Discovering Spatio-Temporal-Individual Coupled Features from Nonstandard Tensors-A Novel Dynamic Graph Mixer Approach Supplementary File

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

I. General Settings (Section V.A)

A. Details of Evaluation Protocol

Let Φ be the testing set, RMSE and MAE are calculated as:

$$RMSE = \sqrt{\left(\sum_{y_{i,j,t}\in\Psi}(y_{i,j,t}-\hat{y}_{i,j,t})^2\right)\Big/|\Phi|},\quad MAE = \left(\sum_{y_{i,j,t}\in\Psi}|y_{i,j,t}-\hat{y}_{i,j,t}|_{abs}\right)\Big/|\Phi|.$$

It should be noted that lower values of RMSE and MAE indicate higher learning performance.

B. Details of Comparison Methods

- **EvolveGCN** [28] is an evolving GCN model, which is designed for dynamic graph representation learning. It captures the spatial patterns using GCNs and learn the temporal patterns using Recurrent Neural Networks (RNNs) to evolve parameters.
- **WD-GCN** [29] is a waterfall dynamic GCN method. It combines the modules of GCN and Long Short-Term Memory (LSTM) network to capture the spatial and temporal patterns within an HDI tensor.
- **SGP** [16] is a scalable GNN-based dynamic predictor. It achieves high efficiency and accuracy by improving the learning scalability based on the combination of GCNs randomized RNNs.
- **JMP-GCF** [30] is a joint multi-gained popularity-aware graph collaborative filter model. It learns multi-gained latent features within HDI data by combining several popularity-varying graph diffusion networks.
- **GRU-GCN** [31] is a novel combination method based on GRUs and GCNs. It first theoretically demonstrates that the method based on “time-then-graph” possesses higher expressivity than the approach on “time-and-graph”.
- **APAN** [32] is an asynchronous propagation attention network. It is designed for higher inference efficiency, which can be used in the tasks of representation learning to continuous and discrete dynamic graphs.
- **TM-GCN** [33] is a dynamic GCN model based on tensor M-product framework. It follows the paradigm of Message Passing Neural Network (MPNN) on an HDI tensor to achieve effective dynamic representation learning.
- **MegaCRN** [34] is a meta-graph convolutional recurrent network-based LFoT model. It implements a meta-node bank for graph structure learning, in which the problems of spatiotemporal heterogeneity and non-stationary are addressed.
- **DDSTGCN** [35] is a dual dynamic spatiotemporal GCN. It generates a dual hypergraph according to the original dynamic graph structure in an HDI tensor, such that the characteristics of nodes and edges are simultaneously learned.
- **CTGCN** [17] is a k-core substructure learning-based temporal GCN method. It applies GCNs to constructed k-core subgraphs for learning local structural information and uses RNNs to capture dynamics.
- **hetGNN-LSTM** [36] is a combination method based on a heterogeneous GNN and an LSTM network. It constructs multiple graphs by considering different types of node interactions, thus learning more comprehensive dynamic spatial patterns.
- **MGDN** [9] is a Markov diffusion network-based LFoT model. It incorporates the idea of distance learning into its Markov diffusion process and uses an MLP network to learn the diffusion features.
- **GraphMixer** [25] is an MLP mixer-based model. It notes that previous complex methods suffer from insufficient representational capacity due to structural redundancy, and thus proposes a simple yet effective approach for dynamic graphs.
- **PGCN** [15] is a progressive GCN-based model for HDI tensor learning. It captures the temporal patterns using the dilated causal convolution and capture the trend similarities with progressive GCNs.
- **DGM** is the dynamic graph mixer-based LFoT model proposed in this paper.

C. Details of Training Settings

We consistently adopt the following settings: All trainable variables of the model are randomly initialized using Xavier initialization, drawing from a normal distribution [26]. Moreover, the parameters are trained by using the Adam optimizer [38].

In individually-built training and validation datasets, the hyperparameters are tuned with care and the achieved best settings are applied to the testing set. For all methods, the dimensionality of latent feature space is fixed at ten, the batch size is set to 2^{15}, and the coefficients controlling L_2 regularization strength and learning rate are tuned among $\{1e{-}5, 1e{-}4, 1e{-}3, 5e{-}3, 1e{-}2\}$. For all GNN-based models, the graph convolutional layer number is tuned in the range of $\{1, 2, 3, 4\}$. For the proposed DGM, we also tune the layer number of TNN module in $\{1, 2, 3, 4\}$. Considering the effects of random factors such as parameter initialization, we apply the ten-fold cross validation to report the averaged results and the corresponding standard deviations. Once the training epochs reach the upper bound, which is set 1000 in our implementations, or the accuracy stops to increase for twenty epochs, we terminate our training process of a model.

II. Supplementary Experimental Tables (Section V.B)

TABLE SI
THE TRAINING TIME IN RMSE AND WIN/LOSS COUNTS OF DGM AND ITS PEERS ON ALL TESTING CASES.

Method	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
EvolveGCN	$10375_{\pm6431.98}$	$4510_{\pm3050.95}$	$3252_{\pm4162.19}$	$5200_{\pm6094.53}$	$8159_{\pm11204.78}$	$5137_{\pm5907.52}$	$10127_{\pm6367.34}$	$1443_{\pm1420.94}$	8/0
WD-GCN	$8325_{\pm4273.51}$	$634_{\pm285.64}$	$28709_{\pm15645.77}$	$10661_{\pm11175.12}$	$17060_{\pm8908.01}$	$2681_{\pm1028.56}$	$7110_{\pm2414.27}$	$438_{\pm121.39}$	8/0
SGP	$2871_{\pm1469.48}$	$1082_{\pm895.18}$	$6668_{\pm4301.42}$	$2017_{\pm1627.00}$	$3839_{\pm975.20}$	$2066_{\pm1509.69}$	$2672_{\pm1901.53}$	$513_{\pm392.47}$	8/0
JMP-GCF	$1078_{\pm51.79}$	$875_{\pm55.33}$	$2296_{\pm70.29}$	$1956_{\pm71.07}$	$1273_{\pm40.28}$	$1001_{\pm27.35}$	$2090_{\pm120.92}$	$1187_{\pm50.05}$	8/0
GRU-GCN	$3377_{\pm1410.47}$	$2649_{\pm1601.13}$	$6782_{\pm2263.72}$	$1440_{\pm1100.78}$	$9420_{\pm3290.96}$	$2145_{\pm1327.81}$	$3253_{\pm1749.59}$	$728_{\pm37.25}$	8/0
APAN	$5036_{\pm2949.75}$	$2173_{\pm700.66}$	$38006_{\pm10153.24}$	$13788_{\pm4612.29}$	$15716_{\pm5151.07}$	$4222_{\pm1950.47}$	$3621_{\pm1935.25}$	$699_{\pm290.10}$	8/0
TM-GCN	$8250_{\pm607.91}$	$3731_{\pm113.68}$	$21367_{\pm1858.38}$	$14151_{\pm1621.55}$	$16105_{\pm1344.87}$	$9022_{\pm618.51}$	$6926_{\pm296.63}$	$1381_{\pm39.66}$	8/0
MegaCRN	$2297_{\pm351.69}$	$923_{\pm292.57}$	$6830_{\pm2574.04}$	$3926_{\pm1022.69}$	$3862_{\pm596.83}$	$1765_{\pm146.03}$	$1517_{\pm265.30}$	$407_{\pm175.14}$	8/0
DDSTGCN	$4185_{\pm668.43}$	$1763_{\pm237.90}$	$21439_{\pm5913.63}$	$9411_{\pm3304.63}$	$11425_{\pm3304.83}$	$2242_{\pm687.74}$	$264_{\pm92.77}$	$47_{\pm28.57}$	8/0
CTGCN	$17207_{\pm8395.54}$	$4405_{\pm1101.33}$	$79398_{\pm29751.29}$	$17600_{\pm17381.99}$	$43956_{\pm2956.75}$	$17460_{\pm12817.65}$	$10422_{\pm2198.65}$	$3553_{\pm676.38}$	8/0
hetGNN-LSTM	$9069_{\pm8403.30}$	$3579_{\pm1661.85}$	$27644_{\pm28841.19}$	$6907_{\pm3930.29}$	$33584_{\pm26999.71}$	$4712_{\pm1963.68}$	$4284_{\pm3458.99}$	$2172_{\pm920.18}$	8/0
MGDN	$3849_{\pm81.67}$	$2068_{\pm51.30}$	$7641_{\pm62.61}$	$4576_{\pm67.52}$	$4665_{\pm45.03}$	$2611_{\pm34.26}$	$2122_{\pm53.86}$	$1149_{\pm41.19}$	8/0
GraphMixer	$27995_{\pm6903.67}$	$17131_{\pm9065.54}$	$79507_{\pm31147.10}$	$33282_{\pm15717.05}$	$58694_{\pm10633.01}$	$26556_{\pm7149.97}$	$7403_{\pm752.61}$	$2799_{\pm111.44}$	8/0
PGCN	$12360_{\pm2677.03}$	$4355_{\pm475.57}$	$60950_{\pm15783.70}$	$19146_{\pm2770.23}$	$33020_{\pm13709.52}$	$12715_{\pm2962.20}$	$5613_{\pm1072.57}$	$1719_{\pm263.47}$	8/0
DGM (Ours)	$\mathbf{220_{\pm38.88}}$	$\mathbf{158_{\pm17.64}}$	$\mathbf{677_{\pm189.01}}$	$\mathbf{420_{\pm65.45}}$	$\mathbf{366_{\pm61.17}}$	$\mathbf{256_{\pm40.34}}$	$\mathbf{69_{\pm31.33}}$	$\mathbf{11_{\pm1.95}}$	—

TABLE SII
THE TRAINING TIME IN MAE AND WIN/LOSS COUNTS OF DGM AND ITS PEERS ON ALL TESTING CASES.

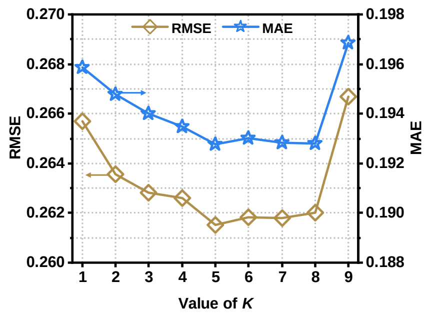
Method	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
EvolveGCN	$10607_{\pm6320.85}$	$4477_{\pm3043.81}$	$3654_{\pm4027.53}$	$5975_{\pm6341.39}$	$8196_{\pm11272.59}$	$5049_{\pm5978.89}$	$10115_{\pm6355.74}$	$1436_{\pm1428.04}$	8/0
WD-GCN	$8325_{\pm4273.52}$	$552_{\pm425.41}$	$26301_{\pm15363.35}$	$7289_{\pm7892.42}$	$15102_{\pm10538.62}$	$2429_{\pm1029.18}$	$3307_{\pm3733.58}$	$286_{\pm150.47}$	8/0
SGP	$2871_{\pm1469.48}$	$1083_{\pm894.28}$	$6548_{\pm4223.00}$	$1855_{\pm1719.66}$	$3887_{\pm1013.18}$	$1945_{\pm1481.41}$	$2670_{\pm1902.67}$	$489_{\pm360.72}$	8/0
JMP-GCF	$1052_{\pm50.91}$	$866_{\pm55.45}$	$2267_{\pm65.36}$	$1935_{\pm67.79}$	$1247_{\pm38.62}$	$984_{\pm26.02}$	$2070_{\pm120.20}$	$1177_{\pm47.82}$	8/0
GRU-GCN	$3455_{\pm1519.39}$	$2850_{\pm1507.78}$	$7112_{\pm2019.30}$	$1473_{\pm1064.06}$	$10364_{\pm3109.95}$	$2368_{\pm1110.01}$	$3267_{\pm1732.10}$	$749_{\pm8.72}$	8/0
APAN	$5855_{\pm2561.99}$	$2425_{\pm1049.63}$	$37644_{\pm10230.38}$	$12232_{\pm2933.88}$	$24436_{\pm6197.51}$	$6928_{\pm2146.25}$	$3925_{\pm1777.44}$	$675_{\pm201.20}$	8/0
TM-GCN	$8318_{\pm492.67}$	$3736_{\pm108.43}$	$21245_{\pm1778.91}$	$14444_{\pm1801.65}$	$16212_{\pm1324.56}$	$8977_{\pm623.70}$	$6905_{\pm301.01}$	$1378_{\pm36.87}$	8/0
MegaCRN	$2279_{\pm370.52}$	$903_{\pm293.70}$	$6750_{\pm2530.14}$	$3744_{\pm958.19}$	$3827_{\pm588.34}$	$1684_{\pm135.39}$	$1120_{\pm127.09}$	$407_{\pm173.74}$	8/0
DDSTGCN	$4077_{\pm639.85}$	$1744_{\pm260.61}$	$21394_{\pm5805.55}$	$8981_{\pm3020.53}$	$10814_{\pm4517.26}$	$2020_{\pm503.11}$	$189_{\pm138.36}$	$45_{\pm29.77}$	8/0
CTGCN	$16944_{\pm8334.81}$	$4437_{\pm1167.77}$	$77888_{\pm29965.56}$	$17422_{\pm16900.06}$	$43883_{\pm3077.13}$	$17634_{\pm12980.10}$	$10154_{\pm1936.71}$	$3521_{\pm658.21}$	8/0
hetGNN-LSTM	$8298_{\pm7987.07}$	$2903_{\pm1469.55}$	$22525_{\pm17819.45}$	$5901_{\pm3488.97}$	$28737_{\pm24056.25}$	$4623_{\pm2245.09}$	$3087_{\pm3749.99}$	$1757_{\pm1216.73}$	8/0
MGDN	$3842_{\pm81.31}$	$2063_{\pm51.47}$	$7630_{\pm62.29}$	$4568_{\pm67.39}$	$4654_{\pm43.14}$	$2605_{\pm33.81}$	$2114_{\pm54.42}$	$1144_{\pm41.21}$	8/0
GraphMixer	$15369_{\pm19035.19}$	$6900_{\pm1799.54}$	$16051_{\pm4641.14}$	$10431_{\pm2966.08}$	$9565_{\pm5712.36}$	$8067_{\pm4581.37}$	$5863_{\pm1331.99}$	$2100_{\pm457.08}$	8/0
PGCN	$8931_{\pm4041.31}$	$3485_{\pm1262.86}$	$58375_{\pm19311.78}$	$15695_{\pm4753.46}$	$24129_{\pm9552.01}$	$7965_{\pm1551.05}$	$2241_{\pm1120.31}$	$785_{\pm364.81}$	8/0
DGM (Ours)	$\mathbf{222_{\pm36.03}}$	$\mathbf{158_{\pm21.58}}$	$\mathbf{698_{\pm200.54}}$	$\mathbf{409_{\pm51.38}}$	$\mathbf{399_{\pm84.45}}$	$\mathbf{277_{\pm40.93}}$	$\mathbf{79_{\pm31.01}}$	$\mathbf{11_{\pm4.49}}$	—

TABLE SIII
WILCOXON SIGNED-RANKS TEST OUTCOMES ON EFFICIENCY. (*For DGM, high R+ value stands for high efficiency. **With the significance level of 0.01, we hypotheses the accepted hypotheses.*)

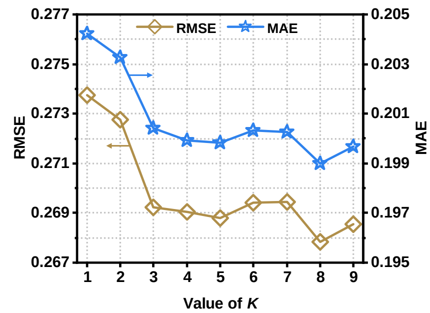
Comparision	R+*	R-	p-value**	Comparision	R+*	R-	p-value**
DGM vs EvolveGCN	136	0	**2.41E-4**	DGM vs MegaCRN	136	0	**2.41E-4**
DGM vs WD-GCN	136	0	**2.41E-4**	DGM vs DDSTGCN	136	0	**2.41E-4**
DGM vs SGP	136	0	**2.41E-4**	DGM vs CTGCN	136	0	**2.41E-4**
DGM vs JMP-GCF	136	0	**2.41E-4**	DGM vs hetGNN-LSTM	136	0	**2.41E-4**
DGM vs GRU-GCN	136	0	**2.41E-4**	DGM vs MGDN	136	0	**2.41E-4**
DGM vs APAN	136	0	**2.41E-4**	DGM vs GraphMixer	136	0	**2.41E-4**
DGM vs TM-GCN	136	0	**2.41E-4**	DGM vs PGCN	136	0	**2.41E-4**

III. Supplementary Experimental Figures (Section V.C)

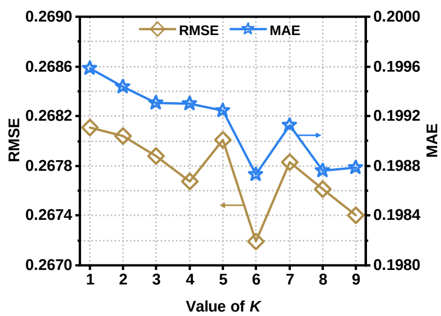
- Fig. S1 (as discussed in Section V.C.1) illustrates how the errors of DGM vary with changes in K.
- Fig. S2 (as discussed in Section V.C.2) illustrates how the errors of DGM vary with changes in L.
- Fig. S3 (as discussed in Section V.C.3) illustrates how the errors of DGM vary with changes in R.



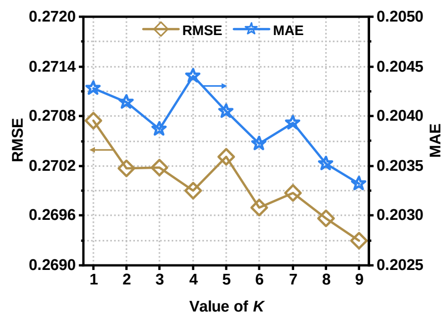
(a) Errors on D1



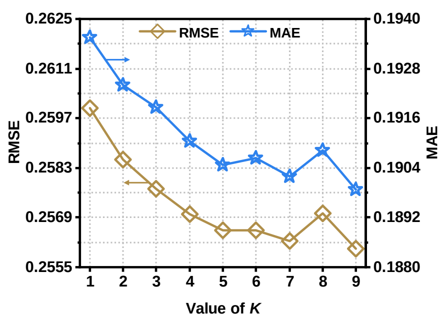
(b) Errors on D2



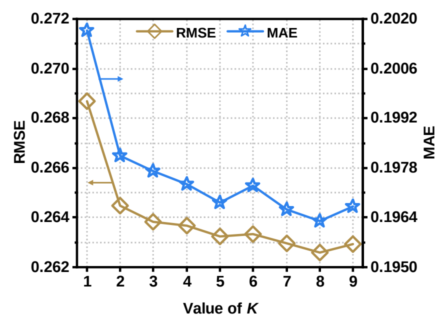
(c) Errors on D3



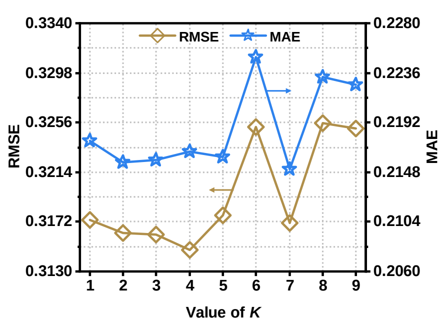
(d) Errors on D4



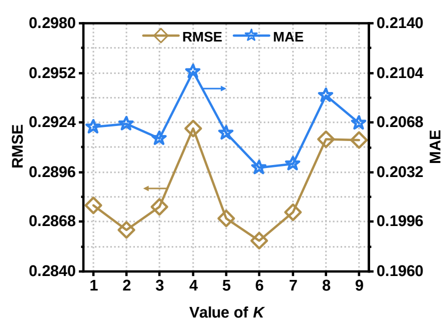
(e) Errors on D5



(f) Errors on D6

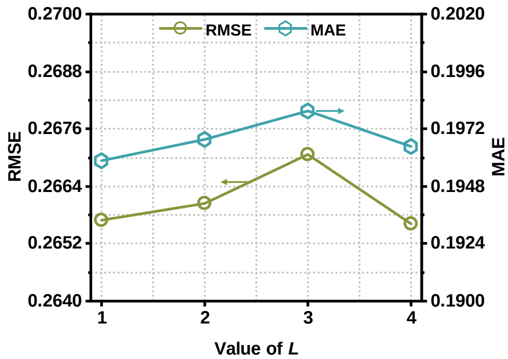


(g) Errors on D7

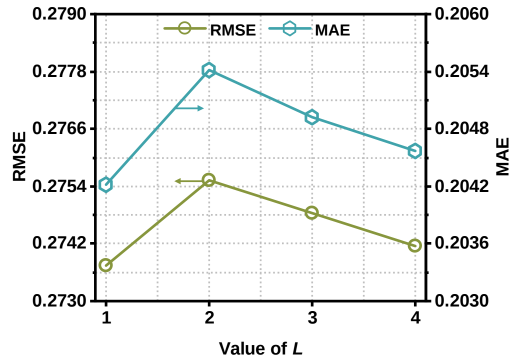


(h) Errors on D8

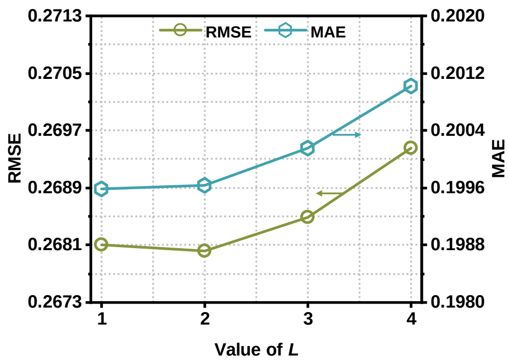
Fig. S1. Errors of DGM as K varies while fixing others on D1-8.



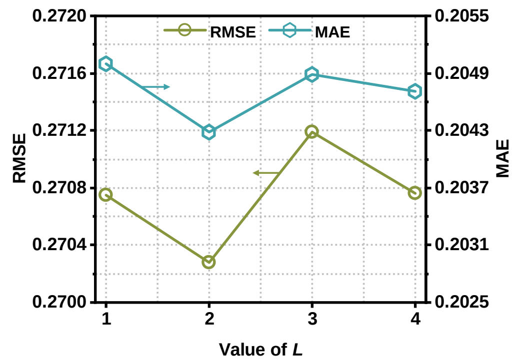
(a) Errors on D1



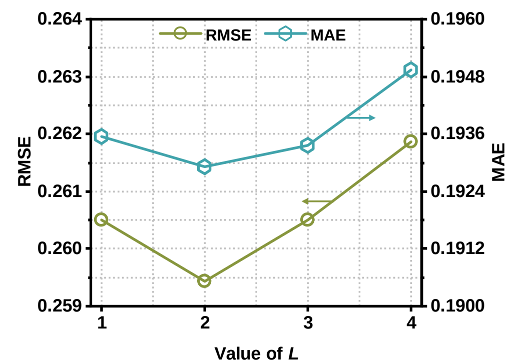
(b) Errors on D2



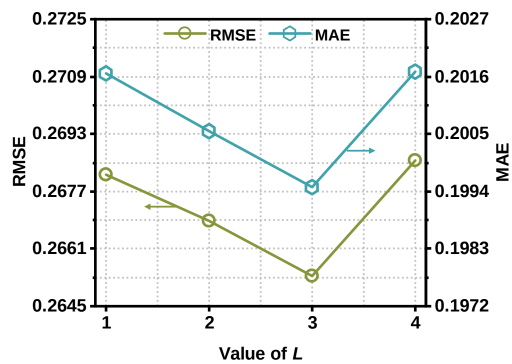
(c) Errors on D3



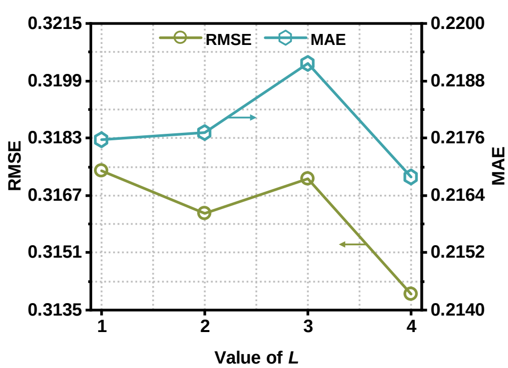
(d) Errors on D4



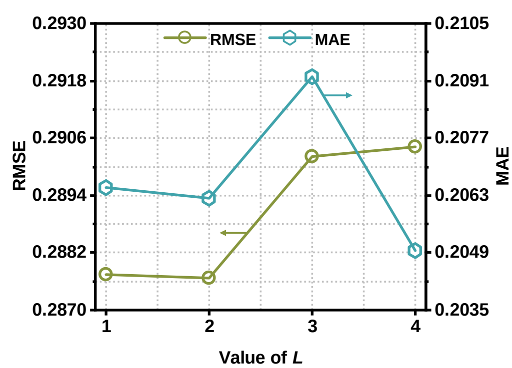
(e) Errors on D5



(f) Errors on D6

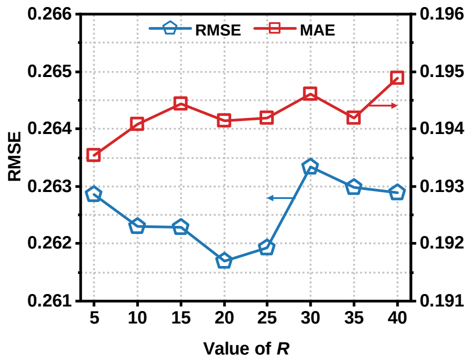


(g) Errors on D7

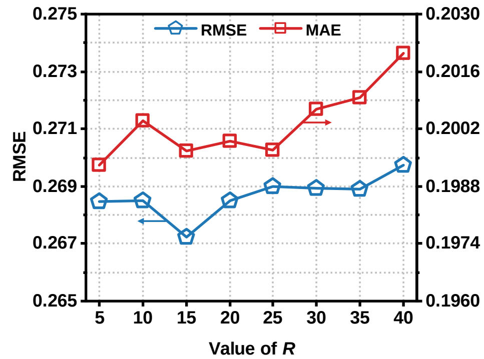


(h) Errors on D8

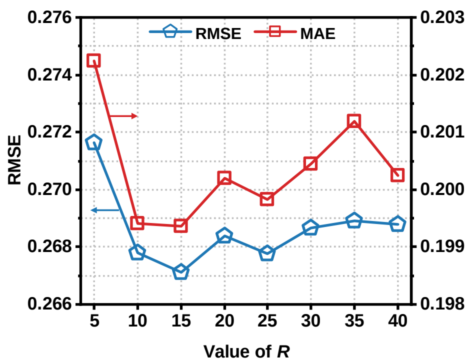
Fig. S2. Errors of DGM as L varies while fixing others on D1-8.



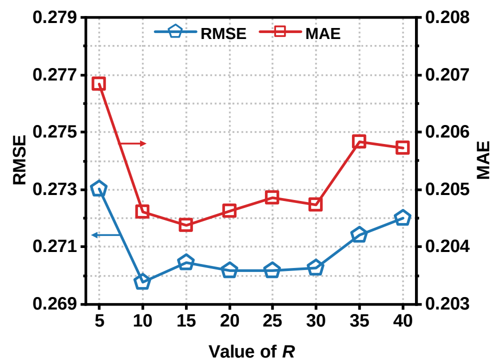
(a) Errors on D1



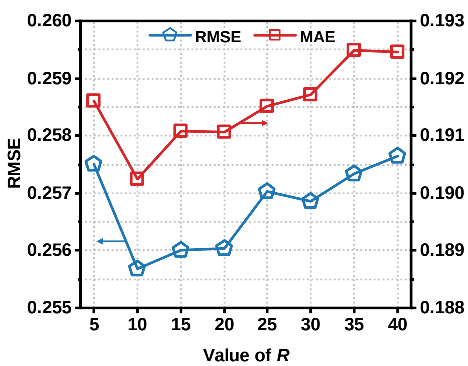
(b) Errors on D2



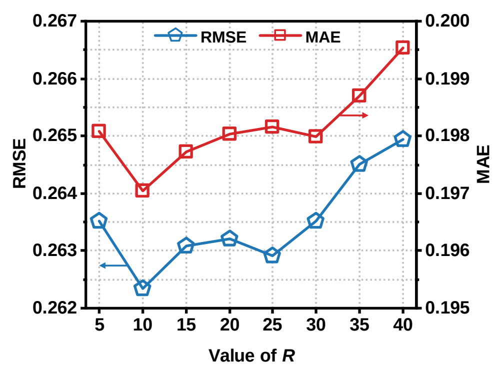
(c) Errors on D3



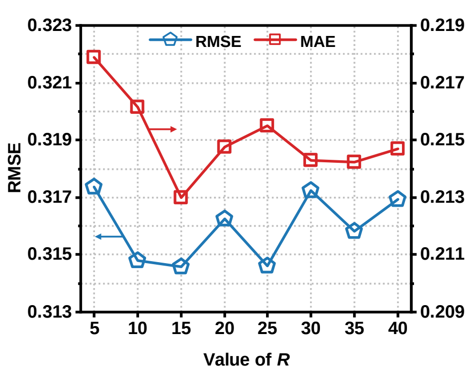
(d) Errors on D4



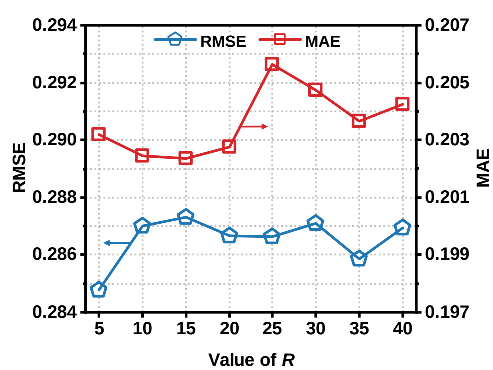
(e) Errors on D5



(f) Errors on D6



(g) Errors on D7



(h) Errors on D8

Fig. S3. Errors of DGM as R varies while fixing others on D1-8.