Supplementary Materials of "A Novel Hybrid Graph Learning Method for Inbound Parcel Volume Forecasting in Logistics System"

Lisha Ye*†‡ Jianfeng Zhou *† Zhe Yin* Kunpeng Han* Haoyuan Hu*
Dongjin Song§‡

1 Introduction

In this documents, we present more detailed ablation experimental results and complexity analysis.

2 More Ablation Experiments

Beyond the results presented in the ablation study of section 5.4.2 in the main text, we conducted further experiments on each module of our method, as well as on combinations and substitutions of these modules. From the individual experiments on the GTD, TA, and TCG modules, it is evident that GTD and TA are the most effective modules. When more modules are combined, such as the integration of GTD and TA, it produces better results than the individual modules alone. Naturally, the combination of all three modules achieves the best performance. Considering the GTD module in isolation, using either the distance matrix or the transition matrix alone does not perform as well as utilizing a combination of both matrices. When we replace the graph diffusion module with a classical GCN [1], there is a significant decrease in performance, demonstrating the effectiveness of the graph diffusion module designed with the intrinsic characteristics of the logistic system.

Table 1: Detailed results from ablation experiments. The performance of our method in this paper are marked in bold. 'sub GTD' refers to the substitution of the modified DTW distance proposed in our model with the original DTW distance. 'sub TA' represents the replacement of the bit-embedding based attention, utilized for integer tokens in the TA module, with the original attention mechanism (same as section 5.4.2 of the main text).

Methods	MAE	MAPE	RMSE
GCN(replace graph diffusion)	9670	0.338	69242
GTD(only distance matrix)	7883	0.261	61099
GTD(only transition matrix)	7978	0.258	61248
GTD	7794	0.248	60910
TA	7854	0.249	60885
TCG	8458	0.271	61561
GTD + TA (w/o.TCG)	7364	0.242	60683
GTD + TCG (w/o.TA)	8030	0.265	60821
TA + TCG (w/o.GTD)	7699	0.245	61273
sub GTD + TA + TCG	7450	0.242	60937
GTD + sub TA + TCG	7470	0.250	60854
GTD+TA+TCG	**7341**	**0.235**	**60358**

*Cainiao Network. Email: {yelisha.yls, zhoujianfeng.zjf, dimon.yz, kunpeng.hkp, haoyuan.huhy}@cainiao.com.
†Both authors contributed equally to this research.
‡Corresponding authors.
§University of Connecticut. Email: dongjin.song@uconn.edu.

3 Complexity Analysis

In Table 2, we outline the time complexity of each component of our method. Here, N denotes the number of nodes, m and n signify the lengths of the normalized historical inbound parcel volume sequences, T_h is the length of the historical input window, and M represents the dimension of the embedding, consistent with the main text. The DTW algorithm, which calculates the similarity between pairings of temporal sequences, has a complexity of $O(mn)$. Since the algorithm processes all N^2 node pairs, the overall complexity of constructing the distance matrix is $O(mnN^2)$. When calculating the transition matrix, the computational overhead arises mainly from the Hadamard product and a MLP layer. The Hadamard product possesses a complexity of $O(N^2)$, necessitated by its execution $O(T_h)$ times. Hence, the time complexity for computing the transition matrix amounts to $O(T_hN^2)$. Graph diffusion convolution, which entails matrix multiplication, carries a complexity of $O(N^3)$ due to its operations involving the distance and transition matrices. The TD component's complexity is primarily a result of an MLP layer, with a computational cost of $O(T_hMN)$. Overall, the complexity of GTD module is primarily governed by the calculation of the distance matrix. The TA module has a complexity of $O(MN^2)$, deriving from the attention calculations performed for pairwise node comparisons. Lastly, the TCG module exhibits a linear complexity in terms of the number of nodes or the length of the input sequence, $O(T_hN)$, due to its reliance on dilated convolutions and GRU.

Table 2: The time complexity of each module/sub module of our method

Module	Sub module	Time complexity
GTD	Distance matrix	$O(mnN^2)$
	Transition matrix	$O(T_hN^2)$
	Diffusion Conv	$O(N^3)$
	TD	$O(T_hMN)$
TA	Self-attention	$O(MN^2)$
TCG	Dilated Conv	$O(T_hN)$
	GRU	$O(T_hN)$

References

[1] T. N. Kipf and M. Welling, *Semi-supervised classification with graph convolutional networks*, arXiv preprint arXiv:1609.02907, (2016).