Temporal Multiresolution Graph Neural Networks For Epidemic Prediction

Truong Son Hy [1], Viet Bach Nguyen [2], Long Tran-Thanh [3] and Risi Kondor [1]

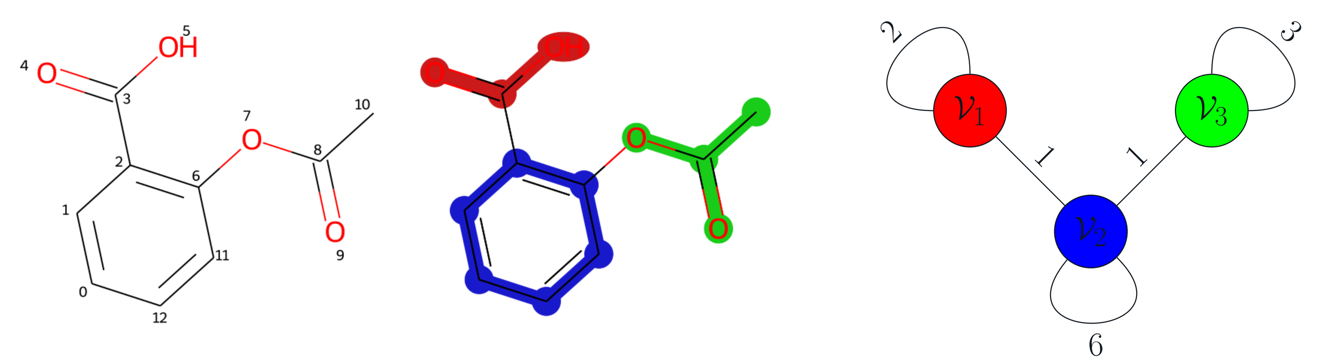
The University of Chicago [1], FPT Software [2], The University of Warwick [3]

Abstract

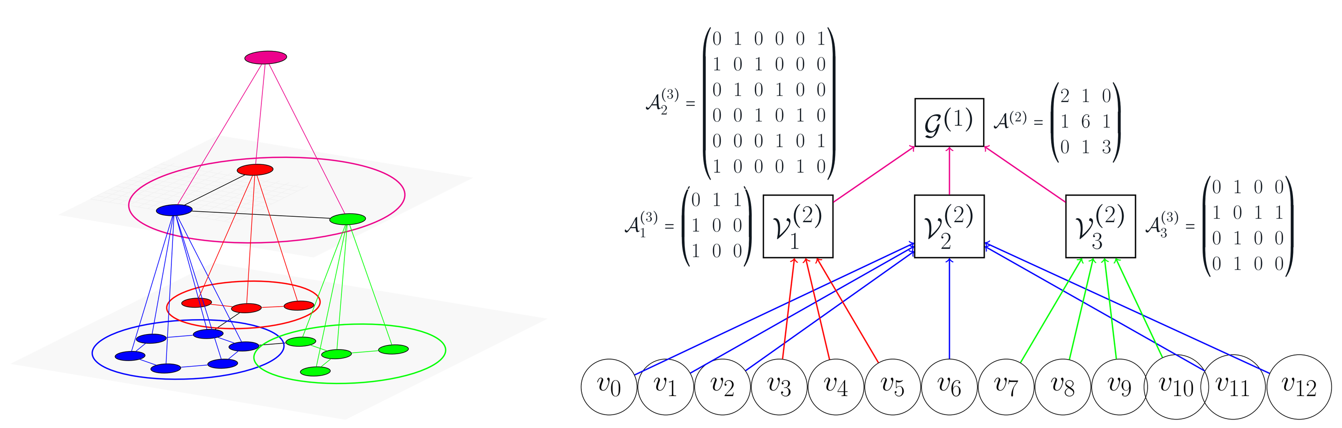
In this paper, we introduce *Temporal Multiresolution Graph Neural Networks* (TMGNN), the first architecture that both learns to construct the multiscale and multiresolution graph structures and incorporates the time-series signals to capture the temporal changes of the dynamic graphs. We have applied our proposed model to the task of predicting future spreading of epidemic and pandemic based on the historical time-series data collected from the actual COVID-19 pandemic and chickenpox epidemic in several European countries, and have obtained competitive results in comparison to other previous state-of-the-art temporal architectures and graph learning algorithms. We have shown that capturing the multiscale and multiresolution structures of graphs is important to extract either local or global information that play a critical role in understanding the dynamic of a global pandemic such as COVID-19 which started from a local city and spread to the whole world. Our work brings a promising research direction in forecasting and mitigating future epidemics and pandemics.

Multiresolution Graph Networks

Previously, (Hy & Kondor, 2021) proposed *Multiresolution Graph Networks* (MGN) that constructs multiple resolutions of the input graph via the learning to cluster algorithm in a data-driven manner.



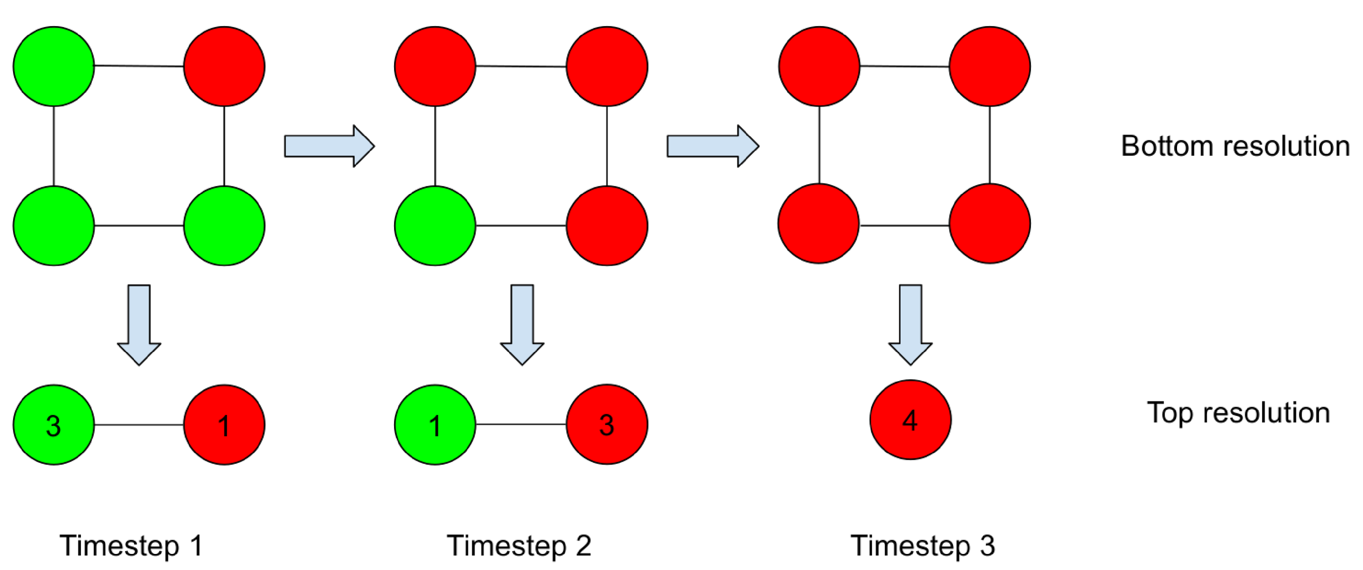
Aspirin $C_9H_8O_4$, its 3-cluster partition and the corresponding coarsen graph.



Hierarchy of 3-level Multiresolution Graph Network on Aspirin molecular graph.

Multiresolution modeling of epidemic dynamics

We apply the coarse-grained and hierarchical graph model (Hy & Kondor, 2021) to learn on the human-interaction network to capture micro to macro (i.e. local to global) information of a pandemic/epidemic.

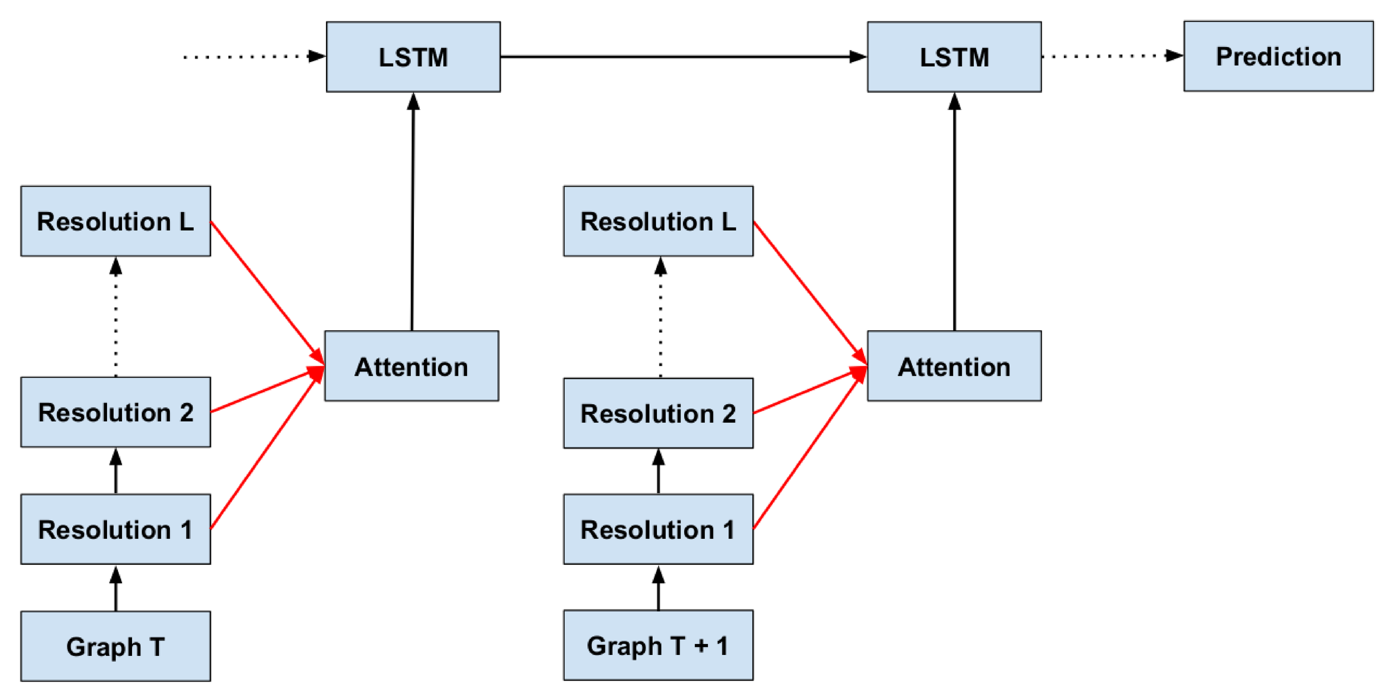


Suppose we are given a interaction network of 4 people as a square graph of 4 nodes and 4 edges. The red color denotes the person gets infected by some virus and the green color denotes otherwise. This figure shows a 2-resolution understanding of the epidemic spread.

Temporal Architecture

Furthermore, we propose:

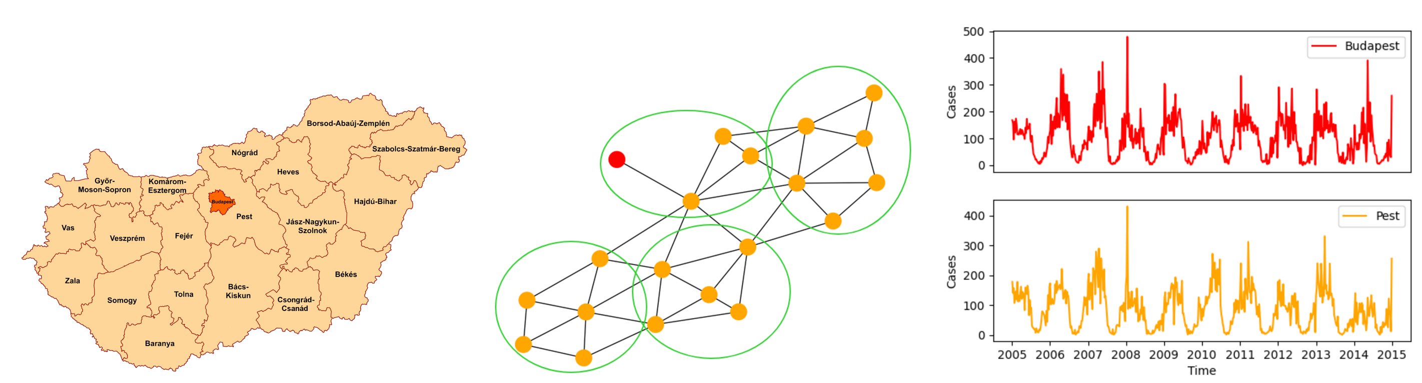
- An attention mechanism to select the right resolution that consequentially predicts the current state or progress of the pandemic/epidemic,
- Temporal architecture incorporating time-series information to make future prediction about the pandemic/epidemic given historical data.



This diagram, snapshotted at time T and $T+1$ in the dynamics, depicts the general architecture of Temporal Multiresolution Graph Neural Networks (TMGNN) with the backbone of Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997). The red arrows denote the attention scores for the multihead attention among resolutions. The black arrows represent the data flow.

Chickenpox epidemic in Hungary

Chickenpox is an infection caused by the varicella-zoster virus (VZV). It is highly contagious to those who have not had the disease or been vaccinated against it. *Chickenpox Cases in Hungary* is a novel spatiotemporal dataset which can be uesd to benchmark the forecasting performance of GNNs architectures, that includes timeseries describing the weekly number of chickenpox cases from 01/2005 to 01/2015 (500 entries for each county).



This is the map of 20 counties in Hungary including the capital city, Budapest, denoted by the red node and the rest of 19 counties denoted by orange nodes. A possible clustering into 4 spatial groups is shown by green circles. The right figure shows the timeseries of chickenpox cases in Budapest and its surrounding county, Pest, are highly correlated.

Hungary	
GConvLSTM (Seo et al., 2018)	1.221 ± 0.010
GConvGRU (Seo et al., 2018)	1.117 ± 0.002
Evolve GCN-O (Pareja et al., 2020)	1.120 ± 0.003
Evolve GCN-H (Pareja et al., 2020)	1.115 ± 0.013
DynGRAE (Taheri et al., 2019)	1.112 ± 0.010
STGCN (Yu et al., 2018)	1.118 ± 0.005
DCRNN (Li et al., 2018)	1.119 ± 0.002
TMGNN (ours)	**0.990 ± 0.003**

The average test mean squared error with standard deviations obtained over 40 weeks long forecasting horizons in Hungary calculated from a 10 experimental runs with different random seeds. The baseline results are taken from (Rozemberczki et al., 2021). Our TMGNN model outperforms the SOTAs by **11%**.

COVID-19 pandemic in Europe

COVID-19 is disease caused by severe acute respiratory syndrome coronavirus 2 or SARS-Cov-2, a newly discovered virus that is closely related to bat coronaviruses (Perlman, 2020), pangolin coronaviruses (Zhang et al., 2020) and SARS-CoV (Sun et al., 2020).

England			
Model	**3 days**	**7 days**	**14 days**
AVG	9.75	9.99	10.09
LAST DAY	7.11	7.62	8.66
AVG_WINDOW	6.52	7.34	8.54
LSTM	9.11	8.97	9.10
ARIMA	13.77	14.55	15.65
PROPHET	10.58	12.25	16.24
TL_BASE	9.65	12.30	13.48
MPNN	6.36	6.86	8.13
MPNN+LSTM	6.41	6.67	7.02
MGN	6.68	7.37	8.74
TMGNN (ours)	**6.26**	**6.55**	**6.80**

France			
Model	**3 days**	**7 days**	**14 days**
AVG	8.50	8.55	8.55
LAST DAY	7.47	7.37	8.03
AVG_WINDOW	**6.04**	6.40	7.24
LSTM	8.08	8.13	7.91
ARIMA	10.72	10.53	10.91
PROPHET	10.34	11.56	14.61
TL_BASE	7.67	9.21	12.27
MPNN	6.16	**5.99**	**6.93**
MPNN+LSTM	6.39	7.21	7.36
MGN	6.65	6.61	7.66
TMGNN (ours)	6.39	7.35	7.51

Italy			
Model	**3 days**	**7 days**	**14 days**
AVG	21.38	22.23	23.09
LAST DAY	17.40	18.49	20.69
AVG_WINDOW	15.17	16.81	19.45
LSTM	22.94	23.17	23.12
ARIMA	35.28	37.23	39.65
PROPHET	24.86	27.39	33.07
TL_BASE	19.12	23.44	24.89
MPNN	**14.39**	**15.47**	17.88
MPNN+LSTM	15.56	16.41	17.25
MGN	15.33	16.73	19.22
TMGNN (ours)	15.09	15.62	**16.38**

Spain			
Model	**3 days**	**7 days**	**14 days**
AVG	45.10	45.87	47.63
LAST DAY	33.58	37.06	43.63
AVG_WINDOW	**32.19**	36.06	42.79
LSTM	51.44	49.89	47.26
ARIMA	40.49	41.64	46.22
PROPHET	54.76	62.16	79.42
TL_BASE	42.25	52.29	59.68
MPNN	35.83	38.51	44.25
MPNN+LSTM	33.35	34.47	35.31
MGN	38.85	43.65	52.23
TMGNN (ours)	33.13	**34.12**	**35.04**

In this experiment, the dataset contains measures of human mobility between regions of member states of the European Union (EU) including England, Frace, Italy and Spain. The raw timeseries data, collected and aggregated from mobile phones, includes three recordings per day (i.e., midnight, morning and afternoon) indicating the number of people travelling from on region to another during that period of time in the day. We compare our model TMGNN with the same baselines as in (Panagopoulos et al., 2021). We outperform the state-of-the-arts in **6 out of 12** experiments.

Software

Our PyTorch implementation is publicly available at:

`https://github.com/bachnguyenTE/temporal-mgn`

Reference

Hy, T. S. and Kondor, R. *Multiresolution equivariant graph variational autoencoder*, 2021. URL `https://arxiv.org/abs/2106.00967`.

Hochreiter, S. and Schmidhuber, J. *Long short-term memory*. Neural Computation, 9:1735–1780, 1997.

Rozemberczki, B., Scherer, P., Kiss, O., Sarkar, R., and Ferenci, T. *Chickenpox cases in hungary: a benchmark dataset for spatiotemporal signal processing with graph neural networks*. ArXiv, abs/2102.08100, 2021.

Panagopoulos, G., Nikolentzos, G., and Vazirgiannis, M. *Transfer graph neural networks for pandemic forecasting*. Proceedings of AAAI, 35(6):4838–4845, May 2021.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. *Neural message passing for quantum chemistry*. Proceedings of the 34th International Conference on Machine Learning, volume 70 of PMLR, pp. 1263–1272.