北京航空航天大學
BEIHANG UNIVERSITY

BIGSCITY
Beihang Interest Group on SmartCity

Transferable Graph Structure Learning for Graph-based Traffic Forecasting Across Cities

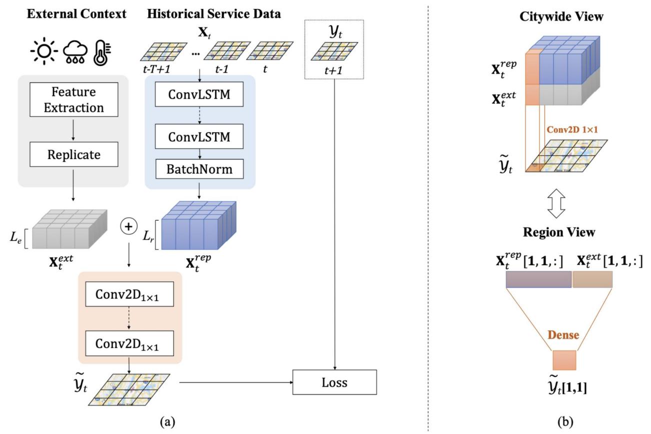
Y. Jin, K. Chen, and Q. Yang, "Transferable graph structure learning for graph-based traffic forecasting across cities," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD ’23)*. New York, NY, USA, 2023, pp. 1032–1043.

交通预测是各种智慧城市应用的基本问题。准确预测未来交通状况是众多智慧城市服务的基础，例如出行规划[21,23,34]、资源管理[5,45,48]、事故预测[13,46]等。交通数据可以通常被建模为时空图，其中传感器对应于节点，节点之间的依赖关系对应于边。

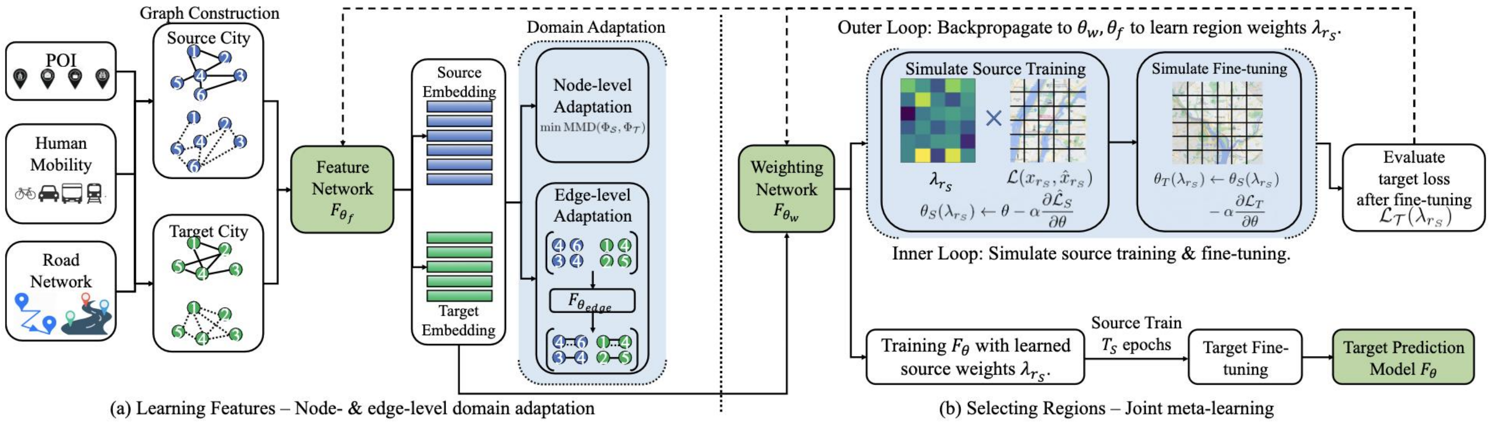
Transfer Learning Methods For Traffic Forecasting

Grid-based Data

Divided into grids with
fixed sizes and spatial relations.



RegionTrans



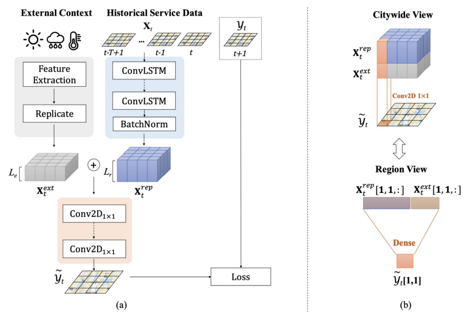
CrossTReS

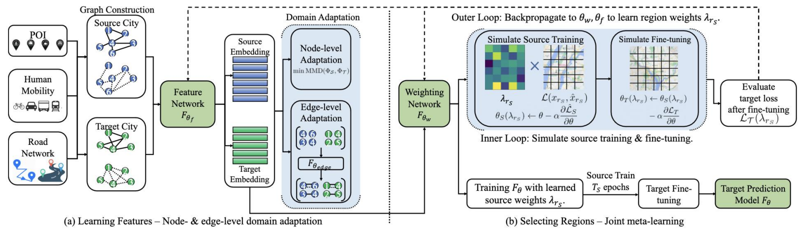
Fail to describe spatial-temporal graphs with irregular and flexible node-wise connections

Transfer Learning Methods For Traffic Forecasting

Grid-based Data

Divided into grids with
fixed sizes and spatial relations.

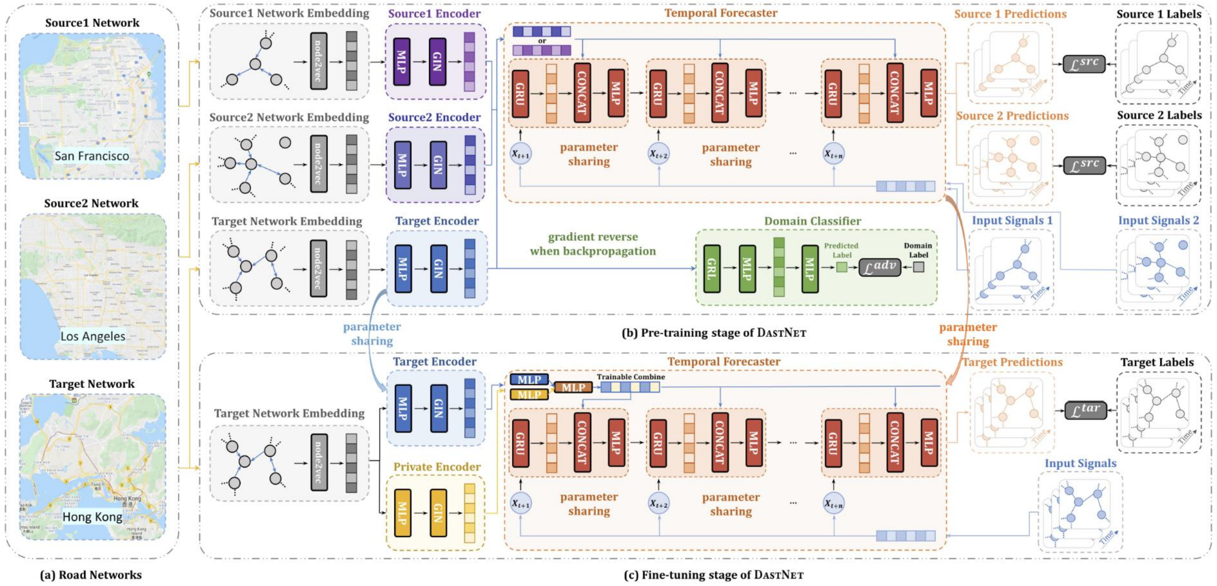




Fail to describe spatial-temporal graphs with
irregular and flexible node-wise connections

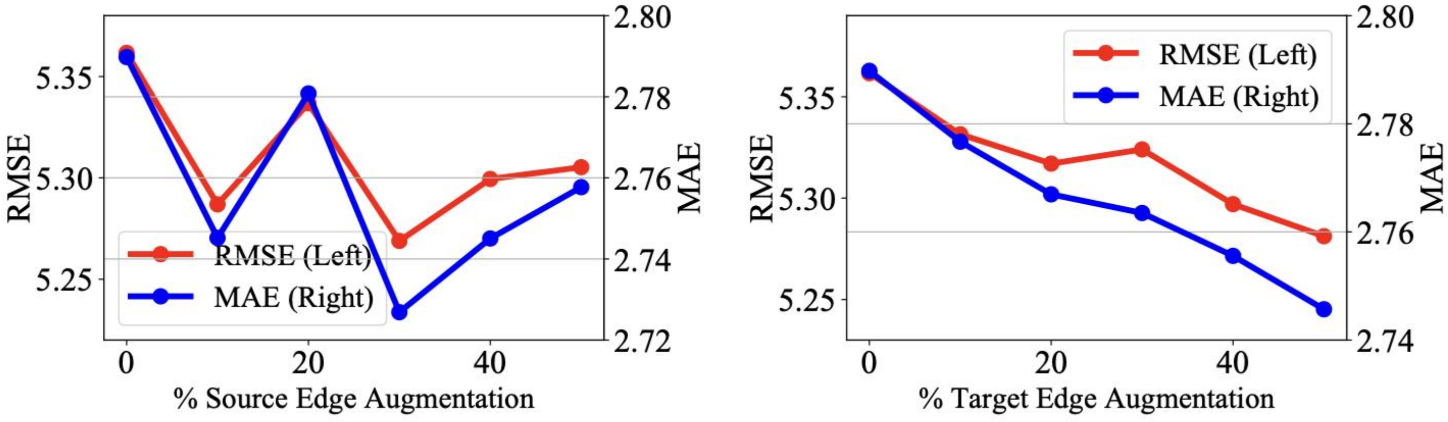
Graph-based Data

Adopt pre-defined graph structures for
knowledge extraction and transfer.



Graph handcrafted with rules, and may thus
be noisy, missing, or biased

Drawback of pre-defined graph



(a) Augmenting the Source Graph

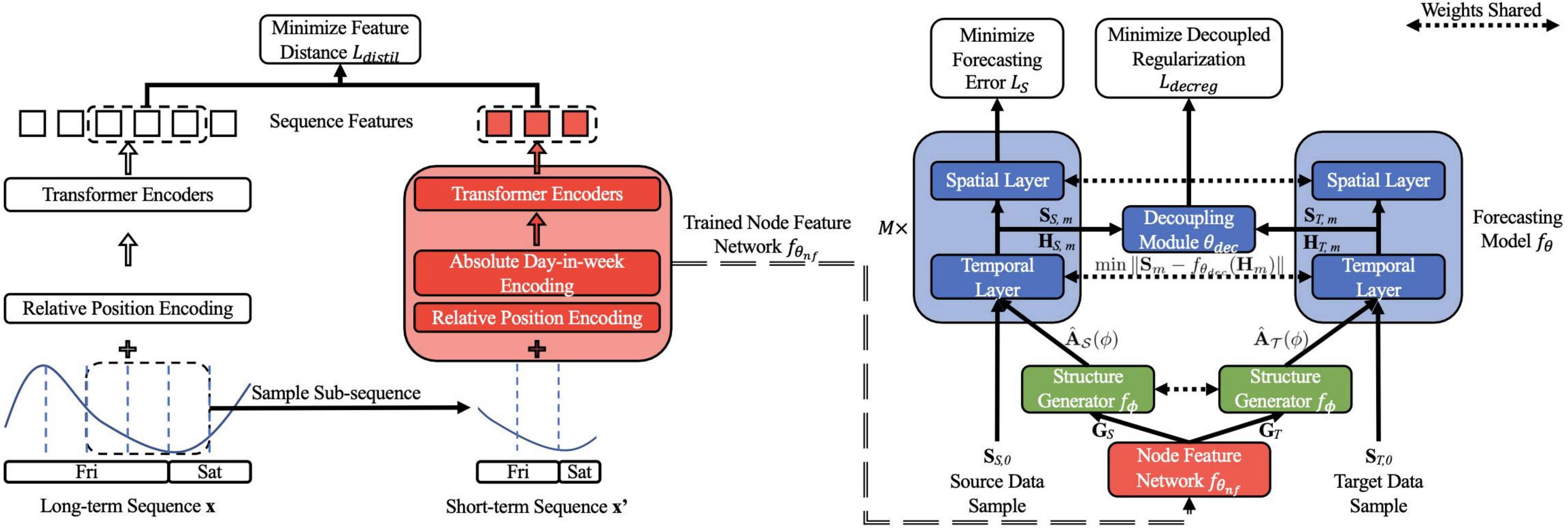
(b) Augmenting the Target Graph

Augment the graph structures of both cities via the triadic closure rule
—— Connect top-� % node pairs with the most common neighbors

Suggest that the pre-defined graph structures, either the source or the target, may not be optimal for knowledge transfer

1. Structure transferred from the source city can better identify helpful node-wise dependencies and learn a more effective target graph.
2. On the other hand, by jointly learning graph structures for both cities, we can narrow the discrepancy between source and target data distributions

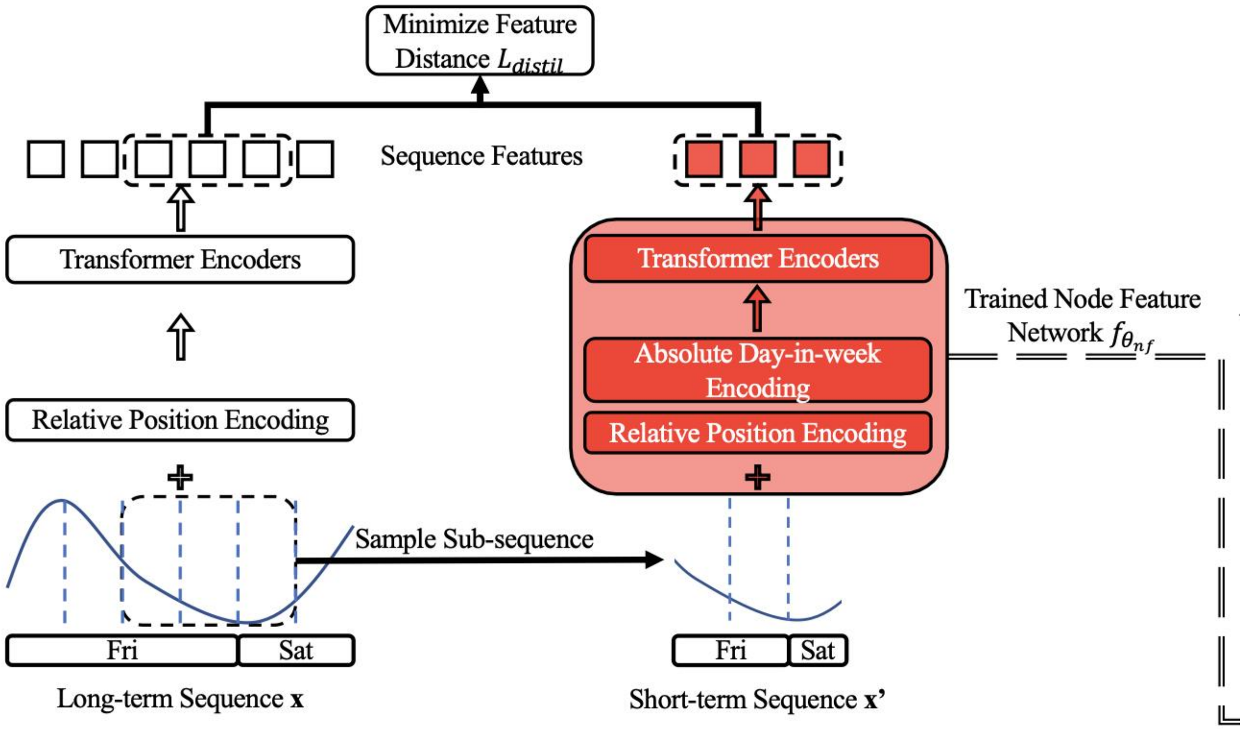
Overview of TransGTR



(a) Learning Node Features with Knowledge Distillation

(b) Learning Graph Structures with Decoupled Regularization

TransGTR - Node Feature Network



(a) Learning Node Features with Knowledge Distillation

Given an input sequence $x \in \mathbb{R}^{L \cdot P}$,

TSFormer $f_{\theta_{nf}}$

- Split x into patches of length P
- Project them into patch embeddings

$x_{emb} \in \mathbb{R}^{L \times n_{emb}}$

- Feed into a series of Transformer blocks

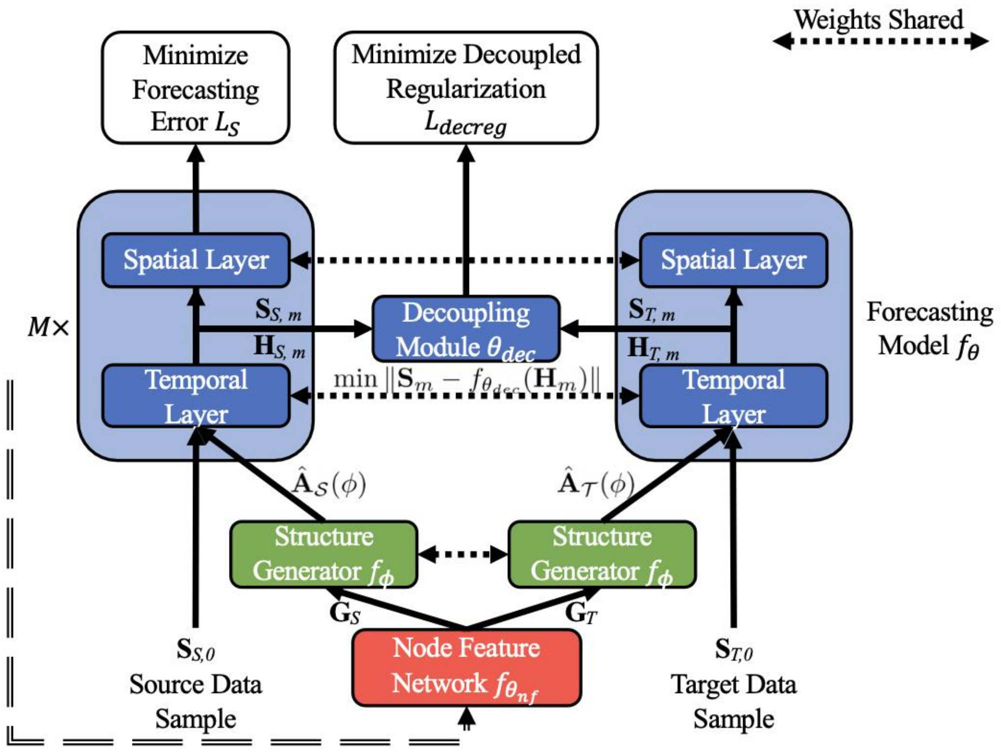
Denote the outputs of the encoder

as $x_{emb} \in \mathbb{R}^{L \times n_{emb}}$

* Enhancing weekly periodicity with day-in-week encodings.

$$\mathbf{pe}_{diw}(\mathbf{x}) = \mathbf{e}_{diw}[\mathbf{t}_{diw}], \tag{4}$$

TransGTR - Structure Generator &Forecasting Model



(b) Learning Graph Structures with Decoupled Regularization

Structure Generator $\mathbf{f_\emptyset}$

- Take the node features learned by $f_{\theta_{nf}}$

$G_S = f_{\theta_{nf}}(X_S)$

$G_T = f_{\theta_{nf}}(X_T)$

- Transform them into graph structures $\widehat{A}_S(\phi)$, $\widehat{A}_T(\phi)$for both cities.

Forecasting Model $\boldsymbol{f_\theta}$

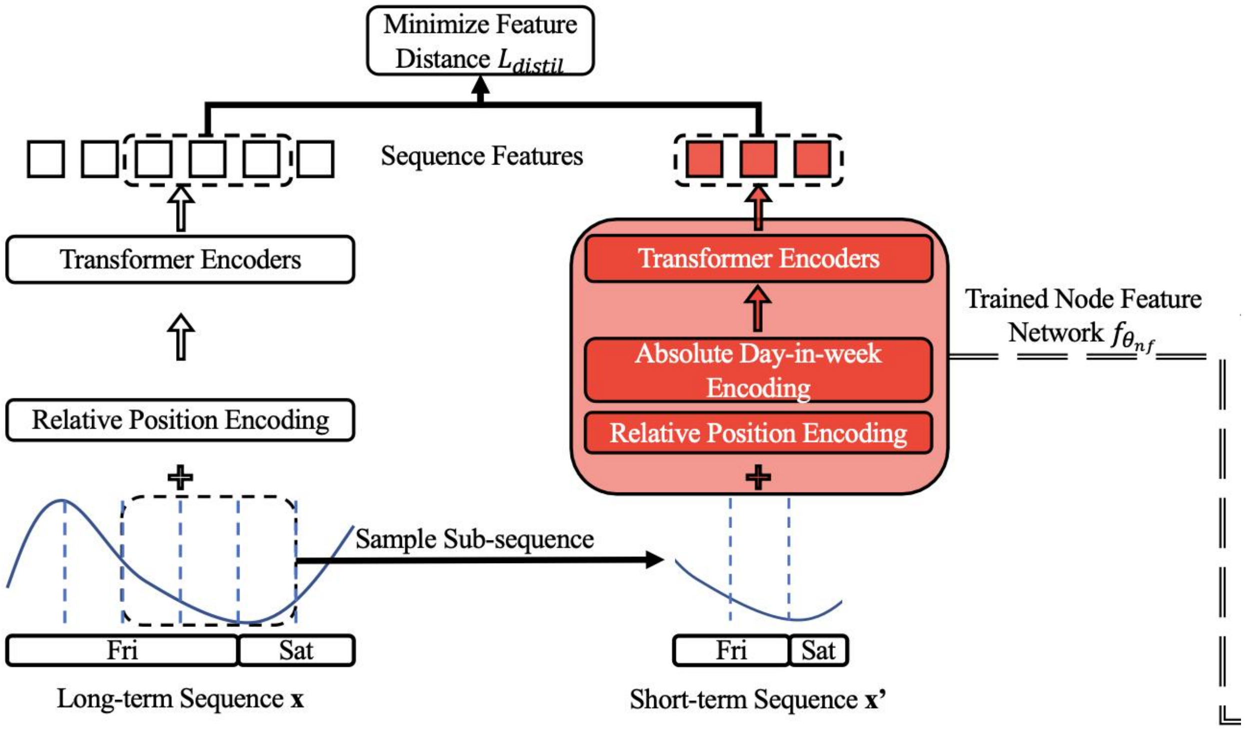
- Given input data and the graph structure $\widehat{A}\ \ (\phi)$
- Model transforms them into predictions

Assume that f_θ consists of � stacked spatial and temporal layers, i.e.

$\mathrm{H}_m = \mathrm{TemporalLayer}_m(\mathrm{S}_{m-1})$,

$\mathrm{S}_m = \mathrm{GNNLayer}_m(\mathrm{H}_m, \widehat{A}\ \ (\phi)), m = 1, \ldots M,$

TransGTR - City-agnostic Node Features



(a) Learning Node Features with Knowledge Distillation

Learning City-agnostic Node Features via Knowledge Distillation

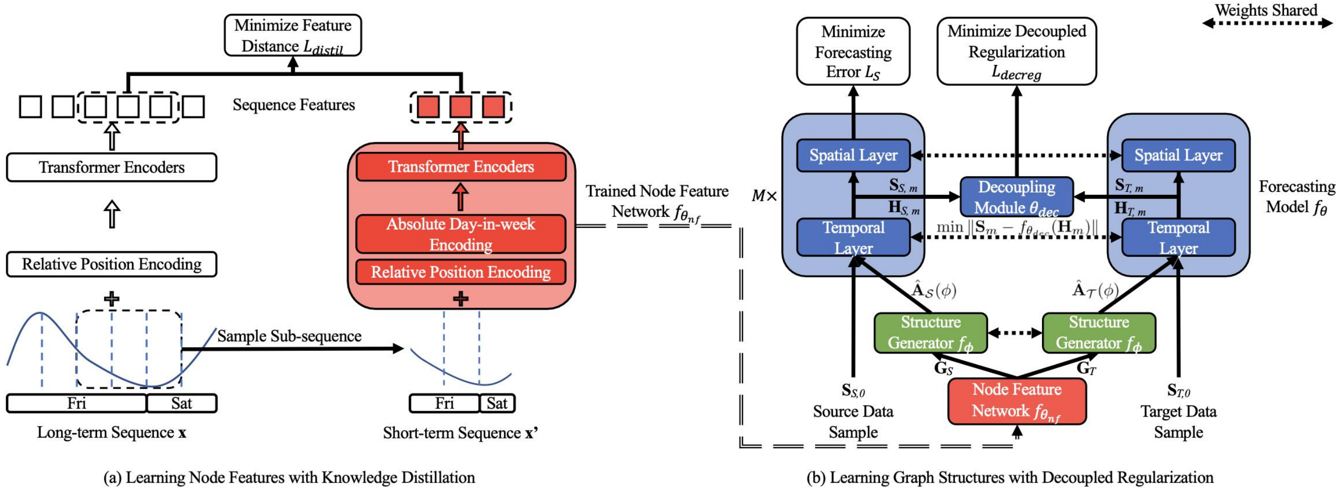
- Follow STEP* to pre-train node feature network
- Distill the rich knowledge encoded
  - Given a long-term sequence, obtain its corresponding short-term sequence $x' \in \mathbb{R}^{L_{short} \cdot P}$

$$\mathbf{x}_{enc} = f_{\theta_{nf,S}}(\mathbf{x}) \in \mathbb{R}^{L \times n_{emb}}, \quad \mathbf{x}'_{enc} = f_{\theta_{nf}}(\mathbf{x}') \in \mathbb{R}^{L_{short} \times n_{emb}}. \tag{8}$$

$$\mathcal{L}_{distil}(\mathbf{x}) = \left\| \mathbf{x}'_{enc} - \mathbf{x}_{enc}[p : p + L_{short}] \right\|^2, \tag{9}$$

Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. 2022. Pre-Training Enhanced Spatial-Temporal Graph Neural Network for Multivariate Time Series Forecasting. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. ACM, New York, NY, USA, 1567–1577.

TransGTR - Decoupled Regularization



(a) Learning Node Features with Knowledge Distillation

(b) Learning Graph Structures with Decoupled Regularization

Learning Graph Structures via Temporal Decoupled Regularization.

- Propose a spatial feature regularization term to minimize the distance between spatial features

$$\mathcal{L}_{reg} = \sum_{m=1}^{M} d\left(\mathbf{S}_{\mathcal{S},m}, \mathbf{S}_{\mathcal{T},m}\right), \quad (11)$$

- Robust Regularization via Temporal Decoupling (For I.I.D)
  - Data samples from different time steps do not necessarily follow the same distribution.
  - Train the decoupling modules to reconstruct Sm with its preceding temporal features Hm

$$\min_{\{\theta_{dec,m}\}_{m=1,\ldots M}} \mathcal{L}_{recons} = \sum_{m=1}^{M} \left\|\mathbf{S}_m - f_{\theta_{dec,m}}(\mathbf{H}_m)\right\|^2. \quad (12)$$

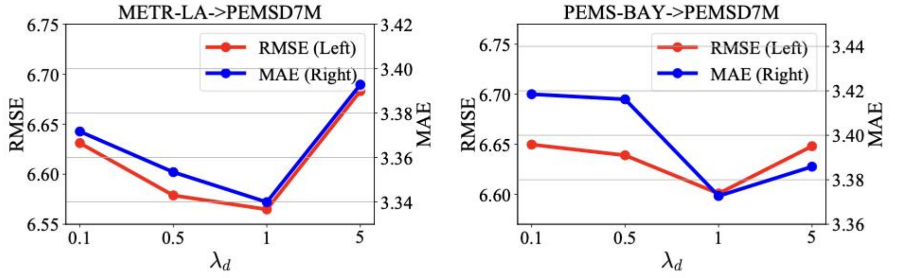
Experiments

Table 1: Comparative evaluation results with PEMSD7M and HKTD as target cities. LA and BAY stand for METR-LA and PEMS-BAY as source cities, respectively. In each column, the best result is presented in bold and the second best is underlined.

Target City	Baselines	Target Data	7-day								3-day							
		Horizon	30 mins				60 mins				30 mins				60 mins			
		Metrics	RMSE		MAE		RMSE		MAE		RMSE		MAE		RMSE		MAE	
PEMSD7M	Target Only	ARIMA	6.525		3.682		8.942		5.426		6.526		3.698		8.946		5.453	
		GWN	5.748		2.999		7.279		3.824		6.053		3.126		7.994		4.162	
		GTS	5.639		2.988		7.071		3.746		5.831		3.111		7.508		4.014	
	Transfer	**Source**	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY
		FT-GWN	5.645	5.771	2.913	2.970	7.038	7.128	3.636	3.690	5.873	5.935	3.045	3.077	7.349	7.596	3.845	3.978
		FT-GTS	5.685	5.651	2.908	2.901	6.952	6.899	3.526	3.543	5.946	5.986	3.024	3.078	7.203	7.205	3.736	3.749
		RegionTrans	5.654	5.702	2.909	2.935	6.986	7.077	3.597	3.659	5.868	5.948	3.046	3.073	7.376	7.545	3.862	3.963
		DASTNet	5.659	5.633	2.901	2.905	6.976	6.954	3.553	3.599	5.839	5.908	3.031	3.078	7.245	7.294	3.774	3.811
		ST-GFSL	5.647	5.642	2.941	2.927	6.937	6.931	3.535	3.541	5.840	5.912	3.012	3.071	7.219	7.218	3.738	3.744
		TransGTR	**5.461**	**5.454**	**2.800**	**2.802**	**6.565**	**6.601**	**3.340**	**3.373**	**5.627**	**5.679**	**2.960**	**2.958**	**6.922**	**6.931**	**3.604**	**3.599**
		Std. Dev.	0.024	0.015	0.019	0.007	0.041	0.022	0.028	0.008	0.029	0.040	0.017	0.016	0.040	0.053	0.026	0.020
HKTD	Target Only	ARIMA	6.648		3.816		8.249		4.843		6.650		3.822		8.253		5.863	
		GWN	6.062		3.386		7.206		4.000		6.333		3.477		7.727		4.333	
		GTS	6.052		3.380		6.954		3.903		6.252		3.453		7.249		4.011	
	Transfer	**Source**	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY	LA	BAY
		FT-GWN	5.755	5.792	3.237	3.264	6.552	6.551	3.682	3.728	5.939	5.949	3.351	3.373	6.984	6.927	3.922	3.980
		FT-GTS	5.792	5.796	3.242	3.253	6.420	6.496	3.633	3.682	5.999	5.982	3.369	3.351	6.784	6.849	3.854	3.862
		RegionTrans	5.696	5.728	3.216	3.228	6.424	6.456	3.654	3.683	5.935	5.943	3.342	3.345	6.870	6.894	3.880	3.933
		DASTNet	5.690	5.704	3.200	3.221	6.411	6.442	3.619	3.655	**5.905**	5.921	3.379	3.361	6.786	6.798	3.881	3.862
		ST-GFSL	5.704	5.739	3.225	3.231	6.477	6.435	3.624	3.638	5.960	5.993	3.392	3.388	6.847	6.821	3.869	3.878
		TransGTR	**5.666**	**5.661**	**3.141**	**3.140**	**6.205**	**6.232**	**3.441**	**3.455**	5.928	**5.877**	**3.305**	**3.290**	**6.622**	**6.589**	**3.693**	**3.697**
		Std. Dev.	0.018	0.016	0.007	0.018	0.026	0.022	0.013	0.017	0.019	0.018	0.010	0.011	0.031	0.025	0.011	0.010

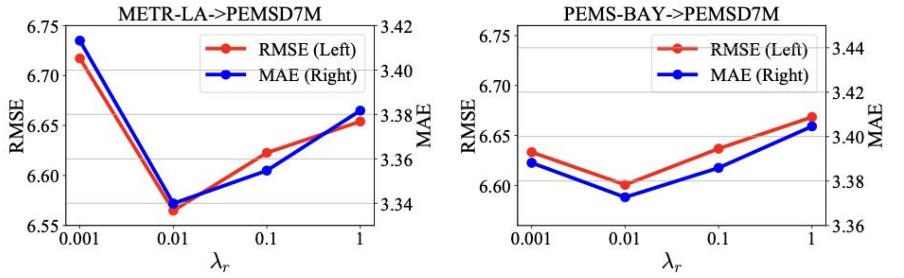
Table 2: Results of Model Analysis. The target city is chosen as PEMSD7M with 7-day data.

Analyzed Component	Source	METR-LA				PEMS-BAY			
	Horizon	30 mins		60 mins		30 mins		60 mins	
	Metric	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Node Feature Learning	TransGTR-NoDistil	5.573	2.869	6.742	3.439	5.545	2.846	6.761	3.474
	TransGTR-NoWP	5.519	2.837	6.671	3.382	5.559	2.833	6.715	3.423
Graph Structure Learning	TransGTR-NoSL	5.645	2.913	7.038	3.636	5.771	2.970	7.128	3.690
	TransGTR-NoReg	5.591	2.860	6.764	3.434	5.591	2.873	6.778	3.460
	TransGTR-NoDec	5.552	2.843	6.693	3.415	5.529	2.840	6.729	3.428
	TransGTR	**5.461**	**2.800**	**6.565**	**3.340**	**5.454**	**2.802**	**6.601**	**3.373**



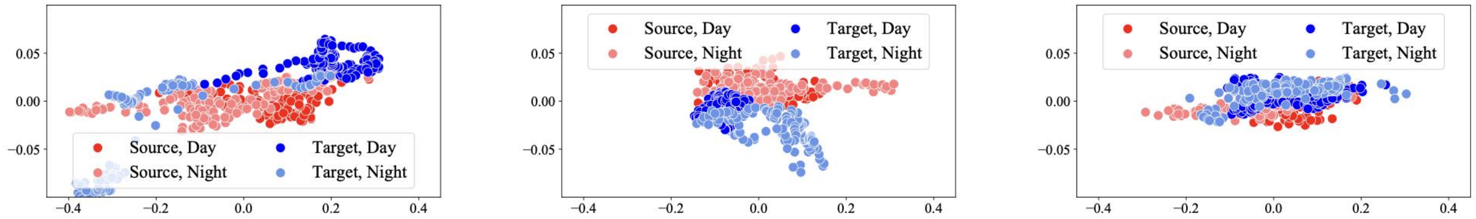
(a) METR-LA to PEMSD7M (b) PEMS-BAY to PEMSD7M

Figure 3: Results of parameter analysis on λ_d, from both METR-LA and PEMS-BAY to PEMSD7M. The reported metrics are evaluated with a forecasting horizon of 60 minutes.



(a) METR-LA to PEMSD7M (b) PEMS-BAY to PEMSD7M

Figure 4: Results of parameter analysis on λ_r, from both METR-LA and PEMS-BAY to PEMSD7M. The reported metrics are evaluated with a forecasting horizon of 60 minutes.



(a) $\lambda_r = 0$, spatial features

(b) $\lambda_r = 0$, residual spatial features

(c) $\lambda_r = 0.01$, residual spatial features

Figure 5: Visualization of spatial features S_M and residual spatial features $\tilde{S}_M$ obtained from TransGTR with $\lambda_r = 0$ and $\lambda_r = 1$. Red dots represent source features, while blue dots represent target features. In addition, dark dots represent day-time features, while light dots represent night-time features. The axes in all sub-figures are of the same range.

Table 3: Mean RMSE and MAE (± std. dev. *within the graph*) between connected node pairs in different graph structures.

Graph Structures	RMSE	MAE
Random	14.90±4.63	10.00±3.49
Pre-defined	13.86±5.21	9.27±3.79
Target-only	13.80±4.19	9.21±3.12
TransGTR	12.85±4.17	8.43±3.22

Thanks.