on D1



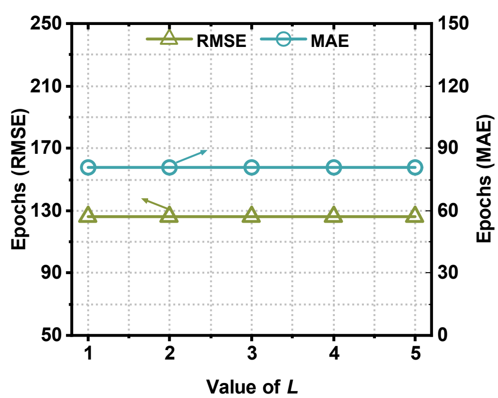
(b) Epochs on D2



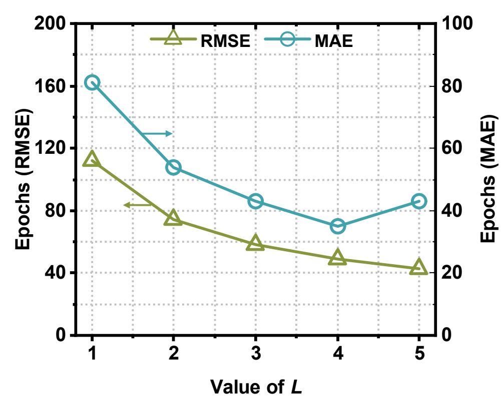
(c) Epochs on D3



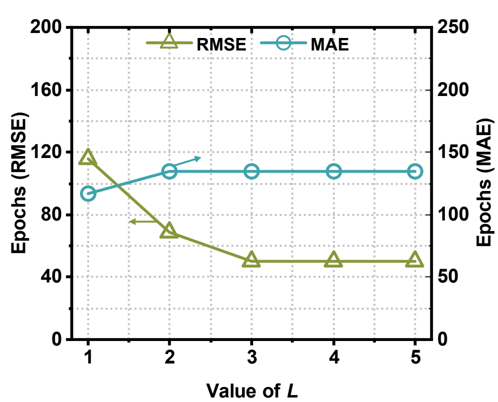
(d) Epochs on D4



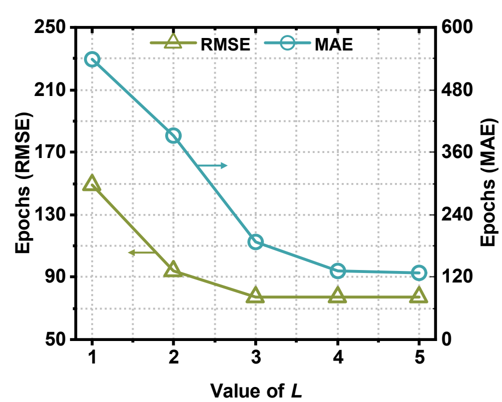
(e) Epochs on D5



(f) Epochs on D6

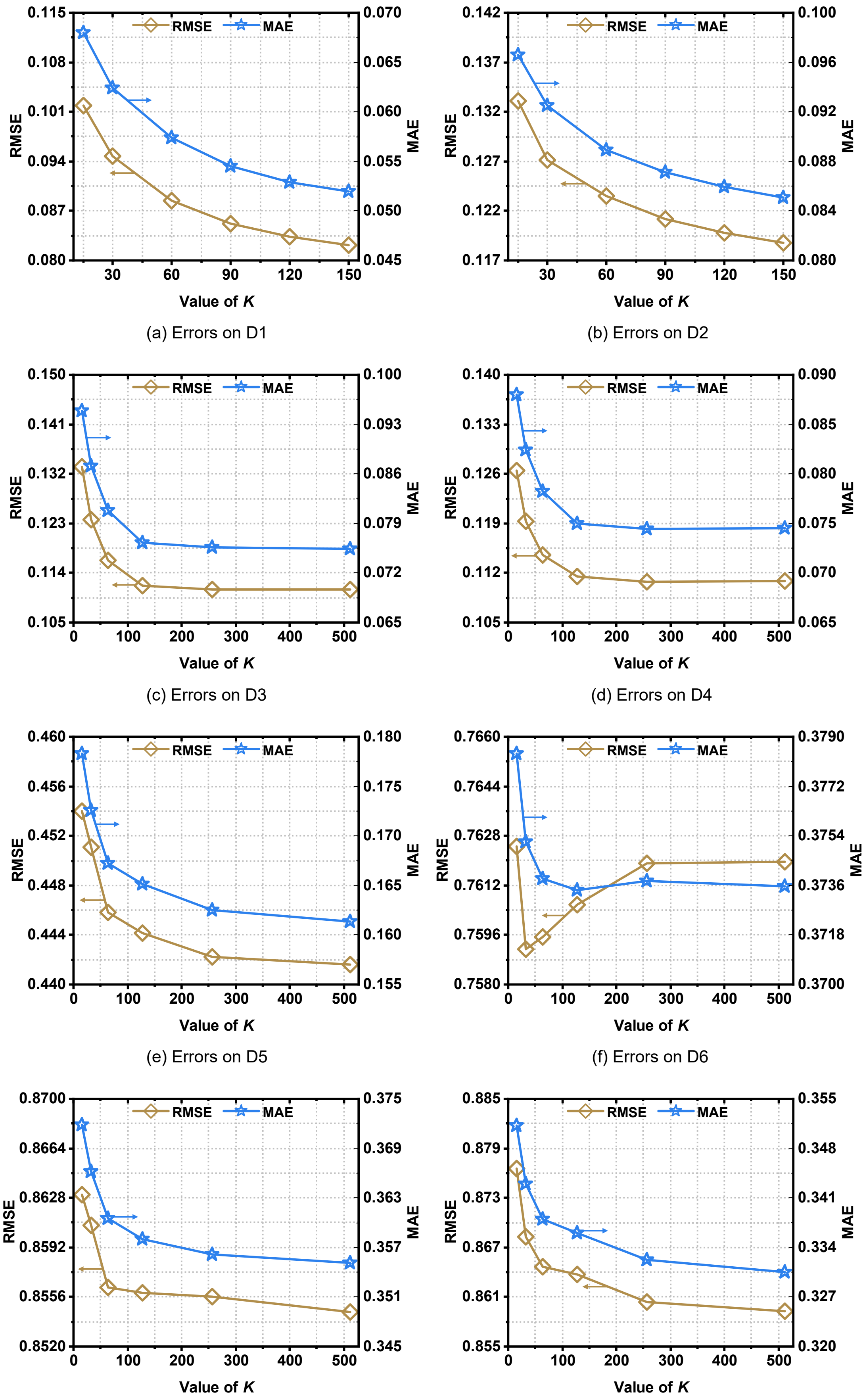


(g) Epochs on D7



(h) Epochs on D8

Fig. S2. Training epochs of GL^{2}FA 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

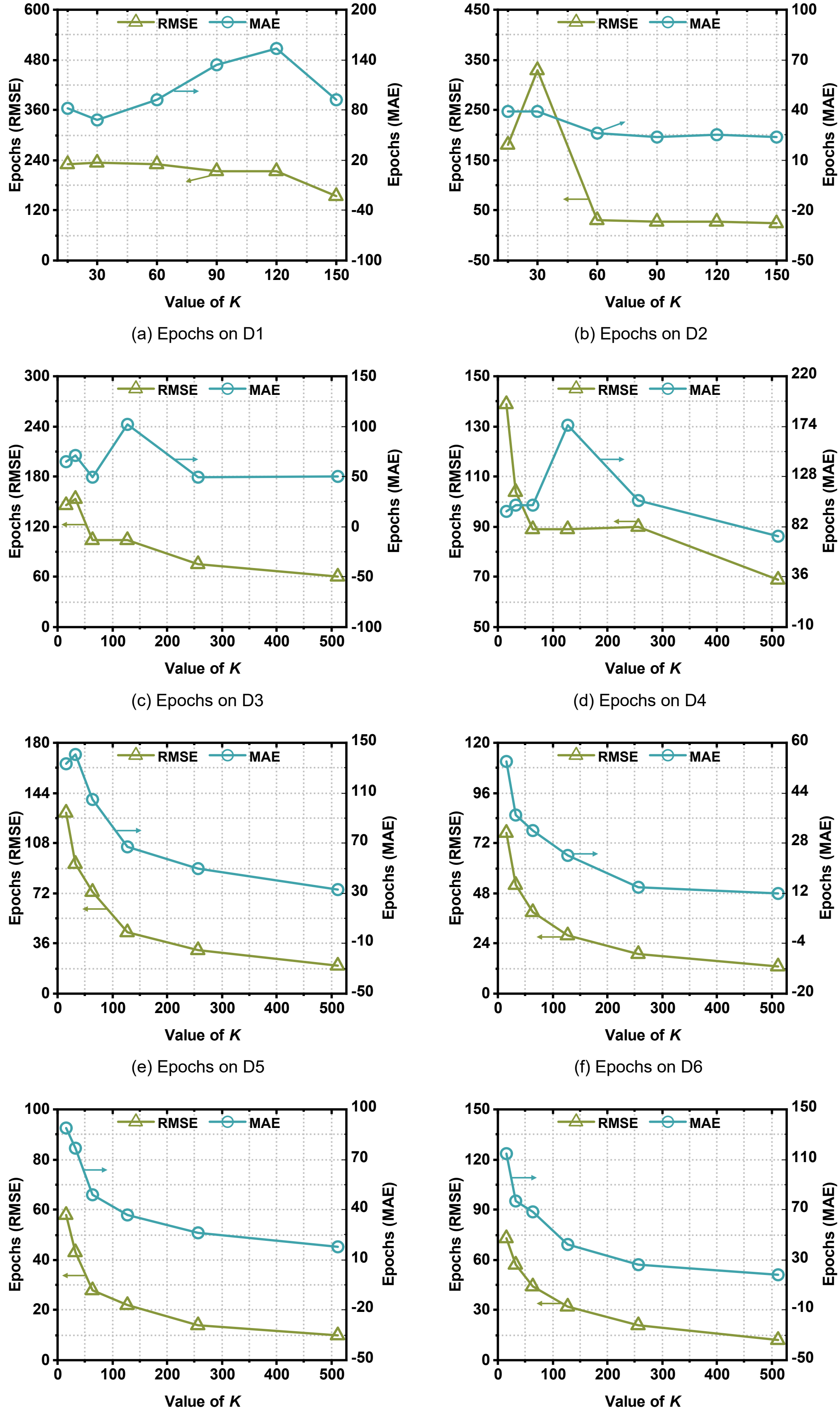
(e) Errors on D5

(f) Errors on D6

(g) Errors on D7

(h) Errors on D8

Fig. S3. Errors of GL^{2}FA as K varies while other hyperparameters are being fixed on D1-8.



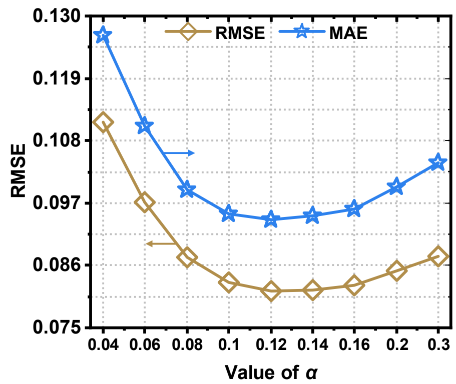
(a) Epochs on D1 (b) Epochs on D2

(c) Epochs on D3 (d) Epochs on D4

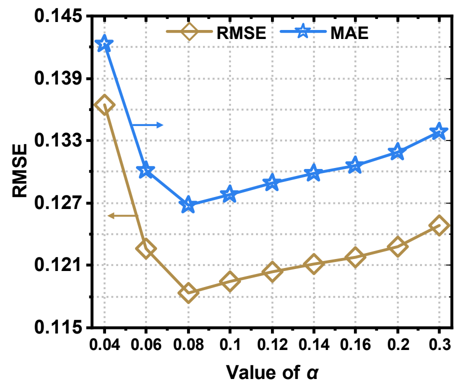
(e) Epochs on D5 (f) Epochs on D6

(g) Epochs on D7 (h) Epochs on D8

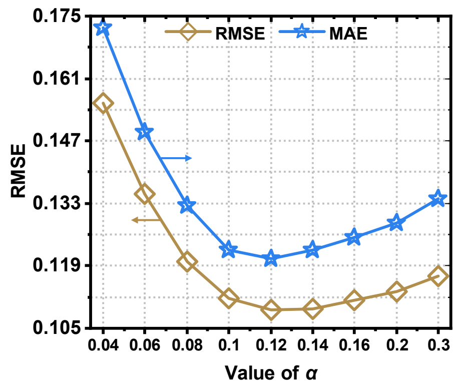
Fig. S4. Training epochs of GL^{2}FA as K varies while other hyperparameters are being fixed on D1-8.



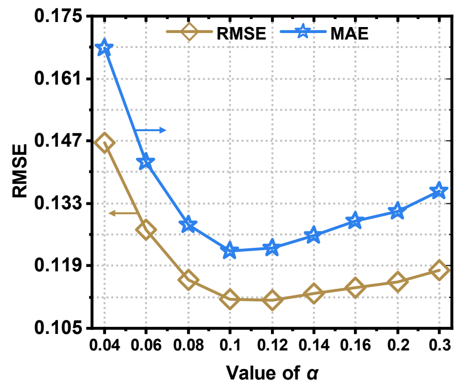
(a) Errors on D1



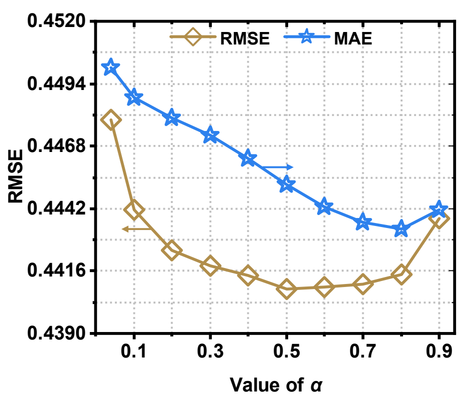
(b) Errors on D2



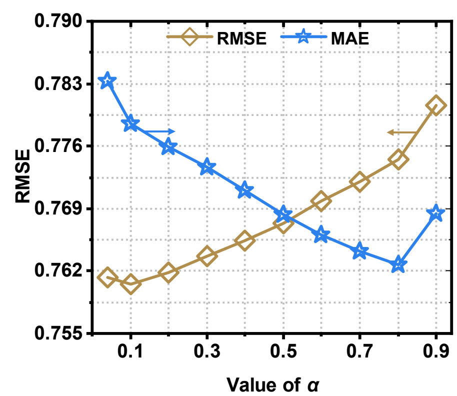
(c) Errors on D3



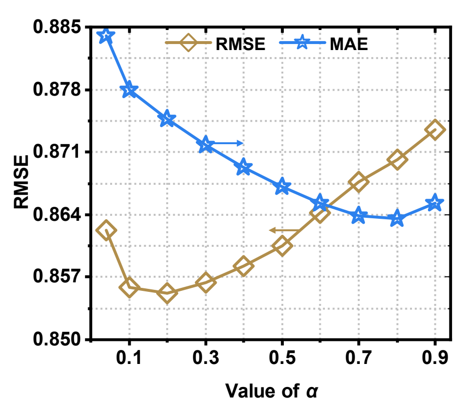
(d) Errors on D4



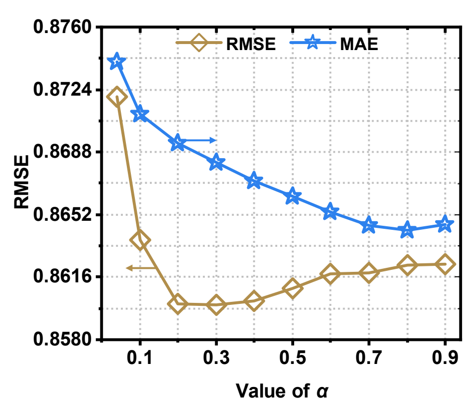
(e) Errors on D5



(f) Errors on D6

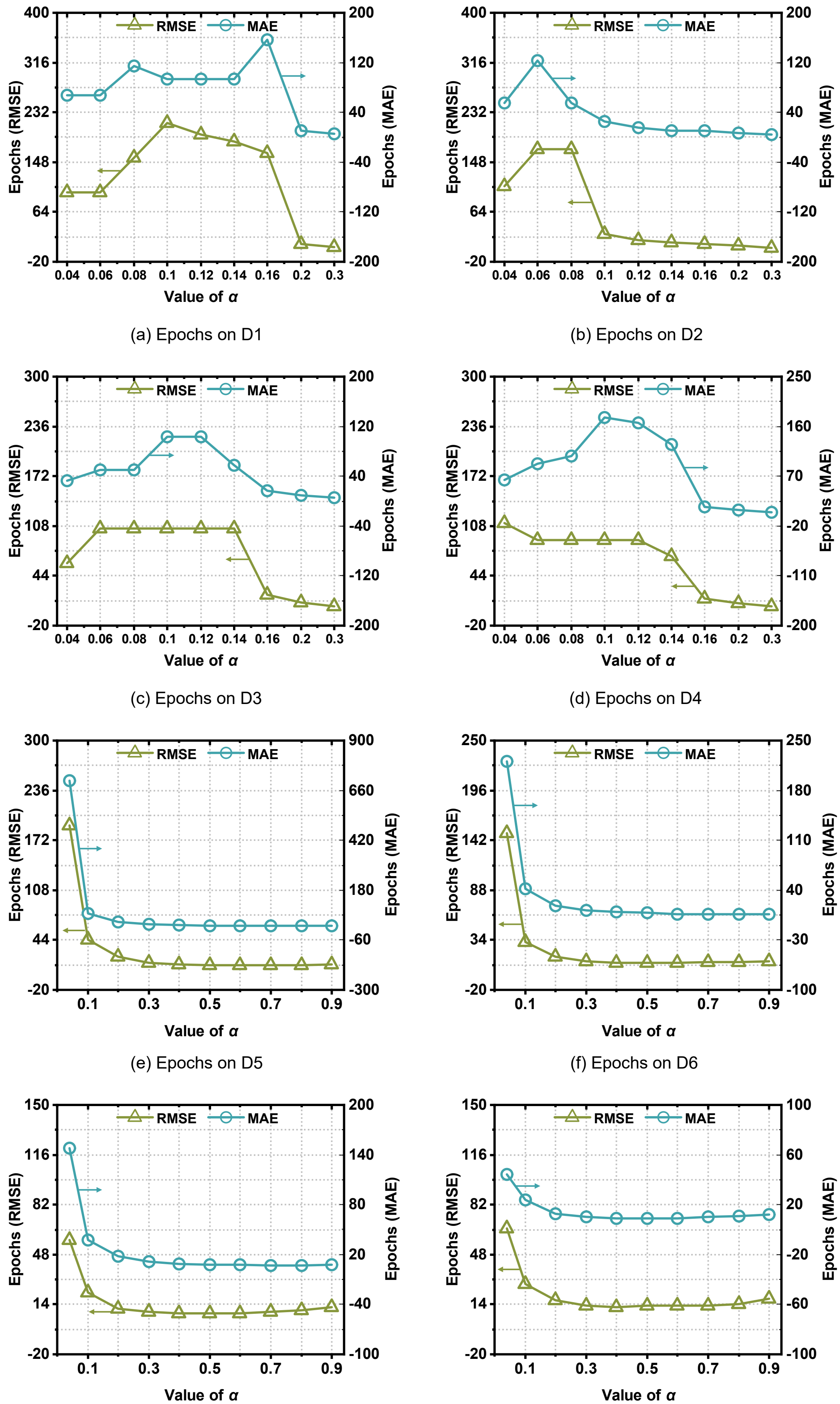


(g) Errors on D7



(h) Errors on D8

Fig. S5. Errors of GL^{2}FA 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 GL²FA as α varies while other hyperparameters are being fixed on D1-8.