AMERICAN UNIVERSITY OF BEIRUT

TRANSFERABILITY OF GRAPH NEURAL NETWORKS FOR TIME SERIES APPLICATIONS

by
ZAHRAA MALEK AL SAHILI

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Electrical and Computer Engineering
of the Maroun Semaan Faculty of Engineering and Architecrure
at the American University of Beirut

Beirut, Lebanon
September 2022

AMERICAN UNIVERSITY OF BEIRUT

TRANSFERABILITY OF GRAPH NEURAL NETWORKS FOR TIME SERIES APPLICATIONS

by
ZAHRAA MALEK AL SAHILI

Approved by:

MARIETTE AWAD

Dr. Mariette Awad, Associate Professor — Advisor
Electrical and Computer Engineering

Joseph Costantine, Associate Professor — Member of Committee
Electrical and Computer Engineering

Wassim El Hajj, Professor — Member of Committee
Computer Science

Date of thesis defense: September 29, 2022

AMERICAN UNIVERSITY OF BEIRUT

THESIS RELEASE FORM

Student Name:	Al Sahili	Zahraa	Malek
	Last	First	Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my thesis; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

- ☐ As of the date of submission
- ☒ One year from the date of submission of my thesis.
- ☐ Two years from the date of submission of my thesis.
- ☐ Three years from the date of submission of my thesis.

zahraa   8/10/2022

Signature   Date

ACKNOWLEDGEMENTS

I would like to thank my advisor Prof. Mariette Awad for her support and the opportunity to work on this thesis topic and my family and friends for all the support in this master's journey. I would like also to thank Andrew Ng for democratizing AI education and giving world-class education and support to the student regardless of their race, gender, and ethnicity. Finally, my sincere thanks to Jure Leskovik and Xavier Bresson for unlocking the graph neural networks knowledge. I promise to be on your path in democratizing AI education and using the acquired knowledge in AI for the good of humanity.

ABSTRACT

OF THE THESIS OF

Zahraa Malek Al Sahili for Master of Engineering
Major: Electrical and Computer Engineering

Title: Transferability of Graph Neural Networks for Time Series Application

Transfer learning enabled machine learning tasks with scarce data to achieve superhuman performance in multiple domains like computer vision and natural language processing. However, knowledge transfer's success was mostly on grid structured data and using convolutional neural networks that assume local, hierarchical, and stationary data. Time series data in several applications, specifically doesn't meet these assumptions. This renders traditional transfer learning irrelevant with the potential leading to negative transfer. After achieving superior performance on high-dimensional data like social networks and recommender systems, graph neural networks are currently applied to time series data. In this thesis, we investigate the transferability of graph neural networks on time series data compared to traditional time series algorithms. We also explore a new graph similarity approach and compare its effect on time series algorithms pretraining and negative transfer for pandemic time series forecasting.

TABLE OF CONTENTS

ILLUSTRATIONS

Figure

TABLES

Table

ABBREVIATIONS

CNN: Convolutional Neural Networks

DTW : Dynamic time warping

GNN: Graph Neural Networks

GAT: Graph Attention Networks

GCN: Graph Convolution Network

NBEATS: Neural basis expansion analysis for interpretable time series forecasting

MAE: Mean absolute error

MSE: mean squared error

PV: Photovoltaic

RMSE: Root mean squared error

STGNN: Spatio-temporal Graph Neural Networks

CHAPTER I

INTRODUCTION

Time series applications have become a very active research area [1]. Various deep learning algorithms were able to achieve very high accuracy in multiple time series forecasting and classification fields. Unfortunately, these algorithms' success is based on sufficient labeled data availability. To achieve high performance in time series applications with scarce data, knowledge transfer from similar data should be adopted [2].

Knowledge transfer using convolutional neural networks achieved superhuman performance in computer vision tasks with a limited amount of data like face recognition and medical imaging [3]. In addition, transfer learning reduces computational costs and increases the speed of deep neural network training [3]. Convolutional neural networks work efficiently on local, hierarchal, and stationary data. On the other hand, time series data may lack the assumptions, which causes traditional transfer learning to fail thus leading to a "negative transfer" outcome [4]. For example, transferring the forecast of covid cases from Uganda to Italy will result in a drop in the deep model accuracy instead of improving the model's performance [4].

Convolutional neural networks were extended to high dimensional unstructured data through graph neural networks (GNN). After achieving state of the art performance on high-dimensional graph applications like social networks and recommender systems, GNNs are currently applied to various time-serious applications [5-7]. From multivariate time series forecasting to time series classification, graph neural network algorithms were able to outperform counterpart algorithms [5-7].

Another challenge for transferring knowledge in time series applications is data similarity. To select the appropriate source data for the target data, multiple metrics were proposed for graphs in general and for time series data specifically. These metrics include cosine similarity, correlation, and dynamic time rapping. However, further experimentations are needed to ensure the negative transfer mitigation.

In this research, we aim to investigate transfer learning for time series applications. First, we propose using the first and second derivatives as a similarity metric to accompany correlation, cosine similarity, and dynamic time wrapping. We also investigate the effect of the first and second derivatives on fine-tuning models. Moreover, we research the transferability of various graph neural networks on time series forecasting and classifications. To the best of our knowledge, our research is the first to investigate the transferability of time series data through graph neural networks and the first to propose an attention-based spatio-temporal graph neural network.

The thesis consists of six chapters. Chapter I provides a brief introduction. In Chapter II, a survey on spatio-temporal graph neural networks is presented focusing on the algorithms and the applications. Chapter III includes a literature review on transfer learning for time series and on transfer learning in graph neural networks research. Then Chapter IV, discusses our contribution related to similarity metrics based on the hurricane model, from theory to experimental results. After that, Chapter V examines the transferability of spatio-temporal graph neural networks for pandemic forecasting from proposing new algorithms to comparing the transferability of these algorithms. Finally, Chapter VI concludes our thesis while discussing future research plans.

CHAPTER II

SPATIO-TEMPORAL GRAPH NEURAL NETWORKS

A. Introduction

Graph Neural Networks for time-varying graph data are used in well-known applications varying from multivariate time series data to social networks and audiovisuals and so-called Spatio-temporal graph neural networks. In this chapter, we provide a comprehensive review of spatio-temporal graph neural network algorithms while proposing a new taxonomy for introducing time to graph neural networks.

B. Algorithms

Spatio-temporal graph neural networks can be classified from an algorithmic perspective as spectral-based and spatial-based. Another classification category is the method time variant introduced: weather using another machine learning algorithm or defining time within the graph structure.

1. Hybrid Spatio-Temporal Graph Neural Networks

Hybrid Spatio-temporal graph neural networks constitute two main components: a spatial component and a time component (Figure 1).

- Spatial Module

In hybrid Spatio-temporal graph neural networks, graph neural network algorithms are used to model the spatial dependencies in the data.

-Spectral Graph Neural Networks

Spectral GNNs are based on the spectral definition of the convolution operation. Early Spatio-temporal GNNs heavily relied on this spectral definition. For example, Yu et al. [8], used Chebyshev GNN in the STGCN algorithm. In addition, Cao et al.[9] used Spectral graph convolution to model the space domain in his StemGNN. Recently, Simeunivic et al [10] used spectral GCNs in his both algorithms: GCLSTM and CGTransfo.

-Spatial Graph Neural Networks

With the advances in spatial graph neural networks research, various researchers used spatial GNNs to model the spatial domain in spatio-temporal GNNs.Chen et al[11], used a recurrent graph neural network(RGNN) with skip connection to model spatial dependencies in traffic forecasting. However, Wu et al. [12] used Graph Convolution neural networks GCNs with skip connection in his MTGNN algorithm. Additionally, GCN was used in the Structural RNN algorithm [13].

Graph neural networks with attention mechanism (GAT) were used in [14,15]. In A2GNN,[14] Huang et al. used the GAT with an autograph learner to improve the forecasting performance. Moreover, Kan et al. [16] used GAT cascaded with a graph transformer and a hierarchical pooling mechanism in the HST-GNN implementation.

More advanced spatiotemporal GNN algorithms were used for spatial modeling in [17,18]. Oreshkin et al. used Gated graph neural networks in the FG-GAGA algorithm. In contrast, the Graph Isomorphism network was used by Kim et al. to model brain connectivity in brain graph representation.

-Graph Transformers

Two algorithms relied on graph transforms to model spatial dependencies: TransMOT and d Forecaster [19,20].In addition, Kan et al.[16] accompanied his GAT with a Graph transformer in his HST-GNN architecture.

- Temporal Module

To model the time domain, various machine learning algorithms can be involved.

-1D-CNN

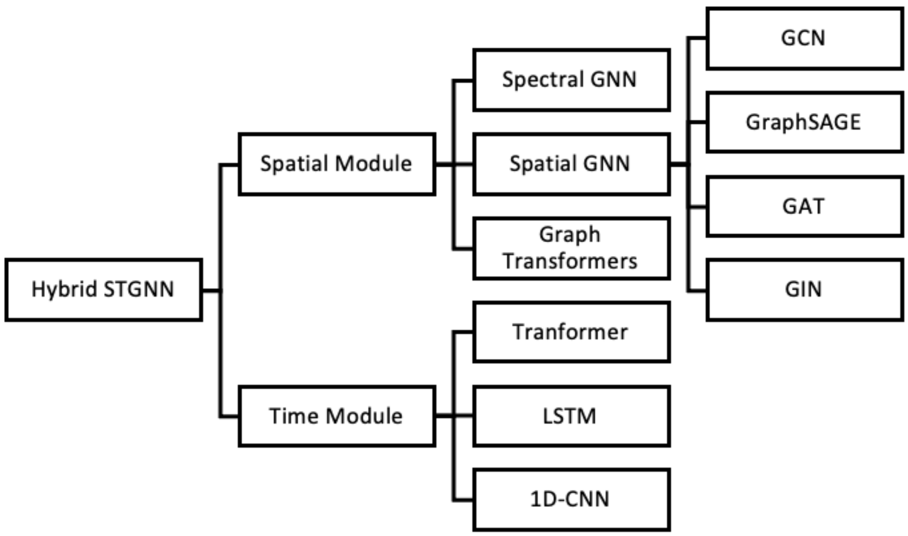
Yu et al.[8] used a 1D-CNN to account for the time domain in his STGCN algorithm. Moreover, Wu et al [12], used an inception layer in the MTGNN implementation. Also, Cao et al. [9] used 1D CNNs with GLU units for the temporal module.

-Recurrent Neural Networks

Recurrent neural networks and their variants as Gated Recurrent units (GRUs) and Long Short Terms Memory units (LSTMs) were widely adopted in hybrid Spatio-temporal GNNs to model the time domain. Jain et al [34] used RNNs in the structural RNN algorithm. On the other side, Oreshkin et al. used GRUs in FG-GAGA GNN while Chen et al. [11] used both GRUs and LSTMs in the MResGNN algorithm. In addition, [7,8,12] all used LSTMs as time modules. In the HST-GNN algorithm, Kan et al. [16] used 2 LSTMs with an attention mechanism within a wider encoder-decoder architecture.

-Transformers

Recently, a huge focus was imposed on transformer architectures that were used to accompany the time domain. The transformer was used by [14, 20] in the TransMOT, Forecaster, STAGIN, and GCTransfo respectively.



1.Hybrid-based Spatio-Temporal Graph Neural Networks

2. Solo-Graph Neural Networks

Another method to model time in spatio-temporal graph neural networks is to define the time frame within the GNN itself. Multiple approaches were proposed including: defining time as an edge, inputting time as a signal to the GNN, time modeled as a subgraph, and sandwiching other machine learning architectures inside the GNN (Figure 2).

- Time as Edge

Kapoor et al. [19] used spatial GCN with skip connections to forecast covid. In his algorithm, time was defined as an edge and locations as graph nodes. Additionally, time was defined as an edge in USTGCN algorithms [20] which modified the space adjacency matrix to a space-time adjacency matrix.

- Time as Signal

Time as an input signal was used in GNN pure-based spatio-temporal GNNs. Zhang et al. [21] used temporal hierarchy modeling to input time to the GAT. The algorithm also included graph trimming and convolution diffusion to improve the performance. Moreover, Shen et al. [22] used a gated dilated casual block for the temporal input. The output of this block was inputted to a dynamic GCN. in parallel with the output of a similar double block for the spatial domain. Time was also inputted as a signal in the CasualGNN [23] algorithm. The algorithm is based on dynamic graph neural network with an attention mechanism and a casual module.

- Time as Subgraph

Li et al [24] modeled time as a subgraph within a graph isomorphism network (GIN). Moreover, Shao et al. [25] used a temporal similarity graph to account for the temporal domain, which was added to other spatial graphs to form a multigraph set that constructed the ASTGCN framework.

- Time using Sandwiching

Karimi et al. [26] used two 1D-CNNs to model time. In his architecture, the 1D-CNNs were sandwiched inside the GCN architecture as sub-modules.

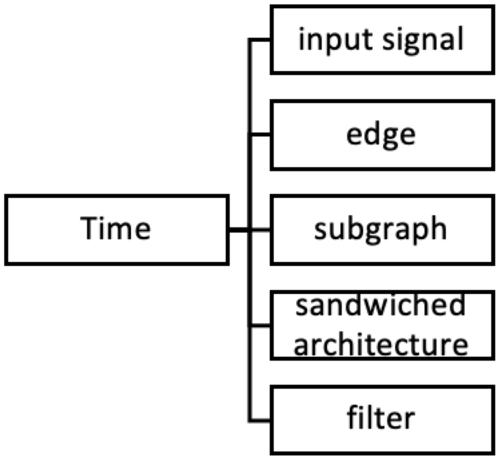
Table 1.Summary of Spatio-temporal Graph Neural Networks

Author	Name	Hybrid Algorithm	GNN only	Spectral Based	Spatial Based	Spatial Module	Time Module
Yu et al.	STGCNGCN	√		√		Chebychev GNN	1D-CNN
Nicoliviou et al.	RSTGCN	√			√	Custom (3D CNN)	LSTM
Chen et al.	MResGNN	√			√	RGNN	GRU & LSTM
Kapoor et al	-		√		√	GCN; time added as an edge	
Wu et al.	MTGNN	√			√	GCN	Inception layer
Li et al.	Unified GNN		√		√	GIN; time added as a subgraph module	
Cao et al.	StemGNN	√		√		Spectral GCN	1D-CNN
Oreshkin et al.	FG-GAGACN	√			√	GGCN	GRU
Jain et al.	Structural RNN	√			√		RNN
Karimi et al.	St-GNN		√		√	GCN with 1DCNN sandwiched in the GNN	
Chu et al.	TransMOT	√			√	Graph transformer	Transformer
Kim et al.	Forecaster	√			√	Graph Transformer	Transformer
Kim et al.	STAGIN	√			√	GIN	Transformer

Zhang et al.	ST-GDN		√		√	GAT; time as input using temporal herirachly modeling	
Shao et al.	ASTGCN		√		√	Multi graph set;Time as Temporal similarity graph	
Huang et al.	A2GNN	√			√	GAT	LSTM
Simeunovic et al.	GCLSTM	√		√		Spectral GCN	LSTM
Simeunovic et al.	GCTransfo	√		√		Spectral GCN	Transformer
Kan et al.	HST-GNN	√			√	Graph Transformer &GAT	LSTM
Hadous et al.	Space-Time GNN		√		√	GCN; time as a filter	
Shen et al.	T2GNN		√		√	Dynamic GCN: time as input signal	
Wang et al.	CasualGNN		√		√	GAT; time as a signal	
Roy et al.	USTGCN		√		√	GCN; time as an edge	

- Time as Filter

In Space-time Graph Neural Network [27], both time and space were introduced as multivariate integral Lipschitz filters inside the GCN.



2.Time Modelling in GNN only algorithm

C. Applications

1.Multivariate Time series Forecasting

Motivated by the power of GNNs in handling relational dependencies [28], spatio-temporal GNNs were widely applied in multivariate time series forecasting. Applications include traffic forecasting, Covid forecasting, PV power consumption, RSU communication, and seismic applications.

- Traffic

Transportation is considered a very important factor in every person's life [28]. Based on a study conducted in 2015, U.S. drivers spend a daily average of 48 minutes behind the wheel [28]. Thus, an accurate real-time forecast of traffic conditions is of dominant importance for road users, private sectors, and governments. However, traditional machine learning forecast systems fail to satisfy accuracy conditions due to the high nonlinearity and complexity of traffic flow [28]. In contrast and based on the power of GNNs in handling nonlinearities, spatiotemporal graph neural networks were widely applied in traffic forecasting in both aspects: long-term and short-term predictions.

- Pandemic Forecasting

In a state of pandemic, the ability to accurately forecast the caseload is extremely important to the country level or the individual level [19]. With conventional algorithms considering forecasting pandemic cases as a closed loop based on previous cases and considering the spatial dependencies between neighborhoods in effecting pandemics, spatio-temporal graph neural networks were used to accompany both space and time in pandemics[19]. Several spatio-temporal graph neural network algorithms were proposed and found to achieve state of art COVID forecasting in the United States, United Kingdom, Germany, and worldwide.

- PV

Due to the rapid increase in the installation of commercial PV power plants, the operation and planning for reliable performance of PV systems is a crucial challenge [29]. Ensuring reliable performance includes monitoring the slow loss of electricity output and effective planning based on the PV power output. This reliability can be achieved by accurate power forecasting. Based on the ability of GNNs in capturing spatial and temporal dependencies, spatio-temporal graph neural networks were widely adopted to forecast PV power [30] and were able to achieve achieved superior performance over other forecasting algorithms.

- RSU communication

As a special type of base station, Road Side Units (RSU) can be deployed at a low cost and effectively alleviate the communication burden of regional Vehicular Ad-hoc Networks (VANETs) [31]. Unfortunately, due to the limited energy storage and peak hour communication demands in VANETs, RSUs must adjust their participation in communication according to the requirements and allocate energy reasonably to balance the workload. [31] proposed a spatio-temporal graph neural network algorithm that forecasts RSU network load through inputting the historical information around RSU and the topological relationship between RSU.

- Human Object Interaction

Learning in the space-time domain remains a very challenging problem in machine learning and computer vision. The main challenge is how to model interactions between objects and higher-level concepts, within the large spatio-temporal context [32]. In such a difficult learning task it is critical to efficiently model the spatial relationships, the local appearance, and the complex interactions and changes that take place over time. [32] introduced a spatio-temporal graph neural network model, recurrent in space and time, suitable for capturing both the local appearance and the complex higher-level interactions of different entities and objects within the changing world scene.

2. Dynamic Graph Representation

Temporal graph representation learning has been considered a very important aspect of graph machine learning [33]. With limitations of existing methods in capturing powerful representations due to reliance on discrete snapshots of the temporal graph, [33] proposed a dynamic graph representation learning method using spatio-temporal graph neural networks. Moreover, [33] used spatio-temporal GNNs today dynamically represent brain graphs.

3. Multiple object tracking

Tracking multiple objects in videos heavily depends on modeling the spatial-temporal interactions between objects. [34] proposed a spatio-temporal graph neural network algorithm that models spatial and temporal interactions among the objects.

4. Sign Language Translation

Sign languages, which engage visual-manual modalities to convey meanings, are the primary communication tools for the deaf and hard-of-hearing community [35]. To reduce the communication gap between spoken language and sign language users, machine learning is involved. Traditionally, neural machine translation has been heavily adopted while more advanced methods are needed to capture the spatial properties in sign languages. [35] presented a spatio-temporal graph neural network-based translation system, that is powerful in capturing spatial and temporal structures of sign language which led to state of art performance compared to traditional neural machine translation methods.

5. Technology Growth Ranking

Understanding the growth rate of technologies is a core key to the technology sector's business strategy. In addition, predicting the growth rate of technologies and their relations to each other informs business decision-making in terms of product definition, marketing strategies, and research and development [36]. [36] proposed a methodology to predict technology growth ranking from social networks using spatio-temporal graph neural networks.

6. Knowledge Graphs and Social Networks

Real-world graphs like social networks and knowledge graphs are dynamic. For example, in a social network, new users join over time and users interact with each other through messages and post reactions. In addition, new events appear with time in knowledge graphs. To account for the evolving dynamic properties in graphs, [37] introduced a temporal graph neural network that can handle billions of nodes and edges and can jointly learn the temporal, structural, and contextual relationships on dynamic graphs.

7. Audio Visuals and Emotion Perception

Effective dimension prediction from multi-modal data is becoming an increasingly challenging and important research area. For example, discriminative features from multiple modalities are critical to accurately recognizing emotional states. Motivated by their spatial and temporal power, [38] investigated spatio-temporal graph neural networks in audiovisuals. The framework achieved superior performance compared to traditional deep learning frameworks when experimented on emotional recognition applications. In addition, [38] proposed a spatio-temporal graph neural network that leverages emotion perception.

CHAPTER III

RELATED WORK

In this chapter, we present an overview of the previous work related to transferring learning on time series data and the transferability of graph neural networks.

A. Transfer Learning on Time Series Data

Fawaz et. al investigated transfer learning on time series classification tasks through convolutional neural networks [4]. [4] applied extensive experiments on UCR 85 datasets which is the current benchmark for time series classification tasks [4]. Experiments included pre-training models each on one of the UCR datasets and then fine-tuning models on the other 84 datasets resulting in 7140 different deep neural networks [4]. The fine-tuned model performance was relative to the source data selected [4]. When a bad source dataset is used for pretraining, the optimization algorithm can be stuck in a local optimum resulting in negative transfer. To mitigate negative transfer, [4] suggests the usage of the Dynamic Time Warping method that measures inter-datasets similarities for source and target dataset selection [4]. Additionally, Ye et al. proposed a deep transfer learning framework resorting to convolutional neural networks (DTr-CNN)[2]. The selection of Space and target datasets was based on both Dynamic Time Warping (DTW) and Jensen-Shannon (JS) divergence metrics [2]. The framework was evaluated on target datasets from real-world scenarios with limited labeled data [2].

Gupta et. al [39] researched the transferability of recurrent neural networks on medical time series data. [39] compared two transfer learning pipelines: Domain adaptation through a universal time series extractor (TimeNet) and task adaptation through pretraining deep recurrent networks (HealthNet)[39]. Both pipelines were evaluated on phenotyping and mortality prediction and achieved classification performance improvement while minimizing dependence on hand-crafted features [39].

Forecasting hierarchical time series is a challenging and time-consuming process. To mitigate this problem, transfer learning was applied [40]. First, the time series is trained at the bottom level of the hierarchy using the proposed deep LSTM auto-encoder (DLSTM-AE) architecture [40]. Then, transfer learning is applied to the upper levels of the hierarchy [40]. The proposed transfer learning pipeline was compared with state of art approaches in two energy and tourism case studies and outperformed all counterpart models [40].

Domain selection is considered one of the main challenges in time series transfer learning research. Naive selection of source and target domains can lead to negative transfer learning. To mitigate negative transfer in pandemic time series forecasting, we propose using first and second derivatives as similarity metric. The metrics heavily rely on the hurricane model that is extensively discussed in chapter IV.

	DTW	Shanon Divergence	First and Second Derivatives	Classification	Forecasting	Transfer Learning

Fawaz et. al	x			x		x
Gupta et al.	x	x			x	x
Jaffer et al.			x		x	
Ours			x		x	x

B. Transfer Learning in Graph Neural Networks

Lee et. al was the first to investigate transfer learning on graph-structured data. The proposed transfer learning framework is based on spectral graph neural networks and thus lacks the fine-tuning capability [41]. An extensive set of experiments was accomplished on four corpora (AG, DBP, YELP, and AMAZ) [41]. The knowledge transfer between task domains was most effective when the source and target domains possess high cosine similarity and correlation, while low similarity and correlation resulted in negative transfer [41]. Ruiz et al. introduced graphon NNs as limit objects of graph neural networks [41]. Graphons were experimented with as graph similarity metrics on graph neural network transfer learning schemes. Experiments concluded the existence of a tradeoff between discriminability and transferability of GNNs [42].

Yoa et. al introduced a new framework graph few shot learning (GFL)to improve the effectiveness of semi-supervised node classification by transferring knowledge learned from auxiliary graphs to a new target graph [42]. GFL integrates both local node-level and global graph-level knowledge to learn a transferable metric space [42]. Extensive experiments on four real-world graph datasets (Cola, PubMed, Reddit, Cita.) demonstrated the superiority of GFL over other algorithms from this research line (i.e., Matchingnet, Protonet) while achieving better performance than MAML [42].

Dai et. al proposed an adversarial domain adaptation framework classification [43]. The framework aims to leverage the label information in a partially labeled source network to assist node classification in a completely unlabeled or partially labeled target network [43]. The knowledge transfer is done through AdaGCN which combines adversarial domain adaptation and graph convolution. First, a semi-supervised learning component learns class discriminative node representations with given label information of the source and target networks [43]. Then, the adversarial domain adaptation component mitigates the distribution divergence between the source and target domains to facilitate knowledge transfer [43].

Hu et al. presented a transfer learning approach for graph data that is based on both pre-training graph neural networks at the level of individual nodes and the level of the entire graph [3]. This approach avoids negative transfer and improves generalization significantly across downstream tasks [3]. The proposed method was evaluated on node classification datasets for protein function prediction leading up to 9.4% absolute improvements in ROC-AUC compared to state of art non-pretrained models. Pretraining was less expressive for GCN, GraphSAGE, and GAT compared to pretraining using the most expressive GNN algorithm Graph GIN [3].

Han et al. introduced a transfer learning paradigm that leverages self-supervised tasks as auxiliary tasks to help the target task [44]. [44] proposed combing different auxiliary tasks adaptively for fine tuning the target task. Meta-learning was implemented to weigh the target model. Extensive multitask learning and finetuning experiments on OAG_CS, Reddit, Lat-FM, and Bookkeeping datasets we accomplished for the proposed approach and achieved significantly better performance than existing state of art methods [44].

The second challenge is transfer learning for time series data is the algorithms' transferability. In this thesis, we investigate the transferability of Spatio-temporal graph neural networks for pandemic forecasting using a meta learning based framework.

	Spectral GNN	Spatial GNN	Transfer Learning	Meta Learning	ST-GNNs	Transferability	Modified GAT
Lee et al.	x		x				
Hu et al.		x	x			x	
Han et al.		x	x	x			
		x	x	x	x		
Ours		x	x	x	x	x	x

CHAPTER IV

DOMAIN SELECTION

A challenge for transferring knowledge in non-Euclidian data including time series applications is the data similarity and domain selection. To select the appropriate source data for the target data, multiple metrics were proposed for graphs in general and for time series data specifically. These metrics include cosine similarity, correlation, and dynamic time rapping. In this thesis, we propose a new similarity metric to mitigate negative transfer in pandemic forecasting based on the Hurricane forecasting model.

A. Similarity Metrics in Literature

Previously, three metrics for domain selection in time series applications were proposed: correlation, cosine similarity, and dynamic time warping.

1.Correlation and Cosine Similarity

Correlation and Cosine similarity are the most common metrics used for graph similarity.

Similarity(A,B)= $\frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=0}^{n} A_i \times B_i}{\sqrt{\sum_{i=0}^{n} A_i^2} \times \sqrt{\sum_{i=0}^{n} B_i^2}}$(1)

Corr(A,B)= $\frac{\sum_{i=0}^{n} (A_i - \bar{A}) \times (B_i - \bar{B})}{\sqrt{\sum_{i=0}^{n} (A_i - \bar{A})^2} \times \sqrt{\sum_{i=0}^{n} (B_i - \bar{B})^2}} = Similarity((A - \bar{A}), (B - \bar{B}))$(2)

2.Dynamic Time Warping

Dynamic time warping (DTW) is an algorithm used to measure similarity between two sequences which may vary in time or speed [45]. First, the two-time series are divided into equal points. Second, the Euclidean distance between the first point in the first series and every point in the second series is calculated. Third, the time wrap stage is reached by storing the minimum distance computed. Then, movement to the second point is done and the second step is repeated. Finally, the process is repeated step by step along with the points till all points are exhausted. The same process is recurred but with the second series as a reference point. To compute DTW, all the minimum distances that were stored are added [45].

B. First and Second Derivatives

Following the hurricane model, we propose using the first derivative and second derivative for the target and source datasets selection. The derivatives will be computed at each time point using the below definitions:

First derivative: $g_i(t) = \frac{A_i(t)-A_i(t-\Delta t)}{\Delta t}(4)$

Second Derivative: $a_i(t) = g_i(t) - g_i(t-\Delta t)(5)$

C. Fractional Derivatives

A fractional derivative is a derivative of any arbitrary order, real or complex, and was first proposed by Leibniz et al. in 1695. The proposition relied on the similarity between the binomial theorem and the Leibniz rule for the fractional derivative of a product of two functions. After that, fractional calculus was introduced by one Abel et al. including the idea of fractional-order integration and differentiation, the mutually inverse relationship between them, the understanding that fractional-order differentiation and integration can be considered as the same generalized operation, and even the unified notation for differentiation and integration of arbitrary real order [46].

Caputo fractional derivative of order α:

Let $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be an element of $C^{+\infty}([a, x])(-\infty < a < x < +\infty)$, with $\alpha \geq 0$ and $n = [\alpha] + 1$, with $[\alpha]$ being the integer part of α. Then, the Caputo fractional derivative of order α of $f(x)$ is defined as follows:

$$(cD_a^{\alpha})f(x) = \begin{cases} \frac{1}{\Gamma(n-a)} \int_a^x \frac{d^n f(t)}{dt^n} \frac{dt}{(x-t)^{\alpha-n+1}}, & \alpha \notin \mathbb{N}, \\ \frac{d^{n-1} f(x)}{dx^{n-1}}, & \alpha = n-1 \in \mathbb{N} \cup \{0\}. \end{cases} \quad (6)$$

Riemann-Liouville fractional derivative of order α:

Let $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be an element of $L^1([a, x])(-\infty < a < x < +\infty)$, with $\alpha \geq 0$ and $n = [\alpha] + 1$, with $[\alpha]$ being the integer part of α. Then, the Riemann-Liouville fractional derivative of order α of $f(x)$ is defined as

$$\left(D_{a^+}^{\alpha}\right)f(x) = \begin{cases} \frac{1}{\Gamma(n-a)} \frac{d^n}{dx^n} \int_a^x \frac{f(t)}{(x-t)^{\alpha-n+1}} dt, & \alpha \notin \mathbb{N}, \\ \frac{d^{n-1} f(x)}{dx^{n-1}}, & \alpha = n-1 \in \mathbb{N} \cup \{0\}. \end{cases} \quad (7)$$

D. Covid Dataset

The dataset used is the OWID covid dataset [47]. The dataset includes the confirmed cases and deaths per country from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). The cases & deaths dataset is updated daily. It also includes the Hospitalizations and intensive care unit (ICU) admissions collected from official sources [47].

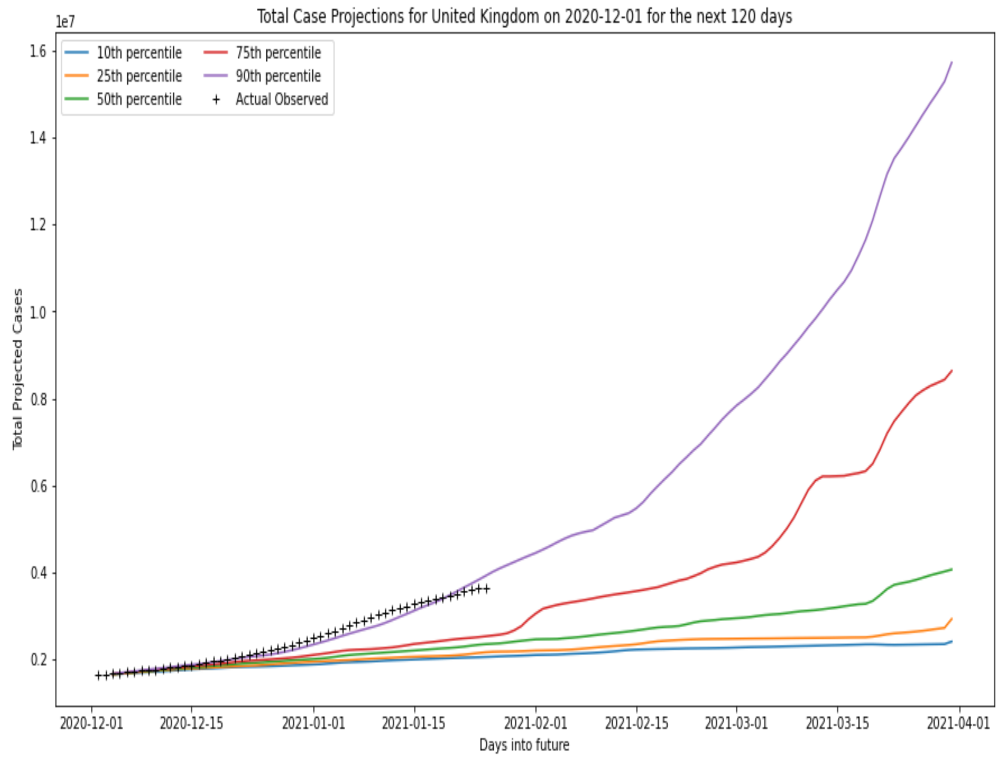
E. Hurricane Model

The hurricane model is the pandemic forecasting model proposed by Jaffer et al. (2022) [48]for pandemic forecasting. The model forecasts the covid cases in a country x through computing:

First derivative: $g_i(t) = \frac{A_i(t) - A_i(t-\Delta t)}{\Delta t}$ (4)

Second Derivative: $a_i(t) = g_i(t) - g_i(t-\Delta t)$(5)

After that, the model recalls the forecasts from other countries having the same first a second derivatives [48]. The forecast of covid cases in country x at time t is the weighted average of the countries having the same first a second derivatives (Figure 3)

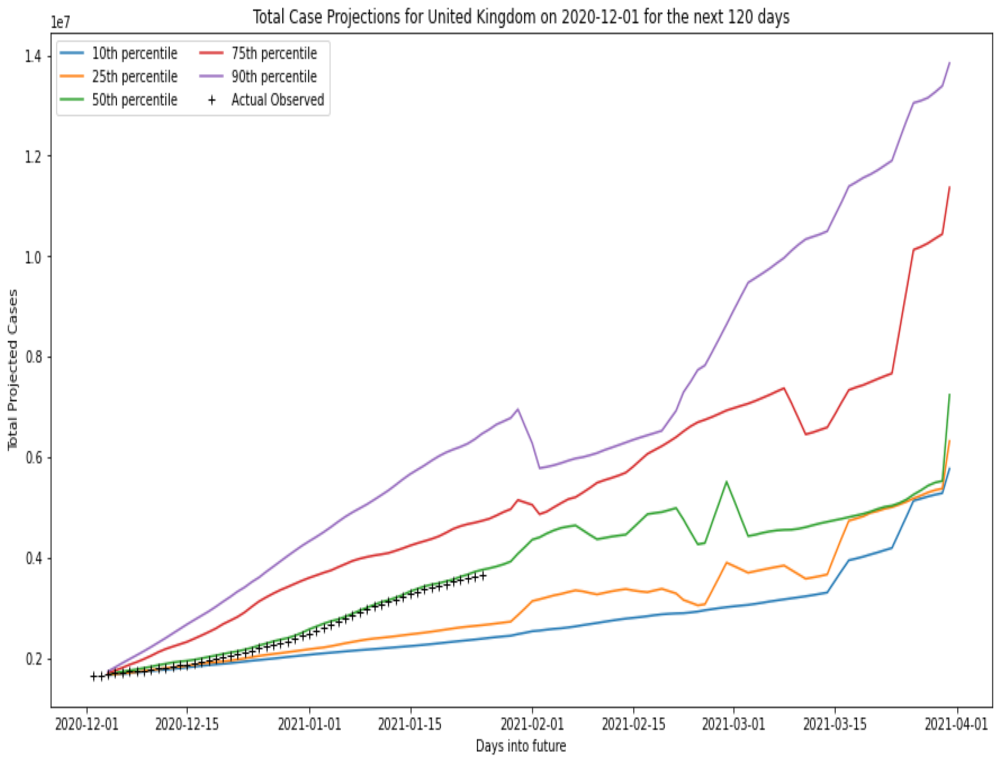


3.Vanilla Hurricane Model with epsD1 = 0.005 and epsD2 = 0.0005

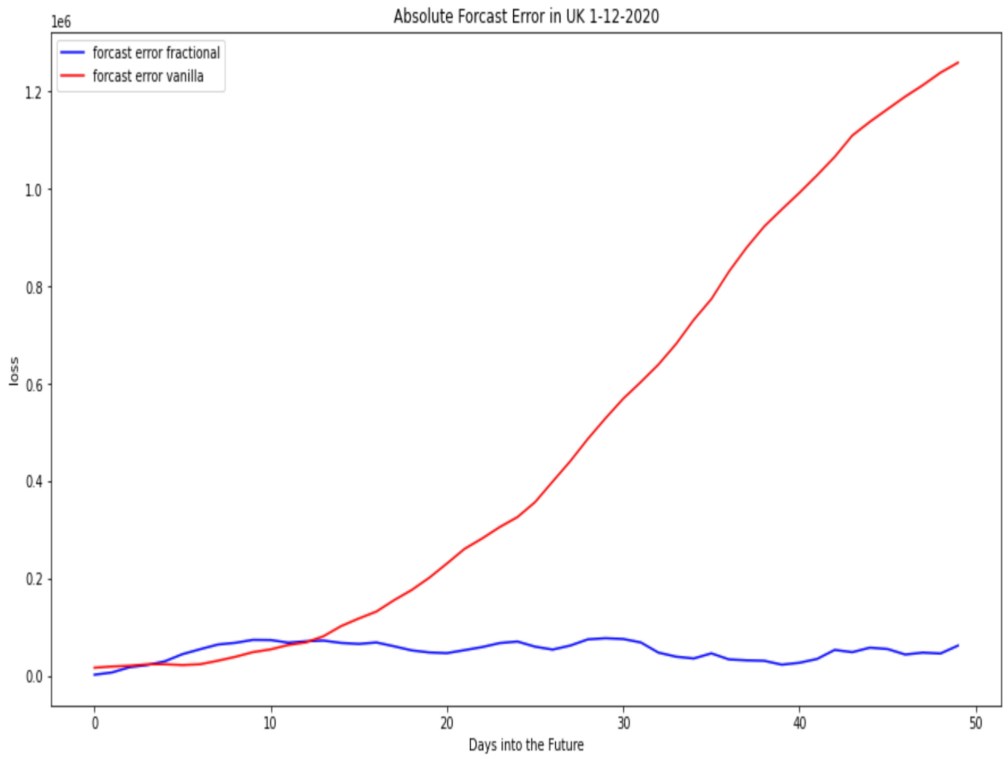
F. Fractional Hurricane Model

The fractional hurricane model is a modified version of the vanilla hurricane forecasting model. The modified model uses fractional derivatives instead of first and second derivatives.

The experimental results using the fractional Hurricane model for UK pandemic forecasting surpassed the vanilla hurricane in both short and long-term forecasting. In Figure 4, we can see how the prediction (green) aliens with the actual observations (black) for the next 2 weeks. On the contrary, the predictions of the vanilla hurricane are far away from the actual cases. In addition, the fractional hurricane model achieved low forecast error for the next 60 days compared to the vanilla hurricane resulting in degraded performance after two weeks of forecast (Figure 5).



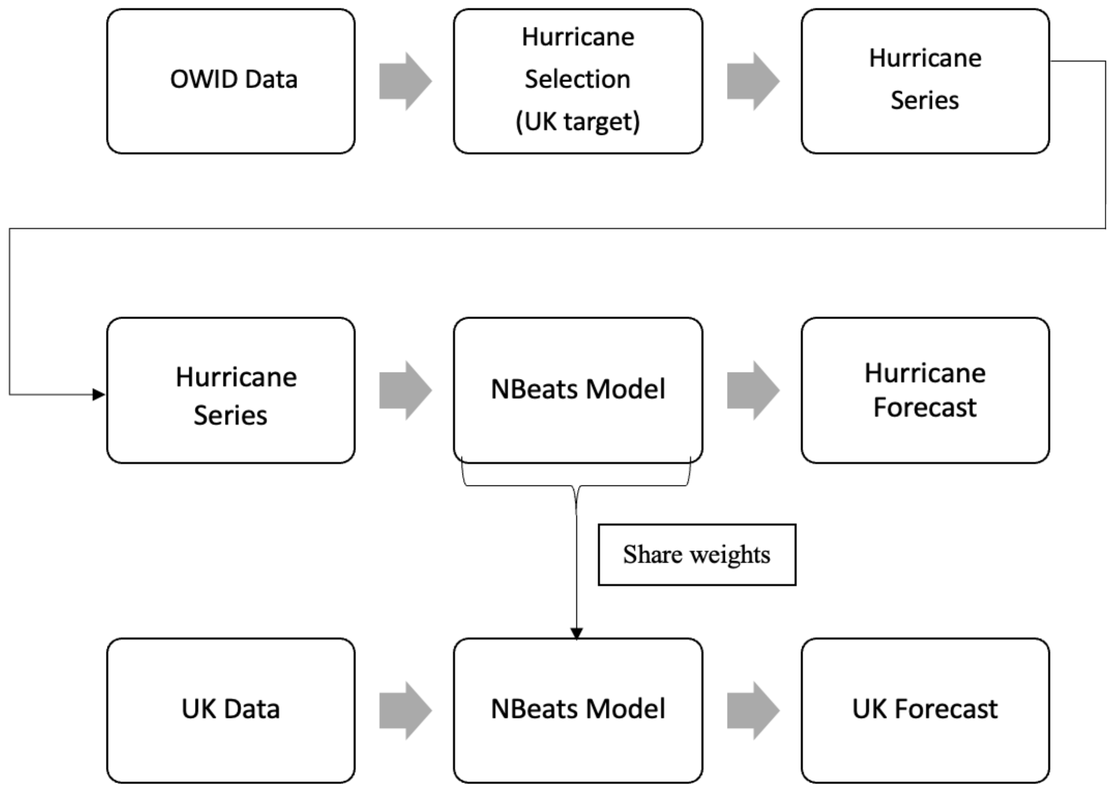
4.Fractional Hurricane Model with fractional orders 0.8 and 1.6



5.Absolute Forecast Error in UK

G. Transfer Hurricane Model

Following the superior performance of hurricane pandemic forecasting, we propose using first and second derivatives as metrics for transfer learning. First, the covid dataset is inputted to the hurricane model to select similar covid time series data. Then, the selected time series are used to pretrain a neural network. Finally, the pretrained model is used to fine tune the neural network that is targeted to forecast covid cases in the targeted country. We evaluated the proposed method on a neural basis expansion analysis for interpretable time series forecasting **(**NBEATS) [49] based transfer learning on the bias, MAE, MSE, and RMSE.



6.Schematic of the proposed transfer learning pipeline

H. Experimental Results

Table 2.Domain Selection Experimental Results for 180 days

	United Kingdom			
	Bias	MAE	MSE	RMSE
ARIMA	**-156**	**2367**	**13811049**	**3716**
NBEATS	-1445	2714	15678842	3959

Transfer Hurricane	812	2421	15165708	3894

ARIMA achieved the highest bias, MAE, MSE, and RMSE of -156, 2367. 13811094, 3716 respectively. However, we were able to achieve positive transfer using NBEATS following the transfer hurricane approach, where transfer hurricane achieved 812, 2421, 15165708, and 3894 compared to -1445, 2714, 15678842, and 3959 errors for NBEATS without transfer learning.

I. Analysis and Discussion

The transfer learning approach using NBEATS architecture was successful. Using the proposed metrics, transfer learning improved the NBEATS performance and resulted in less bias, MAE, MSE, and RMSE. This concludes that the proposed metrics are sufficient to mitigate the negative transfer phenomenon for pandemic forecasting. However, ARIMA surpassed both NBEATS and transfer hurricane, which urges further advances in our pipeline, including the intervention of fractional derivatives motivated by the successes of the fractional hurricane model.

CHAPTER V

TRANSFER LEARNING FOR SPATIO-TEMPORAL GRAPH NEURAL NETWORKS

A. Covid Dataset

The dataset relies on Facebook's human mobility between administrative NUTS33 regions dataset for the spatial domain. The spatial data is collected directly from mobile phones that have the Facebook application installed and the Location History setting enabled. The spatio-temporal dataset focuses on 4 European countries: Italy, Spain, France, and England. The number of cases in the different regions of the 4 considered countries was gathered from owid dataset[50].

B. Spatio-Temporal Graph Neural Networks

1.Graph Convolutional Neural Networks (GCN)

Let graph G = (V, E) be of nodes V and edges E where n = |V | denotes the number of nodes. Then, given a country, a series of graphs are created, each corresponding to a specific date t, i. e., G (1), . . ., G(T). A single date's mobility data is transformed into a weighted, directed graph whose vertices represent the NUTS3 regions and edges capture the mobility patterns. For example, the weight A (t) v,u of the edge (v, u) from vertex v to vertex u denotes the total number of people that moved from region v to region u at time t. The GCN vanilla and the meta-based transfer learning approach heavily rely on [50].

2.Graph Attention Networks

The Graph attention network (GAT) is a GCN network that introduces self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations. The attention weights α_{ij} are computed based on the nodes' neighborhood following the equations below [51].

$$z_i^{(l)} = W^{(l)} h_i^{(l)} , \tag{8}$$

$$e_{ij}^{(l)} = Leaky\ ReLU\left(\vec{a}^{(l)^T}\left(z_i^{(l)} \| z_j^{(l)}\right)\right) , \tag{9}$$

$$\alpha_{ij}^{(l)} = \frac{\exp\left(e_{ij}^{(l)}\right)}{\sum_{k \in N(i)} \exp\left(e_{ik}^{(l)}\right)} , \tag{10}$$

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in N(i)} \alpha_{ij}^{(l)} z_j^{(l)}\right) , \tag{11}$$

3.Modified Graph Attention Networks

In GAT, every node attends to its neighbors given its representation as the query. In addition, GATs use a static attention mechanism, thus it cannot express a controlled problem, from even fitting the training data. To remove this limitation, GATv2 introduces a simple fix by modifying the order of operations resulting in a dynamic graph attention variant that is strictly more expressive than GAT [52].

In this thesis, we propose two new STGNNs relying on GAT and GATv2 algorithms. We experiment with the proposed algorithms on random samples for short-term forecasting and achieved superior performance over the GCN vanilla. In addition, the state-of-art performer over random sub-samples is the modified GATv2 achieving an error of 4.67 only compared to 5.21 for GAT and 6.41 for GCN vanilla.

C. Research Methodology

Considering the algorithmic transferability challenge discussed in the literature, we investigate the transferability of three spatio-temporal graph neural network algorithms (GCN, GAT, and GATv2). The transfer learning pipeline heavily relies on a meta-learning approach, MAML [52].

MAML, or Model-Agnostic Meta-Learning, is a model and task-agnostic algorithm for meta-learning that trains a model's parameters such that a small number of gradient updates will lead to fast learning on a new task.

In meta training, set aside a portion of D as a test task. T

$$\mathcal{D}_{meta-train} = \{(\mathcal{D}_1^{tr}, \mathcal{D}_1^{test}), \dots, (\mathcal{D}_n^{tr}, \mathcal{D}_n^{test})\}$$

The goal is to approximate a learning problem $p(\emptyset_i|\mathcal{D}_i^{tr}, \theta)$ using a neural network. Point estimation is used as such. $\emptyset_i = f_\theta(\mathcal{D}_i^{tr})$. In adapting to the new task, θ is updated using

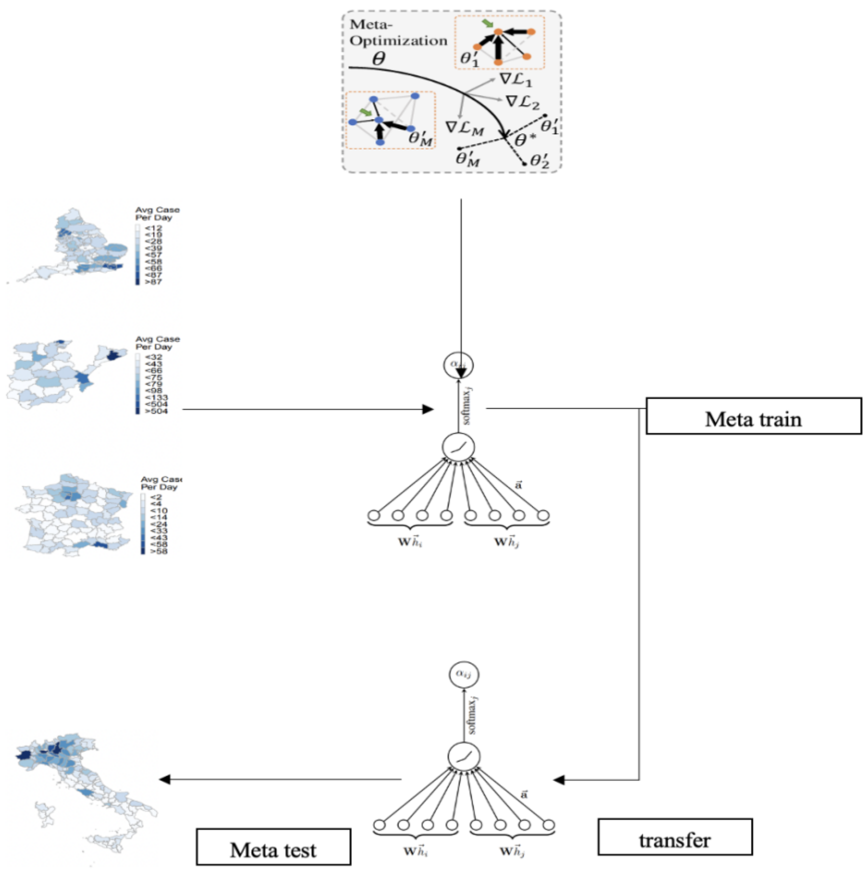
$$\emptyset_i = \theta - \alpha \nabla_\theta \mathcal{L}\left(\theta, \mathcal{D}^{tr}_{\mathcal{T}_i}\right) (12)$$

Where $\mathcal{D}^{tr}_{\mathcal{T}_i}$ is training data for the task and α is the step size. The meta objective is.

$$\min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}\left(\emptyset_i, \mathcal{D}^{test}_{\mathcal{T}_i}\right) = \min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}\left(\theta - \alpha \nabla_\theta \mathcal{L}\left(\theta, \mathcal{D}^{tr}_{\mathcal{T}_i}\right), \mathcal{D}^{test}_{\mathcal{T}_i}\right) (13)$$

Finally, the model parameter θ is meta updated using SGD.

$$\theta \leftarrow \theta - \beta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}\left(\emptyset, \mathcal{D}^{test}_{\mathcal{T}_i}\right) (14)$$

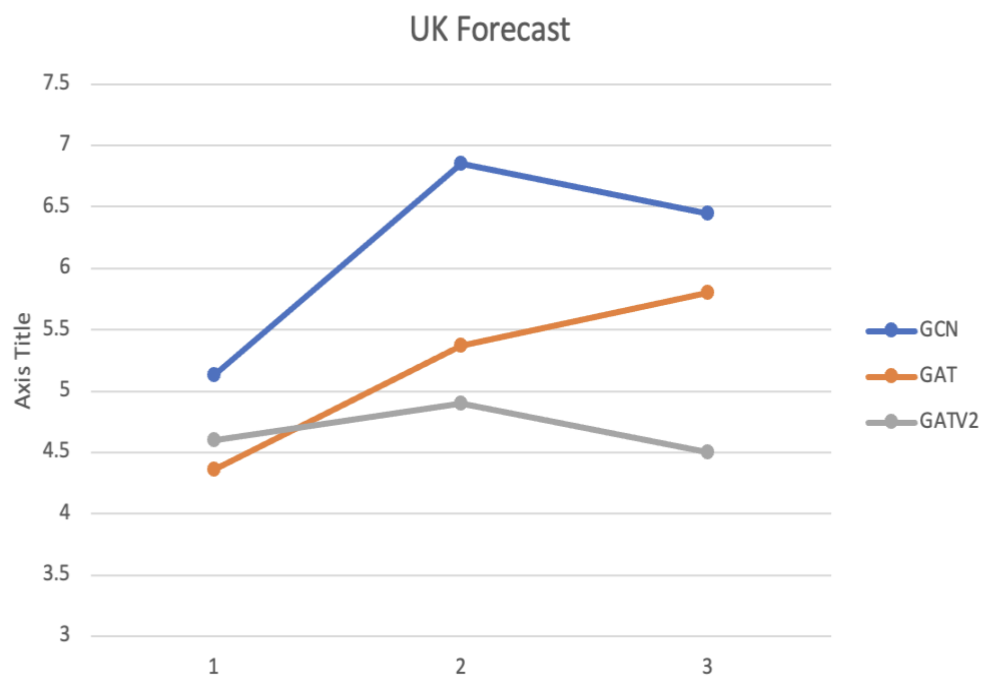


7. Meta-learning transfer approach

D. Experiments

1.Attention-Based Pandemic forecasting

The GNNs were trained using pytorch geometric framework and NVIDIA P100 GPU or a maximum of 500 epochs with early stopping after 50 epochs of patience. Early stopping was introduced from the 100th epoch and onward. The batch size was set to 8 and the optimizer used is Adam the batch size to 8. The learning rate and meta learning rate were set to 0.01. The evaluation is based on Mean absolute error on the test set: error $= |\hat{y}(t) - y(t)|$ over time.



8.MAE for 3 Days Forecast for STGNN forecasting in UK

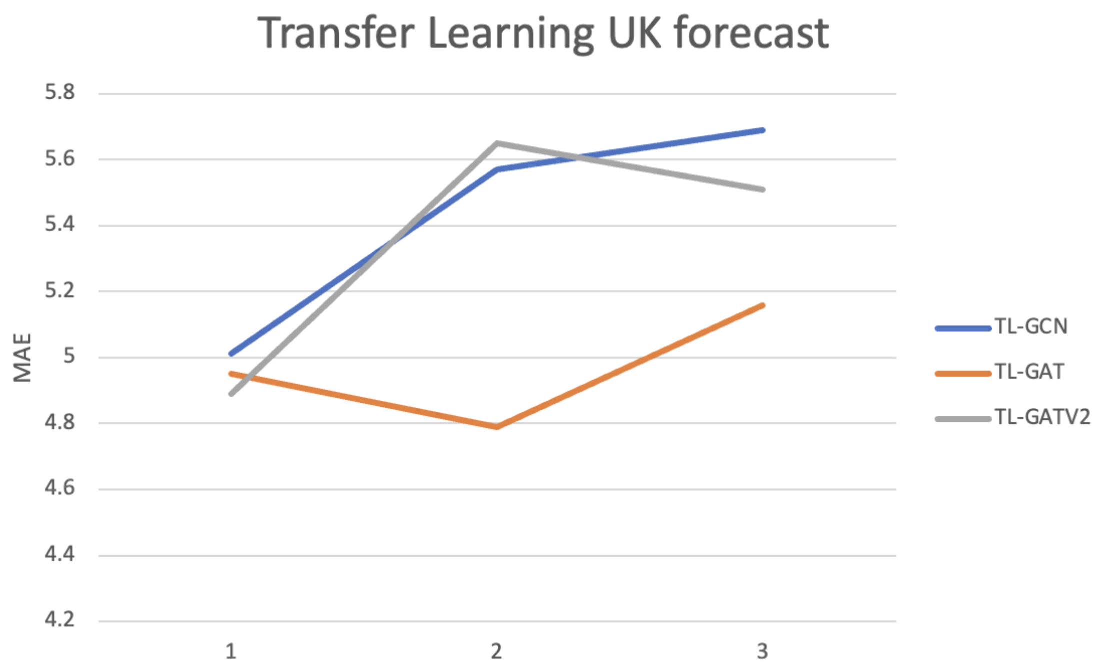
Table 3 Mean of MAE for STGNN forecasting

	Mean Error (3 days forecast)
GCN	6.41
GAT	5.21
GATv2	**4.67**

GATv2 achieved superior performance as a forecasting algorithm compared to GAT and GCN. Using average MAE, GATv2 had 4.67 error for the 3 days forecast compared to 5.21 for GAT and 6.41 for GCN over randomized selected samples (Table 3). In addition, the GATv2 forecast error was more stable compared to dramatically increasing error for GAT and GCN as the forecasting days increases (Figure 8).

2. Transferability of Spatio-temporal GNNs on Subsamples

The experiments were conducted on TL-GCN, TL-GATv2, and TL-GAT over 3 randomly selected test samples. The GNNs were trained for a maximum of 300, 500, 800, and 1000 epochs with early stopping after 50 epochs of patience. Early stopping was introduced from the 100^{th} epoch and onward. The batch size was set to 8 and the optimizer used is Adam the batch size to 8. The learning rate and meta learning rate were varied between 0.01 and 0.001. The evaluation is based on Mean absolute error on the test set: error = $|\hat{y}(t) - y(t)|$ over time.



9.MAE for 3 Days Forecast for TL-STGNN forecasting in UK on submsamples

Table 4.Mean of MAE for UK forecasting for 3 days on Subsamples

Model	Up to next 3 days
GCN	6.41
GAT	5.21
GAT_v2	**4.66**
TL-GCN	5.42
TL-GAT	4.97
TL-GATv2	5.27

GATv2 achieved superior performance as a forecasting algorithm compared to GAT and GCN. However, on samples, transfer learning using GAT resulted in better performance compared to GATv2 and GCN. TL-GAT had an MAE of 4.967error for the

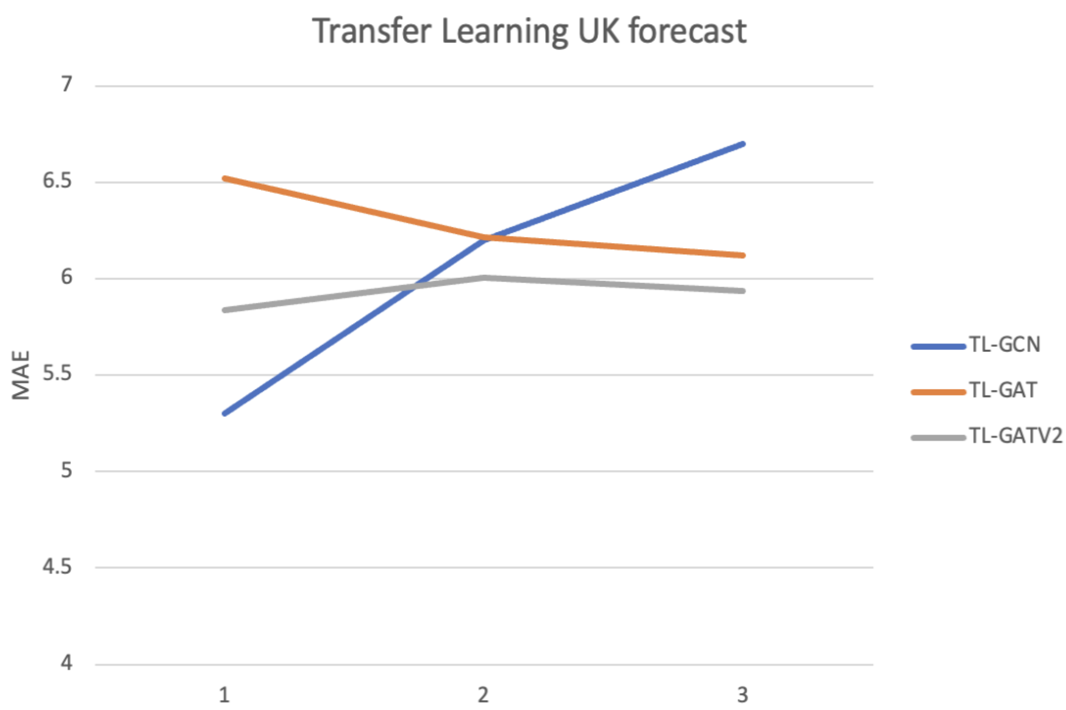
3 days forecast compared to 5.27 for TL-GATv2 and 5.42 for TL-GCN over randomized selected samples (Table 4).

3. Transferability of Spatio-temporal GNNs

The experiments were conducted on ARIMA, GCN, GCN-LSTM, TL-GCN, TL-GATv2, and TL-GAT over 40 test samples. The GNNs were trained for a maximum of 300, 500, 800, and 1000 epochs with early stopping after 50 epochs of patience. Early stopping was introduced from the 50^{th} ,100^{th}, 150^{th} ,300^{th} epoch and onward. The batch size was set to 8 and the optimizer used is Adam the batch size to 8. The learning rate and meta learning rate were varied between 0.01 and 0.001. The evaluation is based on Mean absolute error on the test set: error = $|\hat{y}(t) - y(t)|$ over time.

Table 5.Optimal Hyper-parameters

Hyper-parameter	Optimal value
Number of epochs	500
Early stopping	150
Batch size	8
Meta-learning rate	0.001
Learning rate	0.001
optimizer	Adam



10.MAE for 3 Days Forecast for TL-STGNN forecasting in UK

Table 6. Mean of MAE for UK forecasting for 3 days

Model	MAE
ARIMA	13.77
GCN	6.36
GCN_LSTM	6.41
TL-GCN	6.05
TL-GAT	6.28
TL-GATv2	**5.92**

Transfer learning using GATv2 over 40 samples achieved the state of art performance compared to GAT and GCN. Using average MAE, TL-GATv2 had 5.92 error for the 3 days forecast compared to 6.28 for TL-GAT and 6.05 for TL-GCN over randomized selected samples (Table 6). In addition, the TL-GATv2 and TL-GAT forecast error was more stable compared to dramatically increasing error for TL-GCN (Figure 11).

E. Discussion

Transfer learning on the modified attention-based graph neural networks resulted in a better performance compared to the vanilla graph convolutional neural networks and vanilla graph attention networks. Considering that it is the most expressive graph neural network, this validated our assumption on the relationship between transferability and expressivity for time varying graphs. In addition, the proposed approach surpassed traditional forecasting models like ARIMA. Furthermore, the modified GAT attention network that achieved 5.92 MAE surpassed the graph attention network that achieved an error of 6.28. This assures the importance of the recently introduced adaptive attention, especially when compared to the experimental results on randomized subsamples. To add, the modified version was more robust to negative transfer compared to vanilla GAT which resulted in the abruptly improve in performance of GATv2 compared to degraded performance of GAT when trained on larger samples. However, GNN-based models are very computationally expensive and time-consuming. ARIMA-based models need a few minutes of training compared to hours of training using GNNs in the same settings. Nevertheless, pandemic applications are not designed for online settings and are used by governments that don't restrict computational needs. Finally, we suspect with more data availability we expect to have better performance through connecting our similarity metric approach with the transferability of GNNs research direction.

CHAPTER VI

CONCLUSION

A. Thesis Contribution

The contribution is two folded. We attempted to improve the transferability of spatio-temporal graph neural networks for time series applications. First, we diagnosed the challenges in transfer learning in both time series and graph neural networks: domain selection and transferability of algorithms. Second, we surveyed the spatio-temporal graph neural networks while proposing a new taxonomy for the corresponding algorithms. Moving to the domain selection challenge, we proposed a new similarity metric for mitigating negative transfer in pandemic forecasting based on the Hurricane model. We then tackled the second challenge, the transferability of GNNs by proposing a new transformer-based spatio-temporal graph neural network and evaluating its transferability using a meta-learning-based pipeline referenced to other graph neural networks.

B. Future Work

In the future, we plan to investigate further transfer learning research. The priority will be to run extensive experiments to validate our assumptions. Next, we plan to expand the similarity metric to fractional calculus relying upon the fractional hurricane model. In addition, we will investigate the effect of modern meta learning algorithms like reptile on transferability of STGNNs. Finally, we explore more the transferability of spatio-temporal GCNs on more complex time-varying structures.

REFERENCES

1. José F. Torres, Dalil Hadjout, Abderrazak Sebaa, Francisco Martínez-Álvarez, and Alicia Troncoso.Big Data.Feb 2021.3-21.http://doi.org/10.1089/big.2020.0159

2. Rui Ye, Qun Dai(2021), Implementing transfer learning across different datasets for time series forecasting, Pattern Recognition,Volume 109,107617,doi.org/10.1016/j.patcog.2020.107617.

3. Hu, W., Liu, B., Gomes, J., Zitnik, M., Liang, P., Pande, V.S., & Leskovec, J. (2019). Pre-training Graph Neural Networks. ArXiv, abs/1905.12265.

4. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. (2018). Transfer learning for time series classification. 2018 IEEE International Conference on Big Data (Big Data), 1367-1376.

5. Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

6. Demir, A., Koike-Akino, T., Wang, Y., Haruna, M., & Erdoğmuş, D. (2021). EEG-GNN: Graph Neural Networks for Classification of Electroencephalogram (EEG) Signals. ArXiv, abs/2106.09135.

7. Deng, A., & Hooi, B. (2021). Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. ArXiv, abs/2106.06947.

8. Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.

9. Cao, D., Wang, Y., Duan, J., Zhang, C., Zhu, X., Huang, C., ... & Zhang, Q. (2020). Spectral temporal graph neural network for multivariate time-series forecasting. Advances in neural information processing systems, 33, 17766-17778

10. Simeunović, Jelena, et al. "Spatio-temporal graph neural networks for multi-site PV power forecasting." IEEE Transactions on Sustainable Energy 13.2 (2021): 1210-1220.

11. Chen, C., Li, K., Teo, S.G., Zou, X., Wang, K., Wang, J., & Zeng, Z. (2019). Gated Residual Recurrent Graph Neural Networks for Traffic Prediction. AAAI.

12. Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020, August). Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 753-763).

13. Karimi, A. M., Wu, Y., Koyuturk, M., & French, R. H. (2021). Spatiotemporal Graph Neural Network for Performance Prediction of Photovoltaic Power Systems. Proceedings of the AAAI Conference on Artificial Intelligence, 35(17), 15323-15330. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/17799

14. Huang, L., Wu, L., Zhang, J., Bian, J., & Liu, T. Y. (2022). Dynamic Relation Discovery and Utilization in Multi-Entity Time Series Forecasting. arXiv preprint arXiv:2202.10586.

15. Simeunović, J., Schubnel, B., Alet, P. J., & Carrillo, R. E. (2021). Spatio-temporal graph neural networks for multi-site PV power forecasting. IEEE Transactions on Sustainable Energy, 13(2), 1210-1220.

16. Kan, J., Hu, K., Hagenbuchner, M., Tsoi, A. C., Bennamoun, M., & Wang, Z. (2022). Sign Language Translation with Hierarchical Spatio-Temporal Graph Neural Network. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3367-3376).

17. Oreshkin, B. N., Amini, A., Coyle, L., & Coates, M. (2021, May). FC-GAGA: Fully connected gated graph architecture for spatio-temporal traffic forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 9233-9241).

18. Kim, B. H., Ye, J. C., & Kim, J. J. (2021). Learning dynamic graph representation of brain connectome with spatio-temporal attention. Advances in Neural Information Processing Systems, 34, 4314-4327.

19. Chu, P., Wang, J., You, Q., Ling, H., & Liu, Z. (2021). TransMOT: Spatial-Temporal Graph Transformer for Multiple Object Tracking. ArXiv, abs/2104.00194.

20. Li, Y., & Moura, J. M. (2019). Forecaster: A graph transformer for forecasting spatial and time-dependent data. arXiv preprint arXiv:1909.04019.

21. Zhang, X., Huang, C., Xu, Y., Xia, L., Dai, P., Bo, L., ... & Zheng, Y. (2021, May). Traffic flow forecasting with spatial-temporal graph diffusion network. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 17, pp. 15008-15015).

22. Shen, Y., Li, L., Xie, Q., Li, X., Xu, G. (2022). A Two-Tower Spatial-Temporal Graph Neural Network for Traffic Speed Prediction. In: Gama, J., Li, T., Yu, Y., Chen, E., Zheng, Y., Teng, F. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2022. Lecture Notes in Computer Science(), vol 13280. Springer, Cham. https://doi.org/10.1007/978-3-031-05933-9_32

23. Wang, L., Adiga, A., Chen, J., Sadilek, A., Venkatramanan, S., & Marathe, M. (2022). CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12191-12199. https://doi.org/10.1609/aaai.v36i11.21479

24. Li, Y. F., Gao, Y., Lin, Y., Wang, Z., & Khan, L. (2020). Time Series Forecasting Using a Unified Spatial-Temporal Graph Convolutional Network. In Proceedings of Preregister Workshop in 34th Conference on Neural Information Processing Systems.

25. Shao, W., Jin, Z., Wang, S., Kang, Y., Xiao, X., Menouar, H., ... & Salim, F. (2022). Long-term Spatio-temporal Forecasting via Dynamic Multiple-Graph Attention. arXiv preprint arXiv:2204.11008.

26. Karimi, A. M., Wu, Y., Koyuturk, M., & French, R. H. (2021). Spatiotemporal Graph Neural Network for Performance Prediction of Photovoltaic Power Systems. Proceedings of the AAAI Conference on Artificial Intelligence,

35(17), 15323-15330. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/17799

27. Hadou, S., Kanatsoulis, C. I., & Ribeiro, A. (2021). Space-time graph neural networks. arXiv preprint arXiv:2110.02880.

28. Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020, August). Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 753-763).

29. Wenlong Liao, Birgitte Bak-Jensen, Jayakrishnan Radhakrishna Pillai, Zhe Yang, Kuangpu Liu,Short-term power prediction for renewable energy using hybrid graph convolutional network and long short-term memory approach,Electric Power Systems Research,Volume 211,2022,108614,ISSN 0378-7796,https://doi.org/10.1016/j.epsr.2022.108614.

30. Simeunović, J., Schubnel, B., Alet, P. J., & Carrillo, R. E. (2021). Spatio-temporal graph neural networks for multi-site PV power forecasting. IEEE Transactions on Sustainable Energy, 13(2), 1210-1220.

31. Zheng, Hang & Ding, Xu & Wang, Yang & Zhao, Chong. (2021). Attention Based Spatial-Temporal Graph Convolutional Networks for RSU Communication Load Forecasting. 10.1007/978-3-030-92635-9_7.

32. Nicolicioiu, A.L., Duta, I., & Leordeanu, M. (2019). Recurrent Space-time Graph Neural Networks. NeurIPS.

33. Wen, Z., & Fang, Y. (2022, April). TREND: TempoRal Event and Node Dynamics for Graph Representation Learning. In Proceedings of the ACM Web Conference 2022 (pp. 1159-1169).

34. Jain, A., Zamir, A. R., Savarese, S., & Saxena, A. (2016). Structural-rnn: Deep learning on spatio-temporal graphs. In Proceedings of the ieee conference on computer vision and pattern recognition (pp. 5308-5317).

35. Kan, J., Hu, K., Hagenbuchner, M., Tsoi, A. C., Bennamoun, M., & Wang, Z. (2022). Sign Language Translation with Hierarchical Spatio-Temporal Graph Neural Network. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3367-3376).

36. D., Brahmaroutu, A., Nassar, M., & Ahmed, N.K. (2022). Technology Growth Ranking Using Temporal Graph Representation Learning.

37. Bloemheuvel, S., Hoogen, J. V. D., Jozinović, D., Michelini, A., & Atzmueller, M. (2022). Multivariate Time Series Regression with Graph Neural Networks. arXiv preprint arXiv:2201.00818.

38. Bhattacharya, U., Mittal, T., Chandra, R., Randhavane, T., Bera, A., & Manocha, D. (2020, April). Step: Spatial temporal graph convolutional networks for emotion perception from gaits. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 02, pp. 1342-1350).

39. Gupta, P., Malhotra, P., Narwariya, J., Vig, L., & Shroff, G.M. (2020). Transfer Learning for Clinical Time Series Analysis Using Deep Neural Networks. Journal of Healthcare Informatics Research, 4, 112-137.

40. Sagheer, A., Hamdoun, H., & Youness, H. (2021). Deep LSTM-Based Transfer Learning Approach for Coherent Forecasts in Hierarchical Time Series. Sensors, 21(13), 4379. doi:10.3390/s21134379

41. Lee, J., Kim, H., Lee, J., & Yoon, S. (2017). Transfer Learning for Deep Learning on Graph-Structured Data. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1). Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/10904

42. Yao, H., Zhang, C., Wei, Y., Jiang, M., Wang, S., Huang, J., Chawla, N., & Li, Z.J. (2020). Graph Few-shot Learning via Knowledge Transfer. AAAI.

43. Dai, Q., Shen, X., Wu, X., & Wang, D. (2019). Network Transfer Learning via Adversarial Domain Adaptation with Graph Convolution. ArXiv, abs/1909.01541.

44. Han, X., Huang, Z., An, B., & Bai, J. (2021). Adaptive Transfer Learning on Graph Neural Networks. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining.

45. Bird, Jordan & Faria, Diego & Manso, Luis & Ekárt, A. & Buckingham, Christopher. (2019). A Deep Evolutionary Approach to Bioinspired Classifier Optimisation for Brain-Machine Interaction.