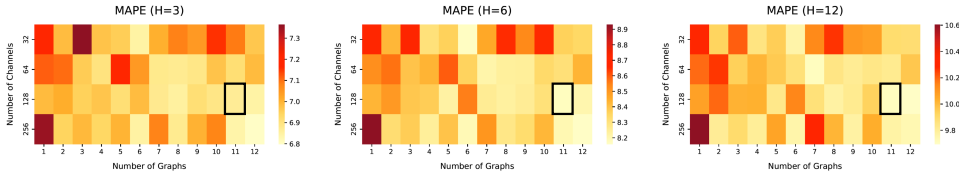
Appendix A. Additional Experimental Results

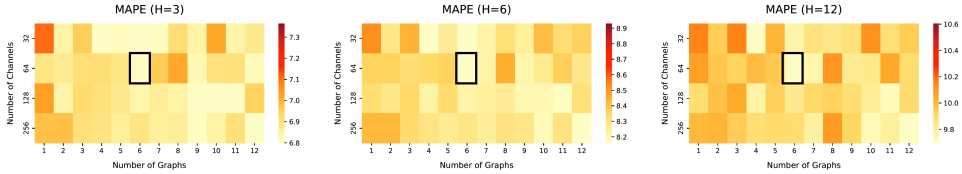
Data	Models			15 min ($3H$)			30 min ($6H$)			60 min ($12H$)		
				MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA	Statistical		HA	4.16	7.80	13.0	4.16	7.80	13.0	4.16	7.80	13.0
			AR	3.81	8.80	9.13	4.94	11.14	12.17	6.30	12.91	16.72
			VAR	4.42	7.89	10.2	5.41	9.13	12.7	6.52	10.11	15.8
			ARIMA	3.99	8.21	9.60	5.15	10.45	12.70	6.90	13.23	17.40
			SVR	3.99	8.45	9.3	5.05	10.87	12.1	6.72	13.76	16.7
	DNN-based		FNN	3.99	7.94	9.9	4.23	8.17	12.9	4.49	8.69	14.0
			FC-LSTM	3.44	6.30	9.60	3.77	7.23	10.90	4.37	8.69	13.20
			WaveNet	2.99	5.89	8.04	3.59	7.28	10.25	4.45	8.93	13.62
	GNN-based	w/o GL	GW	2.72	5.23	7.00	3.11	6.27	8.52	3.58	7.39	10.26
			DCRNN	2.77	5.38	7.30	3.15	6.45	8.80	3.60	7.60	10.50
			ASTGCN	3.01	5.85	8.16	3.53	7.14	10.16	4.25	8.60	12.80
			STSGCN	3.31	7.62	8.06	4.13	9.77	10.29	5.06	11.66	12.91
		with GL	GW	2.77	5.36	7.33	3.14	6.30	8.80	3.55	7.28	10.35
			GTS	2.88	5.55	7.58	3.27	6.54	9.16	3.69	7.53	10.74
			NRI	2.99	5.89	7.91	3.59	7.24	10.23	4.44	8.89	13.76
			DDGCRN†	3.68	9.47	8.27	4.76	12.14	10.30	6.12	14.85	12.57
			MAD-GL^2	2.70 ±0.01	5.19 ±0.01	6.91 ±0.05	**3.06** ±0.01	**6.19** ±0.03	8.32 ±0.11	**3.49** ±0.01	7.26 ±0.05	9.93±0.18
			MAD-GL^{2*}	**2.67** ±0.00	**5.14** ±0.01	**6.76** ±0.04	3.08 ±0.01	**6.14** ±0.03	**8.09** ±0.07	**3.49** ±0.01	**7.18** ±0.03	**9.64** ±0.07
PEMS-BAY	Statistical		HA	2.88	5.59	6.8	2.88	5.59	6.8	2.88	5.59	6.8
			AR	1.59	3.41	3.27	2.15	4.87	4.77	2.97	6.65	7.03
			VAR	1.74	3.16	3.6	2.32	4.25	5.0	2.93	5.44	6.5
			ARIMA	1.62	3.30	3.50	2.33	4.76	5.40	3.38	6.50	8.30
			SVR	1.85	3.59	3.8	2.48	5.18	5.5	3.28	7.08	8.0
	DNN-based		FNN	2.20	4.42	5.19	2.30	4.63	5.43	2.46	4.98	5.89
			FC-LSTM	2.05	4.19	4.80	2.20	4.55	5.20	2.37	4.96	5.70
			WaveNet	1.39	3.01	2.91	1.83	4.21	4.16	2.35	5.43	5.87
			STTN	1.36	2.87	2.89	1.67	3.79	3.78	1.95	4.50	4.58
	GNN-based	w/o GL	GW	1.32	2.77	2.77	1.66	3.76	3.79	1.97	4.56	4.76
			DCRNN	1.38	2.95	2.90	1.74	3.97	3.90	2.07	4.74	4.90
			ASTGCN	1.52	3.13	3.22	2.01	4.27	4.48	2.61	5.42	6.00
			STSGCN	1.44	3.01	3.04	1.83	4.18	4.17	2.26	5.21	5.40
		with GL	GW	1.34	2.83	2.79	1.68	3.80	3.81	2.01	4.57	4.81
			GTS	1.39	2.88	2.91	1.74	3.86	3.96	2.05	4.57	4.88
			NRI	1.40	3.01	2.98	1.84	4.22	4.26	2.37	5.45	6.00
			DDGCRN†	1.35	2.91	2.87	1.68	3.84	3.83	1.99	4.59	4.74
			MAD-GL^2	1.32 ±0.00	2.78 ±0.01	**2.75** ±0.02	**1.64** ±0.01	**3.70** ±0.02	**3.69** ±0.04	**1.93** ±0.01	**4.41** ±0.03	**4.54** ±0.03
			MAD-GL^{2*}	**1.30** ±0.00	**2.73** ±0.01	**2.72** ±0.01	**1.62** ±0.01	**3.66** ±0.02	**3.65** ±0.02	**1.92** ±0.01	**4.40** ±0.03	**4.54** ±0.04

Table A.1: The results of multi-step forecasting task on traffic domain datasets. GNN-based methods without graph learning (w/o GL) utilize an auxiliary graph given in the datasets. Graph learning methods (with GL) does not utilize the auxiliary graph along with the learned graphs, except for our method marked with $*$. We use the same prediction module for all methods except for the method marked with †, which uses its own proposed prediction module. For our methods, we report the mean and standard deviation over ten trials. For graph learning methods, we use the same prediction module, stacked dilated causal convolutions proposed by **?** for a fair comparison. We utilized the official code for implementing the graph learning methods (with GL), with the exception of DDGCRN, for which we employed the same prediction module. **Best results** across each dataset are in bold, while the second-best results are underlined.

For multi-step forecasting task on traffic domain datasets (METR-LA and PEMS-BAY), we also compare our method with conventional methods and deep neural network (DNN)-based methods. As the conventional methods, we use historical average (HA) which models the traffic flow as a seasonal process and uses weighted average of previous seasons as the prediction, simple linear autoregressive (AR) model, vector autoregressive (VAR) model,



(a) Performance without auxiliary graph



(b) Performance with auxiliary graph

Figure B.1: Experiments with different number of graphs and channels with METR-LA.

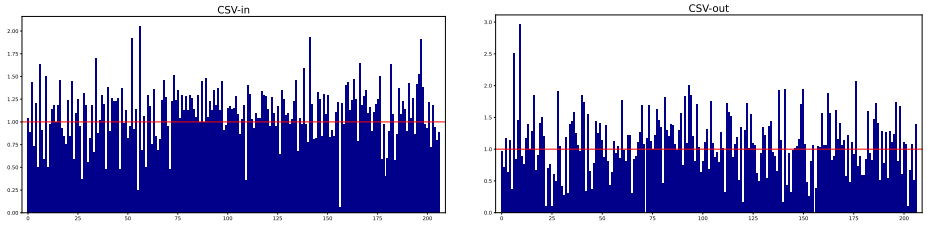
autoregressive integrated moving average (ARIMA) model with Kalman filter, and support vector regressor (SVR) with linear kernel. For DNN-based methods, we use feed forward neural network (FNN) with two hidden layers and L2 regularization, recurrent neural network with fully connected LSTM (FC-LSTM) and CNN for sequential data (WaveNet) (**?**). Table A.1 shows the experimental results of the multi-step forecasting task on traffic domain datasets, METR-LA and PEMS-BAY, with all the baseline methods.

Appendix B. Optimal Number of Multimodal Graphs

We performed additional experiments with different number of graphs and channels in convolutional layer of GTI module to evaluate their impact on the performance, using METR-LA dataset.

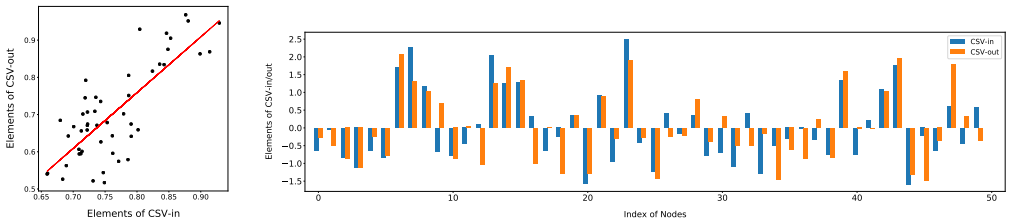
Figure B.1 shows the difference of MAPE with different horizons depending on the use of auxiliary graph. We can see that using multimodal graphs leads to a better performance over all horizons, compared to a single graph. Also, we can see that the optimal[1] number of learned multimodal graphs are less when trained with the auxiliary graph, compared to the case without the auxiliary graph. More specifically, the optimal number of multimodal graphs are eleven when trained without the auxiliary graph, only six graphs are needed when trained with the auxiliary graph. This tells us that the

[1]Optimal number of graphs and channels are selected with the criterion of lowest average validation loss across all horizons.



(a) Barplot of $\text{CSV}_{\texttt{in}}$ (b) Barplot of $\text{CSV}_{\texttt{out}}$

Figure C.2: Visualization of $\text{CSV}_{\texttt{in}}$ and $\text{CSV}_{\texttt{out}}$



(a) $\text{CSV}_{\texttt{in}}$ vs. $\text{CSV}_{\texttt{out}}$ (b) Visualization of (normalized) $\text{CSV}_{\texttt{in}}$ and $\text{CSV}_{\texttt{out}}$

Figure C.3: Inward/Outward connectivity strength of nodes in NASDAQ50 dataset.

some of the graphs learned without using the auxiliary graph captures the relationships hidden in the auxiliary graph.

Appendix C. Interpretation of Connectivity Strength Matrix

Connectivity strength matrix (CSM) is the outer product of two connectivity strength vectors, $\text{CSV}_{\texttt{in}}$ and $\text{CSV}_{\texttt{out}}$. As elements of the two vectors serve as the unique inward/outward strength of each node, we can examine the general connectivity strength of each node by observing the elements of both vectors. Figure C.2 shows the bar plot of the updated values of two CSVs of 207 sensors in METR-LA dataset, and red vertical line desrcibes the initial value one before the update.

Figure C.3a shows the scatter plot of elements in $\text{CSV}_{\texttt{in}}$ and $\text{CSV}_{\texttt{out}}$ with NASDAQ50 dataset. Each point represents the inward and outward connectivity strengths of certain stock, and we can see the positive correlation between the input and output strengths. Figure C.3b shows the bar plot of elements in $\text{CSV}_{\texttt{in}}$ and $\text{CSV}_{\texttt{out}}$ after z-normalization, where we can relatively compare the inner and outer connectivity strength. From this figure, we can check whether each stock is more likely to affect or to be affected by other

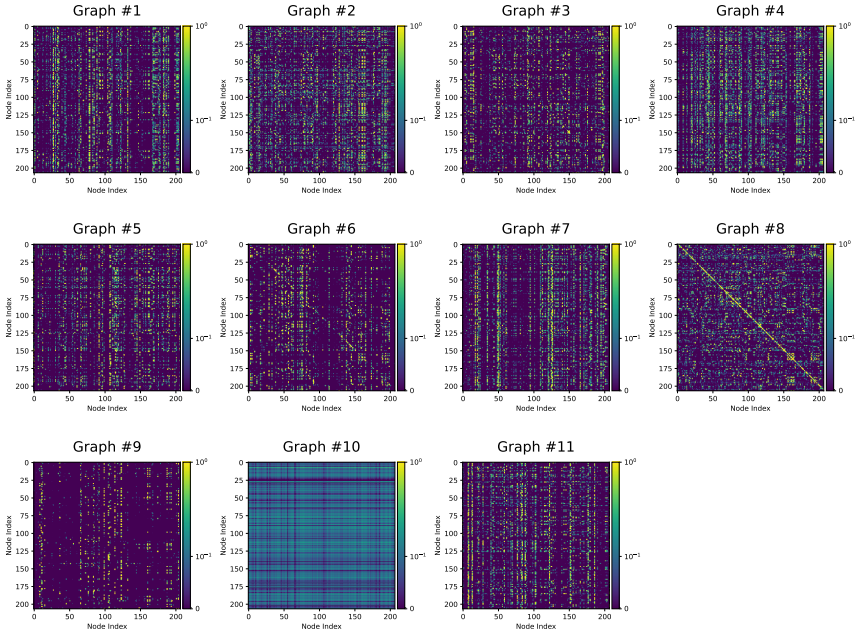
stocks. Both plots in Figure C.3 imply that stocks which are easily affect by other stocks are also more likely to be affected by other stocks.

Appendix D. Visualization of Multimodal Graphs

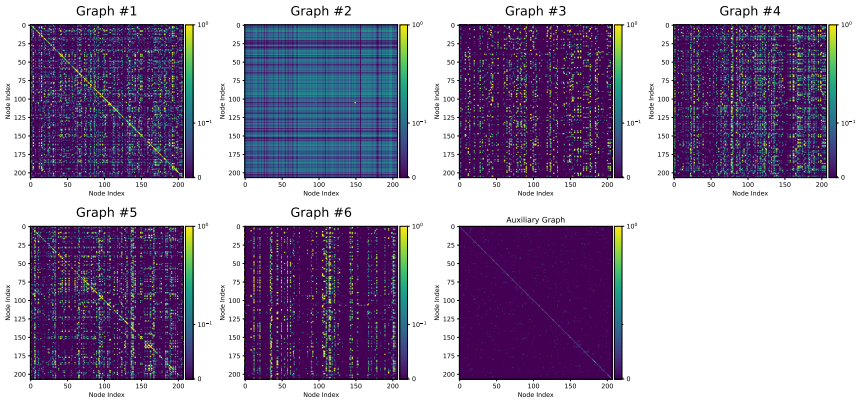
By visualizing the learned multimodal graphs, we can see how our learned graphs captures the multimodal relationships between MTS. Figure D.4 visualizes all the learned multimodal graphs of MAD-GL^2 using METR-LA dataset. Figure D.4a shows eleven multimodal graphs where the model is trained without using an auxiliary graph, and Figure D.4b shows six multimodal graphs and one auxiliary graph, where the model also uses one auxiliary graph along with six learned graphs for prediction module. We can see that all the graphs have their own unique pattern, explaining that each multimodal graph captures different relationships between MTS.

Appendix E. Connectivity of Nodes of Multimodal Graphs

As mentioned earlier, each graph captures different dependency structure between MTS. The column-wise sum and row-wise sum of the graph can be interpreted as the total inward and outward connectivity of each node, respectively. Figure E.5 visualizes the inward and outward connectivity strength of certain nodes of all the learned graphs using METR-LA dataset. For example, the third node has high outward connectivity in the fifth graph, low inward connectivity in the fourth graph, and high inward connectivity in the second graph. This tells us that each node has different role in each graph, showing the multimodality of our learned graphs.



(a) Multimodal Graphs (without auxiliary graph)



(b) Multimodal Graphs (with auxiliary graph)

Figure D.4: Visualization of the learned multimodal graphs with/without auxiliary graph.

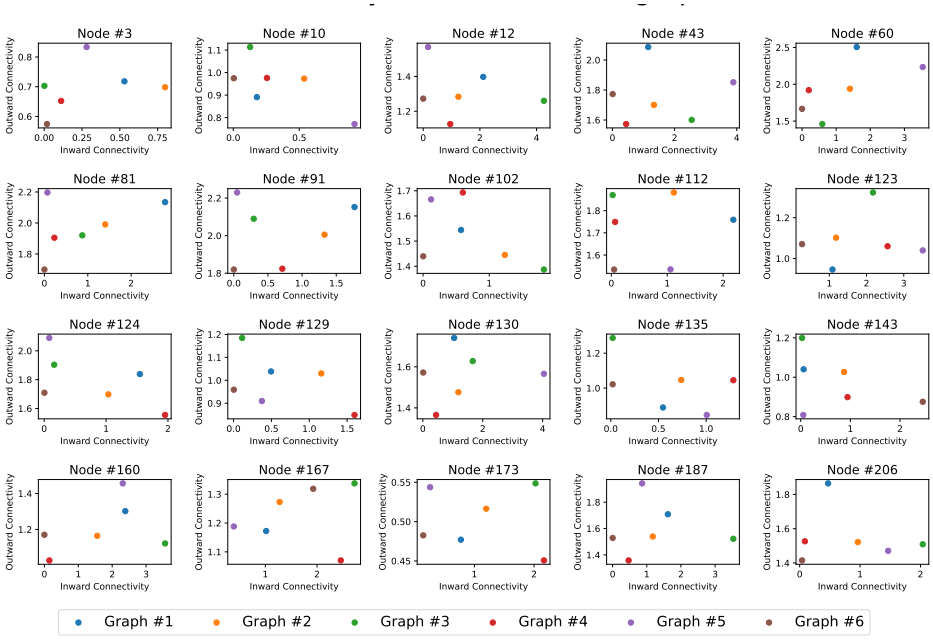


Figure E.5: Visualization of nodes' connectivity of multimodal graphs using METR-LA dataset. In each figure, the x-axis represents the column-wise sum, while the y-axis represents the row-wise sum of the graph. These values can be interpreted as the total inward and outward connectivity of each node, respectively. We can see that each node has different roles in different graphs, capturing different dependency structure between MTS.