Temporal Relations-Aware Nonnegative Latent Factorization of Tensors for Dynamic Directed Graph Representation Supplementary File

Hao Wu, *Member, IEEE*, Xin Luo, *Senior Member, IEEE*, and Xinbo Gao, *Fellow, IEEE*

I. INTRODUCTION

This is the supplementary file for paper entitled "*Temporal Relations-Aware Nonnegative Latent Factorization of Tensors for Dynamic Directed Graph Representation*", which presents the work flow and convergence proof of the proposed TRNL model and the experimental results on eight dynamic directed graphs.

II. WORK FLOW OF TRNL

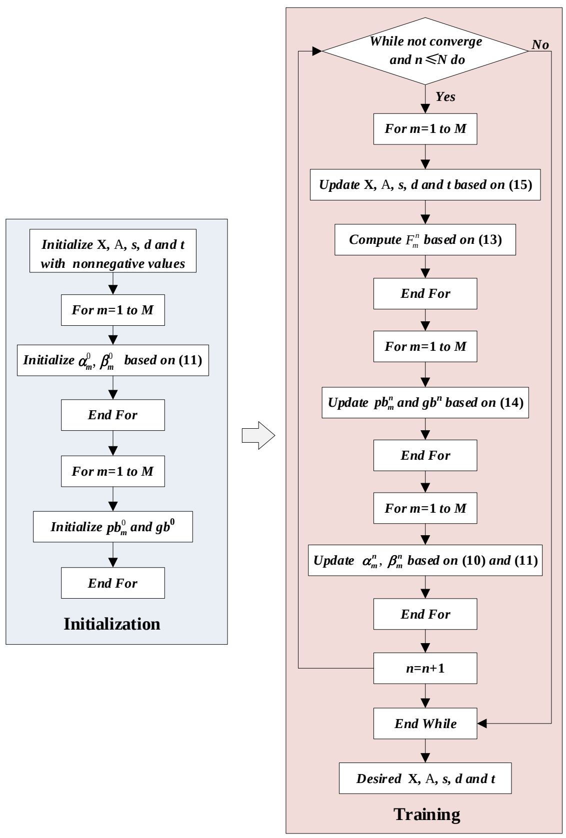

Fig. S1. The workflow of TRNL algorithm

III. CONVERGENCE PROOF OF TRNL

In order to analysis the convergence of the proposed TRNL model, it is first essential to discover the relations between the KKT condition of the learning objective (7) and the SLF-NMUT algorithm [49, 50]. Hence, considering the nonnegative constraints for the LF tensor **X**, LF matrix A and linear biases vectors ***s***, ***d***, and ***t*** we have the Lagrangian function L for (7) as:

$$L=\varepsilon\left(\mathbf{X},\mathrm{A},\boldsymbol{s},\boldsymbol{d},\boldsymbol{t}\right)-\sum_{p=1}^{R}\sum_{q=1}^{R}x_{pqk}\tilde{x}_{pqk}-\sum_{i=1}^{|V|}\sum_{p=1}^{R}a_{ip}\tilde{a}_{ip}-\sum_{j=1}^{|V|}\sum_{q=1}^{R}a_{jq}\tilde{a}_{jq}-\sum_{i=1}^{|V|}s_i\tilde{s}_i-\sum_{j=1}^{|V|}d_j\tilde{d}_j-\sum_{k=1}^{|K|}t_k\tilde{t}_k. \tag{S1}$$

where $\tilde{x}_{pqk}, \tilde{a}_{ip}, \tilde{a}_{jq}, \tilde{s}_i, \tilde{d}_j, \tilde{t}_k$ denote single element in $\tilde{\mathbf{X}}$, Ã, $\tilde{\boldsymbol{s}}$, $\tilde{\boldsymbol{d}}$, and $\tilde{\boldsymbol{t}}$ respectively, which denote Lagrangian multipliers for **X**, A, ***s***, ***d***, and ***t***.

Considering the partial derivatives of L with **X**, A, ***s***, ***d***, and ***t***, they are highly similar for a_{ip}, a_{jp} *and* s_i, d_j, and t_k. Hence, we consider the case of x_{pqk}, a_{ip} and s_i as follows:

$$\begin{cases} \dfrac{\partial L}{\partial x_{pqk}} = \sum\limits_{y_{ijk}\in\Lambda(k)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right) + \lambda\left|\Lambda(k)\right| x_{pqk} - \tilde{x}_{pqk} = 0, \\ \dfrac{\partial L}{\partial a_{ip}} = \sum\limits_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-\sum\limits_{q=1}^{R} x_{pqk}a_{jq}\right)\right) + \lambda\left|\Lambda(i)\right| a_{ip} - \tilde{a}_{ip} = 0, \\ \dfrac{\partial L}{\partial s_i} = \sum\limits_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right) + \lambda_b\left|\Lambda(i)\right| s_i - \tilde{s}_i = 0. \end{cases}$$

$$\Rightarrow \begin{cases} \tilde{x}_{pqk} = \sum\limits_{y_{ijk}\in\Lambda(k)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right) + \lambda\left|\Lambda(k)\right| x_{pqk}, \\ \tilde{a}_{ip} = \sum\limits_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-\sum\limits_{q=1}^{R} x_{pqk}a_{jq}\right)\right) + \lambda\left|\Lambda(i)\right| a_{ip}, \\ \tilde{s}_i = \sum\limits_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right) + \lambda_b\left|\Lambda(i)\right| s_i. \end{cases} \tag{S2}$$

Then, considering the KKT conditions of (S1), i.e., $\forall x_{pqk}$, $\tilde{x}_{pqk}$: $x_{pqk}\,\tilde{x}_{pqk}=0$, $\forall a_{ir}$, $\tilde{a}_{ip}$: $a_{ip}\tilde{a}_{ip}=0$, and $\forall s_i$, $\tilde{s}_i$: $s_i\tilde{s}_i=0$, we have:

$$\begin{cases} x_{pqk}\left(\sum\limits_{y_{ijk}\in\Lambda(k)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right) + \lambda\left|\Lambda(k)\right| x_{pqk}\right) = 0, \\ a_{ip}\left(\sum\limits_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-\sum\limits_{q=1}^{R} x_{pqk}a_{jq}\right)\right) + \lambda\left|\Lambda(i)\right| a_{ip}\right) = 0, \\ s_i\left(\sum\limits_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right) + \lambda_b\left|\Lambda(i)\right| s_i\right) = 0. \end{cases} \tag{S3}$$

With (S3), we can achieve the following parameters equations:

$$\begin{cases} x_{pqk}\sum\limits_{y_{ijk}\in\Lambda(k)} y_{ijk}a_{ip}a_{jq} = x_{pqk}\left(\sum\limits_{y_{ijk}\in\Lambda(k)} \hat{y}_{ijk}a_{ip}a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}\right), \\ a_{ip}\sum\limits_{y_{ijk}\in\Lambda(i)} y_{ijk}\sum\limits_{q=1}^{R} x_{pqk}a_{jq} = a_{ip}\left(\sum\limits_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk}\sum\limits_{q=1}^{R} x_{pqk}a_{jq} + \lambda\left|\Lambda(i)\right| a_{ip}\right), \\ s_i\sum\limits_{y_{ijk}\in\Lambda(i)} y_{ijk} = s_i\left(\sum\limits_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \lambda_b\left|\Lambda(i)\right| s_i\right). \end{cases}$$

$$\Rightarrow \begin{cases} x_{pqk} = x_{pqk}\dfrac{\sum\limits_{y_{ijk}\in\Lambda(k)} y_{ijk}a_{ip}a_{jq}}{\sum\limits_{y_{ijk}\in\Lambda(k)} \hat{y}_{ijk}a_{ip}a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}}, \\ a_{ip} = a_{ip}\dfrac{\sum\limits_{y_{ijk}\in\Lambda(i)} y_{ijk}\sum\limits_{q=1}^{R} x_{pqk}a_{jq}}{\sum\limits_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk}\sum\limits_{q=1}^{R} x_{pqk}a_{jq} + \lambda\left|\Lambda(i)\right| a_{ip}}, \\ s_i = s_i\dfrac{\sum\limits_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum\limits_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \lambda_b\left|\Lambda(i)\right| s_i}. \end{cases} \tag{S4}$$

In particular, with (S4), we can conveniently achieve the parameters update rule given in (8). Therefore, the SLF-NMUT-based learning scheme in the TRNL model is closely connected to the KKT conditions of its learning objective. From this point of view, we theoretically prove the convergence of FRNL in the following two steps:

Step 1: The objective function (7) is non-increasing.

Step 2: Sequences { x_{pqk}^n , a_{ip}^n , a_{jq}^n , s_i^n , d_j^n , t_k^n } converge to an equilibrium point (x_{pqk}^* , a_{ip}^* , a_{jq}^* , s_i^* , d_j^* , t_k^*).

In the Step 1, we aim to prove that objective function (7) is nonincreasing with the SLF-NMUT-based learning scheme (15). To do so, we have:

Theorem 1: (7) is nonincreasing with (15).

In particular, an auxiliary function is essential and vital to prove Theorem 1 [50]. Hence, the following function is defined:

Definition 2: $H(x, x')$ is an auxiliary function of $F(x)$ if

$$H(x, x') \geq F(x),\ H(x, x')=F(x). \tag{S5}$$

Accordingly, we further recall the following lemma [50, 51],

Lemma 1: $F(x)$ keeps nonincreasing with the following rule:

$$x^{n+1}=\mathrm{argmin}_x H(x, x'). \tag{S6}$$

Proof of Lemma 1: With Definition 2, we deduce that

$$F(x^n) = H(x^n, x^n) \geq H(x^{n+1}, x^n) \geq F(x^{n+1}). \tag{S7}$$

Note that we have $F(x^{n+1}) = F(x^n)$ when x^n guarantees a local minimum of $H(x, x^n)$. Hence, $\nabla F(x^n)=0$ holds if $F(x^n)$ is differentiable around x^n. Thus, (S7) can be extended into the following converging sequence to $x_{\min}= \mathrm{argmin}_x F(x)$:

$$F(x_{\min}) \leq \cdots \leq F(x^{n+1}) \leq F(x^n) \leq \cdots \leq F(x_1) \leq F(x_0). \tag{S8}$$

Next, we aim to achieve that (7) for is exactly consistent with that in (S6) with a specifically designed H. Considering $x_{pqk}\in\mathbf{X}$, let Fx_{pqk} be the partial loss from (7) $\varepsilon(\mathbf{X}, \mathrm{A}, \boldsymbol{s}, \boldsymbol{d}, \boldsymbol{t})$ related to x_{pqk} only,

$$F_{x_{pqk}} = \frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)^2 + \lambda\left(\sum_{p=1}^{R}\sum_{q=1}^{R}x_{pqk}^2 + \sum_{p=1}^{R}a_{ip}^2 + \sum_{q=1}^{R}a_{jp}^2\right) + \lambda_b\left(s_i^2 + d_j^2 + t_k^2\right)\right). \tag{S9}$$

As a result, the first-order and second-order derivatives of Fx_{pqk} with respect to x_{pqk} can be obtained as:

$$\begin{aligned} F'_{x_{pqk}} &= \frac{\partial\varepsilon}{\partial x_{pqk}} = \sum_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right) + \lambda\left|\Lambda(k)\right| x_{pqk}, \\ F''_{x_{pqk}} &= \frac{\partial^2\varepsilon}{\partial\left(x_{pqk}\right)^2} = \sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^2 + \lambda\left|\Lambda(k)\right|. \end{aligned} \tag{S10}$$

According to (S8)-(S10), we obtain the following proposition:

Proposition 1: The auxiliary function of Fx_{pqk} is given as:

$$H\left(x, x_{pqk}^n\right) = F_{x_{pqk}}\left(x_{pqk}^n\right) + F'_{x_{pqk}}\left(x_{pqk}^n\right)\left(x - x_{pqk}^n\right) + \frac{1}{2}\left(\left(\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}^n\right)\Big/ x_{pqk}^n\right)\left(x - x_{pqk}^n\right)^2. \tag{S11}$$

With (S11), $H(x, x)= Fx_{pqk}(x)$ holds.

Next, we prove $H\left(x, x_{pqk}^n\right) \geq F_{x_{pqk}}(x)$. To do so, the quadratic approximation to Fx_{pqk} at x_{pqk}^n needs to be first obtained as:

$$F_{x_{pqk}}(x) = F_{x_{pqk}}\left(x_{pqk}^n\right) + F'_{x_{pqk}}\left(x_{pqk}^n\right)\left(x - x_{pqk}^n\right) + \frac{1}{2}F''_{x_{pqk}}\left(x_{pqk}^n\right)\left(x - x_{pqk}^n\right)^2. \tag{S12}$$

By combine (S10)-(S12), we can see that $H\left(x, x_{pqk}^n\right)$ is an auxiliary function of Fx_{pqk} if the following inequality holds:

$$\left(\frac{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}^n}{x_{pqk}^n}\right) \geq \sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^2 + \lambda\left|\Lambda(k)\right|. \tag{S13}$$

Note that $\hat{y}_{ijk}$, a_{ip}, and a_{jq} are nonnegative, thus, (S13) is equal to

$$\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq} \geq x_{pqk}^n\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^2. \tag{S14}$$

Then, we reformulate the left term of (S14) as follows:

$$\begin{aligned} \sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq} &= \sum_{y_{ijk}\in\Lambda(k)}a_{ip}a_{jq}\left(x_{pqk}^n a_{ip}a_{jq} + \sum_{u=1,u\neq p}^{R}\sum_{v=1,v\neq q}^{R}x_{uvk}a_{iu}a_{jv} + s_i + d_j + t_k\right) \\ &= x_{pqk}^n\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^2 + \sum_{y_{ijk}\in\Lambda(k)}a_{ip}a_{jq}\left(\sum_{u=1,u\neq p}^{R}\sum_{v=1,v\neq q}^{R}x_{uvk}a_{iu}a_{jv} + s_i + d_j + t_k\right) \\ &\geq x_{pqk}^n\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^2. \end{aligned} \tag{S15}$$

Note that (S13) holds with (S15), making $H\left(x, x_{pqk}^n\right)$ is an auxiliary function of Fx_{pqk}.

Based on Proposition 1, we achieve the following proof.

Proof of theorem 1: Based on (S6), (S10) and (S11), we have:

$$
\begin{aligned}
x_{pqk}^{n+1} &= \arg\min_x H\left(x, x_{pqk}^n\right) \\
&\Rightarrow F'_{x_{pqk}}\left(x_{pqk}^n\right) + \frac{\sum_{y_{ijk}\in\Lambda(k)} \hat{y}_{ijk} a_{ip} a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}^n}{x_{pqk}^n}\left(x - x_{pqk}^n\right)^2 = 0 \\
&\Rightarrow x_{pqk}^{n+1} \leftarrow x_{pqk}^n \frac{\sum_{y_{ijk}\in\Lambda(k)} y_{ijk} a_{ip} a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)} \hat{y}_{ijk} a_{ip} a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}^n}.
\end{aligned}
\tag{S16}
$$

Based on (S8), it is clearly shown that Fx_{pqk} is nonincreasing with (15). Naturally, $\forall i, j\in V$, $k\in K$, p, $q\in\{1, \ldots, R\}$, (S16) holds. Therefore, Theorem 1 holds.

Following Theorem 1, with a positive initialization, i.e., $x_{pqk}^0 \geq 0, a_{ip}^0 \geq 0, a_{iq}^0 \geq 0, s_i^0 \geq 0, s_j^0 \geq 0, t_k^0 \geq 0$, we have the following recursion:

$$
F\left(x_{pqk}^0\right) \geq F\left(x_{pqk}^n\right) \geq H\left(x_{pqk}^{n+1}, x_{pqk}^n\right) \geq F\left(x_{pqk}^{n+1}\right) \geq 0, \tag{S17}
$$

which demonstrates that a sequence $\left\{F\left(x_{pqk}^n\right)\right\}$ is monotonically nonincreasing and bounded. Therefore, we have:

$$
\lim_{n\to+\infty}\left(F\left(x_{pqk}^{n+1}\right) - F\left(x_{pqk}^n\right)\right) = 0. \tag{S18}
$$

With (S18), we further have the following inference:

$$
\lim_{n\to+\infty}\left|x_{pqk}^{n+1} - x_{pqk}^n\right| = 0. \tag{S19}
$$

Hence, the sequence $\left\{x_{pqk}^n\right\}$ is bounded and convergent. Similarly, the sequence $\left\{a_{ip}^n\right\}$, $\left\{a_{jq}^n\right\}$, $\left\{s_i^n\right\}$, $\left\{d_j^n\right\}$, $\left\{t_k^n\right\}$ are also bounded and convergent.

In the Step 2, we aim to prove that sequences { x_{pqk}^n, a_{ip}^n, a_{jq}^n, s_i^n, d_j^n, t_k^n } obtained by SLF-NMUT-based learning scheme converge to a KKT equilibrium point (x_{pqk}^*, a_{ip}^*, a_{jq}^*, s_i^*, d_j^*, t_k^*) of (7). To prove it, we have:

Theorem 2: Sequences { x_{pqk}^n, a_{ip}^n, a_{jq}^n, s_i^n, d_j^n, t_k^n } by (15) converge to an equilibrium point (x_{pqk}^*, a_{ip}^*, a_{jq}^*, s_i^*, d_j^*, t_k^*) of ε (**X**, A, ***s***, ***d***, ***t***) in (7).

Proof of Theorem 2: From (S19) we see that a sequence $\left\{x_{pqk}^n\right\}$ converges with the update rule (14). Let $\left\{x_{pqk}^*\right\}$ denotes the converging state of $\left\{x_{pqk}^n\right\}$, i.e., $0 \leq x_{pqk}^* = \lim_{n\to+\infty} x_{pqk}^n < +\infty$. Then for the learning objective (7), the following KKT conditions related to $\left\{x_{pqk}^n\right\}$ should be fulfilled if $\left\{x_{pqk}^*\right\}$ is one of its stationary point.

$$
\left.\frac{\partial L}{\partial x_{pqt}}\right|_{x_{pqt}=x_{pqt}^*} = \sum_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk} - \hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right) + \lambda\left|\Lambda(k)\right| x_{pqt}^* - \tilde{x}_{pqt}^* = 0, \tag{S20a}
$$

$$
x_{pqt}^* \cdot \tilde{x}_{pqt}^* = 0, \tag{S20b}
$$

$$
x_{pqt}^* \geq 0, \tag{S20c}
$$

$$
\tilde{x}_{pqt}^* \geq 0. \tag{S20d}
$$

Note that following (S1)-(S3), condition (S20a) is evidently fulfilled with parameter update rule (15), thus, we have the following equation holds:

$$
\tilde{x}_{pqt}^* = \sum_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk} - \hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right) + \lambda\left|\Lambda(k)\right| x_{pqt}^*. \tag{S21}
$$

Next, we mainly analyze condition (S20c) and (S20d). We first construct β_{pqk}^n as:

$$
\beta_{pqk}^n = \frac{\sum_{y_{ijk}\in\Lambda(k)} y_{ijk} a_{ip} a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)} \hat{y}_{ijk} a_{ip} a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}^n}. \tag{S22}
$$

Naturally, (S22) is bounded by non-negative y_{ijk}, a_{ip} and a_{iq}:

$$
0 \leq \beta_{pqk}^* = \lim_{n\to+\infty}\beta_{pqk}^n = \frac{\sum_{y_{ijk}\in\Lambda(k)} y_{ijk} a_{ip} a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)} \hat{y}_{ijk} a_{ip} a_{jq} + \lambda\left|\Lambda(k)\right| x_{pqk}^*}. \tag{S23}
$$

Hence, the update rule of x_{pqk} with SLF-NMU can be written as:

$$
x_{pqk}^{n+1} = x_{pqk}^n \beta_{pqk}^n. \tag{S23}
$$

By combining (S20) and (S23), we have:

$$\lim_{n\to+\infty}\left(x_{pqk}^{n+1}-x_{pqk}^{n}\right)=0\Rightarrow x_{pqk}^{*}\beta_{pqk}^{*}-x_{pqk}^{*}=0. \tag{S24}$$

According to the update rule (15), $x_{pqk}^{*}\geq 0$ with a non-negatively initial hypothesis. Hence, we have the following inferences.

a) **When x_{pqk}^{*} >0.** Based on (S21) and (S24), we have:

$$\lim_{n\to+\infty} x_{pqk}^{*}\beta_{pqk}^{*}-x_{pqk}^{*}=0, x_{pqk}^{*}>0\Rightarrow \beta_{pqk}^{*}=1\Rightarrow \sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqt}^{*}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}a_{ip}a_{jq}=0. \tag{S25}$$

By combing (S20) and (S25), we achieve condition (S20b):

$$\tilde{x}_{pqt}^{*}=\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqt}^{*}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}a_{ip}a_{jq}\Rightarrow \tilde{x}_{pqt}^{*}\cdot x_{pqt}^{*}=0. \tag{S26}$$

Meanwhile, when $\tilde{x}_{pqk}^{*}$ =0 and x_{pqk}^{*} >0, condition (S20c) and (S20d) are naturally fulfilled. Hence, when x_{pqk}^{*} >0, KKT conditions in (S20) are all satisfied.

b) **When x_{pqk}^{*} =0.** The conditions (S20b) and (S20c) naturally holds. Hence, we only need to justify that whether condition (S20d) is fulfilled or not. To do so, we reformulate x_{pqk}^{*} as follows:

$$x_{pqk}^{*}=x_{pqk}^{0}\lim_{n\to+\infty}\prod_{h=1}^{n}\beta_{pqk}^{h}. \tag{S27}$$

Based on (S27) we further have the following deduction:

$$\begin{aligned}
&x_{pqk}^{0}>0, x_{pqk}^{0}\lim_{n\to+\infty}\prod_{h=1}^{n}\beta_{pqk}^{h}=x_{pqk}^{*}=0\Rightarrow \lim_{n\to+\infty}\prod_{h=1}^{n}\beta_{pqk}^{h}=0\\
&\Rightarrow \lim_{n\to+\infty}\beta_{pqk}^{n}=\beta_{pqk}^{*}=\frac{\sum_{y_{ijk}\in\Lambda(k)}y_{ijk}a_{ip}a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{*}}\leq 1\\
&\Rightarrow \tilde{x}_{pqk}^{*}=\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{*}-\sum_{y_{ijk}\in\Lambda(k)}y_{ijk}a_{ip}a_{jq}\geq 0.
\end{aligned} \tag{S28}$$

There, the condition (S20d) holds. Hence, when x_{pqk}^{*} =0, KKT conditions in (S20) are all satisfied. By analogy, the sequences $\left\{a_{ip}^{n}\right\}$, $\left\{a_{jq}^{n}\right\}$, $\left\{s_{i}^{n}\right\}$, $\left\{d_{j}^{n}\right\}$ and $\left\{t_{k}^{n}\right\}$ can also be proven to converge to a stationary point of (7).

As a result, according to Theorems 1-2, it can be proven that TRNL is guaranteed to converge at a KKT stationary point of its learning objective. Therefore, the convergence of TRNL model on a nonnegative HDI tensor is guaranteed.

IV. EXPERIMENTAL RESULTS OF TRNL

TABLE S1. HYPER-PARAMETER SETTINGS OF M1-M9.

Dataset	Hyper-parameter Setting				
D1	M1: Self-adaptation	M2: λ=0.001, η=0.001	M3: λ=0.01, η=0.0005	M4: λ_1=0.05, λ_2=0.05, λ_3=0.0001	M5: λ=0.05, η=0.01
	M6: λ=0.01, η=0.1	M7: λ=0.001, η=0.1			
D2	M1: Self-adaptation	M2: λ=0.001, η=0.001	M3: λ=0.01, η=0.0005	M4: λ_1=0.05, λ_2=0.05, λ_3=0.0001	M5: λ=0.03, η=0.01
	M6: λ=0.01, η=0.1	M7: λ=0.001, η=0.001			
D3	M1: Self-adaptation	M2: λ=0.01, η=0.004	M3: λ=0.001, η=0.0005	M4: λ_1=0.01, λ_2=0.01, λ_3=0.0005	M5: λ=0.01, η=0.02
	M6: λ=0.1, η=0.1	M7: λ=0.01, η=0.3			
D4	M1: Self-adaptation	M2: λ=0.01, η=0.005	M3: λ=0.001, η=0.0005	M4: λ_1=0.005, λ_2=0.01, λ_3=0.0005	M5: λ=0.01, η=0.02
	M6: λ=0.1, η=0.1	M7: λ=0.01, η=0.3			
D5	M1: Self-adaptation	M2: λ=0.0001,η=0.005	M3: λ=0.01, η=0.0001	M4: λ_1=0.01, λ_2=0.001, λ_3=0.05	M5: λ=0.001, η=0.04
	M6: λ=0.1, η=0.03	M7: λ=0.1, η=0.008	M8: $\rho_{\bar{A}n}$=1, $\rho_{\tilde{A}n}$=1, $\rho_{\check{d}n}$=1, μ=0.0001		M9: α=0.0001, β=0.1
D6	M1: Self-adaptation	M2: λ=0.05, η=0.0005	M3: λ=0.01, η=0.0002	M4: λ_1=0.01, λ_2=0.001, λ_3=0.001	M5: λ=0.05, η=0.02
	M6: λ=0.1, η=0.001	M7: λ=0.1, η=0.001	M8: $\rho_{\bar{A}n}$=0.1, $\rho_{\tilde{A}n}$=0.1, $\rho_{\check{d}n}$=0.1, μ=0.0001		M9: α=0.001, β=0.1
D7	M1: Self-adaptation	M2:λ=0.00001,η=0.005	M3: λ=0.1, η=0.0005	M4: λ_1=0.01, λ_2=0.001, λ_3=0.01	M5: λ=0.1, η=0.01
	M6: λ=0.01, η=0.3	M7: λ=0.01,η=0.3	M8: $\rho_{\bar{A}n}$=1, $\rho_{\tilde{A}n}$=1, $\rho_{\check{d}n}$=1,μ=0.0001		M9: α=0.0001, β=0.1
D8	M1: Self-adaptation	M2: λ=0.05, η=0.0001	M3: λ=0.1, η=0.0003	M3: λ_1=0.01, λ_2=0.001, λ_3=0.001	M4: λ=0.1, η=0.02
	M6: λ=0.1, η=0.1	M7: λ=0.01, η=0.1	M8: $\rho_{\bar{A}n}$=0.1, $\rho_{\tilde{A}n}$=0.1, $\rho_{\check{d}n}$=0.1, μ=0.00001		M9: α=0.001, β=0.1

TABLE S2. RMSE, WIN/LOSS COUNTS AND FRIEDMAN TEST OF M1-9 ON D1-8.

No.	D1	D2	D3	D4	D5	D6	D7	D8	W/L	F-R
M1	**0.2969±0.0006**	**0.3129±0.0004**	**0.2671±0.0005**	**0.2707±0.0003**	**0.3438**±0.0025	**0.3205**±0.0021	**0.3262**±0.0053	**0.3281**±0.0012	**--**	**1.00**
M2	0.3091±0.0006	0.3221±0.0004	0.2723±0.0005	0.2738±0.0004	0.3588±0.0022	0.3311±0.0026	0.3356±0.0041	0.3382±0.0018	8/0	3.63
M3	0.3016±0.0010	0.3189±0.0007	0.2681±0.0004	0.2722±0.0005	0.3557±0.0027	0.3241±0.0022	0.3353±0.0036	0.3316±0.0018	8/0	2.00
M4	0.3082±0.0041	0.3223±0.0025	0.2920±0.0054	0.2935±0.0043	0.3599±0.0031	0.3315±0.0024	0.3364±0.0048	0.3358±0.0053	8/0	4.38
M5	0.3189±0.0006	0.3342±0.0007	0.2837±0.0005	0.2895±0.0005	0.3584±0.0026	0.3337±0.0019	0.3418±0.0031	0.3457±0.0042	8/0	5.00
M6	0.4158±0.0001	0.4420±0.0001	0.3893±0.0003	0.3908±0.0001	0.4466±0.0023	0.4136±0.0014	0.4062±0.0016	0.4051±0.0073	8/0	7.00
M7	0.4333±0.0001	0.4563±0.0001	0.3960±0.0003	0.4170±0.0001	0.4596±0.0018	0.4152±0.0011	0.4105±0.0014	0.4072±0.0077	8/0	8.00

M8	Intractable	Intractable	Intractable	Intractable	0.3570±0.0021	0.3346±0.0022	0.3376±0.0025	0.3337±0.0027	8/0	6.38
M9	Intractable	Intractable	Intractable	Intractable	0.3695±0.0028	0.3529±0.0025	0.3406±0.0036	0.3506±0.0015	8/0	7.63

TABLE S3. MAE, WIN/LOSS COUNTS AND FRIEDMAN TEST OF M1-9 ON D1-8.

No.	D1	D2	D3	D4	D5	D6	D7	D8	W/L	F-R
M1	**0.2202±0.0004**	**0.2332±0.0003**	**0.1830±0.0007**	**0.1862±0.0008**	**0.2310**±0.0009	**0.2153**±0.0008	**0.2215**±0.0018	**0.2239**±0.0015	**--**	**1.00**
M2	0.2205±0.0004	0.2347±0.0003	0.1833±0.0003	0.1868±0.0004	0.2383±0.0012	0.2176±0.0009	0.2231±0.0021	0.2262±0.0011	8/0	2.25
M3	0.2247±0.0009	0.2382±0.0008	0.1867±0.0003	0.1919±0.0003	0.2389±0.0011	0.2219±0.0008	0.2253±0.0012	0.2332±0.0013	8/0	4.00
M4	0.2215±0.0010	0.2340±0.0010	0.1834±0.0014	0.1871±0.0005	0.2387±0.0015	0.2183±0.0010	0.2233±0.0016	0.2259±0.0021	8/0	2.75
M5	0.2375±0.0005	0.2499±0.0005	0.1938±0.0003	0.1997±0.0006	0.2391±0.0015	0.2229±0.0006	0.2302±0.0013	0.2355±0.0021	8/0	5.00
M6	0.3116±0.0001	0.3306±0.0001	0.2836±0.0001	0.2853±0.0018	0.3153±0.0009	0.3067±0.0008	0.2981±0.0010	0.3067±0.0044	8/0	7.00
M7	0.3296±0.0001	0.3445±0.0001	0.3083±0.0003	0.3083±0.0002	0.3264±0.0054	0.3077±0.0010	0.3072±0.0009	0.3071±0.0046	8/0	8.00
M8	Intractable	Intractable	Intractable	Intractable	0.2595±0.0018	0.2426±0.0012	0.2386±0.0011	0.2420±0.0023	8/0	7.25
M9	Intractable	Intractable	Intractable	Intractable	0.2723±0.0022	0.2570±0.0019	0.2485±0.0033	0.2569±0.0015	8/0	7.75

TABLE S4. TOTAL TIME COST (SEC.) IN RMSE, WIN/LOSS COUNTS AND FRIEDMAN TEST OF M1-9 ON D1-8.

No.	D1	D2	D3	D4	D5	D6	D7	D8	W/L	F-R
M1	**40750±4892**	**47114±3699**	**2606±588**	**1839±449**	**218±15**	**182±23**	**48±11**	**28±8**	**--**	**1.13**
M2	456754±17214	480636±59978	115811±4468	103307±2279	11528±0	4586±0	2331±0	1675±0	8/0	5.38
M3	2093259±0	1386816±0	241908±0	213564±0	6076±0	2548±0	1540±0	874±0	8/0	5.50
M4	272581±31966	178645±51879	32357±4556	16051±4242	824±158	248±75	118±32	106±27	8/0	3.00
M5	137540±8931	108035±7207	12560±237	7432±116	331±11	179±9	77±3	47±7	8/0	1.88
M6	4005976±472323	1940738±785151	65014±2947	49413±848	1701±106	2765±376	6715±946	3698±308	8/0	5.50
M7	1373295±38053	1019474±60628	72846±2115	74764±2041	1665±134	4958±813	12133±483	14543±556	8/0	5.63
M8	Intractable	Intractable	Intractable	Intractable	3752343±0	1361621±0	306960±0	168289±0	8/0	8.75
M9	Intractable	Intractable	Intractable	Intractable	2605424±185862	573207±52971	212264±16781	59052±3538	8/0	8.25

TABLE S5. TOTAL TIME COST IN MAE (SEC.), WIN/LOSS COUNTS AND FRIEDMAN TEST OF M1-9 ON D1-8.

No.	D1	D2	D3	D4	D5	D6	D7	D8	W/L	F-R
M1	**71312±6641**	**22821±1377**	**984±303**	**1944±253**	**254±56**	**16±3**	**6±2**	**5±1**	**--**	**1.00**
M2	591705±114071	680337±67610	159929±7348	136579±13545	11528±0	4586±0	2331±0	1675±0	8/0	5.50
M3	2093259±0	1386816±0	241908±0	213564±0	6076±0	2548±0	1540±0	874±0	8/0	5.63
M4	277078±42004	183501±48302	29455±8977	17122±3606	673±163	248±76	116±35	102±35	8/0	3.25
M5	145820±15565	122025±9769	22125±195	12198±184	328±16	163±9	72±4	41±4	8/0	2.13
M6	4034993±485104	2021613±826462	22165±363	43759±397	926±28	1795±194	5486±748	3074±254	8/0	5.38
M7	1122224±143792	854406±32557	20854±726	36037±89	1387±65	2776±1173	9783±445	12264±479	8/0	5.13
M8	Intractable	Intractable	Intractable	Intractable	3752343±0	1361621±0	306960±0	168289±0	8/0	8.75
M9	Intractable	Intractable	Intractable	Intractable	2579040±137412	569733±63726	219340±17466	59361±3605	8/0	8.25