Graph-based Forecasting with Missing Data through Spatiotemporal Downsampling

Ivan **Marisca**[1] Cesare **Alippi**[1,2] Filippo Maria **Bianchi**[3,4]

[1]IDSIA USI-SUPSI, Università della Svizzera italiana [🇨🇭] [2]Politecnico di Milano [🇮🇹] [3]UiT the Arctic University of Norway [🇳🇴] [4]NORCE, Norwegian Research Centre AS [🇳🇴]

CODE

made with Torch Spatiotemporal

tl;dr

MOTIVATION

- 🙂 Graph Deep Learning excels in modeling spatial dependencies in time series forecasting.
- ☹️ Missing data affect most real-world applications. (e.g., sensors)
- ☹️ Graph-based predictors deal with complete sequences and need imputation as pre-processing.

CHALLENGES

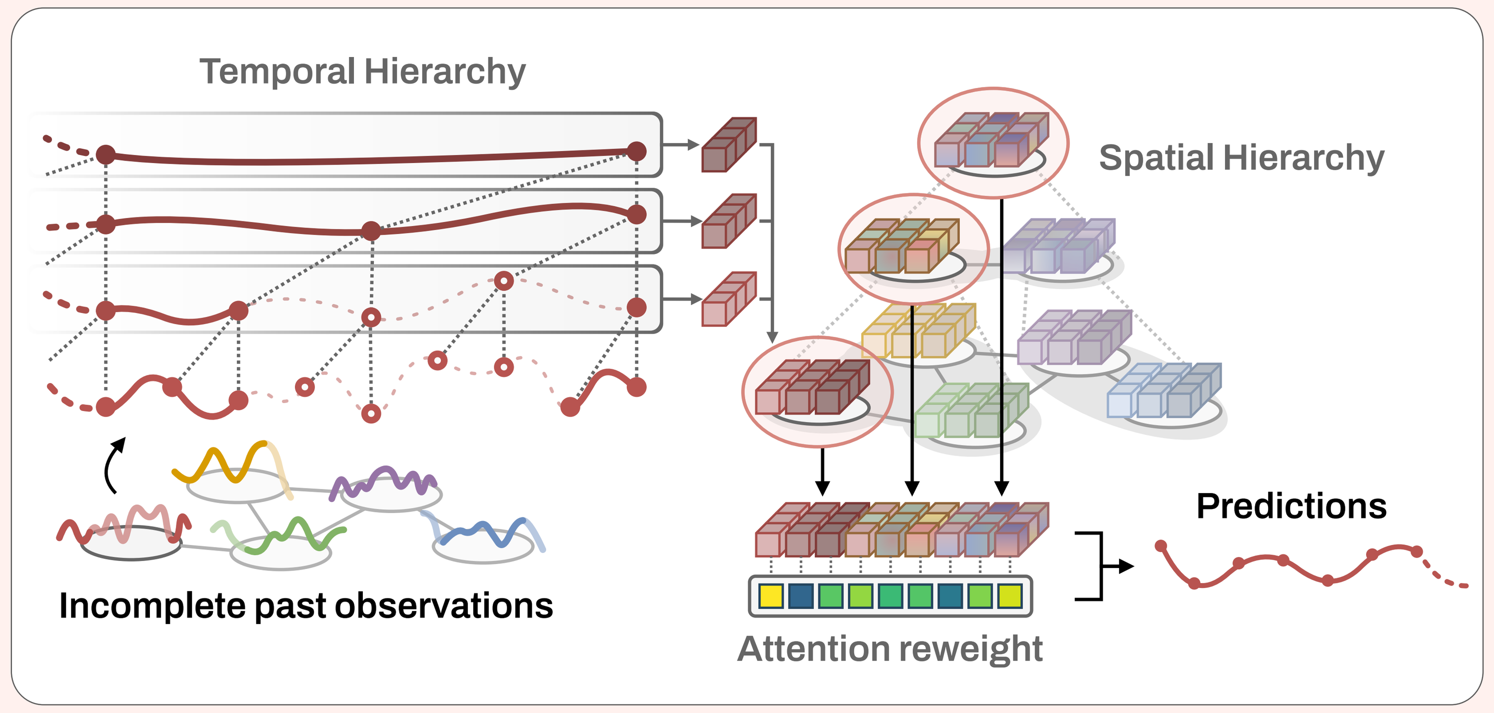
- **Observability.** Missing data **corrupt input dynamics**.
- **Adaptability.** To recover corrupted dynamics, we need **different processing** for **fine-** and **coarse**-grained **scales**.
- **Scalability.** Accounting for **long-range dependencies** in both time and space might cause **efficiency** issues.

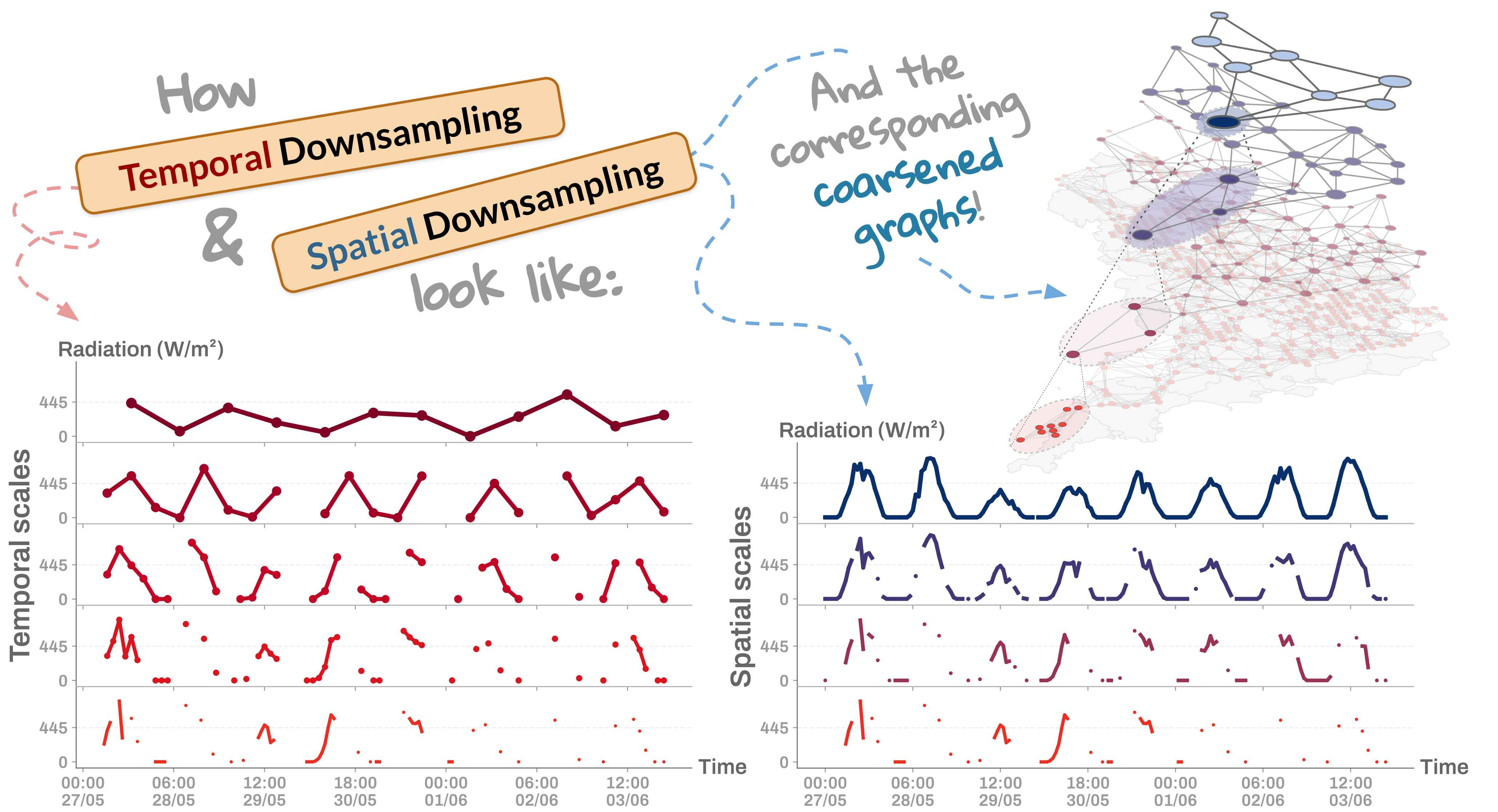
GET RID OF IMPUTATIONS, FORECAST DIRECTLY FROM INCOMPLETE TIME SERIES

CONTRIBUTION

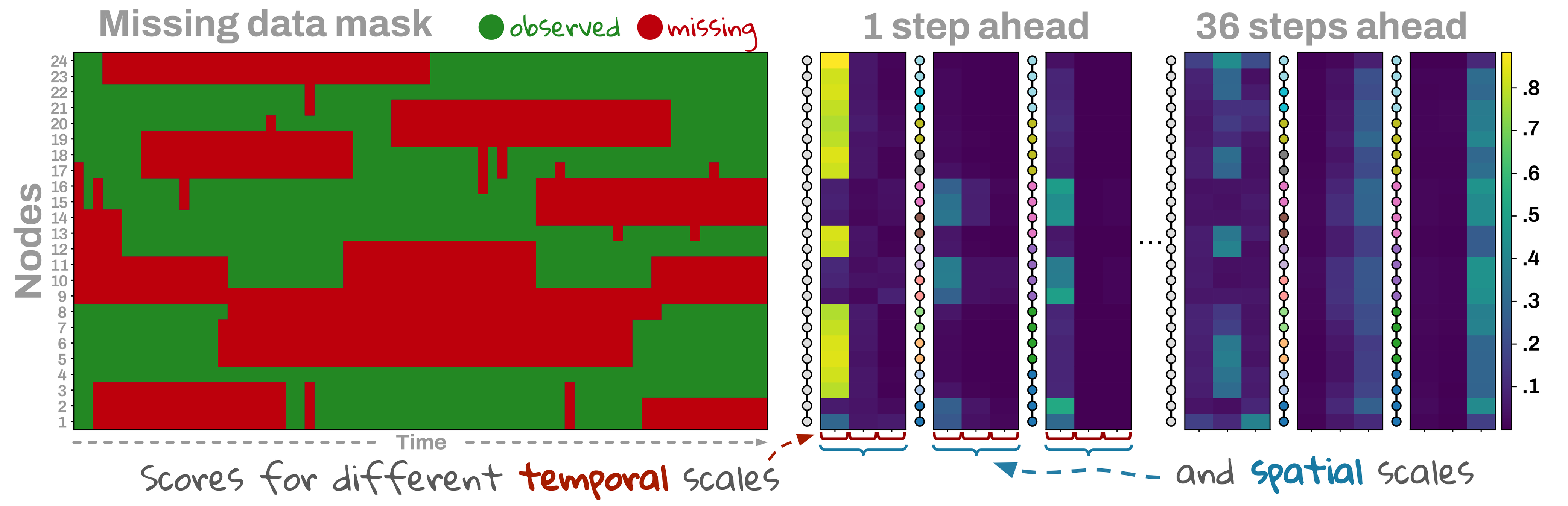
- Obtain a hierarchy of **multi-scale representations**, each accounting for a specific **space-time resolution**.
- **Scalability** given by **factorized time-then-space** processing and precomputed **downsampling** operators.
- **Adaptively weigh** the representations according to the **missing data pattern** in the input.

🔍 **Gain insights** on the scales through the **scores**!

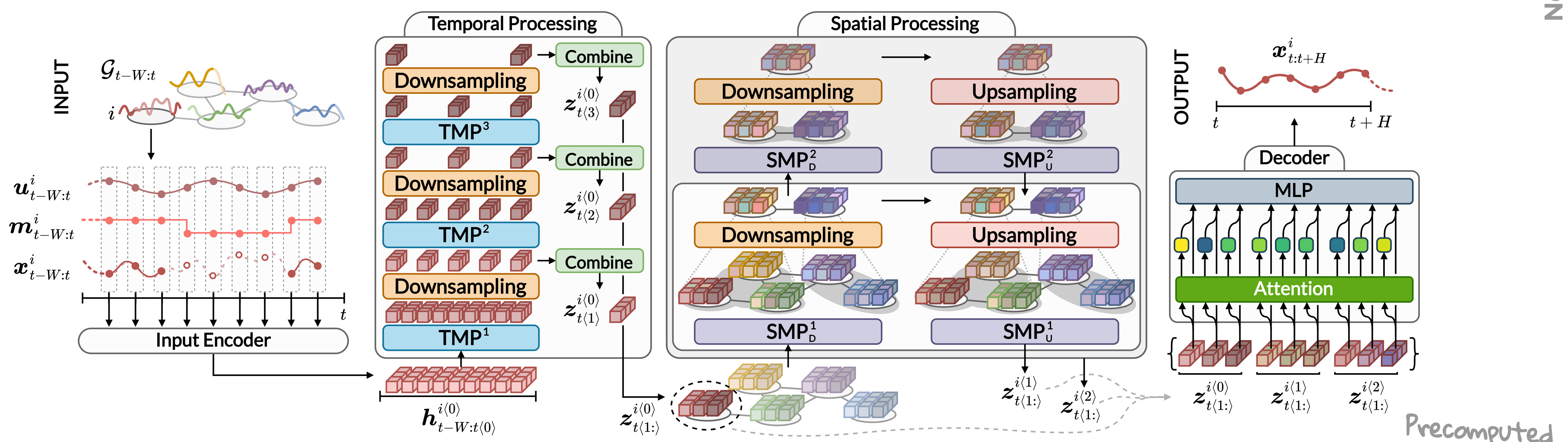




Interpretability of Decoder Weights



HD-TTS : Hierarchical Downsampling Time-Then-Space



Temporal operators

TMP : Temporal Message Passing

RNN, 1D convolution, Transformer

Temporal Downsampling

Dilation or strided operations.

Spatial operators

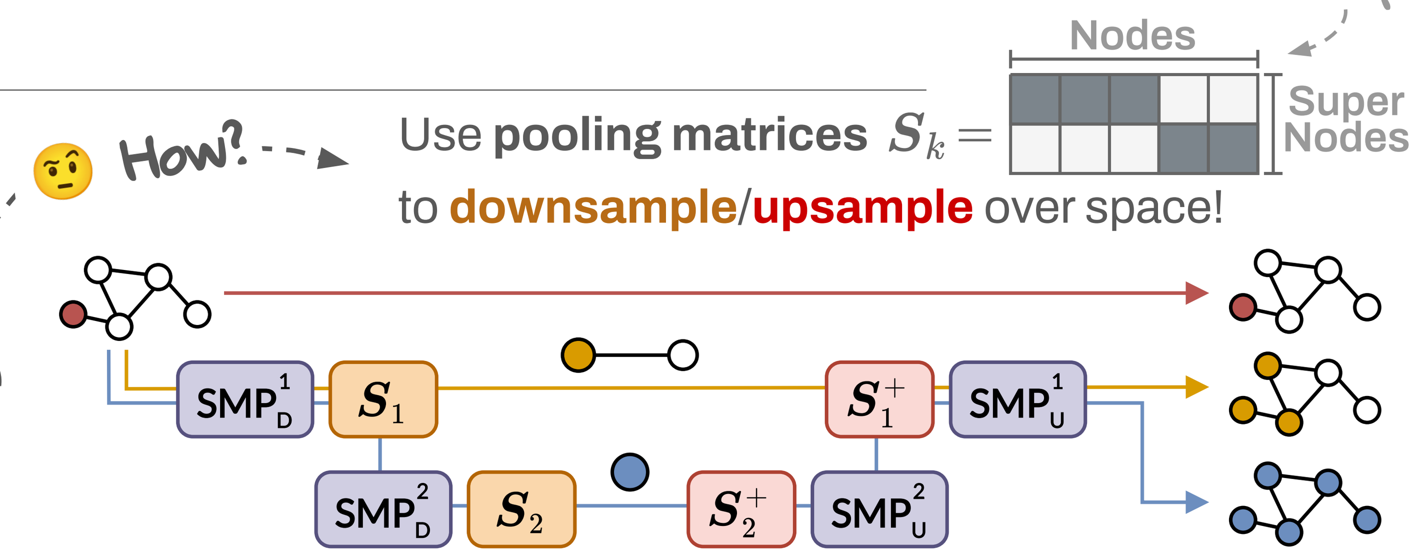
SMP : Spatial Message Passing

Your friendly neighborhood GNN

Spatial Down-/& Up-/sampling

We adopt **graph pooling**.

🤨 How? Use **pooling matrices** $S_k =$ to **downsample**/**upsample** over space!



Some Empirical Results

MAE on real-world datasets with different missing data distributions.

	AQI		EngRAD (NEW DATASET!)		PV-US			
Model	**Original**	**+ Point**	**Block–T**	**Block–ST**	**Block–T**	**Block–ST**	**Batch/s**	**GPU RAM**
GRU	18.17±0.03	19.19±0.06	5.30±0.03	5.42±0.02	3.98±0.02	4.14±0.02	11.59±0.04	12.01 GB
DCRNN	16.99±0.09	17.51±0.08	5.14±0.06	5.33±0.05	3.54±0.01	3.76±0.00	1.36±0.01	19.72 GB
AGCRN	17.19±0.06	17.92±0.05	4.84±0.01	5.10±0.06	4.06±0.01	4.20±0.04	1.15±0.01	23.40 GB
GRIN-P	16.85±0.05	17.59±0.06	4.91±0.04	5.05±0.00	3.62±0.02	3.85±0.07	1.52±0.00	17.28 GB
GWNet	15.89±0.04	16.39±0.14	4.59±0.04	4.76±0.03	**3.48**±0.05	3.71±0.03	2.12±0.00	16.02 GB
T&S-IMP	16.54±0.03	17.13±0.05	4.98±0.01	5.15±0.03	3.60±0.02	3.82±0.03	2.68±0.00	7.03 GB
T&S-AMP	16.15±0.02	16.58±0.10	4.93±0.02	5.11±0.05	N/A	N/A	N/A	N/A
TTS-IMP	16.25±0.01	16.90±0.26	4.81±0.07	5.08±0.04	3.50±0.01	3.66±0.02	18.84±0.14	12.81 GB
TTS-AMP	15.63±0.06	16.15±0.05	4.70±0.00	4.81±0.06	**3.46**±0.03	3.65±0.05	14.26±0.08	12.81 GB
HD-TTS-IMP	15.50±0.07	15.94±0.10	**4.48**±0.01	**4.64**±0.03	**3.47**±0.01	**3.62**±0.02	7.11±0.03	10.86 GB
HD-TTS-AMP	**15.35**±0.01	**15.76**±0.07	4.53±0.03	**4.65**±0.04	**3.47**±0.02	**3.61**±0.02	6.21±0.02	10.86 GB

- 🙂 Improvements in **forecasting accuracy** and **computational efficiency**.
- 😉 Especially in challenging settings with **blocks of missing values** in space and time.