Spatiotemporal Graph Neural Network-Incorporated Latent Factorization of Tensors for Dynamic QoS Estimation Supplementary File

Fanghui Bi, Tiantian He, *Member, IEEE*, Xin Luo*, *Fellow, IEEE*

I. Proofs (Sec. IV)

A. Expressiveness Proof of Vanilla GCNs on Dynamic User-Service Graphs (Proof of Theorem 1)

To establish ***Theorem* 1**, both directions of the iff conditions must be taken into account. First, given the conditions that $\Gamma_{1,t} = \Gamma_{2,t} = \Gamma_t$ and $\tau \cdot \sum_{w=y,w\in Y_{1,t}} \varsigma_{e_{1,t},w} \cdot \sqrt{N(e_{1,t})} = \sum_{w=y,w\in Y_{2,t}} \varsigma_{e_{2,t},w} \cdot \sqrt{N(e_{2,t})}$, for $\tau = \sqrt{N(e_{2,t})/N(e_{1,t})}$ and $y \in \Gamma_t$, we need to prove $f(e_{1,t}, Y_{1,t}) = f(e_{2,t}, Y_{2,t})$. Based on Eq. (23), the aggregation function in the countable feature space can be represented as:

$$f\left(e_{i,t}, Y_{i,t}\right) = \sum_{y\in Y_{i,t}} \varsigma_{e_{i,t},y}, \quad \varsigma_{e_{i,t},y} = \frac{1}{\sqrt{N\left(e_{i,t}\right) \cdot N(y)}}. \tag{S1}$$

With the above equations, we achieve the following inferences for target center nodes:

$$\begin{cases} f\left(e_{1,t}, Y_{1,t}\right) = \sum\limits_{y\in Y_{1,t}} \varsigma_{e_{1,t},y} \cdot y = \sum\limits_{y\in Y_{1,t}} \frac{y}{\sqrt{N(e_{1,t})\cdot N(y)}}, \\ f\left(e_{2,t}, Y_{2,t}\right) = \sum\limits_{y\in Y_{2,t}} \varsigma_{e_{2,t},y} \cdot y = \sum\limits_{y\in Y_{2,t}} \frac{y}{\sqrt{N(e_{2,t})\cdot N(y)}}. \end{cases} \tag{S2}$$

And considering the given condition that $\Gamma_{1,t} = \Gamma_{2,t} = \Gamma_t$, we derive:

$$\begin{aligned} f\left(e_{1,t}, Y_{1,t}\right) - f\left(e_{2,t}, Y_{2,t}\right) &= \sum_{y\in\Gamma_t} \left(\sum_{w=y,w\in Y_{1,t}} \varsigma_{e_{1,t},w} - \sum_{w=y,w\in Y_{2,t}} \varsigma_{e_{2,t},w} \right) \cdot y \\ &= \sum_{y\in\Gamma_t} \left(\sum_{w=y,w\in Y_{1,t}} \frac{1}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}} \frac{1}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} \right) \cdot y. \end{aligned} \tag{S3}$$

Further, by substituting $\tau \cdot \sum_{w=y,w\in Y_{1,t}} \varsigma_{e_{1,t},w} \cdot \sqrt{N(e_{1,t})} = \sum_{w=y,w\in Y_{2,t}} \varsigma_{e_{2,t},w} \cdot \sqrt{N(e_{2,t})}$ and $\tau = \sqrt{N(e_{2,t})/N(e_{1,t})}$ into the above formula, we can intuitively observe that $f(e_{1,t}, Y_{1,t}) - f(e_{2,t}, Y_{2,t}) = 0$. Hence, the first direction is proven.

For the second direction of the iff conditions, given $f(e_{1,t}, Y_{1,t}) = f(e_{2,t}, Y_{2,t})$, we need to prove $\Gamma_{1,t} = \Gamma_{2,t} = \Gamma_t$ and $\tau \cdot \sum_{w=y,w\in Y_{1,t}} \varsigma_{e_{1,t},w} \cdot \sqrt{N(e_{1,t})} = \sum_{w=y,w\in Y_{2,t}} \varsigma_{e_{2,t},w} \cdot \sqrt{N(e_{2,t})}$, for $\tau = \sqrt{N(e_{2,t})/N(e_{1,t})}$ and $y \in \Gamma_t$. To achieve this, we show the contradictions while they do not simultaneously hold. Since $f(e_{1,t}, Y_{1,t}) = f(e_{2,t}, Y_{2,t})$, we directly infer:

$$f\left(e_{1,t}, Y_{1,t}\right) - f\left(e_{2,t}, Y_{2,t}\right) = \sum_{y\in Y_{1,t}} \frac{y}{\sqrt{N\left(e_{1,t}\right)\cdot N(y)}} - \sum_{y\in Y_{2,t}} \frac{y}{\sqrt{N\left(e_{2,t}\right)\cdot N(y)}} = 0. \tag{S4}$$

We first assume that $\Gamma_{1,t} \neq \Gamma_{2,t}$ for all $Y_{1,t}, Y_{2,t} \in \mathcal{Y}$, such that the following derivation holds:

$$\begin{aligned} f\left(e_{1,t}, Y_{1,t}\right) - f\left(e_{2,t}, Y_{2,t}\right) = &\sum_{y\in\Gamma_{1,t}\cap\Gamma_{2,t}} \left(\sum_{w=y,w\in Y_{1,t}} \frac{y}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}} \frac{y}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} \right) \\ &+ \sum_{y\in\Gamma_{1,t}\setminus\Gamma_{2,t}} \sum_{w=y,w\in Y_{1,t}} \frac{y}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{y\in\Gamma_{2,t}\setminus\Gamma_{1,t}} \sum_{w=y,w\in Y_{2,t}} \frac{y}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} = 0. \end{aligned} \tag{S5}$$

Note that any y confirms to Eq. (S5), thus a function $g(\cdot)$ can be defined as:

$$y = \begin{cases} g(y), & \text{for } y \in \Gamma_{1,t} \cap \Gamma_{2,t}; \\ g(y) - 1, & \text{for } y \in \Gamma_{1,t}\setminus\Gamma_{2,t}; \\ g(y) + 1, & \text{for } y \in \Gamma_{2,t}\setminus\Gamma_{1,t}. \end{cases} \tag{S6}$$

Based on the assumption that $\Gamma_{1,t} \neq \Gamma_{2,t}$, Eq. (S5) holds, indicating that the following inference also holds:

$$\begin{aligned}
&f\left(e_{1,t}, Y_{1,t}\right) - f\left(e_{2,t}, Y_{2,t}\right) \\
=&\sum_{y\in\Gamma_{1,t}\cap\Gamma_{2,t}}\left(\sum_{w=y,w\in Y_{1,t}}\frac{g(y)}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}}\frac{g(y)}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}}\right) \\
&+\sum_{y\in\Gamma_{1,t}\setminus\Gamma_{2,t}}\sum_{w=y,w\in Y_{1,t}}\frac{g(y)}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{y\in\Gamma_{2,t}\setminus\Gamma_{1,t}}\sum_{w=y,w\in Y_{2,t}}\frac{g(y)}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} \\
=&\sum_{y\in\Gamma_{1,t}\cap\Gamma_{2,t}}\left(\sum_{w=y,w\in Y_{1,t}}\frac{y}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}}\frac{y}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}}\right) \\
&+\sum_{y\in\Gamma_{1,t}\setminus\Gamma_{2,t}}\sum_{w=y,w\in Y_{1,t}}\frac{(y+1)}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{y\in\Gamma_{2,t}\setminus\Gamma_{1,t}}\sum_{w=y,w\in Y_{2,t}}\frac{(y-1)}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} = 0.
\end{aligned} \tag{S7}$$

Since Eq. (S5) is equivalent to Eq. (S7), we obtain the inference as follows:

$$\begin{aligned}
&f\left(e_{1,t}, Y_{1,t}\right) - f\left(e_{2,t}, Y_{2,t}\right) \\
=&\sum_{y\in\Gamma_{1,t}\setminus\Gamma_{2,t}}\sum_{w=y,w\in Y_{1,t}}\frac{1}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} + \sum_{y\in\Gamma_{2,t}\setminus\Gamma_{1,t}}\sum_{w=y,w\in Y_{2,t}}\frac{1}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} = 0.
\end{aligned} \tag{S8}$$

It is apparent that the terms in Eq. (S8) are positive, which signifies the above formula is not valid. Hence, the assumption that $\Gamma_{1,t} \neq \Gamma_{2,t}$ is not true, i.e., $\Gamma_{1,t} = \Gamma_{2,t}$. Now we can assume $\Gamma_{1,t} = \Gamma_{2,t} = \Gamma_t$, and have the following inference:

$$\sum_{y\in\Gamma_{1,t}\setminus\Gamma_{2,t}}\sum_{w=y,w\in Y_{1,t}}\frac{y}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{y\in\Gamma_{2,t}\setminus\Gamma_{1,t}}\sum_{w=y,w\in Y_{2,t}}\frac{y}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} = 0. \tag{S9}$$

Eq. (S9) indicates that Eq. (S5) can be rewritten as:

$$\begin{aligned}
&f\left(e_{1,t}, Y_{1,t}\right) - f\left(e_{2,t}, Y_{2,t}\right) \\
=&\sum_{y\in\Gamma_t}\left(\sum_{w=y,w\in Y_{1,t}}\frac{1}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}}\frac{1}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}}\right)\cdot y = 0.
\end{aligned} \tag{S10}$$

From Eq. (S10), it is observed that all terms in its summation are zero, thus we have:

$$\sum_{w=y,w\in Y_{1,t}}\frac{1}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}}\frac{1}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} = 0. \tag{S11}$$

And we further infer:

$$\begin{aligned}
&\sum_{w=y,w\in Y_{1,t}}\frac{1}{\sqrt{N\left(e_{1,t}\right)\cdot N(w)}} - \sum_{w=y,w\in Y_{2,t}}\frac{1}{\sqrt{N\left(e_{2,t}\right)\cdot N(w)}} \\
=&\frac{1}{\sqrt{N\left(e_{1,t}\right)}}\sum_{w=y,w\in Y_{1,t}}\varsigma_{e_{1,t},w}\cdot\sqrt{N\left(e_{1,t}\right)} - \frac{1}{\sqrt{N\left(e_{2,t}\right)}}\sum_{w=y,w\in Y_{2,t}}\varsigma_{e_{2,t},w}\cdot\sqrt{N\left(e_{2,t}\right)} = 0.
\end{aligned} \tag{S12}$$

By setting $\tau = \sqrt{N(e_{2,t})/N(e_{1,t})}$, Eq. (S12) is represented as:

$$\tau\cdot\sum_{w=y,w\in Y_{1,t}}\varsigma_{e_{1,t},w}\cdot\sqrt{N\left(e_{1,t}\right)} - \sum_{w=y,w\in Y_{2,t}}\varsigma_{e_{2,t},w}\cdot\sqrt{N\left(e_{2,t}\right)} = 0. \tag{S13}$$

Based on the above inferences, we prove that $\Gamma_{1,t} = \Gamma_{2,t} = \Gamma_t$ and $\tau\cdot\sum_{w=y,w\in Y_{1,t}}\varsigma_{e_{1,t},w}\cdot\sqrt{N(e_{1,t})} = \sum_{w=y,w\in Y_{2,t}}\varsigma_{e_{2,t},w}\cdot\sqrt{N(e_{2,t})}$, for $\tau = \sqrt{N(e_{2,t})/N(e_{1,t})}$ and $y \in \Gamma_t$. In conclusion, ***Theorem* 1** holds.□

B. Expressiveness Proof of SGLFT (Proof of Corollary 1)

In accordance with ***Theorem* 1**, the aggregator of Eq. (13) in the countable feature space ($\mathcal{Y}$) can be denoted as $f\left(e_t, Y\right) = \sum_{y\in Y}\xi_{e_t,y}\cdot y$, which is specifically presented as:

$$f\left(e_{i,t}, Y_i\right) = \sum_{y\in Y_i}\xi_{e_{i,t},y}\cdot y, \quad \xi_{e_{i,t},y} = \sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{i,j},y}\cdot\varsigma_{e_{i,j},y}, \tag{S14}$$

where $a_{e_{i,j},y}$ denotes the adjacency status between two nodes at the j-th timestamp. Specifically, if there is a connection between the two nodes at time j, the value of $a_{e_{i,j},y}$ is 1; otherwise, it is 0. Let $Y=\{Y_1,Y_2,...,Y_{t-1},Y_t,...,Y_{|T|-1},Y_{|T|}\}\in\mathcal{Y}$. For a more intuitive comparison, it can be represented as $Y=\{Y_t,Y_{T-t}\}$. Thus, Eq. (S14) is rewritten as:

$$\begin{aligned}
&f\left(e_{i,t},Y_i\right)\\
&=\sum_{y\in Y_{i,t}}\xi_{e_{i,t},y}\cdot y+\sum_{y\in Y_{i,T-t}}\xi_{e_{i,t},y}\cdot y\\
&=\sum_{y\in Y_{i,t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{i,j},y}\cdot\varsigma_{e_{i,j},y}\cdot y+\sum_{y\in Y_{i,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{i,j},y}\cdot\varsigma_{e_{i,j},y}\cdot y\\
&=\sum_{y\in Y_{i,t}}\omega_{t,t}^{(0)}\cdot a_{e_{i,t},y}\cdot\varsigma_{e_{i,t},y}\cdot y+\sum_{y\in Y_{i,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{i,j},y}\cdot\varsigma_{e_{i,j},y}\cdot y+\sum_{y\in Y_{i,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{i,j},y}\cdot\varsigma_{e_{i,j},y}\cdot y.
\end{aligned}\tag{S15}$$

Then, we have the following inference:

$$\begin{cases}
f\left(e_{1,t},Y_1\right)=\sum_{y\in Y_1}\xi_{e_{1,t},y}\cdot y\\
\qquad=\sum_{y\in Y_{1,t}}\omega_{t,t}^{(0)}\cdot a_{e_{1,t},y}\cdot\varsigma_{e_{1,t},y}\cdot y+\sum_{y\in Y_{1,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},y}\cdot\varsigma_{e_{1,j},y}\cdot y+\sum_{y\in Y_{1,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},y}\cdot\varsigma_{e_{1,j},y}\cdot y,\\
f\left(e_{2,t},Y_2\right)=\sum_{y\in Y_2}\xi_{e_{2,t},y}\cdot y\\
\qquad=\sum_{y\in Y_{2,t}}\omega_{t,t}^{(0)}\cdot a_{e_{2,t},y}\cdot\varsigma_{e_{2,t},y}\cdot y+\sum_{y\in Y_{2,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},y}\cdot\varsigma_{e_{2,j},y}\cdot y+\sum_{y\in Y_{2,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},y}\cdot\varsigma_{e_{2,j},y}\cdot y.
\end{cases}\tag{S16}$$

Thus, we easily obtain:

$$\begin{aligned}
&f\left(e_{1,t},Y_1\right)-f\left(e_{2,t},Y_2\right)\\
&=\sum_{y\in Y_{1,t}}\omega_{t,t}^{(0)}\cdot a_{e_{1,t},y}\cdot\varsigma_{e_{1,t},y}\cdot y-\sum_{y\in Y_{2,t}}\omega_{t,t}^{(0)}\cdot a_{e_{2,t},y}\cdot\varsigma_{e_{2,t},y}\cdot y\\
&+\sum_{y\in Y_{1,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},y}\cdot\varsigma_{e_{1,j},y}\cdot y-\sum_{y\in Y_{2,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},y}\cdot\varsigma_{e_{2,j},y}\cdot y\\
&+\sum_{y\in Y_{1,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},y}\cdot\varsigma_{e_{1,j},y}\cdot y-\sum_{y\in Y_{2,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},y}\cdot\varsigma_{e_{2,j},y}\cdot y.
\end{aligned}\tag{S17}$$

Following ***Theorem* 1**, we denote $Y_{1,t}=\{\Theta_{1,t},\mu_{1,t}\}$, $Y_{2,t}=\{\Theta_{2,t},\mu_{2,t}\}$. It is worth noting that Γ_t is included in both $\Theta_{1,t}$ and $\Theta_{2,t}$, although $\Theta_{1,t}$ and $\Theta_{2,t}$ may not be equal. In ***Theorem* 1**, we prove that the aggregation function in Eq. (23) fails to distinguish a certain graph structure with the conditions that $\Gamma_{1,t}=\Gamma_{2,t}=\Gamma_t$ and $\tau\cdot\sum_{w=y,w\in Y_{1,t}}\varsigma_{e_{1,t},w}\cdot\sqrt{N(e_{1,t})}=\sum_{w=y,w\in Y_{2,t}}\varsigma_{e_{2,t},w}\cdot\sqrt{N(e_{2,t})}$, for $\tau=\sqrt{N(e_{2,t})/N(e_{1,t})}$ and $y\in\Gamma_t$. First, given $\Gamma_{1,t}=\Gamma_{2,t}=\Gamma_t$, we achieve:

$$\begin{aligned}
&f\left(e_{1,t},Y_1\right)-f\left(e_{2,t},Y_2\right)\\
&=\sum_{y\in\Gamma_t}\left(\sum_{w=y,w\in Y_{1,t}}\omega_{t,t}^{(0)}\cdot a_{e_{1,t},w}\cdot\varsigma_{e_{1,t},w}-\sum_{w=y,w\in Y_{2,t}}\omega_{t,t}^{(0)}\cdot a_{e_{2,t},w}\cdot\varsigma_{e_{2,t},w}\right)\cdot y\\
&+\sum_{y\in\Theta_{1,t}\backslash\Gamma_t}\omega_{t,t}^{(0)}\cdot a_{e_{1,t},y}\cdot\varsigma_{e_{1,t},y}\cdot y-\sum_{y\in\Theta_{2,t}\backslash\Gamma_t}\omega_{t,t}^{(0)}\cdot a_{e_{2,t},y}\cdot\varsigma_{e_{2,t},y}\cdot y\\
&+\sum_{y\in\Gamma_t}\left(\sum_{w=y,w\in Y_{1,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},w}\cdot\varsigma_{e_{1,j},w}-\sum_{w=y,w\in Y_{2,t}}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},w}\cdot\varsigma_{e_{2,j},w}\right)\cdot y\\
&+\sum_{y\in\Theta_{1,t}\backslash\Gamma_t}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},y}\cdot\varsigma_{e_{1,j},y}\cdot y-\sum_{y\in\Theta_{2,t}\backslash\Gamma_t}\sum_{j=1,j\neq t}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},y}\cdot\varsigma_{e_{2,j},y}\cdot y\\
&+\sum_{y\in Y_{1,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{1,j},y}\cdot\varsigma_{e_{1,j},y}\cdot y-\sum_{y\in Y_{2,T-t}}\sum_{j=1}^{|T|}\omega_{t,j}^{(0)}\cdot a_{e_{2,j},y}\cdot\varsigma_{e_{2,j},y}\cdot y.
\end{aligned}\tag{S18}$$

By considering the adjacent statuses at time t, Eq. (S18) can be further simplified as:

$$\begin{aligned}
&f\left(e_{1,t}, Y_1\right)-f\left(e_{2,t}, Y_2\right) \\
&=\sum_{y \in \Gamma_t}\left(\sum_{w=y, w \in Y_{1,t}} \varsigma_{e_{1,t}, w}-\sum_{w=y, w \in Y_{2,t}} \varsigma_{e_{2,t}, w}\right) \cdot \omega_{t,t}^{(0)} \cdot y \\
&+\sum_{y \in \Gamma_t}\left(\sum_{w=y, w \in Y_{1,t}} \sum_{j=1, j \neq t}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{1,j}, w} \cdot \varsigma_{e_{1,j}, w}-\sum_{w=y, w \in Y_{2,t}} \sum_{j=1, j \neq t}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{2,j}, w} \cdot \varsigma_{e_{2,j}, w}\right) \cdot y \\
&+\sum_{y \in \Theta_{1,t} \backslash \Gamma_t} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{1,j}, y} \cdot \varsigma_{e_{1,j}, y} \cdot y-\sum_{y \in \Theta_{2,t} \backslash \Gamma_t} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{2,j}, y} \cdot \varsigma_{e_{2,j}, y} \cdot y \\
&+\sum_{y \in Y_{1,T-t}} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{1,j}, y} \cdot \varsigma_{e_{1,j}, y} \cdot y-\sum_{y \in Y_{2,T-t}} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{2,j}, y} \cdot \varsigma_{e_{2,j}, y} \cdot y .
\end{aligned} \tag{S19}$$

Then, only based on the condition that $\tau \cdot \sum_{w=y, w \in Y_{1,t}} \varsigma_{e_{1,t}, w} \cdot \sqrt{N(e_{1,t})}=\sum_{w=y, w \in Y_{2,t}} \varsigma_{e_{2,t}, w} \cdot \sqrt{N(e_{2,t})}$, for $\tau=\sqrt{N(e_{2,t}) / N(e_{1,t})}$ and $y \in \Gamma_t$, we have the following conclusion:

$$\begin{aligned}
&f\left(e_{1,t}, Y_1\right)-f\left(e_{2,t}, Y_2\right) \\
&=\sum_{y \in \Gamma_t}\left(\sum_{w=y, w \in Y_{1,t}} \sum_{j=1, j \neq t}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{1,j}, w} \cdot \varsigma_{e_{1,j}, w}-\sum_{w=y, w \in Y_{2,t}} \sum_{j=1, j \neq t}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{2,j}, w} \cdot \varsigma_{e_{2,j}, w}\right) \cdot y \\
&+\sum_{y \in \Theta_{1,t} \backslash \Gamma_t} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{1,j}, y} \cdot \varsigma_{e_{1,j}, y} \cdot y-\sum_{y \in \Theta_{2,t} \backslash \Gamma_t} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{2,j}, y} \cdot \varsigma_{e_{2,j}, y} \cdot y \\
&+\sum_{y \in Y_{1,T-t}} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{1,j}, y} \cdot \varsigma_{e_{1,j}, y} \cdot y-\sum_{y \in Y_{2,T-t}} \sum_{j=1}^{|T|} \omega_{t,j}^{(0)} \cdot a_{e_{2,j}, y} \cdot \varsigma_{e_{2,j}, y} \cdot y \neq 0 .
\end{aligned} \tag{S20}$$

From Eq. (S20), we observe that the tensor product-driven graph convolution is not limited to capturing the information at time t like vanilla GCNs do, but also captures the graph structural information and temporal state information of nodes from other time slots. And according to the above inferences, the topological properties mentioned in ***Theorem* 1** that are not successfully distinguished previously may now be correctly differentiated by the aggregator in Eq. (13). To summarize, ***Corollary* 1** holds.□

II. General Settings (Sec. V.A)

A. Details of Models for Comparison

The details of all comparison models in this paper are presented as follows:

- **EvolveGCN** [15]: An evolving GCN-based model for dynamic QoS estimation, which utilizes a GCN to learn spatial information and evolves the GCN parameters with a RNN for the dynamism of state graph sequence.
- **CIGCN** [11]: A channel-independent GCN-based QoS estimator, which uses a diagonal parameter matrix in each graph convolutional layer for better feature transformation, but fails to capture the dynamics in a tensor.
- **WD-GCN** [40]: A waterfall dynamic GCN-based approach, which combines GCNs and Long Short-Term Memory (LSTM) networks for jointly exploit the graph structure and temporal information in a dynamic graph.
- **TM-GCN** [25]: A dynamic GCN-based LFT model using a tensor algebra framework, which accounts for the temporal and spatial message passing based on MPNN framework.
- **JMP-GCF** [8]: A GCN-based QoS estimator, which is aware of joint multi-grained popularity in graphs by simultaneously learning these diverse granularities of popularity feature signals.
- **HMLET** [23]: A nonlinear GCN-based approach to learning latent features, which improves the light GCN process by adaptively determining nonlinear or linear propagation rules for each node based on a gating module.
- **APAN** [41]: An asynchronous propagation network-based model for continuous-time dynamic graph representation, which decouples graph computation and model inference for millisecond-level inference tasks.
- **TeDGaN** [42]: A tensor decomposition-based approach, which adopts graph Laplacian regularization to capture the spatial information and uses an LSTM network for desired time information.
- **CGTF** [43]: A coupled graph-tensor factorization model for dynamic graph embedding, which can overcome the missing slab problem and develops an alternating direction of multipliers method.
- **CTGCN** [33]: A k-core based temporal GCN model for QoS estimation, which uses RNNs to capture graph dynamics, and accounts for both global structural similarity and local connective proximity.

- **DDSTGCN** [18]: A dual dynamic spatiotemporal GCN-based LFT model, which simultaneously capture the dynamic properties of nodes and edges with transformed dual hypergraphs.
- **GRU-GCN** [16]: A GRU and GCN combined method, which highlights the expressivity advantage of time-then-graph modeling compared to time-and-graph one in representing a temporal graph.
- **MegaCRN** [14]: A meta-graph convolutional recurrent network-based QoS estimator, which plugs a meta-graph learner using the module labeled meta-node bank into its learning structure.
- **SGP** [19]: A scalable graph predictor for dynamic QoS, which aims at improving the scalability of spatiotemporal representation learning with a randomized RNN along with a GCN.
- **MGDN** [10]: A Markov graph diffusion network-based QoS estimator, which uses an untrainable Markov process for distance learning and whose learning process can be viewed as the construction for context features of vertices.
- **PGCN** [34]: A progressive GCN-based dynamic graph representation learning model, which combines GCN module with the dilated causal convolution extracting temporal patterns.
- **SGLFT**: The model presented in this study.

B. Details of Training Settings

The following settings are applied to each involved model in the experiments: We employ the Xavier method [5] to randomly initialize all trainable variables, and adopt the Adam optimizer [38] to train them. The latent feature space dimension, i.e., F is uniformly fixed at ten for objective comparison; the learning rate is tuned in $\{1e-5, 1e-4, 1e-3, 5e-3, 1e-2, 5e-2\}$; the L_2 regularization coefficient is tuned in $\{1e-5, 1e-4, 1e-3, 1e-2, 1e-1\}$; the size of mini-batch is uniformly set at 2^{15}; for SGLFT, the sliding window size, i.e., S is tuned in $\{5, 10, 15, 20, 25, 30, 35, 40\}$; for all GCN-based models, the number of graph convolutional layer is tuned in $\{1, 2, 3, 4\}$; and for other certain structures or unique hyperparameters of a target model, the suggested settings of its authors are adopted with care. We tune the fundamental hyperparameters of each model in individually-built cases Υ and Π to achieve its best performance, and then apply the obtained settings to test the QoS estimation accuracy on Ξ. Ten-fold cross validation is used to mitigate random effects: the above process of data partitioning, model inference, and performance test is conducted ten times, such that the averaged results along with standard deviations are recorded. For each model, we terminate the training process if: a) its epochs meet a preset upper bound, i.e., 1000; or b) its estimation accuracy for dynamic QoS continuously deteriorates for 30 epochs.

III. Supplementary Tables (Sec. V.B)

- Tables SI-SII report the training time in RMSE and MAE of all models on D1.1-2.4.
- Tables SIII-SV summarize the statistical outcomes from the Friedman test and the Wilcoxon signed-ranks test for all models.

TABLE SI
The Training Time in RMSE of ALL Models on D1.1-2.4.

Method	D1.1	D1.2	D1.3	D1.4	D2.1	D2.2	D2.3	D2.4
EvolveGCN	$84880_{\pm 19976.46}$	$42495_{\pm 27852.09}$	$61033_{\pm 16456.91}$	$44135_{\pm 29099.11}$	$61542_{\pm 37656.20}$	$8321_{\pm 35264.11}$	$19769_{\pm 2705.08}$	$25918_{\pm 12168.90}$
CIGCN	$4035_{\pm 588.94}$	$10268_{\pm 1197.53}$	$10692_{\pm 937.35}$	$10300_{\pm 613.12}$	$6161_{\pm 667.05}$	$12199_{\pm 891.84}$	$11998_{\pm 508.02}$	$11422_{\pm 515.23}$
WD-GCN	$35831_{\pm 13013.23}$	$18218_{\pm 8214.25}$	$17695_{\pm 8060.95}$	$37354_{\pm 25827.23}$	$31300_{\pm 5633.68}$	$35670_{\pm 16513.64}$	$30366_{\pm 22626.27}$	$11014_{\pm 4319.49}$
TM-GCN	$52858_{\pm 31052.79}$	$46913_{\pm 12995.81}$	$43689_{\pm 7402.54}$	$70233_{\pm 34177.10}$	$37859_{\pm 29825.21}$	$24849_{\pm 12480.50}$	$34075_{\pm 16390.77}$	$56305_{\pm 18938.79}$
JMP-GCF	$5628_{\pm 204.25}$	$8410_{\pm 214.77}$	$9348_{\pm 405.14}$	$22294_{\pm 913.23}$	$6575_{\pm 470.81}$	$7955_{\pm 186.57}$	$16158_{\pm 330.06}$	$32964_{\pm 705.67}$
HMLET	$21852_{\pm 517.96}$	$38889_{\pm 1135.18}$	$39540_{\pm 670.20}$	$37217_{\pm 654.53}$	$32928_{\pm 1037.48}$	$43219_{\pm 481.45}$	$42796_{\pm 266.93}$	$39777_{\pm 1209.61}$
APAN	$13685_{\pm 1522.41}$	$21943_{\pm 5612.63}$	$10506_{\pm 1463.47}$	$21067_{\pm 11782.67}$	$26562_{\pm 2971.16}$	$47516_{\pm 17734.95}$	$46893_{\pm 13619.94}$	$28148_{\pm 9797.44}$
TeDGaN	$26172_{\pm 484.96}$	$10353_{\pm 1829.91}$	$8592_{\pm 1998.69}$	$6088_{\pm 560.40}$	$16670_{\pm 178.40}$	$9427_{\pm 184.64}$	$8103_{\pm 456.43}$	$9914_{\pm 3649.57}$
CGTF	$14894_{\pm 4092.48}$	$12409_{\pm 4314.35}$	$11260_{\pm 3113.23}$	$8799_{\pm 3467.79}$	$13136_{\pm 5600.79}$	$15800_{\pm 4075.58}$	$12037_{\pm 1997.45}$	$9293_{\pm 1147.48}$
CTGCN	$63176_{\pm 32982.29}$	$90539_{\pm 18518.53}$	$104040_{\pm 25726.27}$	$98168_{\pm 25075.94}$	$49546_{\pm 21770.02}$	$113250_{\pm 13714.01}$	$92152_{\pm 17919.54}$	$72489_{\pm 16833.06}$
DDSTGCN	$46882_{\pm 8296.31}$	$70376_{\pm 29143.92}$	$74186_{\pm 18444.76}$	$50860_{\pm 11029.80}$	$44957_{\pm 24205.82}$	$78111_{\pm 18733.56}$	$74948_{\pm 13021.14}$	$54779_{\pm 25222.31}$
GRU-GCN	$13763_{\pm 699.35}$	$26755_{\pm 4411.50}$	$67629_{\pm 1564.93}$	$25143_{\pm 10873.56}$	$10390_{\pm 3163.50}$	$48200_{\pm 39247.17}$	$14723_{\pm 8052.21}$	$37686_{\pm 20093.32}$
MegaCRN	$19851_{\pm 1575.14}$	$34239_{\pm 13253.26}$	$32235_{\pm 13064.07}$	$28647_{\pm 11466.32}$	$28882_{\pm 7112.08}$	$40461_{\pm 16326.12}$	$39728_{\pm 17109.12}$	$29679_{\pm 10260.03}$
SGP	$7125_{\pm 5009.47}$	$16650_{\pm 13471.94}$	$13338_{\pm 4715.62}$	$8062_{\pm 5844.78}$	$17563_{\pm 6784.06}$	$8858_{\pm 5066.47}$	$10880_{\pm 3089.83}$	$7376_{\pm 3864.71}$
MGDN	$12770_{\pm 795.65}$	$19270_{\pm 713.68}$	$16912_{\pm 636.86}$	$15925_{\pm 619.93}$	$14920_{\pm 934.66}$	$19007_{\pm 1361.42}$	$20455_{\pm 753.23}$	$20386_{\pm 1172.33}$
PGCN	$20589_{\pm 2471.59}$	$24905_{\pm 2474.59}$	$36015_{\pm 4679.79}$	$49866_{\pm 12741.13}$	$15091_{\pm 3443.66}$	$22808_{\pm 10693.89}$	$22442_{\pm 10369.52}$	$28393_{\pm 10155.73}$
SGLFT (Ours)	$\mathbf{1937_{\pm 39.21}}$	$\mathbf{1621_{\pm 56.40}}$	$\mathbf{1613_{\pm 20.61}}$	$\mathbf{1623_{\pm 30.44}}$	$\mathbf{3243_{\pm 148.77}}$	$\mathbf{2672_{\pm 106.34}}$	$\mathbf{2405_{\pm 69.51}}$	$\mathbf{2376_{\pm 152.93}}$

TABLE SII
THE TRAINING TIME IN RMSE OF ALL MODELS ON D1.1-2.4.

Method	D1.1	D1.2	D1.3	D1.4	D2.1	D2.2	D2.3	D2.4
EvolveGCN	$84538_{\pm 20142.85}$	$35093_{\pm 25262.91}$	$58340_{\pm 15689.28}$	$40656_{\pm 27933.17}$	$58806_{\pm 35809.91}$	$9798_{\pm 34848.18}$	$19319_{\pm 3154.78}$	$24661_{\pm 12341.47}$
CIGCN	$4027_{\pm 589.60}$	$10260_{\pm 1196.58}$	$10683_{\pm 936.05}$	$10289_{\pm 611.16}$	$6153_{\pm 66.43}$	$12190_{\pm 890.60}$	$11988_{\pm 507.25}$	$11411_{\pm 512.25}$
WD-GCN	$24291_{\pm 14191.35}$	$5129_{\pm 10016.92}$	$15056_{\pm 6901.48}$	$35467_{\pm 24079.55}$	$18673_{\pm 15738.59}$	$18945_{\pm 20859.38}$	$23956_{\pm 24767.50}$	$8136_{\pm 1550.09}$
TM-GCN	$46109_{\pm 33639.89}$	$26475_{\pm 15873.67}$	$33303_{\pm 11543.58}$	$49882_{\pm 39525.67}$	$29018_{\pm 16701.13}$	$16737_{\pm 8158.82}$	$28267_{\pm 17214.76}$	$42964_{\pm 13201.51}$
JMP-GCF	$5617_{\pm 204.96}$	$8397_{\pm 214.70}$	$9333_{\pm 404.88}$	$22277_{\pm 912.55}$	$6565_{\pm 470.06}$	$8203_{\pm 645.05}$	$16721_{\pm 1213.93}$	$32947_{\pm 706.47}$
HMLET	$21843_{\pm 518.51}$	$38881_{\pm 1134.90}$	$39531_{\pm 669.94}$	$37207_{\pm 654.49}$	$32920_{\pm 1037.32}$	$43210_{\pm 481.49}$	$42787_{\pm 266.76}$	$39767_{\pm 1209.17}$
APAN	$15329_{\pm 5076.96}$	$11582_{\pm 1001.44}$	$10765_{\pm 123.86}$	$6960_{\pm 2326.40}$	$20484_{\pm 7310.11}$	$16339_{\pm 7285.46}$	$46902_{\pm 18248.91}$	$28148_{\pm 11543.13}$
TeDGaN	$26602_{\pm 463.54}$	$11679_{\pm 902.26}$	$8069_{\pm 1505.95}$	$6086_{\pm 904.48}$	$16138_{\pm 466.92}$	$9634_{\pm 1284.01}$	$8297_{\pm 682.53}$	$10621_{\pm 2751.03}$
CGTF	$14229_{\pm 3365.60}$	$12656_{\pm 4279.88}$	$11277_{\pm 2820.20}$	$8542_{\pm 2529.83}$	$12678_{\pm 4961.09}$	$15432_{\pm 3503.00}$	$12584_{\pm 2298.57}$	$10202_{\pm 1754.05}$
CTGCN	$61758_{\pm 34233.25}$	$90864_{\pm 18419.57}$	$103337_{\pm 25104.45}$	$97308_{\pm 25695.62}$	$47155_{\pm 22042.09}$	$112161_{\pm 23179.47}$	$91281_{\pm 27542.00}$	$71994_{\pm 17369.16}$
DDSTGCN	$47240_{\pm 8435.04}$	$69057_{\pm 30532.77}$	$70978_{\pm 17323.84}$	$47920_{\pm 9939.55}$	$44777_{\pm 22807.33}$	$77140_{\pm 20233.68}$	$74404_{\pm 13279.02}$	$54178_{\pm 24308.50}$
GRU-GCN	$13700_{\pm 762.17}$	$26299_{\pm 4235.04}$	$67702_{\pm 1637.49}$	$24361_{\pm 9736.12}$	$20738_{\pm 10456.95}$	$52076_{\pm 37049.04}$	$17417_{\pm 8073.00}$	$37925_{\pm 20005.00}$
MegaCRN	$17236_{\pm 3604.35}$	$33660_{\pm 13247.59}$	$32066_{\pm 12533.35}$	$28652_{\pm 11473.37}$	$23460_{\pm 7671.24}$	$40259_{\pm 16250.02}$	$39665_{\pm 16372.89}$	$29056_{\pm 11178.89}$
SGP	$2039_{\pm 1818.11}$	$3369_{\pm 4537.92}$	$7062_{\pm 6420.65}$	$\mathbf{927_{\pm 1020.40}}$	$11466_{\pm 8164.18}$	$6566_{\pm 3961.92}$	$3778_{\pm 2882.37}$	$5397_{\pm 3670.52}$
MGDN	$12760_{\pm 795.62}$	$19261_{\pm 712.82}$	$16901_{\pm 635.89}$	$15914_{\pm 618.38}$	$14912_{\pm 933.85}$	$18997_{\pm 1360.57}$	$20445_{\pm 751.90}$	$20374_{\pm 1169.97}$
PGCN	$18341_{\pm 3853.74}$	$22522_{\pm 3518.37}$	$30483_{\pm 6009.65}$	$42600_{\pm 13046.46}$	$12753_{\pm 3080.75}$	$21103_{\pm 8371.49}$	$19353_{\pm 9329.32}$	$20770_{\pm 7420.22}$
SGLFT (Ours)	$\mathbf{1953_{\pm 42.90}}$	$\mathbf{1639_{\pm 46.64}}$	$\mathbf{1635_{\pm 35.07}}$	$1661_{\pm 57.55}$	$\mathbf{3273_{\pm 195.84}}$	$\mathbf{2679_{\pm 94.06}}$	$\mathbf{2427_{\pm 83.92}}$	$\mathbf{2426_{\pm 210.65}}$

TABLE SIII
FRIEDMAN TEST OUTCOMES REGARDING ESTIMATION ACCURACY AND COMPUTATIONAL EFFICIENCY. (*Higher F-rank represents lower error. **Higher F-rank represents lower training time cost.*)

Method	Accuracy*	Efficiency**	Method	Accuracy*	Efficiency**	Method	Accuracy*	Efficiency**
EvolveGCN	15.03	11.94	APAN	12.16	8.82	MegaCRN	6.88	11.25
CIGCN	12.56	4.38	TeDGaN	2.81	5.318	SGP	11.81	3.44
WD-GCN	11.38	9.19	CGTF	3.81	5.69	MGDN	7.00	7.63
TM-GCN	13.56	13.38	CTGCN	9.13	16.75	PGCN	9.31	10.00
JMP-GCF	13.25	5.25	DDSTGCN	2.38	15.50	**SGLFT (Ours)**	**1.00**	**1.06**
HMLET	7.56	13.00	GRU-GCN	13.38	10.44	—	—	—

TABLE SIV
WILCOXON SIGNED-RANKS TEST OUTCOMES FOR ESTIMATION ERRORS RECORDED IN TABLES IV AND V. (*For SGLFT, higher R+ value corresponds to higher estimation accuracy. **At a significance level of 0.01, the accepted hypotheses are emphasized.*)

Comparision	R+*	R-	p-value**	Comparision	R+*	R-	p-value**
SGLFT vs EvolveGCN	136	0	**2.41E-4**	SGLFT vs CGTF	136	0	**2.41E-4**
SGLFT vs CIGCN	136	0	**2.41E-4**	SGLFT vs CTGCN	136	0	**2.41E-4**
SGLFT vs WD-GCN	136	0	**2.41E-4**	SGLFT vs DDSTGCN	136	0	**2.41E-4**
SGLFT vs TM-GCN	136	0	**2.41E-4**	SGLFT vs GRU-GCN	136	0	**2.41E-4**
SGLFT vs JMP-GCF	136	0	**2.41E-4**	SGLFT vs MegaCRN	136	0	**2.41E-4**
SGLFT vs HMLET	136	0	**2.41E-4**	SGLFT vs SGP	136	0	**2.41E-4**
SGLFT vs APAN	136	0	**2.41E-4**	SGLFT vs MGDN	136	0	**2.41E-4**
SGLFT vs TeDGaN	136	0	**2.41E-4**	SGLFT vs PGCN	136	0	**2.41E-4**

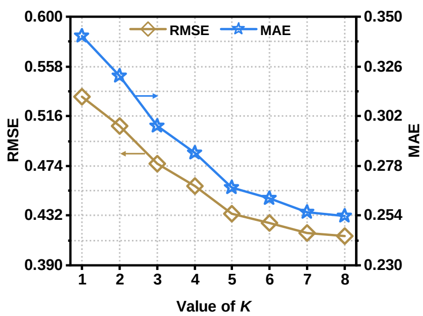
TABLE SV
WILCOXON SIGNED-RANKS TEST OUTCOMES FOR TRAINING TIME RECORDED IN TABLES SI AND SI. (*For SGLFT, higher R+ value corresponds to higher computational efficiency. **At a significance level of 0.01, the accepted hypotheses are emphasized.*)

Comparision	R+*	R-	p-value**	Comparision	R+*	R-	p-value**
SGLFT vs EvolveGCN	136	0	**2.41E-4**	SGLFT vs CGTF	136	0	**2.41E-4**
SGLFT vs CIGCN	136	0	**2.41E-4**	SGLFT vs CTGCN	136	0	**2.41E-4**
SGLFT vs WD-GCN	136	0	**2.41E-4**	SGLFT vs DDSTGCN	136	0	**2.41E-4**
SGLFT vs TM-GCN	136	0	**2.41E-4**	SGLFT vs GRU-GCN	136	0	**2.41E-4**
SGLFT vs JMP-GCF	136	0	**2.41E-4**	SGLFT vs MegaCRN	136	0	**2.41E-4**
SGLFT vs HMLET	136	0	**2.41E-4**	SGLFT vs SGP	134	2	**3.53E-4**
SGLFT vs APAN	136	0	**2.41E-4**	SGLFT vs MGDN	136	0	**2.41E-4**
SGLFT vs TeDGaN	136	0	**2.41E-4**	SGLFT vs PGCN	136	0	**2.41E-4**

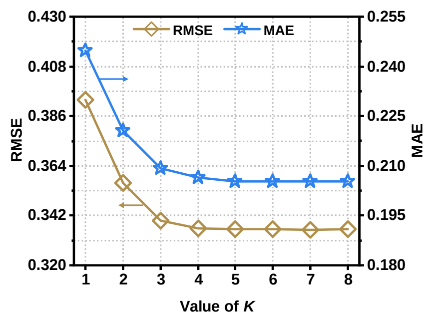
IV. Supplementary Figures (Secs. V.C and V.D)

A. Results of Hyperparameter Sensitivity Test (Sec. V.C)

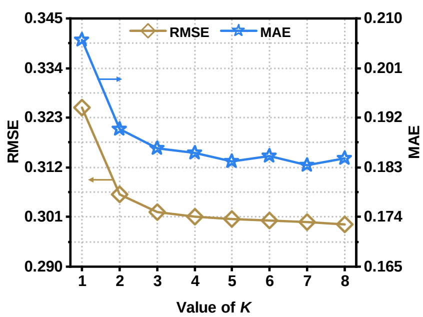
- Fig. S1 (as discussed in Sec. V.C.i) illustrates how the errors of SGLFT vary with changes in K.
- Fig. S2 (as discussed in Sec. V.C.ii) illustrates how the errors of SGLFT vary with changes in S.
- Fig. S3 (as discussed in Sec. V.C.iii) illustrates how the errors of SGLFT vary with changes in F.



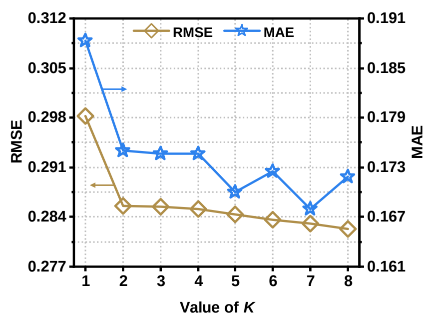
(a) Errors on D1.1



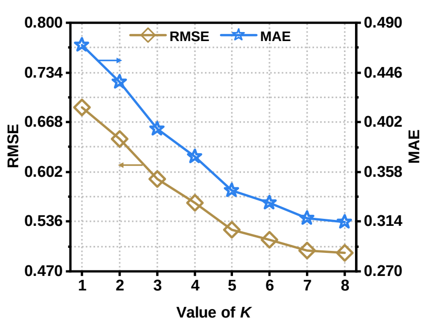
(b) Errors on D1.2



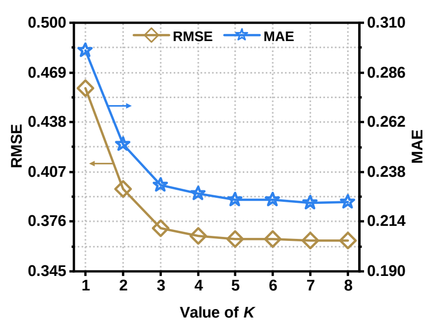
(c) Errors on D1.3



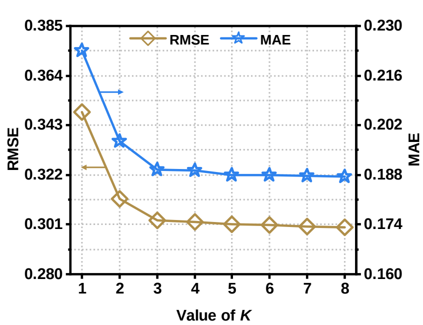
(d) Errors on D1.4



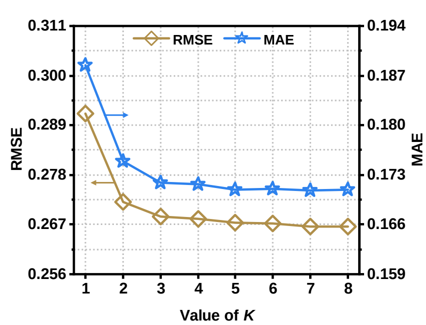
(e) Errors on D2.1



(f) Errors on D2.2

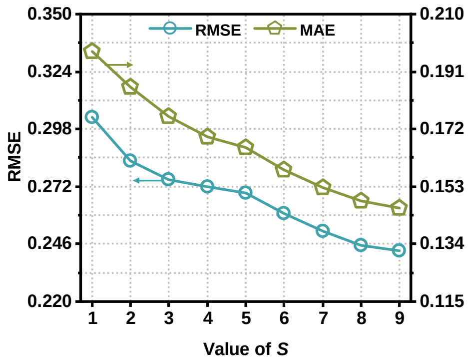


(g) Errors on D2.3

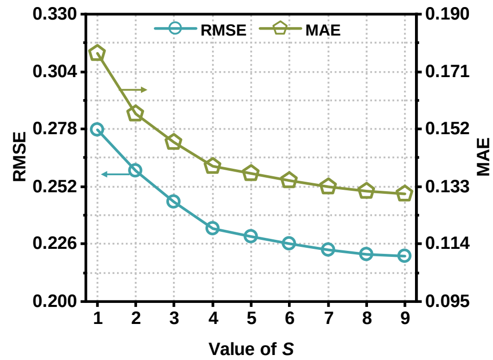


(h) Errors on D2.4

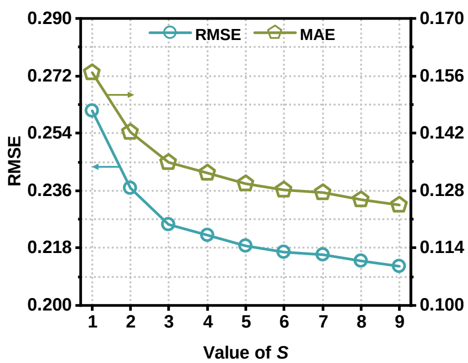
Fig. S1. Errors of SGLFT for different numbers of spatiotemporal graph convolutional layers.



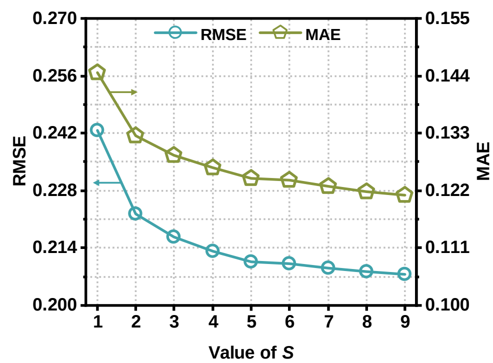
(a) Errors on D1.1



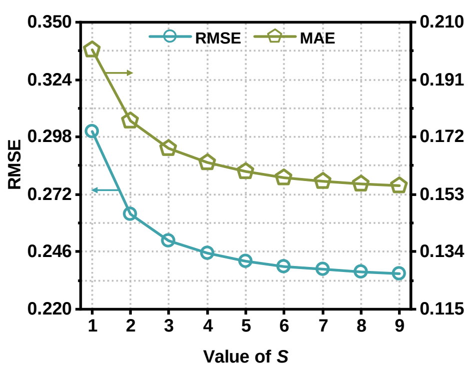
(b) Errors on D1.2



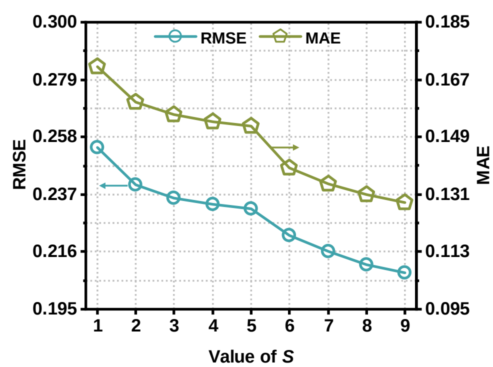
(c) Errors on D1.3



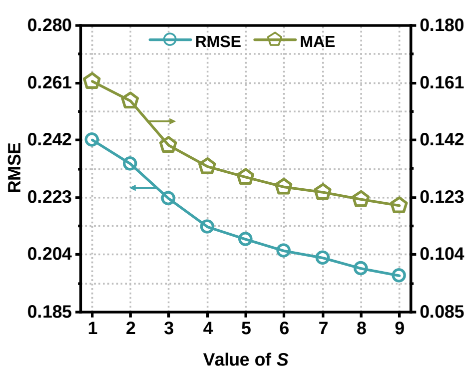
(d) Errors on D1.4



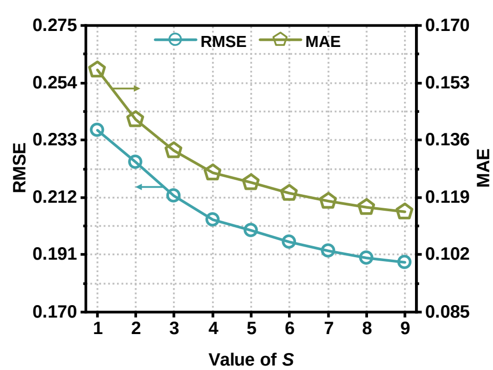
(e) Errors on D2.1



(f) Errors on D2.2

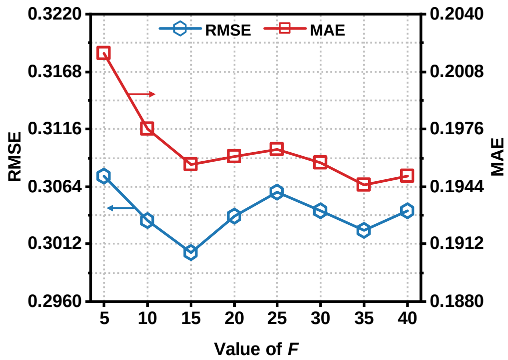


(g) Errors on D2.3

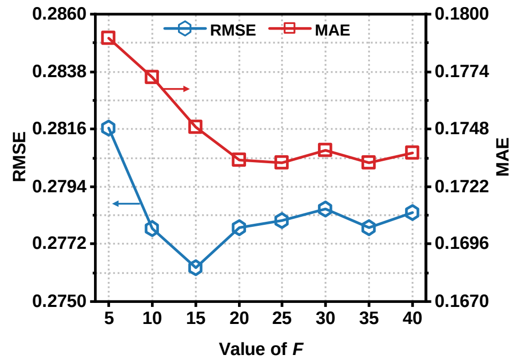


(h) Errors on D2.4

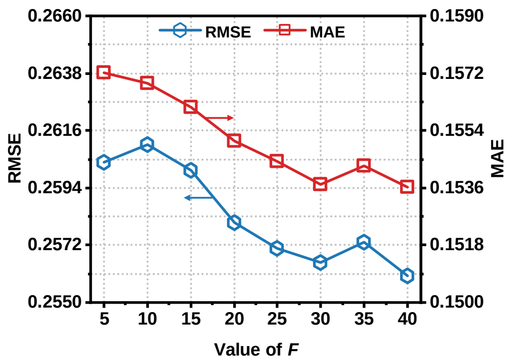
Fig. S2. Errors of SGLFT for different sliding window sizes.



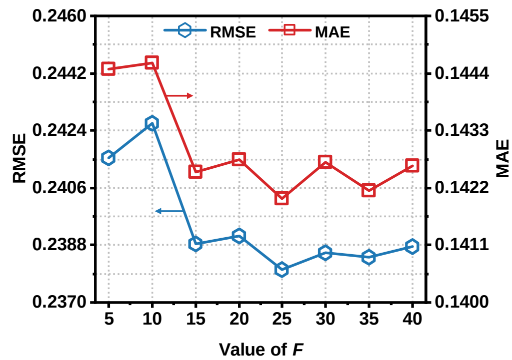
(a) Errors on D1.1



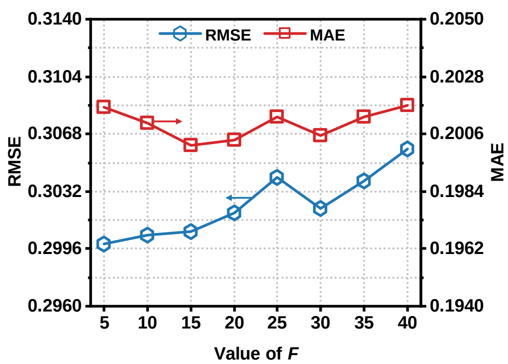
(b) Errors on D1.2



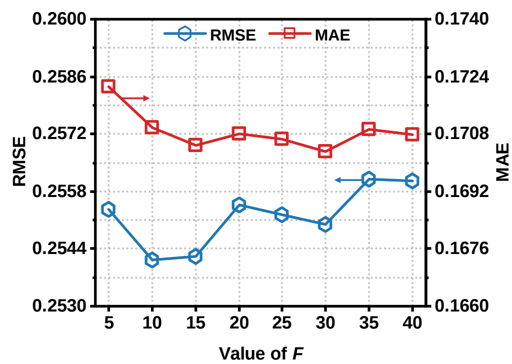
(c) Errors on D1.3



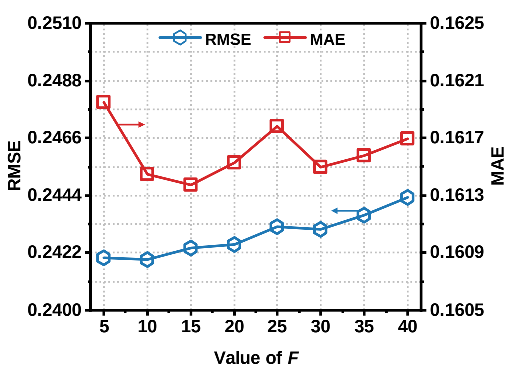
(d) Errors on D1.4



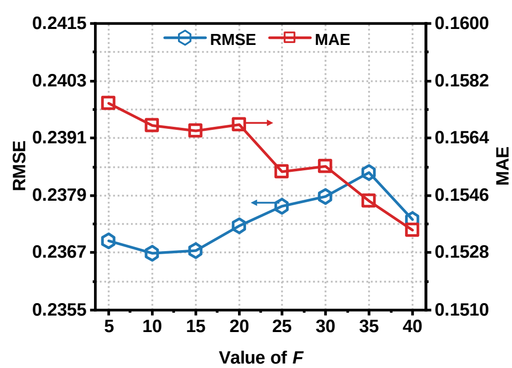
(e) Errors on D2.1



(f) Errors on D2.2



(g) Errors on D2.3

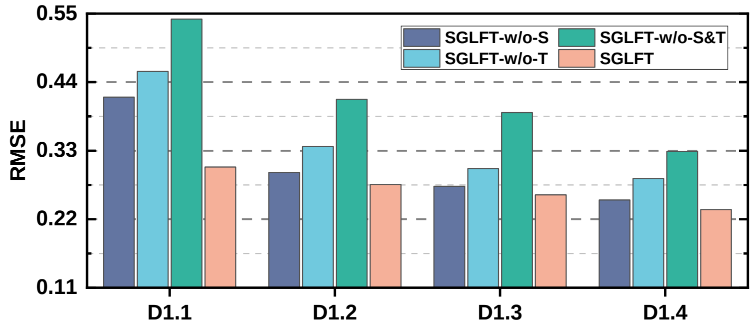


(h) Errors on D2.4

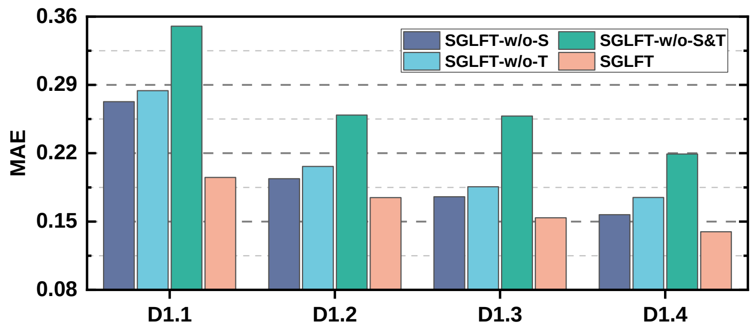
Fig. S3. Errors of SGLFT for different latent feature dimensions.

B. Results of Ablation Studies (Sec. V.D)

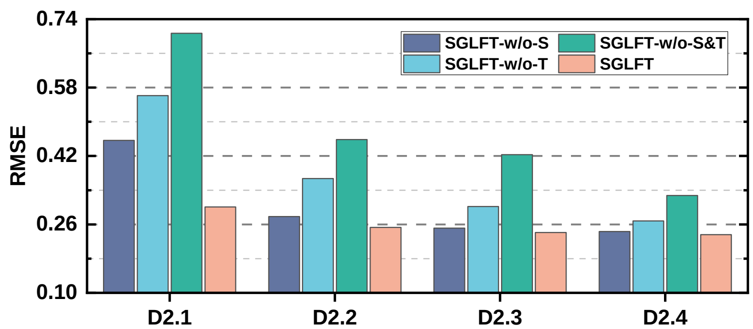
- Fig. S4 (as discussed in Sec. V.D.i) plots the errors of SGLFT and its variants including SGLFT-w/o-S, SGLFT-w/o-T, and SGLFT-w/o-S&T on all testing cases.
- Fig. S5 (as discussed in Sec. V.D.ii) plots the errors of SGLFT and its variants including SGLFT-M, SGLFT-C, SGLFT-S, SGLFT-L, and SGLFT-N on all testing cases.



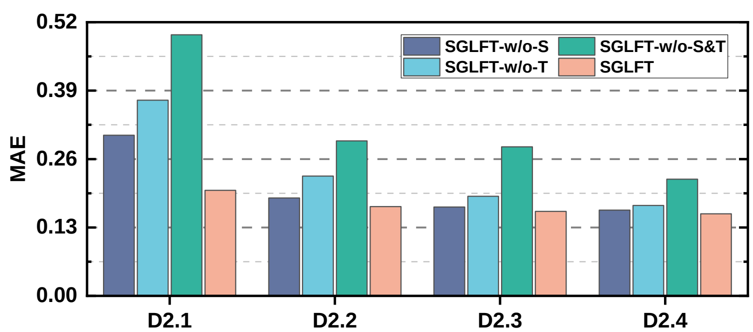
(a) RMSE on D1.1-1.4.



(b) MAE on D1.1-1.4.

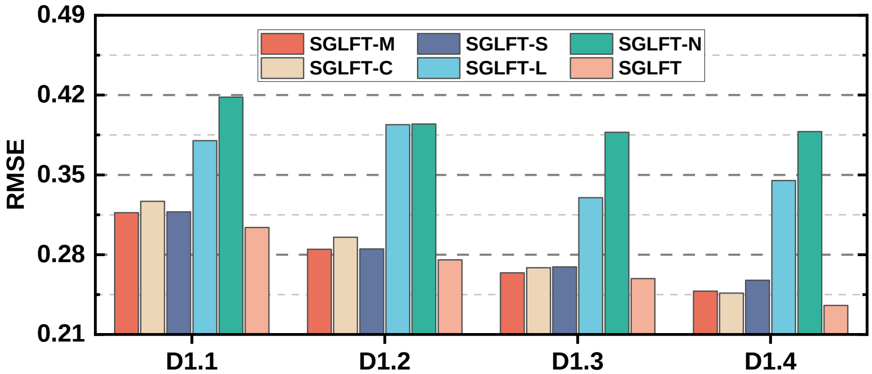


(c) RMSE on D2.1-2.4.

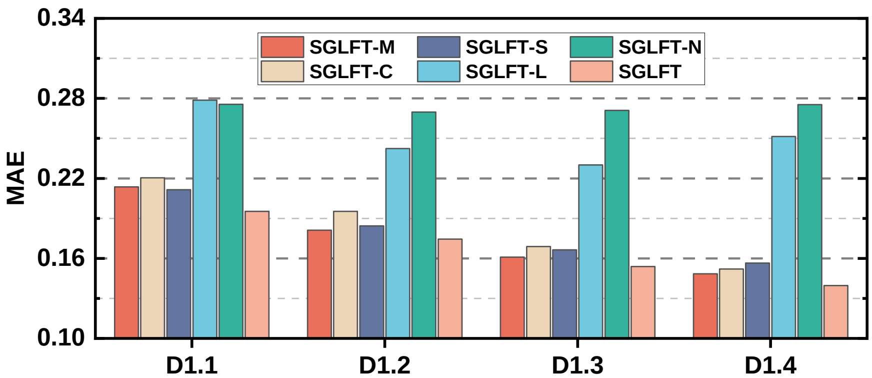


(d) MAE on D2.1-2.4.

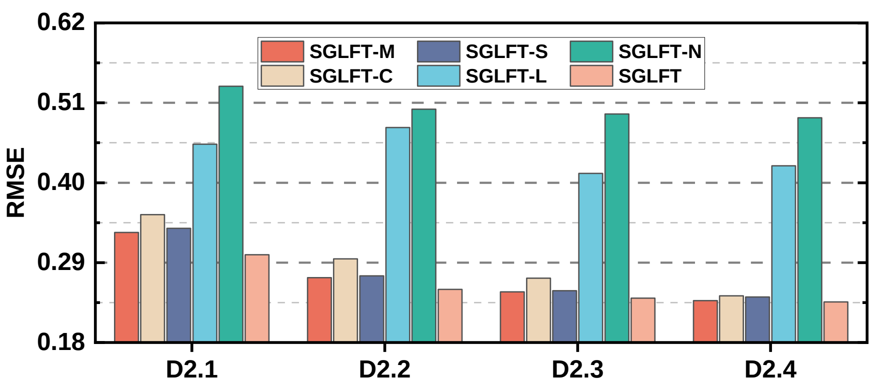
Fig. S4. Errors of SGLFT and its variants omitting the capture of spatial information (SGLFT-w/o-S), temporal information (SGLFT-w/o-T), and both (SGLFT-w/o-S&T) on all testing cases.



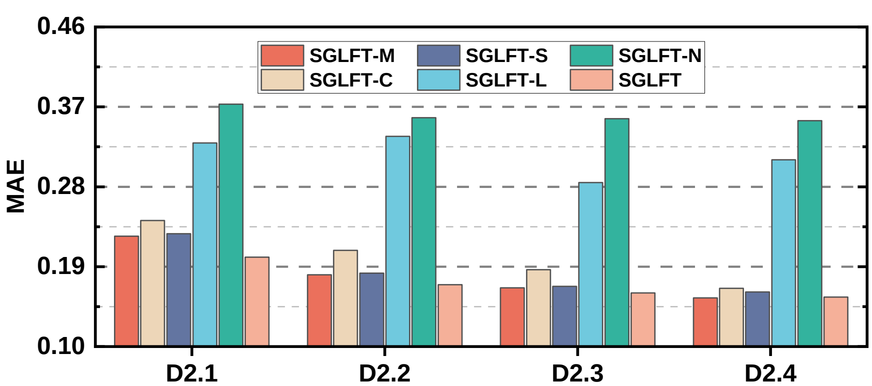
(a) RMSE on D1.1-1.4.



(b) MAE on D1.1-1.4.



(c) RMSE on D2.1-2.4.



(d) MAE on D2.1-2.4.

Fig. S5. Errors of SGLFT and its variants using different layer combination methods on all testing cases.