A Proximal-ADMM-incorporated Nonnegative Latent-Factorization-of-Tensors Model for Representing Dynamic Cryptocurrency Transaction Network Supplementary File

Anonymous

I. Introduction

This is the supplementary file for paper entitled *A Proximal-ADMM-incorporated Nonnegative Latent-Factorization-of-Tensors for Dynamic Cryptocurrency Transaction Network Embedding*. The convergence proof of PNL, supplementary procedure, experimental results, and related work are put into this file.

II. Convergence Proof of PNL

Given $i\in I$, $j\in J$, and $k\in K$, the PNL model's convergence proof is presented as follows:

(a) Proof of Step 1: Note that we present the proof procedure for variable $\hat{u}_{ir}$, u_{ir}, and d_{ir}, and the similar variables also applies to the same conclusion.

***Lemma* 1**. With (11), $\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2$, $\left(h_{jr}^{t+1}-h_{jr}^{t}\right)^2$, and $\left(l_{kr}^{t+1}-l_{kr}^{t}\right)^2$ are bounded as:

$$\begin{aligned}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2 \le\, & 8\left(\left(\eta-1\right)^2\alpha_i^2+\rho^2\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2+8\alpha_i^2\left(\left(\eta-1\right)^2+1\right)\left(s_{ir}^{t+1}-s_{ir}^{t}\right)^2\\ &+8\rho^2\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)^2+8\alpha_i^2\left(u_{ir}^{t}-u_{ir}^{t-1}\right)^2+4\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)^2=\varphi_d\end{aligned} \tag{S1}$$

Where Δ_{ir}^{t+1} is defined as:

$$\Delta_{ir}^{t+1}=\sum_{y_{ijk}\in\Lambda}\left(y_{ijk}-\left(\sum_{f_1=1}^{r-1}\hat{u}_{if_1}^{t+1}\hat{v}_{jf_1}^{t+1}\hat{w}_{kf_1}^{t+1}+\hat{u}_{ir}^{t+1}\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}+\sum_{f_2=r+1}^{R}\hat{u}_{if_2}^{t}\hat{v}_{jf_2}^{t}\hat{w}_{if_2}^{t}+\hat{a}_i^t+\hat{c}_j^t+\hat{e}_k^t\right)\right)\left(-\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right). \tag{S2}$$

***Proof* 1**. Note that (7) is non-convex and its zero-gradient points, such as local/global optimum and saddle point, should be regarded as a feasible solution. Therefore, assuming that $\hat{u}_{ir}^{t+1}$ is the solution to $\hat{u}_{ir}$ by (11), the following condition is fulfilled:

$$\Delta_{ir}^{t+1}+\alpha_i\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t}+\frac{d_{ir}^{t}}{\alpha_i}\right)+\rho\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)=0. \tag{S3}$$

By substituting the update rule in (11) and (15) into (S3), the following equation is achieved:

$$d_{ir}^{t+1}=\left(\eta-1\right)\alpha_i\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t+1}\right)-\alpha_i\left(u_{ir}^{t+1}-u_{ir}^{t}\right)-\rho\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)-\Delta_{ir}^{t+1}. \tag{S4}$$

Further, the difference between d_{ir}^{t+1} and d_{ir}^{t} is given as:

$$\begin{aligned}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2=\,&(\left(\eta-1\right)\alpha_i\left(\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t+1}\right)-\left(\hat{u}_{ir}^{t}-u_{ir}^{t}\right)\right)-\alpha_i\left(\left(u_{ir}^{t+1}-u_{ir}^{t}\right)-\left(u_{ir}^{t}-u_{ir}^{t-1}\right)\right)\\ &-\rho\left(\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)-\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)\right)-\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right))^2.\end{aligned} \tag{S5}$$

With the inequality $\left(a-b-c-d\right)^2\le 4\left(a^2+b^2+c^2+d^2\right)$, we have:

$$\begin{aligned}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2\le\,&4\left(\eta-1\right)^2\alpha_i^2\left(\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t+1}\right)-\left(\hat{u}_{ir}^{t}-u_{ir}^{t}\right)\right)^2+4\alpha_i^2\left(\left(u_{ir}^{t+1}-u_{ir}^{t}\right)-\left(u_{ir}^{t}-u_{ir}^{t-1}\right)\right)^2\\ &+4\rho^2\left(\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)-\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)\right)^2+4\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)^2.\end{aligned} \tag{S6}$$

With (S6), we implement (S1) by using the inequality $\left(a-b\right)^2\le 2\left(a^2+b^2\right)$. Note that applying the same principle, we can get $\left(h_{jr}^{t+1}-h_{jr}^{t}\right)^2\le\varphi_h$ and $\left(l_{kr}^{t+1}-l_{kr}^{t}\right)^2\le\varphi_l$. Hence, ***Lemma* 1** stands.

(b) Proof of Step 2: ***Lemma* 1** has been proved, then we perform Step 2. For simplicity, we first introduce seven functions and intermediate variables to express similar structures as follows:

$$F_1\left(\hat{u}_{ir},u_{ir},\alpha_i\right)=\left(\frac{8\left((\eta-1)^2\alpha_i^2+\rho^2\right)}{\eta\alpha_i}-\frac{1}{2}\left(\sum_{i\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^2+\alpha_i+\rho\right)\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2+\left(\frac{8\rho^2}{\eta\alpha_i}\right)\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)^2=F_1^I. \tag{S7}$$

With (S7), we define a function expression for $\{\hat{u}_{ir},u_{ir},\alpha_i\}$. Note that the above three variables are related to the I-dimension of the tensor, and we can get F_1^J and F_1^K for $\{\hat{v}_{jr},v_{jr},\beta_j\}$ in J-dimension and $\{\hat{w}_{kr},w_{kr},\delta_k\}$ in K-dimension by adopting the similar expression. Note that for the term $\sum_{i\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^{2}$ in *I*-dimension, its expression for the *J*-dimension and *K*-dimension are $\sum_{j\in\Lambda(j)}\left(\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}\right)^{2}$ and $\sum_{k\in\Lambda(k)}\left(\hat{u}_{ir}^{t}\hat{v}_{jr}^{t}\right)^{2}$, respectively. Hence, we define the first intermediate variable $A_1 = F_1^I + F_1^J + F_1^K$.

Next the second function expression is as follows:

$$F_2\left(\hat{u}_{ir},u_{ir},\alpha_i,\Delta_{ir}\right)=\left(\left(\frac{\alpha_i}{2}-\frac{8\alpha_i\left((\eta-1)^2+1\right)}{\eta}\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2-\frac{8\alpha_i\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)^2}{\eta}-\frac{4\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)^2}{\eta\alpha_i}\right)=F_2^I. \tag{S8}$$

Similarly, with (S8), we get F_2^I, F_2^J, F_2^K and thus the second intermediate variable $A_2 = F_2^I + F_2^J + F_2^K$.

The third function expression is given as follows:

$$F_3\left(\hat{u}_{ir},u_{ir},\alpha_i\right)=\sum_{i\in I}\alpha_i\left(\sum_{f_1=1}^{r-1}\left(\hat{u}_{if_1}^{t+1}-u_{if_1}^{t+1}\right)^2+\sum_{f_2=r}^{R}\left(\hat{u}_{if_2}^{t}-u_{if_2}^{t}\right)^2\right)=F_3^I. \tag{S9}$$

With (S9), the third intermediate variable is defined as $A_3 = F_3^I + F_3^J + F_3^K$. Similarly, the fourth function expression is:

$$F_4\left(\hat{u}_{ir}\right)=\frac{\rho}{2}\left(\sum_{i\in I}\left(\sum_{f_1=1}^{r-1}\left(\hat{u}_{if_1}^{t+1}-\hat{u}_{if_1}^{t}\right)^2+\sum_{f_2=r}^{R}\left(\hat{u}_{if_2}^{t}-\hat{u}_{if_2}^{t-1}\right)^2\right)\right)=F_4^I. \tag{S10}$$

Correspondingly, the fourth intermediate variable is $A_4 = F_4^I + F_4^J + F_4^K$. The next fifth function expression is given as follows:

$$\begin{aligned}F_5\left(\hat{u}_{ir},u_{ir},\alpha_i,\Delta_{ir}\right)=&\sum_{i\in I}\left(\sum_{f_1=2}^{r-1}\left(\alpha_i\left(u_{if_1}^{t+1}-u_{if_1}^{t}\right)+\Delta_{if_1}^{t+1}+\rho\left(\hat{u}_{if_1}^{t+1}-\hat{u}_{if_1}^{t}\right)\right)\left(u_{if_1}^{t+1}-\hat{u}_{if_1}^{t+1}\right)\right)\\&+\sum_{i\in I}\left(\sum_{f_2=r}^{R}\left(\alpha_i\left(u_{if_2}^{t}-u_{if_2}^{t-1}\right)+\Delta_{if_2}^{t}+\rho\left(\hat{u}_{if_2}^{t}-\hat{u}_{if_2}^{t-1}\right)\right)\left(u_{if_2}^{t}-u_{if_2}^{t}\right)\right)=F_5^I.\end{aligned} \tag{S11}$$

With (S11), the fifth intermediate variable is $A_5 = F_5^I + F_5^J + F_5^K$.

A functional expression is given as:

$$F_6\left(\hat{u}_{ir},u_{ir},d_{ir}\right)=\sum_{i\in I}\left(\sum_{f_1=1}^{r-1}d_{if_1}^{t+1}\left(\hat{u}_{if_1}^{t+1}-u_{if_1}^{t+1}\right)+\sum_{f_2=r}^{R}d_{if_2}^{t}\left(\hat{u}_{if_2}^{t}-u_{if_2}^{t}\right)\right)=F_6^I. \tag{S12}$$

With (S12), we can get the *J*-dimension and *K*-dimension versions, F_6^J and F_6^K. Hence, the sixth intermediate variable is $A_6 = F_6^I + F_6^J + F_6^K$. Finally, a functional expression for the bias is given as:

$$F_6\left(\hat{u}_{ir},u_{ir},d_{ir}\right)=\sum_{i\in I}\left(\sum_{f_1=1}^{r-1}d_{if_1}^{t+1}\left(\hat{u}_{if_1}^{t+1}-u_{if_1}^{t+1}\right)+\sum_{f_2=r}^{R}d_{if_2}^{t}\left(\hat{u}_{if_2}^{t}-u_{if_2}^{t}\right)\right)=F_6^I. \tag{S13}$$

With (S13), it defines a function expression for $\{\hat{a}_i,a_i,\sigma_i,o_i\}$ and is related to the *I*- dimension. We can obtain expressions about the *J*-dimension and *K*-dimension by similar expressions for $\{\hat{c}_j,c_j,\phi_j,s_j\}$ and $\{\hat{e}_k,e_k,\psi_k,z_k\}$, i.e., F_7^J and F_7^K. Therefore, we get the last intermediate variable as $A_7 = F_7^I + F_7^J + F_7^K$.

***Lemma* 2**. If the following inequality is satisfied:

$$A_1 \le A_2. \tag{S14}$$

Then the following inequalities holds:

$$L_p\left(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t+1}\right)-L_p\left(\mathcal{D}_1^{t}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}\right)\le 0. \tag{S15}$$

Note that we have:

$$L_p\left(\mathcal{D}_1^{t}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}\right)\ge 0. \tag{S16}$$

With the following condition:

$$\eta \ge \frac{1}{2}-\frac{A_4+A_5+A_7}{A_3}. \tag{S17}$$

***Proof* 2**. By expanding the second-order Taylor expansion of L_p at the point $\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}$, we can get the following equality:

$$\begin{aligned}L_p\left(D_1^{t+1}\cup D_2^{t}\cup D_3^{t}\right)-L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)\overset{(11)}{=}&-\frac{1}{2}\left(\sum_{y_{ik}\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^2+\alpha_i+\rho\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2\\-\frac{1}{2}\left(\sum_{y_{ik}\in\Lambda(j)}\left(\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}\right)^2+\beta_j+\rho\right)\left(\hat{v}_{jr}^{t+1}-\hat{v}_{jr}^{t}\right)^2&-\frac{1}{2}\left(\sum_{y_{ik}\in\Lambda(k)}\left(\hat{u}_{ir}^{t}\hat{v}_{jr}^{t}\right)^2+\delta_k+\rho\right)\left(\hat{w}_{kr}^{t+1}-\hat{w}_{kr}^{t}\right)^2.\end{aligned} \tag{S18}$$

Note that the equality (S18) holds when the first-order terms in (9) are equal zero. Therefore, the difference between $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t})$ and $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t})$ is given as:

$$L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t}\right)-L_p\left(D_1^{t+1}\cup D_2^{t}\cup D_3^{t}\right)\overset{(9)}{\le}-\frac{\alpha_i}{2}\left(u_{ir}^t-u_{ir}^{t+1}\right)^2-\frac{\beta_j}{2}\left(v_{jr}^t-v_{jr}^{t+1}\right)^2-\frac{\delta_k}{2}\left(w_{kr}^t-w_{kr}^{t+1}\right)^2. \tag{S19}$$

Note that the inequality (S19) considers the optimal condition of (11) and the projection rule of (14), the first-order terms equal to or less than zero are omitted. Therefore, the difference between $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t+1})$ and $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t})$ is:

$$\begin{aligned}L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t+1}\right)-L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t}\right)&\overset{(11),(14)}{=}\frac{\left(d_{ir}^{t+1}-d_{ir}^t\right)^2}{\eta\alpha_i}+\frac{\left(h_{jr}^{t+1}-h_{jr}^t\right)^2}{\eta\beta_j}+\frac{\left(l_{kr}^{t+1}-l_{kr}^t\right)^2}{\eta\delta_k}\\&\overset{(S1)}{\le}\frac{\varphi_d}{\eta\alpha_i}+\frac{\varphi_h}{\eta\beta_j}+\frac{\varphi_l}{\eta\delta_k},\end{aligned} \tag{S20}$$

where the equality depends on (11) and (15), and the inequality follows ***Lemma* 1**. With (S18)-(S20), we have:

$$L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t+1}\right)-L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)\le A_1-A_2. \tag{S21}$$

Therefore, with (S14), the following inequality evidently holds:

$$L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t+1}\right)-L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)\le 0. \tag{S22}$$

Hence, the proximal-incorporated augmented Lagrangian of (7) is non-increasing. Moreover, after the *t*-th iteration, (7) is formulated as:

$$L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)=\frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(y_{ijk}-\sum_{f_1=1}^{r-1}\hat{u}_{if_1}^{t+1}\hat{v}_{jf_1}^{t+1}\hat{w}_{kf_1}^{t+1}-\sum_{f_2=r}^{R}\hat{u}_{if_2}^{t}\hat{v}_{jf_2}^{t}\hat{w}_{kf_2}^{t}-\hat{a}_i^t-\hat{c}_j^t-\hat{e}_k^t\right)^2+\frac{A_3}{2}+A_4+A_6+A_7. \tag{S23}$$

By substituting (S4) into (S23), the following deduction is achieved:

$$L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)=\frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(y_{ijk}-\sum_{f_1=1}^{r-1}\hat{u}_{if_1}^{t+1}\hat{v}_{jf_1}^{t+1}\hat{w}_{kf_1}^{t+1}-\sum_{f_2=r}^{R}\hat{u}_{if_2}^{t}\hat{v}_{jf_2}^{t}\hat{w}_{kf_2}^{t}-\hat{a}_i^t-\hat{c}_j^t-\hat{e}_k^t\right)^2+\frac{(2\eta-1)A_3}{2}+A_4+A_5+A_7. \tag{S24}$$

With (S17) and (S24), (S16) is fulfilled, i.e., (7) is lower-bounded. Therefore, ***Lemma* 2** holds.

(c) Proof of Step 3: Considering $\mathrm{U}\ge 0$, $\mathrm{V}\ge 0$, and $\mathrm{W}\ge 0$, the proximal augmented Lagrangian of (7) is extended as:

$$L_p^{\#}=L_p-tr(\mathrm{GU})-tr(\mathrm{MV})-tr(\mathrm{PW})=L_p-\sum_{i\in I}\sum_{r=1}^{R}g_{ir}u_{ir}-\sum_{j\in J}\sum_{r=1}^{R}m_{jr}v_{jr}-\sum_{k\in K}\sum_{r=1}^{R}p_{kr}w_{kr}, \tag{S25}$$

where $tr(\cdot)$ calculates the trace of an involved matrix. G, M, and P denote Lagrangian multiplier for PNL's nonnegative constraint. Then, ***Theorem* 1** is presented:

***Theorem* 1**. If the following conditions hold:

$$\begin{cases}\dfrac{8\left((\eta-1)^2(\alpha_i)^2+\rho^2\right)}{\eta\alpha_i}\ne-\dfrac{1}{2}\left(\displaystyle\sum_{y_{ijk}\in\Lambda(i)}\left(\hat{v}_{jr}^t\hat{w}_{kr}^t\right)^2+\alpha_i+\rho\right)\\ \dfrac{8\left((\eta-1)^2(\beta_j)^2+\rho^2\right)}{\eta\beta_j}\ne-\dfrac{1}{2}\left(\displaystyle\sum_{y_{ijk}\in\Lambda(j)}\left(\hat{u}_{ir}^t\hat{w}_{kr}^t\right)^2+\beta_j+\rho\right)\\ \dfrac{8\left((\eta-1)^2(\delta_k)^2+\rho^2\right)}{\eta\delta_k}\ne-\dfrac{1}{2}\left(\displaystyle\sum_{y_{ijk}\in\Lambda(k)}\left(\hat{u}_{ir}^t\hat{v}_{jr}^t\right)^2+\delta_k+\rho\right)\\ \dfrac{\alpha_i}{2}\ne\dfrac{8\alpha_i\left((\eta-1)^2+1\right)}{\eta},\dfrac{\beta_j}{2}\ne\dfrac{8\beta_j\left((\eta-1)^2+1\right)}{\eta},\dfrac{\delta_k}{2}\ne\dfrac{8\delta_k\left((\eta-1)^2+1\right)}{\eta}\\ \rho\ne 0,\eta\ne 0,\alpha_i\ne 0,\beta_j\ne 0,\delta_k\ne 0.\end{cases} \tag{S26}$$

With (11), the equilibrium point $\mathcal{D}_1^*\cup\mathcal{D}_2^*\cup\mathcal{D}_3^*$ of $\mathcal{D}_1\cup\mathcal{D}_2\cup\mathcal{D}_3$ is a KKT stationary point, and the following KKT conditions holds:

$$\hat{u}_{ir}^*-u_{ir}^*=0,\hat{v}_{jr}^*-v_{jr}^*=0,\hat{w}_{kr}^*-w_{kr}^*=0, \tag{S27a}$$

$$\left.\frac{\partial L_p^{\#}}{\partial\hat{u}_{ir}}\right|_{\hat{u}_{ir}=\hat{u}_{ir}^*}=\Delta_{ir}^*+d_{ir}^*=0;\left.\frac{\partial L_p^{\#}}{\partial\hat{v}_{jr}}\right|_{\hat{v}_{jr}=\hat{v}_{jr}^*}=\Delta_{jr}^*+h_{jr}^*=0;\left.\frac{\partial L_p^{\#}}{\partial\hat{w}_{kr}}\right|_{\hat{w}_{kr}=\hat{w}_{kr}^*}=\Delta_{kr}^*+l_{kr}^*=0, \tag{S27b}$$

$$\left.\frac{\partial L_p^{\#}}{\partial u_{ir}}\right|_{u_{ir}=u_{ir}^*}=-\alpha_i\left(\hat{u}_{ir}^*-u_{ir}^*+\frac{d_{ir}^*}{\alpha_i}\right)-g_{ir}^*=0;\left.\frac{\partial L_p^{\#}}{\partial v_{jr}}\right|_{v_{jr}=v_{jr}^*}=-\beta_j\left(\hat{v}_{jr}^*-v_{jr}^*+\frac{h_{jr}^*}{\beta_j}\right)-m_{jr}^*=0;\left.\frac{\partial L_p^{\#}}{\partial w_{kr}}\right|_{w_{kr}=w_{kr}^*}=-\delta_k\left(\hat{w}_{kr}^*-w_{kr}^*+\frac{l_{kr}^*}{\delta_k}\right)-p_{kr}^*=0, \quad \text{(S27c)}$$

$$g_{ir}^*u_{ir}^*=0, m_{jr}^*v_{jr}^*=0, p_{kr}^*w_{kr}^*=0, \quad \text{(S27d)}$$

$$u_{ir}^*\geq 0, v_{jr}^*\geq, w_{kr}^*\geq 0, \quad \text{(S27e)}$$

$$g_{ir}^*\geq 0, m_{jr}^*\geq 0, p_{kr}^*\geq 0. \quad \text{(S27f)}$$

***Proof* 3**. With ***Lemma* 2**, the following inequality holds since $t\to\infty$:

$$L_p\left(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t+1}\right)-L_p\left(\mathcal{D}_1^{t}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}\right)\leq A_1-A_2\to 0. \quad \text{(S28)}$$

Based on (S24) and (S28), we can get:

$$\begin{cases}\lim\limits_{t\to\infty}\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)\to 0, \lim\limits_{t\to\infty}\left(u_{ir}^{t+1}-u_{ir}^{t}\right)\to 0, \lim\limits_{t\to\infty}\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)\to 0\\ \lim\limits_{t\to\infty}\left(\hat{v}_{jr}^{t+1}-\hat{v}_{jr}^{t}\right)\to 0, \lim\limits_{t\to\infty}\left(v_{jr}^{t+1}-v_{jr}^{t}\right)\to 0, \lim\limits_{t\to\infty}\left(\Delta_{jr}^{t+1}-\Delta_{jr}^{t}\right)\to 0\\ \lim\limits_{t\to\infty}\left(\hat{w}_{kr}^{t+1}-\hat{w}_{kr}^{t}\right)\to 0, \lim\limits_{t\to\infty}\left(w_{kr}^{t+1}-w_{kr}^{t}\right)\to 0, \lim\limits_{t\to\infty}\left(\Delta_{kr}^{t+1}-\Delta_{kr}^{t}\right)\to 0.\end{cases} \quad \text{(S29)}$$

With (S1), we infer that:

$$\lim_{t\to\infty}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)\to 0; \lim_{t\to\infty}\left(h_{jr}^{t+1}-h_{jr}^{t}\right)\to 0; \lim_{t\to\infty}\left(l_{kr}^{t+1}-l_{kr}^{t}\right)\to 0. \quad \text{(S30)}$$

According to (15) and (S30), (S27a) holds. (11) can be reconstructed as:

$$\begin{cases}\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t+1}\right)\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}+\alpha_i+\rho\right)=\Delta_{ir}^{t+1}+\alpha_i\left(\hat{u}_{ir}^{t}-u_{ir}^{t}\right)+d_{ir}^{t}\\ \left(\hat{v}_{jr}^{t}-\hat{v}_{jr}^{t+1}\right)\left(\sum\limits_{y_{ijk}\in\Lambda(j)}\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}+\beta_j+\rho\right)=\Delta_{jr}^{t+1}+\beta_j\left(\hat{v}_{jr}^{t}-v_{jr}^{t}\right)+h_{jr}^{t}\\ \left(\hat{w}_{kr}^{t}-\hat{w}_{kr}^{t+1}\right)\left(\sum\limits_{y_{ijk}\in\Lambda(k)}\hat{u}_{ir}^{t}\tilde{v}_{jr}^{t}+\delta_k+\rho\right)=\Delta_{kr}^{t+1}+\delta_k\left(\hat{w}_{kr}^{t}-w_{kr}^{t}\right)+l_{kr}^{t}.\end{cases} \quad \text{(S31)}$$

Note that (S27b) holds via (S27a), (S29), and (S31). Further, the following inference holds by applying the partial derivation of $L_p^{\#}$ to u_{ir}, v_{jr}, and w_{kr}:

$$\begin{cases}\dfrac{\partial L_p^{\#}}{\partial u_{ir}}=-\alpha_i\left(\hat{u}_{ir}-u_{ir}+\dfrac{d_{ir}}{\alpha_i}\right)-g_{ir}=0\\ \dfrac{\partial L_p^{\#}}{\partial v_{jr}}=-\beta_j\left(\hat{v}_{jr}-v_{jr}+\dfrac{h_{jr}}{\beta_j}\right)-m_{jr}=0\\ \dfrac{\partial L_p^{\#}}{\partial w_{kr}}=-\delta_k\left(\hat{t}_{kr}-w_{kr}+\dfrac{l_{kr}}{\delta_k}\right)-p_{kr}=0\end{cases}\Rightarrow\begin{cases}g_{ir}=-\alpha_i\left(\hat{u}_{ir}-u_{ir}+\dfrac{d_{ir}}{\alpha_i}\right)\\ m_{jr}=-\beta_j\left(\hat{v}_{jr}-v_{jr}+\dfrac{h_{jr}}{\beta_j}\right)\\ p_{kr}=-\gamma_k\left(\hat{w}_{kr}-w_{kr}+\dfrac{l_{kr}}{\delta_k}\right).\end{cases} \quad \text{(S32)}$$

With (S32), g_{ir}, m_{jr}, and p_{kr} are implicitly updated and generate their limits g_{ir}^*, m_{jr}^*, p_{kr}^*. Considering (S26)'s KKT conditions that $\forall g_{ir},u_{ir}:g_{ir}u_{ir}=0$, $\forall m_{jr},v_{jr}:m_{jr}v_{jr}=0$, and $\forall p_{kr},w_{kr}:p_{kr}w_{kr}=0$, we have:

$$\begin{cases}-\alpha_i u_{ir}\left(\hat{u}_{ir}-u_{ir}+\dfrac{d_{ir}}{\alpha_i}\right)=0\\ -\beta_j v_{jr}\left(\hat{v}_{jr}-v_{jr}+\dfrac{h_{jr}}{\beta_j}\right)=0\\ -\delta_k w_{kr}\left(\hat{w}_{kr}-w_{kr}+\dfrac{l_{kr}}{\delta_k}\right)=0\end{cases}\Rightarrow\begin{cases}u_{ir}=\hat{u}_{ir}+\dfrac{d_{ir}}{\alpha_i}\\ v_{jr}=\hat{v}_{jr}+\dfrac{h_{jr}}{\beta_j}\\ w_{kr}=\hat{w}_{kr}+\dfrac{l_{kr}}{\delta_k}.\end{cases} \quad \text{(S33)}$$

Hence, we can update u_{ir}, v_{jr}, and w_{kr} by (S33). Additionally, the nonnegative truncation is applied to u_{ir}, v_{jr}, and w_{kr} to ensure its nonnegativity and (S26)'s KKT conditions, it's given as:

$$u_{ir}=\max\left(0,\hat{u}_{ir}+\frac{d_{ir}}{\alpha_i}\right); v_{jr}=\max\left(0,\hat{v}_{jr}+\frac{h_{jr}}{\beta_j}\right); w_{kr}=\max\left(0,\hat{w}_{kr}+\frac{l_{kr}}{\delta_k}\right). \quad \text{(S34)}$$

Note that the update rules of (14) and (S34) are equivalent. When $t\to\infty$, (S27c)-(S27e) are hold via (14) and (S32)-(S34).

Hence, considering (S27e)'s conditions that $g_{ir}^* \geq 0$, $m_{jr}^* \geq 0$, and $p_{kr}^* \geq 0$, there are two cases as:

- If $u_{ir}^* = 0$, $v_{jr}^* = 0$, and $w_{kr}^* = 0$, the following inequality holds according to (14):

$$\hat{u}_{ir}^* + \frac{d_{ir}^*}{\alpha_i} \leq 0; \hat{v}_{jr}^* + \frac{h_{jr}^*}{\beta_j} \leq 0; \hat{w}_{kr}^* + \frac{l_{kr}^*}{\delta_k} \leq 0, \tag{S35}$$

which indicates that $g_{ir}^* \geq 0$, $m_{jr}^* \geq 0$, and $p_{kr}^* \geq 0$ by collectively analyzing (S32);

- If $u_{ir}^* > 0$, $v_{jr}^* > 0$, and $w_{kr}^* > 0$, the following equality is inferred according to (14):

$$u_{ir}^* = \hat{u}_{ir}^* + \frac{d_{ir}^*}{\alpha_i}; v_{jr}^* = \hat{v}_{jr}^* + \frac{h_{jr}^*}{\beta_j}; w_{kr}^* = \hat{w}_{kr}^* + \frac{l_{kr}^*}{\delta_k}. \tag{S36}$$

With (S32) and (S36), $g_{ir}^* = 0$, $m_{jr}^* = 0$, and $p_{kr}^* = 0$. Therefore, (S27f) holds and ***Theorem*** **1** holds. Based on the above inferences, the implemented steps 1-3 demonstrate that the PNL's convergence is theoretically guaranteed.

III. Additional Procedures

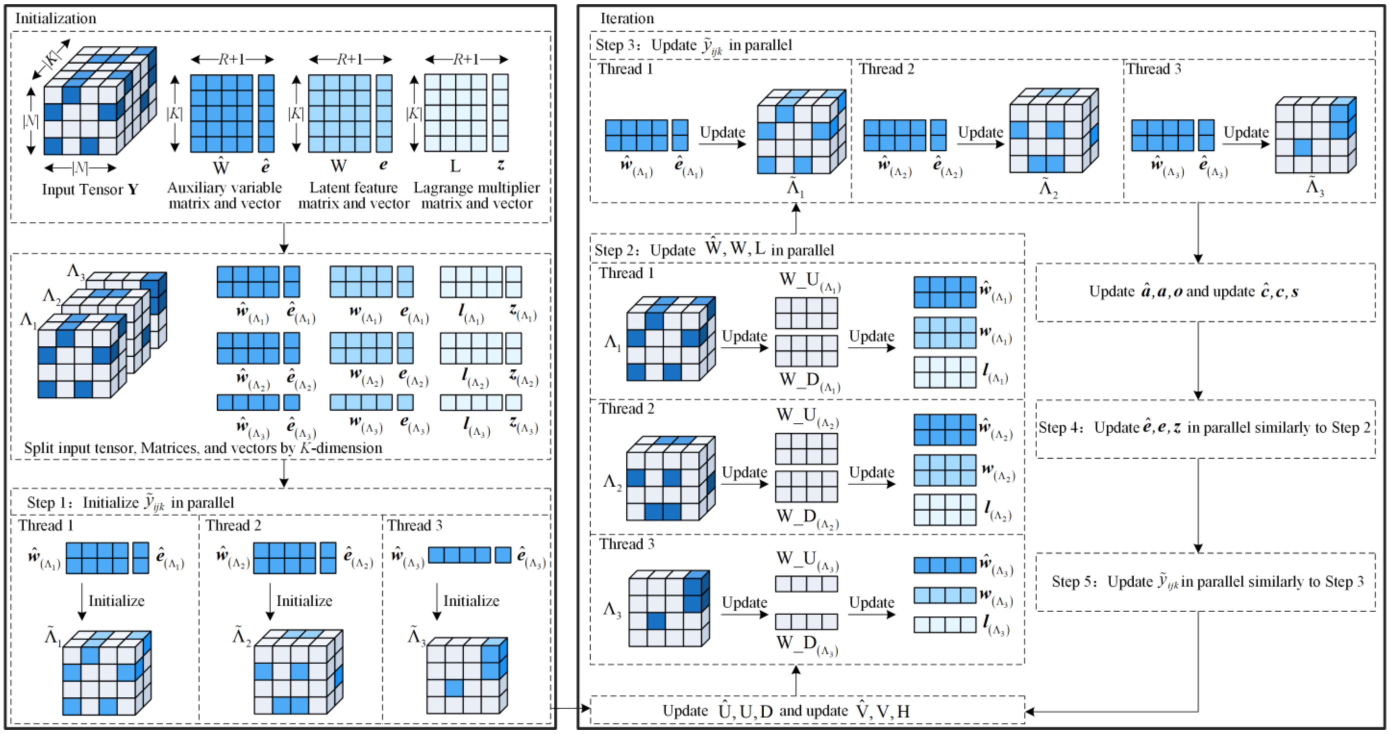


Fig. S1. An example of three threads for PNL's parallel design.

Procedure: Parallel_update_$\hat{\mathrm{W}}$

Input: $\hat{\boldsymbol{u}}_r, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{w}}_r, \boldsymbol{u}_r, \boldsymbol{d}_r, \boldsymbol{\tau}_{(p)}, \Lambda, K, Q$

Output: Updated $\hat{\mathrm{W}}$, W, L

Operation	Cost
1. Init $\mathrm{W\_U}^{\lvert K\rvert\times R}, \mathrm{W\_D}^{\lvert K\rvert\times R}=0$	$\Theta(2\times N\times R)$
2. for each $q\in Q$ **Parallelization**	$\times Q$
3. for each r=1 **to** R **do**	$\times R$
4. for each $y_{ijk}\in\Lambda_q$	$\times\lvert\Lambda_q\rvert$
5. $err = y_{ijk}-\tilde{y}_{ijk}$	$\Theta(1)$
6. $\mathrm{W\_U}_{kr} += \hat{u}_{ir}\hat{v}_{jr}(err+\hat{u}_{ir}\hat{v}_{jr}\hat{w}_{kr})$	$\Theta(1)$
7. $\mathrm{W\_D}_{kr} += (\hat{u}_{ir}\hat{v}_{jr})^2$	$\Theta(1)$
8. for each $k\in K_q$	$\times\lvert K_q\rvert$
9. $\hat{w}_{kr}=\dfrac{\mathrm{W\_U}_{kr}+\lambda_{(p)}\lvert\Lambda_{(k)}\rvert w_{kr}-l_{kr}+\rho_{(p)}\hat{w}_{kr}^{t}}{\mathrm{W\_D}_{kr}+\lambda_{(p)}\lvert\Lambda_{(k)}\rvert+\rho_{(p)}}$	$\Theta(1)$
10. $w_{kr}=\max\left(0,\hat{w}_{kr}+\dfrac{l_{kr}}{\lambda_{(p)}\lvert\Lambda_{(k)}\rvert}\right)$	$\Theta(1)$
11. $l_{kr} += \eta_{(p)}\lambda_{(p)}\lvert\Lambda_{(k)}\rvert(\hat{w}_{kr}-w_{kr})$	$\Theta(1)$

Procedure: Parallel_update_$\hat{e}$	
Input: $\hat{\boldsymbol{e}}, \boldsymbol{e}, \boldsymbol{z}, \boldsymbol{\tau}_{(p)}, \Lambda, K, Q$	
Output: Updated $\hat{\boldsymbol{e}}$, $\boldsymbol{e}$, $\boldsymbol{z}$	
Operation	**Cost**
1. Init $\mathrm{E\_U}^{\vert N\vert} = 0$	$\Theta(N)$
2. for each $q \in Q$ **Parallelization**	$\times Q$
3. **for each** $y_{ijk} \in \Lambda_q$	$\times\vert\Lambda_q\vert$
4. $err = y_{ijk} - \tilde{y}_{ijk}$	$\Theta(1)$
5. $\mathrm{E\_U}_k += err + \hat{e}_k$	$\Theta(1)$
6. **for each** $k \in K_q$	$\times\vert K_q\vert$
7. $\hat{e}_k = \dfrac{\mathrm{E\_U}_k + \lambda_{(p)}\vert\Lambda_{(k)}\vert e_k - z_k + \rho_{(p)}\hat{e}_k^t}{\vert\Lambda_{(k)}\vert + \lambda_{(p)}\vert\Lambda_{(k)}\vert + \rho_{(p)}}$	$\Theta(1)$
9. $e_k = \max(0, \hat{e}_k + \dfrac{z_k}{\lambda_{(p)}\vert\Lambda_{(k)}\vert})$	$\Theta(1)$
10. $z_k += \eta_{(p)}\lambda_{(p)}\vert\Lambda_{(k)}\vert(\hat{e}_k - e_k)$	$\Theta(1)$

IV. EXPERIMENTAL RESULTS

TABLE S1

P^2SO PARAMETERS SETTINGS

Parameters	Settings
P	5
ω	0.724
c_1	2
c_2	2
r_1	random number $\in [0, 1]$
r_2	random number $\in [0, 1]$
μ	1
$[\breve{\rho}, \widehat{\rho}]$	[10, 100]
$[\breve{\lambda}, \widehat{\lambda}]$	[0.1, 1]
$[\breve{\eta}, \widehat{\eta}]$	[0.1, 1]
$[\breve{\upsilon}_\rho, \widehat{\upsilon}_\rho]$	$[-0.2\times(\widehat{\rho}-\breve{\rho}), 0.2\times(\widehat{\rho}-\breve{\rho})]$
$[\breve{\upsilon}_\lambda, \widehat{\upsilon}_\lambda]$	$[-0.2\times(\widehat{\lambda}-\breve{\lambda}), 0.2\times(\widehat{\lambda}-\breve{\lambda})]$
$[\breve{\upsilon}_\eta, \widehat{\upsilon}_\eta]$	$[-0.2\times(\widehat{\eta}-\breve{\eta}), 0.2\times(\widehat{\eta}-\breve{\eta})]$

TABLE S2

DATASET DETAILS

Datasets	**Nodes**	**Time Slots**	**Density**	**Entries**
D1	16428	124	6.85×10^{-7}	22938
D2	36055	149	3.06×10^{-7}	59334
D3	60500	168	1.49×10^{-7}	91978
D4	90971	248	7.85×10^{-8}	161136
D5	98022	360	6.08×10^{-8}	210354
D6	240114	496	1.99×10^{-8}	570820
D7	426840	620	7.88×10^{-9}	891063
D8	460891	720	7.25×10^{-9}	1109218

TABLE S3

THE OPTIMAL HYPER-PARAMETERS BY MANUAL TUNING

Datasets	Optimal Hyper-parameters		
	ρ	λ	η
D1	90	0.3	0.2
D2	100	0.3	0.1
D3	90	0.9	0.1
D4	100	0.1	0.3
D5	80	1.0	0.1
D6	100	0.1	0.3
D7	80	0.4	0.2
D8	90	0.2	0.3

TABLE S4

THE PERFORMANCE OF PNL WITH HYPER-PARAMETER ADAPTATION AND MANUAL TUNING

Datasets		Estimation Accuracy			Iteration Count			Time Cost (Secs)	
		Adaptive	Manual		Adaptive	Manual		Adaptive	Manual
D1	RMSE	**1.0694±0.0156**	1.0711±0.0013	I.RMSE*	**2±0**	10±0	T.RMSE**	**2.756±0.006**	1267.168
	MAE	**0.8760±0.0095**	0.8793±0.0024	I.MAE	**3±0**	14±1	T.MAE	**3.113±0.010**	1744.036
D2	RMSE	**0.7703±0.0030**	0.7736±0.0010	I.RMSE	**8±2**	40±1	T.RMSE	**7.568±1.178**	6200.060
	MAE	**0.6091±0.0020**	0.6097±0.0011	I.MAE	**20±2**	113±4	T.MAE	**19.39±1.343**	17515.170
D3	RMSE	**0.8326±0.0014**	0.8345±0.0007	I.RMSE	**6±1**	31±1	T.RMSE	**11.10±1.578**	5798.217
	MAE	**0.6789±0.0011**	0.6792±0.0004	I.MAE	**17±1**	90±2	T.MAE	**28.07±1.566**	16833.534
D4	RMSE	**0.7583±0.0013**	0.7615±0.0004	I.RMSE	**6±1**	33±1	T.RMSE	**15.83±2.833**	7003.429
	MAE	**0.6113±0.0007**	0.6131±0.0003	I.MAE	**12±2**	71±2	T.MAE	**31.51±3.400**	15067.985
D5	RMSE	**0.7064±0.0011**	0.7083±0.0005	I.RMSE	**9±2**	34±1	T.RMSE	**27.06±6.592**	12229.457
	MAE	**0.5357±0.0008**	0.5359±0.0005	I.MAE	**21±1**	94±2	T.MAE	**64.92±7.405**	33810.853
D6	RMSE	**0.6724±0.0021**	0.6750±0.0003	I.RMSE	**8±2**	54±1	T.RMSE	**60.85±7.778**	63636.940
	MAE	**0.4918±0.0007**	0.4928±0.0002	I.MAE	**30±2**	138±2	T.MAE	**221.6±12.62**	16267.738
D7	RMSE	**0.6019±0.0014**	0.6020±0.0002	I.RMSE	**8±2**	38±1	T.RMSE	**122.2±13.01**	120227.394
	MAE	**0.3966±0.0009**	0.3975±0.0002	I.MAE	**31±3**	119±2	T.MAE	**486.9±21.63**	397994.657
D8	RMSE	**0.6919±0.0020**	0.6940±0.0002	I.RMSE	**6±1**	40±0	T.RMSE	**101.2±15.55**	171180.498
	MAE	**0.4693±0.0006**	0.4706±0.0001	I.MAE	**20±3**	93±1	T.MAE	**309.2±26.89**	397994.657

* Iteration Count in RMSE and also applies to the MAE; ** Time Cost in RMSE and also applies to the MAE.

TABLE S5

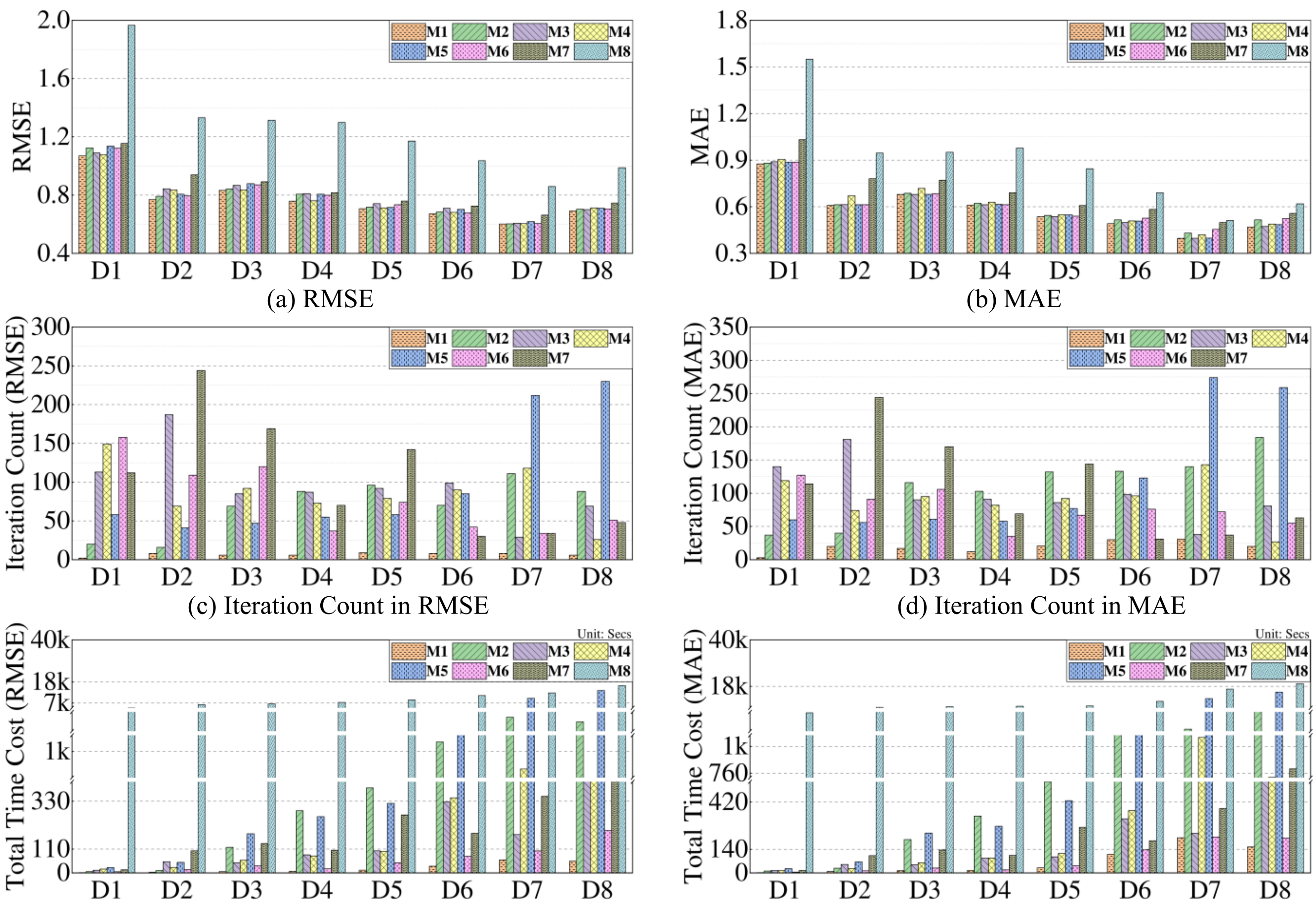
MODEL DETAILS

Models	Descriptions
M1	The proposed model in this paper.
M2	A multi-dimensional tensor decomposition model [14] based on the CPD framework that employs the gradient descent algorithms and alternating least square to train the desired latent features.
M3	A robust tensor model [15] that evaluates the gap between the predicted and true values via Cauchy Loss and provides accurate predictions.
M4	A biased nonnegative tensor factorization model that incorporates biases into the CPD framework and uses a nonnegative update method to maintain data nonnegativity.
M5	An integrated multi-linear algebra model [16] that achieves effective estimation by combining tensor decomposition and reconstruction optimization algorithms.
M6	An Adam-incorporated LFT model that adopts the Adam algorithm to train the desired model parameters.
M7	A temporal-aware latent factor analysis model, its objective function combines temporal effects to effectively describes temporal patterns of dynamic data. Its input is a set of matrices derived from a three-order tensor split by the temporal dimension.
M8	A graph convolutional network (GCN) model [17], which uses a light graph convolutional layer to adjust the normalization process of neighborhood aggregation, thereby achieving a balance between model accuracy and novelty Moreover, it gets graph slices by splitting the DCTN along the temporal dimension and considers them as inputs.

TABLE S6

HYPER-PARAMETERS SETTING OF M1-8

Datasets	M1	M2	M3	M4	M5	M6	M7	M8
D1	P^2SO	$\eta=5\times10^{-2}$	$\eta=10^{-5}$	$\lambda=10^{-2}$	$\eta=5\times10^{-3}$	$\alpha=5\times10^{-3}, \beta_1=0.7$	$\eta=10^{-2}$	$K=3, \eta=1, \lambda=0.5$
	$Q=16$	$\lambda=10^{-2}$	$\gamma=40$	$\lambda_b=10^{-2}$	$\lambda=5\times10^{-3}$	$\beta_2=0.999$	$\lambda=10^{-2}$	mini-batch size=2048
D2	P^2SO	$\eta=5\times10^{-2}$	$\eta=10^{-6}$	$\lambda=10^{-2}$	$\eta=5\times10^{-3}$	$\alpha=5\times10^{-3}\ \beta_1=0.7$	$\eta=10^{-1}$	$K=3, \eta=1, \lambda=1$
	$Q=16$	$\lambda=10^{-2}$	$\gamma=80$	$\lambda_b=10^{-2}$	$\lambda=5\times10^{-3}$	$\beta_2=0.999$	$\lambda=10^{-1}$	mini-batch size=2048
D3	P^2SO	$\eta=10^{-2}$	$\eta=10^{-6}$	$\lambda=0.5$	$\eta=5\times10^{-3}$	$\alpha=5\times10^{-3}, \beta_1=0.7$	$\eta=10^{-1}$	$K=3, \eta=1, \lambda=10^{-1}$
	$Q=16$	$\lambda=10^{-1}$	$\gamma=80$	$\lambda_b=10^{-1}$	$\lambda=10^{-2}$	$\beta_2=0.999$	$\lambda=10^{-1}$	mini-batch size=2048
D4	P^2SO	$\eta=10^{-2}$	$\eta=10^{-5}$	$\lambda=0.5$	$\eta=5\times10^{-3}$	$\alpha=5\times10^{-2}, \beta_1=0.7$	$\eta=10^{-1}$	$K=3, \eta=1, \lambda=1$
	$Q=16$	$\lambda=10^{-1}$	$\gamma=20$	$\lambda_b=10^{-1}$	$\lambda=10^{-2}$	$\beta_2=0.999$	$\lambda=5\times10^{-2}$	mini-batch size=2048
D5	P^2SO	$\eta=10^{-2}$	$\eta=10^{-5}$	$\lambda=0.5$	$\eta=5\times10^{-3}$	$\alpha=5\times10^{-2}, \beta_1=0.5$	$\eta=10^{-1}$	$K=3, \eta=5\times10^{-2}, \lambda=5\times10^{-2}$
	$Q=16$	$\lambda=10^{-1}$	$\gamma=40$	$\lambda_b=10^{-1}$	$\lambda=5\times10^{-3}$	$\beta_2=0.999$	$\lambda=5\times10^{-2}$	mini-batch size=2048
D6	P^2SO	$\eta=10^{-2}$	$\eta=10^{-5}$	$\lambda=0.5$	$\eta=5\times10^{-3}$	$\alpha=5\times10^{-2}, \beta_1=0.5$	$\eta=10^{-1}$	$K=3, \eta=1, \lambda=1$
	$Q=16$	$\lambda=0.5$	$\gamma=40$	$\lambda_b=10^{-1}$	$\lambda=5\times10^{-3}$	$\beta_2=0.999$	$\lambda=5\times10^{-3}$	mini-batch size=2048
D7	P^2SO	$\eta=10^{-2}$	$\eta=10^{-5}$	$\lambda=10^{-1}$	$\eta=10^{-3}$	$\alpha=5\times10^{-3}, \beta_1=0.7$	$\eta=10^{-1}$	$K=3, \eta=10^{-1}, \lambda=10^{-1}$
	$Q=16$	$\lambda=10^{-1}$	$\gamma=40$	$\lambda_b=5\times10^{-2}$	$\lambda=10^{-3}$	$\beta_2=0.999$	$\lambda=5\times10^{-3}$	mini-batch size=2048
D8	P^2SO	$\eta=5\times10^{-3}$	$\eta=10^{-5}$	$\lambda=0.5$	$\eta=10^{-3}$	$\alpha=5\times10^{-3}, \beta_1=0.9$	$\eta=10^{-1}$	$K=3, \eta=10^{-1}, \lambda=10^{-1}$
	$Q=16$	$\lambda=10^{-3}$	$\gamma=40$	$\lambda_b=0.5$	$\lambda=10^{-3}$	$\beta_2=0.999$	$\lambda=5\times10^{-3}$	mini-batch size=2048



(a) RMSE (b) MAE

(c) Iteration Count in RMSE (d) Iteration Count in MAE

(e) Time Cost in RMSE (f) Time Cost in MAE

Fig. **S2**. Performance comparison of M1-8 on D1-8.

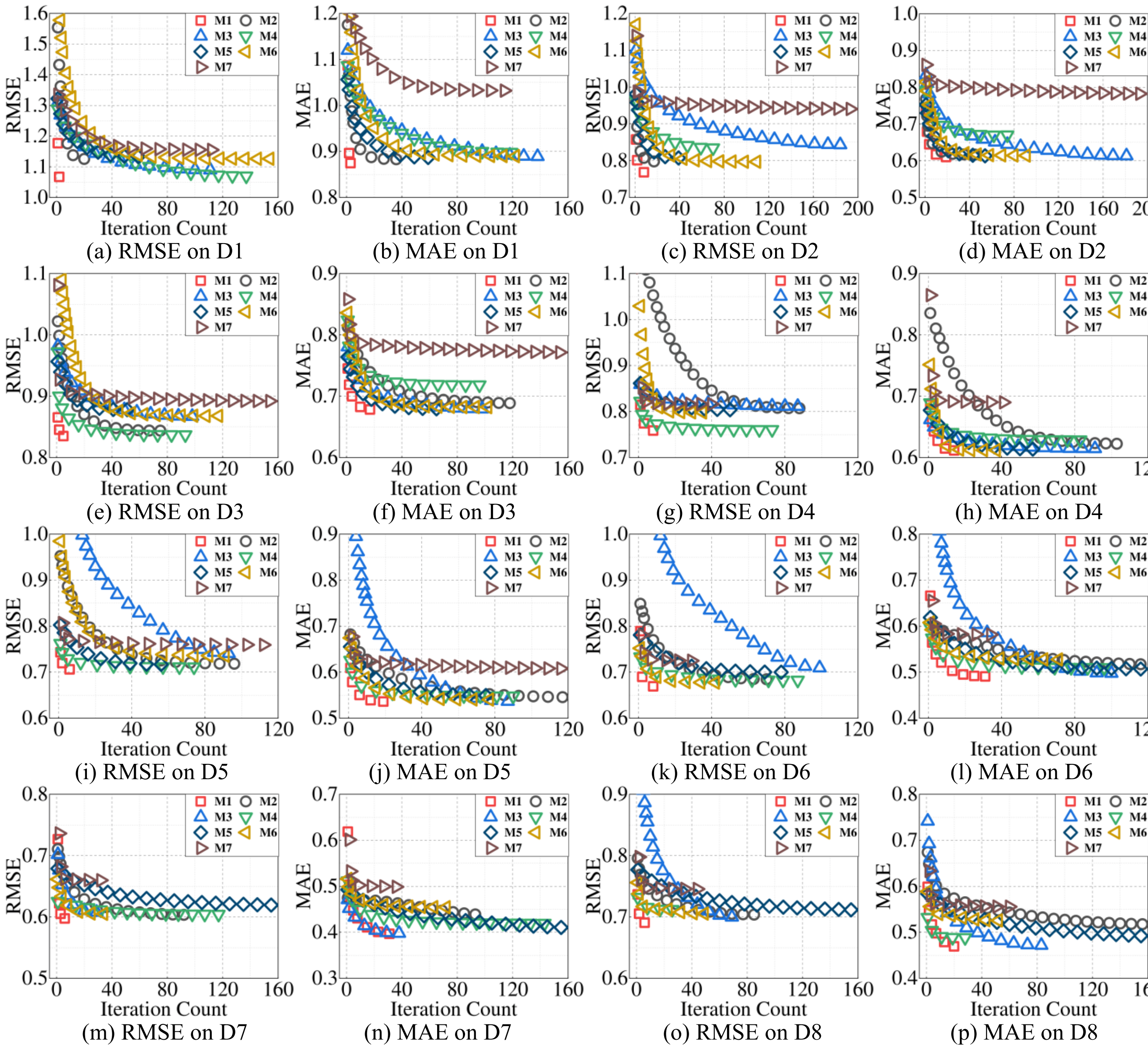


Fig. S3. Training curves of M1-7 in RMSE and MAE on D1-8.

TABLE S7

ESTIMATION ACCURACY OF ALL MODELS FOR RMSE AND MAE ON D1-8, INCLUDING WIN/LOSS COUNTS AND FRIEDMAN TEST

Datasets		M1	M2	M3	M4	M5	M6	M7	M8
D1	RMSE	**1.0694±0.0156**	1.1241±0.0017	1.0895±0.0005	1.0780±0.0008	1.1363±0.0044	1.1222±0.0011	1.1549±0.0001	1.9658±0.0008
	MAE	**0.8760±0.0095**	0.8811±0.0016	0.8893±0.0008	0.9051±0.0008	0.8867±0.0047	0.8869±0.0010	1.0320±0.0001	1.5500±0.0008
D2	RMSE	**0.7703±0.0030**	0.7922±0.0021	0.8431±0.0007	0.8352±0.0004	0.8067±0.0031	0.7942±0.0014	0.9391±0.0006	1.3308±0.0004
	MAE	**0.6091±0.0020**	0.6134±0.0016	0.6136±0.0006	0.6705±0.0002	0.6131±0.0021	0.6127±0.0013	0.7801±0.0007	0.9453±0.0004
D3	RMSE	**0.8326±0.0014**	0.8422±0.0014	0.8672±0.0007	0.8363±0.0006	0.8783±0.0022	0.8701±0.0007	0.8918±0.0004	1.3138±0.0002
	MAE	**0.6789±0.0011**	0.6871±0.0012	0.6796±0.0005	0.7179±0.0005	0.6801±0.0013	0.6840±0.0007	0.7710±0.0008	0.9504±0.0001
D4	RMSE	**0.7583±0.0013**	0.8065±0.0008	0.8095±0.0008	0.7611±0.0004	0.8066±0.0019	0.7994±0.0005	0.8166±0.0007	1.2983±0.0002
	MAE	**0.6113±0.0007**	0.6220±0.0006	0.6131±0.0006	0.6293±0.0003	0.6153±0.0009	0.6132±0.0006	0.6900±0.0010	0.9777±0.0002
D5	RMSE	**0.7064±0.0011**	0.7173±0.0005	0.7410±0.0014	0.7120±0.0007	0.7154±0.0014	0.7343±0.0005	0.7596±0.0005	1.1713±0.0001
	MAE	**0.5357±0.0008**	0.5442±0.0007	0.5366±0.0010	0.5485±0.0004	0.5475±0.0015	0.5388±0.0005	0.6076±0.0009	0.8449±0.0001
D6	RMSE	**0.6724±0.0021**	0.6844±0.0004	0.7107±0.0008	0.6814±0.0004	0.7019±0.0008	0.6764±0.0006	0.7249±0.0014	1.0361±0.0001
	MAE	**0.4918±0.0007**	0.5155±0.0004	0.4985±0.0006	0.5087±0.0003	0.5075±0.0006	0.5265±0.0005	0.5891±0.0021	0.6903±0.0001
D7	RMSE	**0.6019±0.0014**	0.6027±0.0005	0.6072±0.0002	0.6042±0.0001	0.6188±0.0006	0.6053±0.0003	0.6625±0.0049	0.8585±0.0001
	MAE	**0.3966±0.0009**	0.4309±0.0041	0.3975±0.0002	0.4195±0.0001	0.3986±0.0020	0.4543±0.0007	0.4985±0.0098	0.5120±0.0001
D8	RMSE	**0.6916±0.0020**	0.7035±0.0003	0.7005±0.0004	0.7119±0.0001	0.7106±0.0007	0.7051±0.0004	0.7461±0.0017	0.9884±0.0001
	MAE	**0.4693±0.0006**	0.5152±0.0004	0.4730±0.0006	0.4881±0.0001	0.4847±0.0005	0.5251±0.0005	0.5569±0.0026	0.6190±0.0001
Win/Loss		-*	16/0	16/0	16/0	16/0	16/0	16/0	16/0
Rank		**1.00**	3.88	3.75	4.19	4.19	4.00	7.00	8.00

* Not involved

TABLE S8

ITERATION COUNT OF ALL MODELS FOR RMSE AND MAE ON D1-8, INCLUDING WIN/LOSS COUNTS AND FRIEDMAN TEST

Datasets		M1	M2	M3	M4	M5	M6	M7	M8
D1	I.RMSE	**2±0**	20±0	113±2	149±1	58±3	158±3	112±0	-
	I.MAE	**3±0**	37±2	140±2	119±1	60±2	127±5	114±0	-
D2	I.RMSE	**8±2**	16±1	187±2	69±1	41±2	109±4	244±22	-
	I.MAE	**20±2**	40±3	181±2	74±1	56±2	91±4	244±22	-
D3	I.RMSE	**6±1**	69±4	85±2	92±1	46±3	120±3	169±24	-
	I.MAE	**17±1**	116±2	90±2	95±1	61±2	106±12	170±24	-
D4	I.RMSE	**6±1**	88±1	87±2	73±3	55±2	37±3	70±27	-
	I.MAE	**12±2**	103±1	91±2	82±2	58±2	35±7	69±29	-
D5	I.RMSE	**9±2**	96±2	92±2	79±3	58±2	74±3	142±22	-
	I.MAE	**21±1**	132±2	86±2	92±2	77±2	67±5	144±22	-
D6	I.RMSE	**8±2**	70±2	99±1	90±3	85±6	42±2	30±6	-
	I.MAE	**30±2**	133±1	98±1	96±2	123±3	76±3	31±7	-
D7	I.RMSE	**8±2**	111±11	29±1	118±1	212±6	34±1	34±18	-
	I.MAE	**31±3**	140±28	38±2	143±1	274±10	72±4	37±24	-
D8	I.RMSE	**6±1**	88±3	69±1	26±1	230±8	51±2	48±10	-
	I.MAE	**20±3**	184±2	81±1	27±1	259±3	55±2	63±15	-
Win/Loss		-	16/0	16/0	16/0	16/0	16/0	16/0	-
Rank		**1.00**	4.75	5.06	4.56	4.13	3.75	4.75	-

TABLE S9

TIME COST OF ALL MODELS FOR RMSE AND MAE ON D1-8, INCLUDING WIN/LOSS COUNTS AND FRIEDMAN TEST (SECS)

Datasets		M1	M2	M3	M4	M5	M6	M7	M8
D1	T.RMSE	**1.178±0.146**	5.892±0.016	12.10±0.043	17.57±0.092	24.20±0.207	5.702±0.053	14.40±0.057	4533±11.63
	T.MAE	**1.398±0.247**	11.05±0.033	15.00±0.054	14.09±0.075	25.24±0.227	4.571±0.042	14.65±0.058	4727±12.74
D2	T.RMSE	**3.216±0.654**	11.41±0.041	51.76±0.078	25.19±0.056	48.85±0.273	14.97±0.105	101.8±0.464	6188±12.97
	T.MAE	**8.493±0.735**	27.84±0.148	50.21±0.074	26.85±0.059	65.91±0.352	12.43±0.090	101.9±0.465	8081±16.83
D3	T.RMSE	**5.788±0.641**	117.7±3.072	46.55±0.099	58.54±0.123	180.1±0.608	32.36±0.101	135.3±0.801	6580±15.72
	T.MAE	**13.44±0.678**	197.6±5.311	48.95±0.112	60.48±0.118	235.7±0.574	28.77±0.151	136.0±0.802	8319±17.88
D4	T.RMSE	**7.256±1.149**	286.8±0.500	82.89±0.120	78.04±0.170	259.4±0.501	18.56±0.076	104.6±1.615	7343±57.19
	T.MAE	**15.09±1.354**	336.9±0.629	87.29±0.110	88.49±0.142	276.3±0.408	17.65±0.150	103.6±1.732	8527±53.94
D5	T.RMSE	**12.46±1.139**	391.4±1.089	101.8±0.128	99.03±0.229	320.2±0.497	46.87±0.188	267.0±2.256	8712±15.26
	T.MAE	**30.26±1.858**	539.3±1.554	95.35±0.132	115.4±0.201	427.8±0.433	42.06±0.174	270.3±2.254	8669±15.71
D6	T.RMSE	**31.11±4.917**	1058±6.459	326.8±0.353	346.0±0.536	1723±4.288	76.37±0.214	182.4±1.304	10919±57.35
	T.MAE	**110.2±7.594**	2003±10.92	320.9±0.353	369.0±0.443	2485±2.376	136.3±0.422	190.7±1.510	10845±72.71
D7	T.RMSE	**59.52±10.86**	2744±9.479	176.6±0.280	893.4±8.890	9437±122.7	101.1±0.234	352.9±7.289	12291±21.23
	T.MAE	**208.6±14.41**	3455±25.85	235.3±0.410	1078±10.82	12154±158.5	212.1±0.679	381.3±9.851	16648±23.64
D8	T.RMSE	**53.77±9.566**	2696±8.711	489.1±0.282	677.2±7.767	13547±200.0	195.3±0.438	604.8±4.886	6129±39.30
	T.MAE	**153.7±13.08**	5655±15.23	575.4±0.422	724.3±8.340	15250±217.8	207.1±0.473	799.7±7.350	19236±49.87
Win/Loss		-	16/0	16/0	16/0	16/0	16/0	16/0	16/0
Rank		1.00	5.44	3.81	4.25	6.56	2.06	4.88	8.00

TABLE S10

WILCOXON SIGNED RANK TEST RESULTS OF M1 VERSUS M2-8 WITH THE SIGNIFICANCE LEVEL EQUAL TO 0.05

Model Comparison	Estimation Accuracy			Iteration Count			Time Cost		
	w^+*	w^-	*p*-value**	w^+	w^-	p-value	w^+	w^-	*p*-value
M1 vs M2	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**
M1 vs M3	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**
M1 vs M4	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**
M1 vs M5	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**
M1 vs M6	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**
M1 vs M7	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**
M1 vs M8	136	0	**0.000**	136	0	**0.000**	136	0	**0.000**

* w^+ denotes the counts that M1 outperforms its peers, and w^- denotes the opposite; ** *p*-value denotes a probability value, i.e., the hypothesis of M1 outperforms its peers can be accepted when *p*-value<0.05.

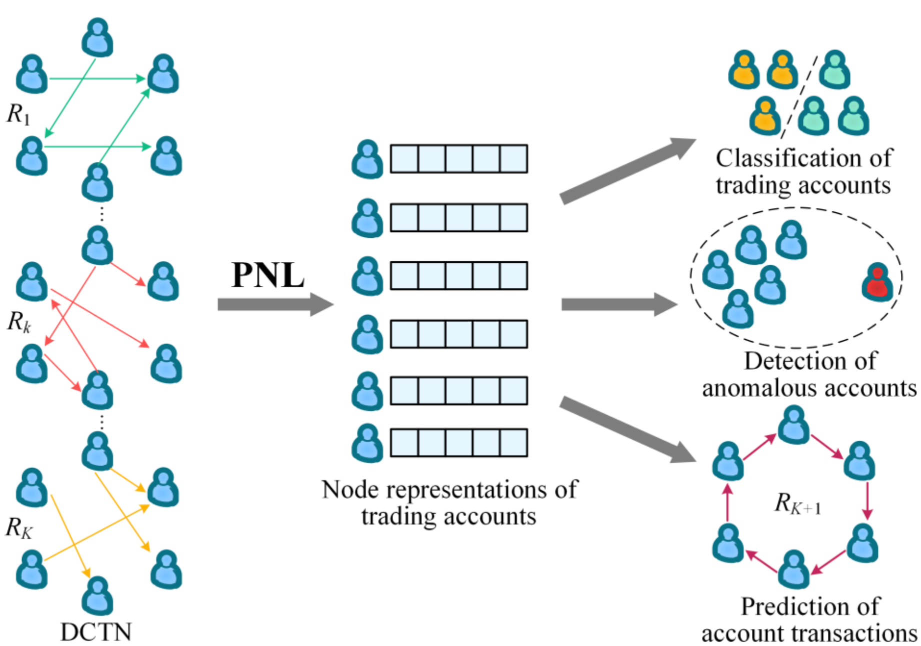


Fig. S4. Industrial Application Instances of PNL Implementing DCTNE.

V. Related Work

So far, researchers developed various DCTNE methods for the extraction of the desired knowledge, such as dynamic random walk models, Recurrent Neural Network (RNN) model, GCN models, etc. Specifically, Lin *et al.* [18] introduce a temporal weighted multi-graph embedding model, employing the random walk approach with temporal walk and edge sampling strategies to efficiently capture the attributes of the trading network. Wang *et al.* [19] present a generative link sequence model, which employs a self-tokenization mechanism to capture network topology information and temporal link formation patterns. Jiao *et al.* [20] design a variational-autoencoder-incorporated dynamic network embedding model, which update node representations via a self-attention mechanism and RNN.

Li *et al.* [21] propose a dynamic GCN model to efficiently extract structural and temporal information in dynamic graphs by performing spatio-temporal convolution operations. Xie *et al.* [22] introduce a graph temporal edge aggregation approach that combines a sequence model and a temporal encoder to learn node representations via graph neural networks. Bonner et al. [23] present a temporal neighborhood aggregation model, which learns temporal evolution patterns in different vertex neighborhoods based on graph convolution to capture topological and temporal information. Dave *et al.* [24] develop an effective GraNite framework, which incorporates a time-preserving node embedding method into graphlet-based time-ordering to solve triangle completion time prediction. Wu *et al.* [25] design an encoder-predictor-decoder framework that employs a recurrent architecture Graph Neural Networks (GNN) to extract the time evolution patterns of dynamic networks.

Pareja *et al.* [26] present an EvolveGCN model, which combines RNN and GNN to acquire the dynamics in the networks. Gao *et al*. propose a network topology transformation-based representations method that flips the edge links of node pairs and learns the node representations from the original and transformed networks. Zhang *et al.* propose a multi-view fuzzy method for multi-view representation learning, which transforms the multi-view data into a high-dimensional fuzzy feature space to explore the common information between views and view-specific information, and preserves the geometric structure of the data through Laplace diagrams. Chang *et al.* [27] design a multivariate timeseries representation learning framework, which employs a CLS token strategy and the instance-contrastive tasks for representation learning to achieve efficient learning of timeseries data.

Xia *et al*. [28] propose a self-supervised framework DiscoGNN to perform graph representation learning. It first replaces some nodes and edges randomly, then pre-trains the GNN to detect and correct the replaced nodes and edges from all nodes and edges, and finally captures the similarity ranking information between graphs. Kwon *et al.* [29] propose an exact decomposition method for irregular tensors, which reduces the output size by unifying all first mode factor matrices into a single matrix and enhances its representation by using the Tucker decomposition principle, thus providing more accurate representation precision. Li *et al.* [30] introduced a novel framework based on dynamic neural dowker networks to capture higher-order topological features of dynamic directed graphs, which employs a source-sink line GNN layer to capture the neighborhood relationships between dynamic edges and utilizes a duality edge fusion mechanism to ensure the duality principle of dowker complexes.

The above models are effective for extracting knowledge in DCTNs. However, they afford the expensive computational and storage overheads, constraining the application and generalization of these models. On the contrary, the PNL model attains high computational efficiency and affordable memory overhead by employing the proximal-incorporated ADMM learning scheme and hyper-parameters adaptation approach. Consequently, the PNL model exhibits elegant handling of DCTNs compared to these DCTNE models.