Universidad Politécnica
de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Máster Universitario en Inteligencia Artificial

Trabajo Fin de Máster

Spatio-Temporal Graph Convolutional Networks (ST-GCN) for predicting International Trade

Autor: Alberto Tomás Martín
Tutor: Damiano Zanardini

Madrid, Julio 2024

Este Trabajo Fin de Máster se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Máster
Máster Universitario en «título del máster»

Título: Spatio-Temporal Graph Convolutional Networks (ST-GCN) for predicting International Trade

Julio 2024

Autor(a): Alberto Tomás Martín
Tutor(a): Damiano Zanardini
Departamento de Inteligencia Artificial
ETSI Informáticos
Universidad Politécnica de Madrid

Resumen

Esta tesis explora la aplicación de Redes de Convolución Espacio-Temporales (ST-GCN) para predecir flujos de comercio internacional. Aprovechando el Índice de Complejidad Económica (ECI) y una matriz de adyacencia refinada, el estudio tiene como objetivo capturar tanto las dimensiones temporales como espaciales de los datos de comercio internacional. El ECI se calcula utilizando los datos de productos SITC desde 1962 hasta 2021, mientras que la matriz de adyacencia considera las distancias bilaterales entre las principales ciudades ponderadas por la participación de la población.

Nuestra investigación evalúa diferentes configuraciones de hiperparámetros y optimizadores, incluyendo AdamW y RMSprop, combinados con programadores de tasa de aprendizaje como StepLR y CosineAnnealingLR. Realizamos experimentos utilizando matrices de distancia regulares e invertidas exponencialmente para evaluar su impacto en el rendimiento del modelo. Los hallazgos revelan que la matriz de distancia invertida exponencialmente mejora significativamente la capacidad del modelo para capturar dependencias espacio-temporales, resultando en una menor pérdida de validación y una mayor precisión predictiva.

Esta tesis contribuye al avance de las aplicaciones de Redes Neuronales de Grafos (GNN) en la previsión económica. La integración de datos temporales y espaciales a través de formaciones de matrices innovadoras demuestra un marco robusto para predecir patrones complejos de comercio, ofreciendo valiosos conocimientos para los responsables de políticas y economistas en un mundo cada vez más interconectado.

Abstract

This thesis explores the application of Spatio-Temporal Graph Convolutional Networks (ST-GCN) to predict international trade flows. By leveraging the Economic Complexity Index (ECI) and a refined adjacency matrix, the study aims to capture both temporal and spatial dimensions of international trade data. The ECI is computed using the SITC product data from 1962 to 2021, while the adjacency matrix considers bilateral distances between major cities weighted by population shares.

Our research evaluates different hyperparameter configurations and optimizers, including AdamW and RMSprop, combined with learning rate schedulers like StepLR and CosineAnnealingLR. We conducted experiments using both regular and exponentially inverted distance matrices to assess their impact on the model's performance. The findings reveal that the exponentially inverted distance matrix significantly enhances the model's ability to capture spatio-temporal dependencies, resulting in lower validation loss and improved predictive accuracy.

This thesis contributes to the advancement of Graph Neural Networks (GNN) applications in economic forecasting. The integration of temporal and spatial data through innovative matrix formations demonstrates a robust framework for predicting complex trade patterns, offering valuable insights for policymakers and economists in an increasingly interconnected world.

Table of Contents

Chapter 1

Introduction

International trade is an integral component of the global economy, facilitating the exchange of goods and services across national boundaries, driving economic growth, fostering specialization, and enabling countries to leverage their comparative advantages. Predicting the evolution of international trade presents formidable challenges due to the multifaceted nature of economic interactions, policy changes, and geopolitical events. Accurate trade forecasts are indispensable for policymakers, economists, and businesses as they inform decisions regarding economic policies, trade agreements, and market strategies [20].

Historically, econometric models have been the cornerstone of international trade forecasting. The Gravity Model of Trade, a prevalent approach, posits that trade volume between two countries is directly proportional to their economic sizes (typically measured by GDP) and inversely proportional to the geographical distance between them [21]. Despite its utility, the Gravity Model relies on simplifying assumptions about variable relationships and requires extensive historical data and manual adjustments, limiting its ability to capture the nonlinear and dynamic nature of international trade [22].

In the quest for more sophisticated predictive techniques, neural networks and deep learning models have emerged as powerful tools. This thesis builds on the foundational work of Marcos Casas Cuadrado, Raúl Gutiérrez Sanchís, and Damiano Zanardini, presented in *Predicting International Trade with Graph Neural Networks* [23]. Their study leveraged the state-of-the-art Spatio-Temporal Graph Convolutional Network (ST-GCN) architecture to predict international trade, marking a significant advancement over traditional models like ARIMA, which are limited to separate time series analyses. A pivotal contribution of their research was the construction of a graph representing country-to-country interactions. In this framework, economic ties between countries are represented as edges in a graph, where nodes signify countries. This graph structure is essential for the application of deep learning architectures like ST-GCNs, which rely on Graph Neural Networks (GNNs) to capture complex interactions and dependencies within data [23].

Temporal series analysis involves examining data points collected or recorded at time intervals to identify patterns, trends, and seasonal effects. Time series data is ubiquitous across various fields, including economics, finance, environmental science, and engineering [24]. Traditional statistical methods, such as autoregressive models,

moving averages, and exponential smoothing, have been fundamental in temporal series analysis [25]. However, the growing complexity and volume of modern datasets necessitate more advanced analytical tools. Machine learning and deep learning techniques have become increasingly vital for temporal series analysis, capable of capturing complex, nonlinear relationships within data. These methods enhance predictive accuracy and reliability, making them indispensable for contemporary time series forecasting [26].

Graph Neural Networks (GNNs) are a class of deep learning algorithms specifically designed to process and analyze graph-structured data. Graphs consist of nodes (vertices) and edges (links) that model pairwise relationships between objects [27]. GNNs iteratively update node representations by aggregating information from their neighbors and the overall graph structure, a process known as message passing or neighborhood aggregation [28]. This capability allows GNNs to learn hierarchical feature representations that encapsulate both local and global information within the graph. The versatility and efficacy of GNNs in modeling relational data have led to their application in diverse domains, including social network analysis, recommendation systems, and bioinformatics [29]. Their ability to capture complex interactions makes them particularly suitable for tasks involving structured and interconnected data, such as international trade prediction [30].

Spatial Temporal Graph Convolutional Networks (ST-GCNs) extend the functionality of conventional GNNs by incorporating temporal dynamics into the graph structure [31]. ST-GCNs operate on graph-structured data where each node and edge can have time-dependent features, enabling the simultaneous modeling of temporal trends and spatial dependencies [32]. This dual capability provides a comprehensive framework for analyzing complex datasets, such as international trade. ST-GCNs leverage graph convolutions to model interactions between nodes, where each node represents a commercial relationship between two countries. This architecture allows ST-GCNs to effectively capture evolving patterns and trends over time, offering a more robust and nuanced approach to forecasting international trade compared to traditional methods [33].

In this thesis, the ST-GCN receives as input a data matrix V with temporal information from the countries for each year, and an adjacency matrix A containing the spatial or invariable-to-time information of each country. Specifically, the data matrix utilizes the SITC Economic Complexity Index (ECI) for years ranging from 1962 to 2021. The adjacency matrix encodes spatial context data based on bilateral distances between the largest cities of each country, weighted by the share of each city's population within the overall country's population. This approach allows the model to capture both the temporal evolution of economic complexity and the spatial dependencies between countries, providing a more accurate prediction framework for international trade.

The Economic Complexity Index (ECI) is a sophisticated measure that captures the diversity and sophistication of a country's economic activities. It is based on the principle that countries with more complex economies produce a wider variety of sophisticated products [34]. The ECI is computed using international trade data, specifically considering the diversity of products a country exports and the ubiquity of those products across different countries [35]. Incorporating the ECI into temporal series analysis provides a nuanced understanding of economic development

and trade patterns. The ECI serves as a temporal feature that reflects changes in a country's economic structure and competitiveness over time. When combined with graph-based methods, the ECI enhances the prediction of trade flows by offering insights into the economic capabilities and dynamics of each country [36].

Geographical distance is a critical factor in international trade, influencing transportation costs, delivery times, and trade relationships [37]. Traditional measures of distance between countries, often represented by the distance between capital cities, can be overly simplistic [38]. A more accurate approach involves calculating bilateral distances between the largest cities in each country, weighted by their population shares. This method accounts for the economic significance of urban agglomerations, providing a more realistic representation of geographical proximity [39].

This thesis aims to advance the field of international trade prediction by exploring the use of Spatial Temporal Graph Convolutional Networks (ST-GCNs). Building on the foundational work of Marcos Casas Cuadrado et al., this research employs the same dataset but introduces new matrix formations to better capture temporal and spatial dimensions [23]. The temporal data matrix leverages the Economic Complexity Index (ECI) to represent the evolving economic landscape of each country, while the spatial context is modeled using an adjacency matrix based on bilateral distances between major cities, weighted by their population shares. By integrating these novel features, the ST-GCN model aims to provide a more accurate and comprehensive prediction of international trade flows.

1.1 State of the Art

1.1.1 Classical Models for Predicting International Trade

International Trade Prediction can be considered as a special case of time-series prediction, making it amenable to be tackled by classical time-series forecasting methods. These classical methods rely on statistical models designed to capture the underlying patterns and dynamics within time series data. They employ historical observations to predict future values, relying on specific assumptions about the characteristics of time series. However, these assumptions may not hold for specific tasks, so that these methods could suffer in such scenarios.

For instance, the widely-used integrated autoregressive moving average model (ARIMA) [2] assumes stationarity, meaning that time series exhibit constant statistical properties like mean and variance over time. ARIMA incorporates autoregressive and moving average components and a differentiating order to handle non-stationarity. Organizations like the National Bank of Spain and the Instituto Nacional de Estadística extensively use ARIMA for predictions.

ARIMA comprises three key elements: autoregressive (AR), integrated (I), and moving average (MA). The AR component regresses the variable of interest on its past values, the I component signifies differentiating, and the MA component relates to error terms. These elements collectively aim to optimize the model's fit to the data.

The Vector Autoregression model (VAR) [3] is an extension of ARIMA designed for multivariate time series. Like ARIMA, VAR features equations for each variable, modeling its temporal evolution; however, it encompasses the past values of the variable and other variables in the model, along with an error term in its equation.

As a conclusion, classical forecasting methods rely on statistical models with specific assumptions about time series characteristics, limiting their ability to accommodate outliers, external factors, and nonlinear relationships. These limitations can impact the accuracy and reliability of their predictions. Consequently, researchers are exploring alternative methods based on Machine Learning and Deep Learning to mitigate some of these shortcomings. The following sections will present some key developments in this growing area of Artificial Intelligence.

1.1.2 Recurrent Neural Models

Among the neural models employed for time series forecasting, recurrent models are conceptually the simplest. They predict a sequence of observations as a sequence of one-step-ahead forecasts. During training, forecasts are generated for the entire sequence, and the error is aggregated over the sequence for model updates. In out-of-sample forecasting, the model recursively updates itself by using its own predictions from previous steps as input.

Salinas et al. [4] introduce DeepAR, a forecasting model based on Recurrent Neural Networks (RNN) employing either LSTM or GRU cells. DeepAR takes as input the prior time points and co-variances, estimating the distribution of the next time point. This is achieved by estimating the parameters of a pre-selected parametric distribution, such as the negative binomial distribution. Both training and prediction follow the general approach of autoregressive models.

Lingxue et al. [5] employ LSTM with Monte Carlo dropout as both the encoder and decoder components. However, unlike models that directly use RNN for generating forecasts, they use the learned embedding at the end of the decoding step as the input to an MLP-type prediction network. This learned embedding is combined with other external features to generate forecasts. Similarly, Laptev et al. [6] use LSTM as a feature extractor. They leverage the extracted features, coupled with external inputs, to generate forecasts using a different LSTM network.

In the realm of sequence-to-sequence models (seq2seq), input sequences are directly mapped to output sequences, which may have varying lengths. This modeling style was initially introduced in machine translation, where the original and translated sequences frequently possess different lengths [7]. Additionally, seq2seq models often need to encode the entire input sequence before generating the output sequence, as they typically require a comprehensive understanding of the input data context to generate the output data. The typical architecture of a seq2seq model falls into the encoder-decoder pattern: the encoder maps the input sequence to a latent state vector (embedding), and the decoder generates the output sequence based on this state vector. Encoders and decoders can, theoretically, be constructed using various neural network architectures.

1.1.3 Attention-Based Models

Until recently, the majority of deep learning models for sequence data were built on Recurrent Neural Networks. However, RNNs have posed practical challenges: they can be hard to train, often requiring extensive parameter tuning to get satisfactory results. Additionally, they struggle with learning long-range dependencies, which is a significant issue for tasks like speech synthesis. Furthermore, their sequential nature

makes them less compatible with highly parallel computing architectures such as GPUs, resulting in slower training.

Recently, new model architectures centered around the attention mechanism have gained prominence, particularly in the context of seq2seq models [8]. These models address the short-term memory limitation of RNNs by enabling the model to discern between essential and non-essential information in a sequence.

One notable development inspired by attention models is the Transformer model [9]. Transformers abandon RNNs entirely, relying solely on the attention mechanism in conjunction with feed-forward neural networks to achieve state-of-the-art results. The core idea of transformers is to extend the encoder-decoder attention mechanism to intra- or self-attention within the encoder and decoder. This allows the model to learn where to focus its attention to obtain effective feature representations. In simplified terms, the encoder of a transformer identifies relevant positions or context within the entire input sequence to compute feature embeddings for each element of the sequence.

Another illustration of the application of transformers is presented by Grigsby [10]. In his work, a novel approach to spatio-temporal forecasting employing long-range transformers (LRTs) is introduced. LRTs represent a variant of the popular Transformer architecture adapted to spatio-temporal data. The authors propose a novel technique called dynamic spatio-temporal attention (DSA), enabling the model to selectively attend to distinct regions of the input space at different time points. This capability allows the model to capture intricate spatio-temporal relationships. Transformers have demonstrated their effectiveness in various sequence forecasting tasks, including those relevant to international trade prediction.

1.1.4 Models Based on Graph Neural Networks

One notable challenge with transformers is their need for large volumes of data in order to learn efficiently. As a matter of fact, data availability is often the primary limitation when it comes to fully exploiting the potential of these models. That is the reason why GNNs are still among the most promising architectures for many relevant tasks.

Besides, the challenge of taking both temporal and spatial information into account has led to many novel neural architectures. In addition to purely temporal models like LSTMs, Yao et al. [11] capture spatial dependencies among nearby regions with convolutional layers in several timesteps, which are later fed to an LSTM network and concatenated with the output of a so-called semantic view. Liu et al. [12] use a context-aware attention mechanism to integrate three different predictions: an instant spatio-temporal module, a short-term module, and a long-term periodic module. Finally, Prado-Rujas et al. [13] design a spatio-temporal model, the Spatio-Temporal Mobility Demand Forecaster (ST-MDF), that is based on the idea of Conv-LSTM [14].

In all these works, the spatial dimension comes as a fixed grid on which some kind of convolution is computed. On the other hand, many tasks require considering space in a more flexible manner; to this end, models that work directly on graph structures have been proposed. The present paper is inspired by one such model: a neural network architecture called Spatio-Temporal Graph Convolutional Network (ST-GCN)

[15], which was introduced to forecast traffic demand. ST-GCN builds upon the premise that traffic patterns inherently possess both spatial and temporal characteristics, emphasizing the importance of capturing spatial relationships between different locations and the temporal evolution of traffic for accurate predictions. ST-GCN adopts a graph-based representation of the road network, where nodes represent locations in the road network, and edges represent spatial connections between locations. The network leverages graph convolutions to learn the spatial relationships between locations, and temporal convolutions to capture the temporal evolution of traffic. This means that the network is able to make predictions by simultaneously exploiting temporal and spatial data.

What distinguishes this approach is the use of a Graph Neural Network, which preserves the regularization properties that enable RNNs to train effectively without the need for vast amounts of data. Furthermore, GNNs possess an inherent structure that permits the efficient modeling of specific relationships, such as spatio-temporal relationships. Such a novel approach has potential when it comes to forecasting tasks, including international trade prediction.

Chapter 2

Development

2.1 STGCN Architecture

In this section, we present the architecture of Spatio-Temporal Graph Convolutional Networks (STGCN), a sophisticated deep learning framework engineered to address the time series prediction problem. Although originally designed for traffic forecasting, the flexibility of STGCN allows its application to a variety of spatio-temporal data domains, including international temporal datasets. Unlike traditional methods that rely on regular convolutional and recurrent units, our approach redefines the problem within the context of graphs, constructing the model exclusively with convolutional structures. To our knowledge, this is the inaugural application of purely convolutional structures to simultaneously extract spatio-temporal features from graph-structured time series, achieving substantial performance gains in prediction accuracy and computational efficiency.

2.1.1 Problem Formulation

Traffic forecasting, a classical example of time-series prediction, involves predicting future traffic measurements such as speed or flow based on historical data from a traffic network. The task can be formally defined as follows:

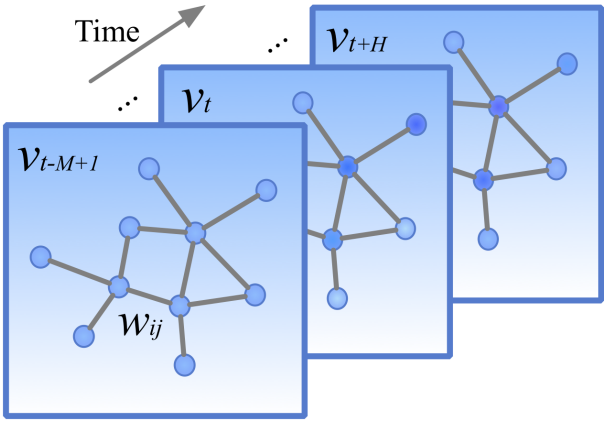

Figure 2.1: Graph-structured traffic data. Each node represents a snapshot of a chosen value at a specific time step, recorded in a graph-structured data matrix.

In our application, the formulation extends beyond traffic data to encompass temporal datasets between countries. The goal is to predict future values of interest based on past observations, structured as graph data where each node corresponds to a specific temporal data point for a country, and edges represent relationships between these points.

2.1.2 Network Structure

The STGCN architecture is depicted in Figure 2.4. The framework comprises multiple Spatio-Temporal Convolutional (ST-Conv) blocks followed by a fully-connected output layer. Each ST-Conv block integrates two temporal gated convolution layers and one spatial graph convolution layer. Residual connections and bottleneck strategies are implemented within each block to enhance feature extraction and propagation efficiency. The input data undergoes uniform processing by the ST-Conv blocks, allowing the model to coherently explore and integrate spatial and temporal dependencies. The output layer amalgamates the comprehensive features to generate the final prediction.

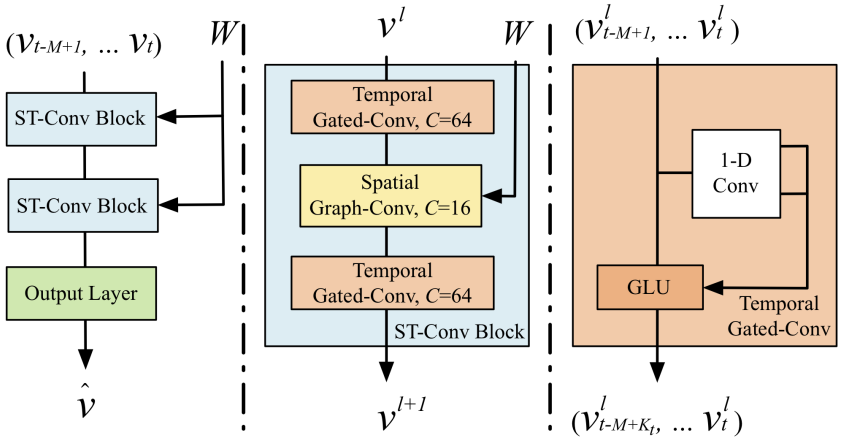

Figure 2.2: Architecture of Spatio-Temporal Graph Convolutional Networks.

2.1.2.1 Graph Convolutions for Spatial Features

Traditional traffic networks can be mathematically represented as graphs, making it intuitive to employ graph convolutional networks (GCNs) to capture spatial dependencies. The graph convolution operation, defined in the spectral domain, allows the model to handle complex graph structures by transforming graph signals through learned filters.

2.1.2.2 Temporal Convolutions for Temporal Features

To address the temporal dimension, STGCN utilizes gated convolutional neural networks (CNNs) on the time axis, replacing the traditional recurrent neural networks (RNNs) approach. This design choice circumvents the iterative training process of RNNs, which is often computationally intensive and prone to error accumulation. By employing temporal convolutions, STGCN achieves faster training times and more robust temporal feature extraction.

2.1.2.3 Spatio-Temporal Convolutional Block

The core of STGCN is the Spatio-Temporal Convolutional Block (ST-Conv block), which fuses spatial and temporal features. Each block contains a sequence of graph convolutions and temporal gated convolutions, structured to maximize the extraction of relevant spatio-temporal patterns. The "sandwich" configuration, with temporal layers surrounding a central spatial layer, facilitates efficient feature processing and dimensionality reduction.

2.2 Data Processing

2.2.1 Original Data

The original dataset for the data matrix was the Harvard International Trade Data, SITC version [18]. This dataset contains trade flows classified via the Standard International Trade Classification (SITC) Revision 2. In comparison to the similar Harvard's dataset Harmonized System (HS) [**?**], SITC data offers a longer time-series but covers fewer goods, as the goods classification names had to remain rather consistent since the 1960s, despite the onset of new goods segments or products such as electronics. The key characteristics of the dataset are:

- Categorizes approximately 700 goods
- Covers years from 1962 to 2020
- Categories break down to 1-, 2-, or 4-digit detail levels

Raw data on trade in goods is provided by the United Nations Statistical Division (COMTRADE). The data is then cleaned by Growth Lab researchers using the Bustos-Yildirim Method, which uses bilateral trade flows to account for inconsistent reporting and provides more reliable accounting. This means that we do not need to perform a data cleaning process such as removing blank or null values.

In addition to trade in goods, the dataset also includes unilateral data on services trade provided by the International Monetary Fund (IMF) and acquired through the World Development Indicators (WDI) of The World Bank.

The dataset fields and their descriptions are detailed in the International Trade Data (SITC, Rev. 2) Data Dictionary:

- **location_id**: CID-assigned numerical country identifier. *Type: Integer. Notes: Can be joined with the location table within the classifications dataset for more detailed information.*
- **partner_id**: CID-assigned numerical country identifier. *Type: Integer. Notes: Can be joined with the location table within the classifications dataset for more detailed information; only applicable to country-partner-product-year tables.*
- **product_id**: CID-assigned numerical product identifier. *Type: Integer. Notes: Can be joined with the sitc_product table within the classifications dataset for more detailed information.*
- **year**: Year of record. *Type: Integer. Notes: SITC records start at 1962.*
- **export_value**: Current USD export value. *Type: Integer.*

- **import_value**: Current USD import value. *Type: Integer.*
- **sitc_eci**: Economic Complexity Index computed using SITC product data. *Type: Float. Notes: Variable applicable to the respective country-year, independent of individual product.*
- **sitc_coi**: Complexity Outlook Index computed using SITC product data. *Type: Float. Notes: Variable applicable to the respective country-year, independent of individual product.*
- **pci**: Product Complexity Index. *Type: Float.*
- **location_code**: 3-character ISO country code. *Type: String. Notes: Follows ISO 3166-1 alpha-3.*
- **partner_code**: 3-character ISO country code. *Type: String. Notes: Follows ISO 3166-1 alpha-3; only applicable to country-partner-product-year tables.*
- **sitc_product_code**: SITC (Rev. 2) product code. *Type: String.*

	location_id	partner_id	product_id	year	export_value	import_value	sitc_eci	sitc_coi	location_code	partner_code	sitc_product_code
0	0	1	102	2018.0	27560.0	0.0	0.800212	0.174964	ABW	AFG	02
1	0	1	102	2019.0	51178.0	0.0	0.935727	-0.282090	ABW	AFG	02
2	7	1	102	2013.0	5530393.0	7302.0	-0.338417	-0.754102	ARE	AFG	02
3	7	1	102	2019.0	3661820.0	0.0	0.321754	0.026674	ARE	AFG	02
4	7	1	102	2021.0	1304571.0	0.0	0.124103	0.087807	ARE	AFG	02

Figure 2.3: SITC Original partners dataset

2.2.2 Matrix Formation

The input for the Spatio-Temporal Graph Convolutional Network (STGCN), as stated, consists of a 'V' data matrix with temporal information and an adjacency matrix 'W' with spatial information, invariant to time. Both matrices must have the same number of columns c. The first matrix contains values for each column c over n number of times representing the time series, and the second matrix is a $c \times c$ square matrix with the spatial relationship between each column c.

2.2.2.1 Adjacency Matrix

For the W adjacency Matrix matrix, the desired relationship is the value between a pair of countries. The chosen value was the distance between countries, as distance is considered a primary indicator and decision value for two countries to engage in trade. The distance value between each pair of countries is given by the $distwces$ calculation provided by the CEPII's distances measures: The GeoDist database [16].

GeoDist is widely cited in gravity literature for providing a comprehensive set of gravity variables developed by Mayer and Zignago (2005), covering all countries. Its main contribution lies in consistently computing both internal and international bilateral distances. Accurate measurement of relative distances is crucial for estimating trade impediments. For instance, the similar GDPs of the UK and Italy have little impact on their internal versus international trade ratios. However, trade is more frequent

within each country due to shorter domestic distances, reducing transport costs compared to international trade. Misjudging distances, such as equating Italy-to-Italy with UK-to-Italy, misattributes internal trade surplus to the border effect. Overestimating the internal/external distance ratio biases this effect. Mayer and Zignago [16] addressed this by calculating distances using city-level data to reflect population distribution, as proposed by Head and Mayer [17].

This dataset contains latitudes, longitudes, and populations data of main agglomerations of all countries available in the World Gazetteer website, which provides current population figures and geographic coordinates for cities, towns, and places of all countries. The general formula developed by Head and Mayer (2002) and used for calculating distances between country i and j is:

$$d_{ij} = \left(\sum_{k \in i} \left(\frac{pop_k}{pop_i} \right) \sum_{\ell \in j} \left(\frac{pop_\ell}{pop_j} \right) d_{k\ell}^{\theta} \right)^{\frac{1}{\theta}},$$

where pop_k designates the population of agglomeration k belonging to country i. The parameter θ measures the sensitivity of trade flows to bilateral distance $d_{k\ell}$. For the distw calculation, θ is set equal to 1. The distwces calculation sets θ equal to -1, which corresponds to the usual coefficient estimated from gravity models of bilateral trade flows.

	year	COM	GUY	LBR	SEN	BRN	JPN	USA	RUS
0	1962	-0.344110	0.068288	-0.442236	-0.203830	0.193716	2.091124	2.072395	0.535564
1	1963	-0.138500	-0.112739	-0.764035	-0.198813	-1.363110	1.917151	2.034126	0.316475
2	1964	-1.421275	-0.130738	-0.604106	-0.484891	-0.469227	1.895972	2.039541	0.726326
3	1965	-1.312219	-0.209335	-0.565352	-0.422038	-1.309534	1.944695	2.066791	0.723283
4	1966	-1.334038	-0.376698	-0.523193	-0.147933	-0.461554	1.881980	2.018157	0.638421

Figure 2.4: SITC Original partners dataset

2.2.2.2 Data Matrix

Because the data matrix must have c columns, identical to the W matrix, and the same actual columns, we need values for each country, not for each pair of countries. Thus, we cannot directly place a value of exports or imports between two countries because there is only room for a single country. The W matrix helps establish the relationship between each pair of countries.

For this value, for each year, the SITC Economic Complexity Index (ECI) is used. SITC ECI is computed using SITC product data, applicable to the respective country-year, independent of individual products.

The SITC ECI value is rank of countries based on how diversified and complex their export basket is. Countries with a great diversity of productive know-how, particu-

larly complex specialized know-how, can produce a wide range of sophisticated products.

The complexity of a country's exports is highly predictive of current income levels. Where complexity exceeds expectations for a country's income level, the country is predicted to experience more rapid growth in the future. ECI therefore provides a useful measure of economic development.

Technical breakout: Economic complexity is calculated from equations for diversity and ubiquity to express the recursion:

$$k_{c,n} = \frac{1}{k_{c,0}} \sum_p \frac{M_{cp}}{k_{p,0}} \sum_{c'} \frac{M_{c'p}}{k_{c',n-2}} = \sum_{c'} k_{c',n-2} \sum_p \frac{k_{c,0} k_{p,0} M_{c'p} M_{cp}}{k_{c,0} k_{p,0}} = \sum_{c'} k_{c',n-2} \tilde{M}^C_{c,c'},$$

where we define

$$\tilde{M}^C_{c,c'} \equiv \sum_p \frac{M_{cp} M_{c'p}}{k_{c,0} k_{p,0}}.$$

Hence, in vector notation, if $\vec{k}_n$ is the vector whose cth element is $k_{c,n}$ then:

$$\vec{k}_n = \tilde{M}^C \times \vec{k}_{n-2},$$

where $\tilde{M}^C$ is the matrix whose (c, c')th element is $\tilde{M}^C_{c,c'}$.

If we take n to infinity, this equation leads to the distribution which remains fixed up to a scalar factor:

$$\tilde{M}^C \times \vec{k} = \lambda \vec{k}.$$

Therefore, $\vec{k}$ is an eigenvector of $\tilde{M}^C$. We define the Economic Complexity Index as the eigenvector corresponding to the second-largest eigenvalue of the $\tilde{M}^C$ matrix.

	AFG	AGO	ALB	ARG	AUS	AUT	BDI	BEL	BEN
0	90.264152	7563.813745	4243.150201	15333.278441	11020.750969	4557.808571	5842.157165	5326.836645	7323.327350
1	7563.813501	67.455484	5699.611253	7961.882044	12879.337364	6423.483550	1888.743190	6809.597763	2307.673135
2	4243.150158	5699.611363	51.529788	11655.905498	15152.402953	840.694865	5048.706435	1606.491842	4097.628784
3	15333.278882	7961.882510	11655.905388	96.149337	12018.272985	11742.894606	9824.102613	11297.558182	8038.747870
4	11020.750741	12879.337827	15152.402925	12018.272781	153.628704	15559.910589	12045.124538	16277.454792	15161.884522

Figure 2.5: Data matrix with ECI values

The data matrix was split as between training, validation and test set and we used the *StandardScaler* to standardize the dataset. The *StandardScaler* transforms the data by removing the mean and scaling to unit variance, which is particularly beneficial for algorithms that rely on gradient descent. The scaler was fitted on the training data and then applied to the validation and test sets to ensure that all subsets are scaled consistently. This approach prevents data leakage and ensures that the scaling parameters are derived solely from the training data.

Chapter 3

Results and Conclusions

This section introduces the methodology employed for training the models on the specific task of international trade forecasting. All experiments were conducted and tested on a desktop computer with the following specifications: Processor - 13th Gen Intel(R) Core(TM) i5-13600KF @ 3.50 GHz and GPU - NVIDIA Gigabyte GeForce RTX 3090 TURBO 24GB GDDR6X.

The optimization process involved different experiments to determine the optimal hyperparameters. The code development is available in the repository [19].

3.1 Hyperparameter Configuration

To ensure the robustness and generalizability of our model, the dataset was divided into three subsets: training, validation, and testing. This division allows us to train the model, validate its performance during training, and finally test its accuracy on unseen data.

The following hyperparameters were considered for our experiments:

- **Epochs**: We started with 150 epochs. An epoch represents one complete pass through the entire training dataset. Starting with a higher number of epochs allows the model sufficient opportunity to learn the data patterns, especially given the complexity of spatio-temporal data.
- **Batch Size**: We fixed the batch size at 24. The batch size refers to the number of training examples utilized in one iteration. Given the nature of our dataset, with 60 items corresponding to 60 years, a batch size of 24 ensures efficient computation while maintaining a balance between the learning stability and the model's generalization capability.
- **Learning Rate**: The initial learning rate was set at 0.01. The learning rate determines the step size at each iteration while moving towards a minimum of the loss function. A starting learning rate of 0.01 is chosen to facilitate rapid learning in the initial stages of training.
- **Optimizer**: We experimented with two optimizers: AdamW and RMSprop.
  - **AdamW**: An adaptive learning rate optimization algorithm designed to handle sparse gradients on noisy problems. AdamW incorporates weight decay

directly into the update rule, which improves the model's generalization capabilities by penalizing large weights.

  - **RMSprop**: An optimizer that adjusts the learning rate for each parameter, which is particularly effective in dealing with non-stationary objectives and alleviating the issues of vanishing and exploding gradients.

- **Scheduler**: Learning rate scheduling was used to adjust the learning rate during training dynamically. Two schedulers were considered:

  - **StepLR**: This scheduler decreases the learning rate by a factor of 0.7 every 15 epochs. The StepLR scheduler is effective in providing a steady decay of the learning rate, helping the model to converge smoothly.

  - **CosineAnnealingLR**: This scheduler reduces the learning rate following the cosine function. It allows the learning rate to decrease gradually and then increase slightly, which can help in escaping local minima and achieving better overall performance.

3.2 Results

This section presents the results of training and evaluating the Spatio-Temporal Graph Convolutional Network (STGCN) model. We provide detailed performance metrics and compare the effectiveness of different training configurations, including variations in optimizers and learning rate schedulers. The evaluation metrics include Test Loss, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). The experiments conducted with the STGCN model demonstrated its effectiveness in capturing the spatio-temporal dependencies in the dataset. The performance metrics evaluated on the test set included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the R-squared score (R^2). These metrics provided a comprehensive understanding of the model's prediction accuracy and its capability to generalize to unseen data.

3.2.1 Regular distance matrix

First we train with the data matrix without any changes, and we show comparisons and evaluate with different hiperparameters, but having in common: 150 epochs, a batch size of 24, an the initial learning rate of 0.01. In 3.1 the evaluation metrics are compared with the different configurations.

Model	**Test Loss**	**MAE**	**MAPE**	**RMSE**
AdamW-CosineAnnealing	4.7072	0.4170	20.78%	0.5426
RMSprop-CosineAnnealing	4.7045	0.4187	19.02%	0.5391
AdamW-StepLR	4.7078	0.4172	20.67%	0.5425
RMSprop-StepLR	4.7085	0.4196	19.19%	0.5425

Table 3.1: Performance Metrics for Regular Distance Configurations

The model with the best performance was the model with RMSDrop as the optimizer and Cosine Annealing as the learning rate scheduler, with a MAPE of 20.67

In Figure 3.5, we present a comparative analysis of the validation loss for each epoch. To enhance the clarity of the progress of the validation loss, we have applied filtering to the loss values. Specifically, we have excluded the minimum loss value to disregard anomalously high values at the start of the initial epochs, which do not contribute meaningful insights for the analysis. Additionally, we have filtered the minimum and maximum epochs to more accurately reflect the overall trends and provide a clearer representation of the loss progression.

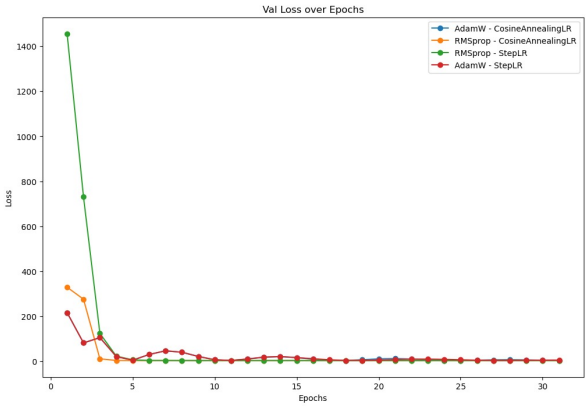

Figure 3.1: regular-1

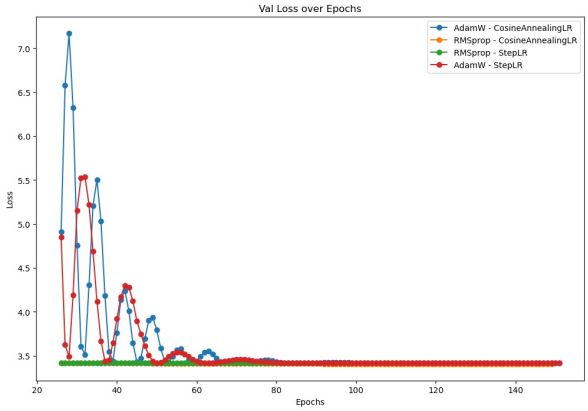

Figure 3.2: regular-2

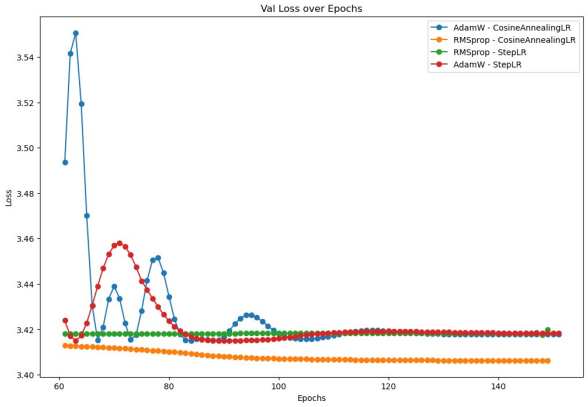

Figure 3.3: regular-3

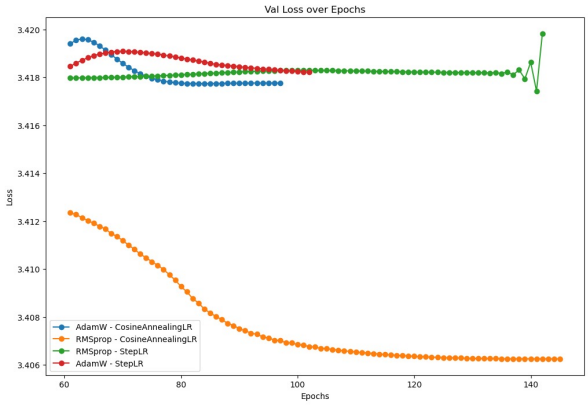

Figure 3.4: regular-4

Figure 3.5: Comparison with regular W

The initial epochs show a rapid decrease in loss values, particularly for the AdamW optimizer with CosineAnnealingLR, which achieves a significant reduction in the loss within the first few epochs. This rapid convergence indicates that AdamW with CosineAnnealingLR is highly effective in optimizing the model early in the training process. In contrast, other configurations demonstrate a more gradual decline, suggesting varying degrees of effectiveness in minimizing the validation loss.

This figure (3.2) depicts the validation loss over 150 epochs for the same set of configurations. The early epochs reveal significant fluctuations, particularly with the RMSprop optimizer combined with CosineAnnealingLR, which exhibits high volatility in the loss values. As training progresses, the loss stabilizes, and we notice that

AdamW with StepLR achieves a more consistent and lower loss compared to the other configurations. The presence of initial fluctuations followed by stabilization indicates the models' ability to adapt and converge over a longer training period.

Figure 3.3 (3.3) focuses on the validation loss over approximately 150 epochs, where: - *AdamW with CosineAnnealingLR* and *RMSprop with StepLR* show initial spikes in loss values, indicating potential instability or sensitivity to certain data points. - *AdamW with StepLR* shows a more consistent decline in loss values, reflecting its robustness and reliability in achieving lower loss over the training period.

This figure highlights the differing convergence behaviors among the configurations, with some showing more significant initial variations while others maintain steadier improvement.

In the final figure (3.4), the validation loss over approximately 150 epochs is shown for the same configurations. This figure emphasizes the long-term performance of the models. We observe that: - *AdamW with CosineAnnealingLR* maintains a lower loss throughout the epochs. - The loss values for *RMSprop with CosineAnnealingLR* and *RMSprop with StepLR* remain higher, indicating less effective convergence.

Overall, this figure underscores the effectiveness of RMSDrop combined with CosineAnnealingLR in maintaining lower loss values consistently over an extended training period for this model. For this reason we only use RMSDrop as the optimizer from now on.

3.2.2 Exponentially inverted distance matrix

Additionally, we compare the base model by altering the adjacency distance matrix using the same formula applied in the original paper of the model, for constructiong the adyancency matrix, and compare results to our base model, and evaluating it.

In PeMSD7, the adjacency matrix of the road graph is computed based on the distances among stations in the traffic network. The weighted adjacency matrix W can be formed as:

$$w_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & \text{if } i \neq j \text{ and } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

where w_{ij} is the weight of the edge, which is determined by d_{ij} (the distance between country i and j). σ^2 and ϵ are thresholds to control the distribution and sparsity of matrix W, assigned to 10 and 0.5, respectively.

For this model, as stated before, we only use RMSDrop. The evaluation values are collected in table 3.6, where it can bee seen that once again, the best value is 'RMSprop-CosineAnnealing', with a MAPE of 17.99%, with a 1.03% decrease compared to the previous model with the regular adyancency matrix.

Model	**Test Loss**	**MAE**	**MAPE**	**RMSE**
RMSprop-CosineAnnealing	4.6457	0.4143	17.99%	0.5373
RMSprop-StepLR	4.6941	0.4175	19.56%	0.5409

Table 3.2: Performance Metrics for Regular Distance Configurations

In Figures 3.6 and 3.7, a comparative analysis of the validation loss for each epoch using the exponentially inverted distance matrix is illustrated. Consistent with previous findings, the RMSprop optimizer paired with the CosineAnnealingLR scheduler exhibits a swift reduction in loss values during the initial epochs. This configuration not only achieves lower loss values but also maintains stability over time, surpassing the performance of the RMSprop with StepLR setup. These results imply that the altered adjacency matrix significantly enhances the model's capability to capture spatio-temporal dependencies, leading to better performance metrics.

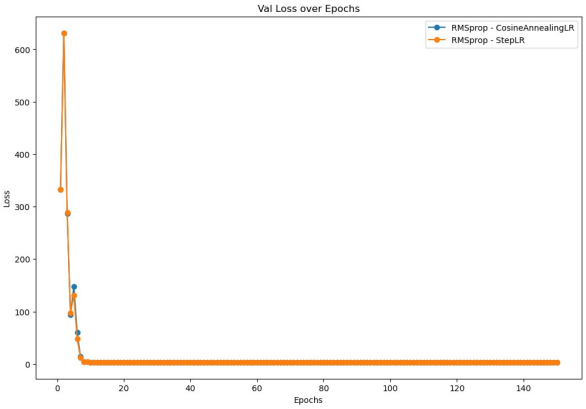

Figure 3.6: exp-1

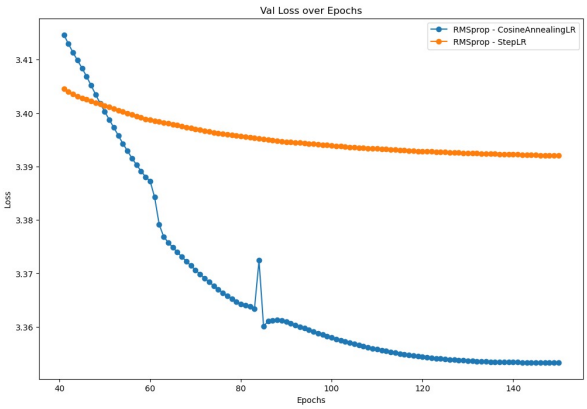

Figure 3.7: exp-2

Figure 3.8: Comparison with exponencially inverted W

3.2.3 Exponentially inverted distance matrix - Longer epochs

Next, we compare the best of both model groups, regular matrix, and exponencially inverted matrix in 3.9 to analyze how both models directly compare to each other their progress of validation loss. The model trained with the exponencially inverted W show a better learning curve. For this, we trained this same models more epochs, with the intention that the validation loss gets smaller.

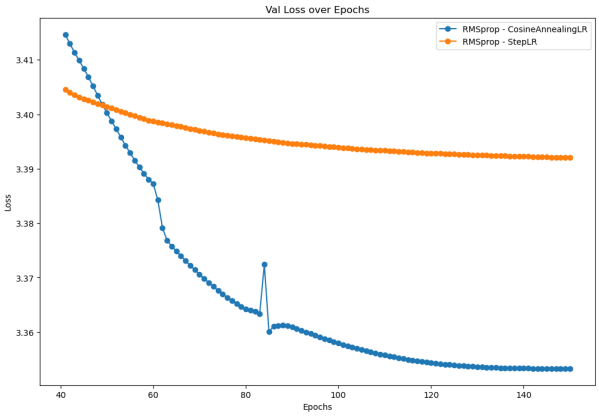

Figure 3.9: Regular W - Exp Inverted W

For this model, we then used the exponencially inverted W , the RMSDrop optimizer and the CosineAnnealingLR scheduler, but for 300 epochs, instead of 150 epochs.

Model	Test Loss	MAE	MAPE	RMSE
RMSprop-CosineAnnealing	4.5890	0.4078	15.69%	0.5333

Table 3.3: Performance Metrics for Regular Distance Configurations

The evaluation results are presented in figure 3.3. This models gets even better results than the previous ones, reaching a 15.69% MAPE, a 2.3% decrease from the previous best model.

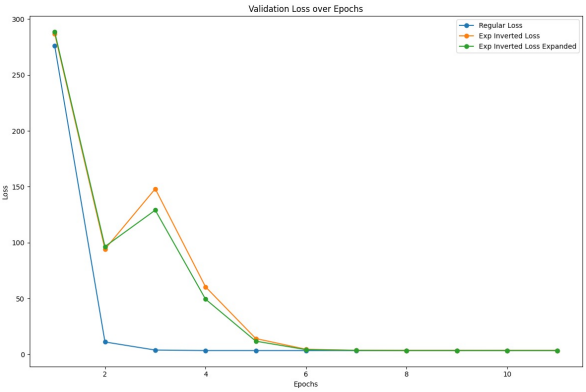

Figure 3.10: inv-v2-1

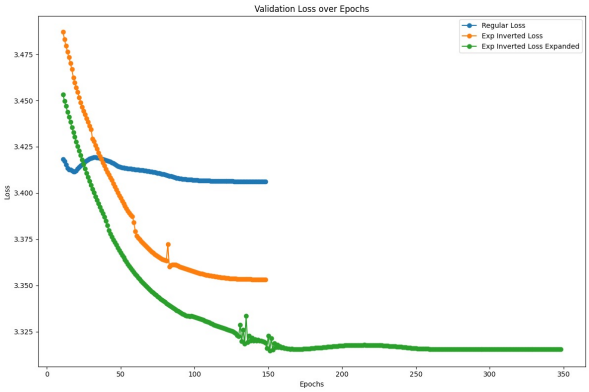

Figure 3.11: inv-v2-2

Figure 3.12: Comparison with exponencially inverted W

The comparison of models using regular and exponentially inverted distance matrices reveals significant differences in their performance, as illustrated in Figures 3.10 and 3.11. The model employing the exponentially inverted W exhibits a more rapid and stable reduction in validation loss across the epochs compared to the model using the regular distance matrix. Specifically, the exponentially inverted matrix configuration demonstrates superior learning efficiency and effectiveness, as evidenced by its lower validation loss values over time. This suggests that the exponentially inverted W matrix enhances the model's capability to capture and leverage spatio-temporal dependencies more effectively than the regular distance matrix. Consequently, the model with the exponentially inverted W not only achieves lower overall error rates but also shows improved generalization performance, indicating its robustness in handling complex spatio-temporal data relationships. The quantitative metrics further support these observations, with the exponentially inverted W configuration achieving a lower Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE), thus underscoring its superiority in predictive accuracy and reliability.

3.3 Interpretation of Results

This sections discusses and explores the predictions obtained by our STGCN best model, with the configuration of RMSDrop, CosineAnnealingLR, and 300 epochs of training. In figures 3.13 and 3.14 we compare the test real values(blue) with the predicted values by the model(orange). Both show similar real values and predicted values. To explore further, we are going to take deeper look into the 2021 year com-

parison.

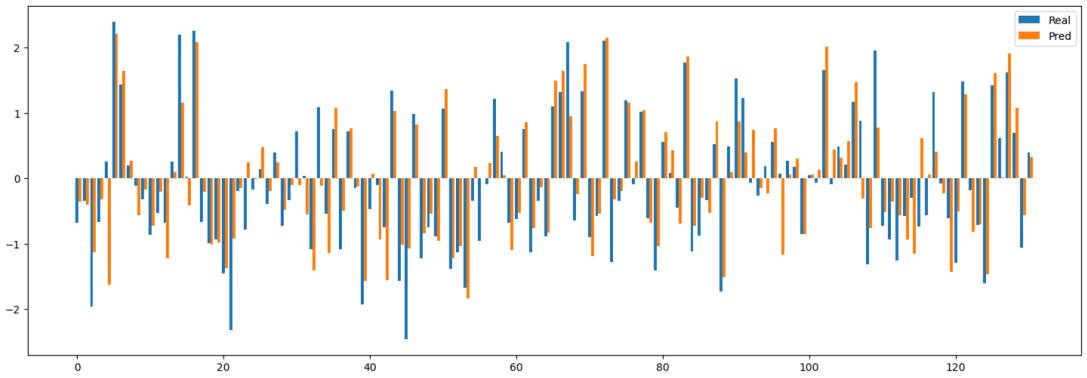

Figure 3.13: Prediction for 2020 for each country

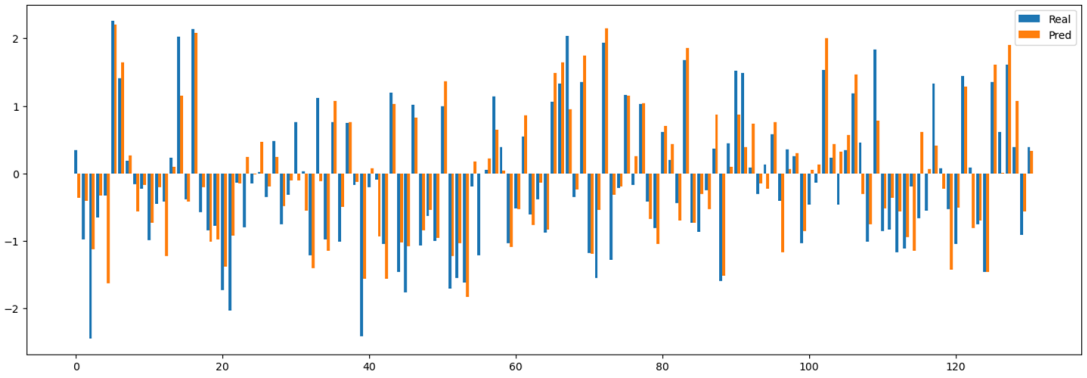

Figure 3.14: Prediction for 2021 for each country

To further investigate the model's predictive accuracy, we focus on the year 2021. Specifically, we analyze the discrepancies between the predicted values and the actual indices for each country. Our findings reveal that Albania exhibited the largest discrepancy, with a difference of 1.3962 in the STIC ECI index. In contrast, the Philippines had the smallest discrepancy, with a difference of only 0.0067, demonstrating the model's high precision in this case.

3.3.1 Detailed Data Analysis

To deepen our understanding of the model's performance, we will compare the data of Albania(ABN) and the Philippines(PHL) , focusing on their relationship with other countries based on weighted distances. This comparison will help identify potential factors contributing to the observed discrepancies and provide insights into the model's strengths and limitations.

First, we compare the variance and the SITC ECI value of both countries in Table 3.6. Variance, in statistical terms, is a measure of the dispersion or spread of a set of values. It quantifies the extent to which each number in the set differs from the mean and, hence, from every other number in the set. Mathematically, it is expressed as the average of the squared differences from the mean. In predictive modeling, a lower variance is typically desirable as it indicates that the predictions are more consistent

and less spread out, which generally corresponds to a lower error rate and higher reliability of the model's performance.

However, as shown in Table 3.6, contrary to the conventional understanding, in our case, a lower difference between the predicted and actual SITC ECI values does not correlate with a lower variance. Instead, the data reveals an inverse relationship: the higher the variance, the lower the loss in the predicted SITC ECI values. This unexpected finding suggests that in certain scenarios, the model's predictive performance might benefit from a higher variance, which could indicate a more complex and nuanced understanding of the underlying data patterns by the model.

Country	Difference	Variance
ALB	1.3962	0.0611
PHL	0.0067	0.1228

Table 3.4: Performance Metrics for Regular Distance Configurations

To further explain and anayze this , in figures 3.15 and 3.16, the SITC ECI values of both countries are represented for each year since 1962 to 2021. In figure 3.15, the SITC ECI values are showed for Albania.

Figure 3.15 shows significant fluctuations in Albania's SITC ECI values, with marked peaks and troughs reflecting economic instability. This variability corresponds to a higher variance of 0.0611 and a greater discrepancy between predicted and actual values.

In contrast, Figure 3.16 demonstrates a consistent upward trend in the Philippines' SITC ECI values, with fewer fluctuations. The stability is indicated by a lower variance of 0.1228 and a minimal discrepancy of 0.0067 between predicted and actual values.

These figures highlight the relationship between economic stability and predictive accuracy. Albania's high variance presents challenges for the model, whereas the Philippines' stable economic progression allows for more accurate predictions.

To consider the effect of the adjacency distance matrix, an analysis is conducted on the five closest and furthest countries to Albania and the Philippines, focusing on their STIC ECI values and variance over the last 10 years. This period is chosen as it provides the input data for predicting the years 2020 and 2021, allowing for a detailed evaluation of the model's performance.

The analysis of Albania's partner countries shows that the closest nations (GRC, BGR, ITA, HUN, MLT) have a low average variance of 0.0191, suggesting consistent and stable economic interactions. Conversely, the most distant countries (NZL, WSM, FJI, NCL, VUT) demonstrate a higher average variance of 0.0949, indicating greater economic diversity and unpredictability, which complicates the model's accuracy in predictions.

The analysis of variance values for the Philippines (PHL) reveals that the closest countries (TWN, HKG, BRN, VNM, KHM) have a higher average variance (0.0715) compared to the furthest countries (BOL, PRY, BRA, PER, GUY) with an average variance of 0.0375. This indicates significant economic variability among geographically close countries, while distant countries exhibit more stable economic conditions.

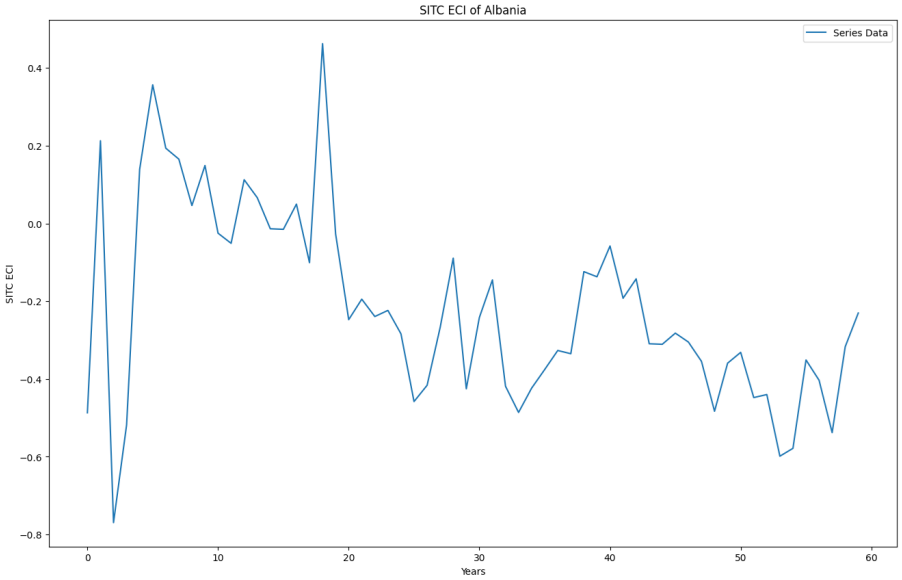

Figure 3.15: SITC ECI value for ABL

The comparison of variance values for partner countries relative to the Philippines (PHL) and Albania (ALB) reveals significant insights into economic stability and predictive accuracy. For PHL, the closest countries exhibit a higher average variance (0.0715) compared to the furthest countries (0.0375), suggesting that geographical proximity does not necessarily correlate with economic stability. Conversely, for ALB, the closest countries have a lower average variance (0.0191) than the furthest countries (0.0949), indicating more consistent economic interactions with nearby nations.

These findings suggest that the distance matrix may not significantly impact the predictive performance of this specific model and data set. Typically, in data science, time series with lower variance are expected to yield more accurate predictions. However, this analysis of country values indicates that lower variance does not necessarily correlate with improved predictive accuracy in this context.

3.4 Conclusion

This thesis extends on predicting international trade using Graph Neural Networks (GNNs), particularly focusing on Spatio-Temporal Graph Convolutional Networks (ST-GCNs). By incorporating novel matrix formations to better capture both temporal and spatial dimensions, this study aimed to enhance the predictive accuracy of international trade flows.

We employed the COEI data matrix, which includes the Economic Complexity Index (ECI) computed using SITC product data to represent temporal dynamics, and a refined adjacency matrix for spatial context. The adjacency matrix calculated bilateral distances between the largest cities of respective countries, weighted by the population share of each city, to emphasize the economic significance of urban agglomerations.

Our experimental results demonstrate that while the ST-GCN model effectively captures spatio-temporal dependencies in international trade data, the impact of the dis-

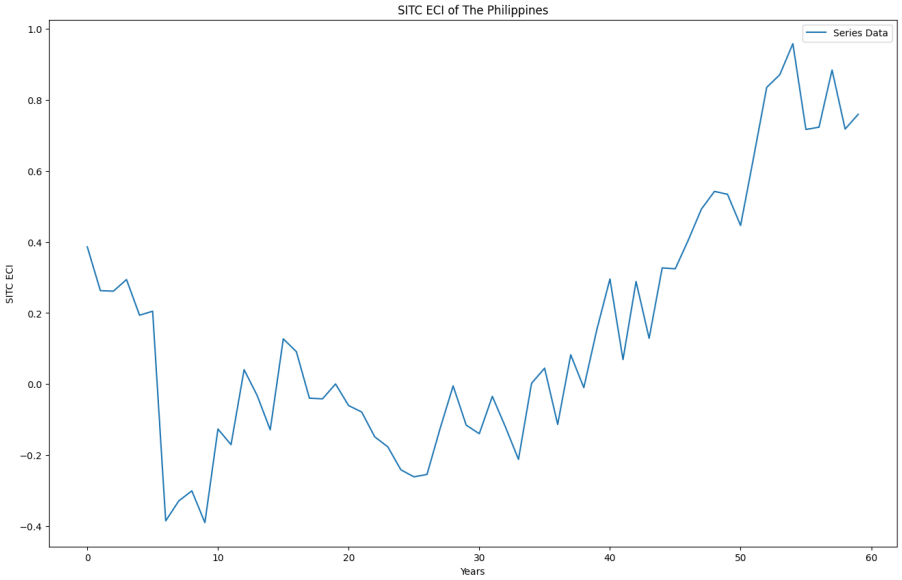

Figure 3.16: SITC ECI value for PLH

tance matrix on predictive performance is not as significant as initially hypothesized. The comparison of regular and exponentially inverted distance matrices revealed that the latter configuration enhances the model's capability to capture spatio-temporal dependencies more effectively, achieving lower error rates and improved generalization performance.

These findings in the detailed analysis suggest that lower variance in time series data does not always correlate with improved predictive accuracy in the context of international trade forecasting. This research contributes to the ongoing evolution of GNN applications in economic forecasting, offering new insights into the integration of spatial and temporal data for enhanced predictive performance. The methodologies and results presented in this thesis provide a robust framework for understanding and forecasting complex trade patterns in an increasingly interconnected world.

3.5 Future Work

The findings and methodologies presented in this thesis open several avenues for future research to further enhance the predictive performance and robustness of Spatio-Temporal Graph Convolutional Networks (ST-GCNs) in the context of international trade forecasting.

One primary direction for future work involves utilizing a larger volume of data. The initial ST-GCN paper demonstrated the model's capabilities using traffic prediction data with thousands of time stamps, whereas our dataset consists of only 60 points, one per year. Increasing the data volume could significantly improve the model's training and predictive accuracy. Techniques such as data augmentation through Linear and Polinomic Iinterpolation, Spline, and Simple Exponential Smoothing can be employed to generate additional data points, thereby enriching the dataset and enabling the model to learn more effectively from a denser time series.

Another promising approach is to explore alternative data matrices that directly rep-

Country	Distance	Variance	Avg Variance
GRC	Closest	0.0425	
BGR	Closest	0.0348	
ITA	Closest	0.0026	
HUN	Closest	0.0094	
MLT	Closest	0.0060	
			0.0191
NZL	Furthest	0.0042	
WSM	Furthest	0.1646	
FJI	Furthest	0.1128	
NCL	Furthest	0.1446	
VUT	Furthest	0.0482	
			0.0949

Table 3.5: Data Analysis of countries related to ALB

Country	Partner	Distance	Variance	Avg Variance
PHL	TWN	Closest	0.0144	
PHL	HKG	Closest	0.0033	
PHL	BRN	Closest	0.2451	
PHL	VNM	Closest	0.0740	
PHL	KHM	Closest	0.0207	
				0.0715
PHL	BOL	Furthest	0.0961	
PHL	PRY	Furthest	0.0176	
PHL	BRA	Furthest	0.0344	
PHL	PER	Furthest	0.0034	
PHL	GUY	Furthest	0.0360	
				0.0375

Table 3.6: Performance Metrics for Regular Distance Configurations

resent the relationships between countries, rather than relying solely on the Economic Complexity Index (ECI). Data that explicitly captures bilateral trade agreements, tariffs, import-export dependencies, and other economic interactions could provide a more detailed and relevant basis for prediction. By integrating such relationship-oriented data, the model may better capture the complexities and nuances of international trade dynamics.

Moreover, future research could investigate the use of different spatial adjacency matrices. Instead of using geographical distances, other economic indicators such as GDP correlations, trade volume similarities, or even qualitative data derived from large language models (LLMs) could be employed. LLMs can provide context and general opinions about a country's economic position, adding a rich layer of semantic information that traditional quantitative measures might miss. This approach could enhance the model's ability to capture complex economic relationships and improve its predictive performance.

In summary, expanding the volume of data, employing advanced data augmentation techniques, exploring alternative data matrices, and utilizing diverse spatial adjacency matrices are key areas for future research. These strategies hold the potential to significantly enhance the robustness and accuracy of ST-GCN models in predicting international trade, contributing to a deeper understanding of global economic patterns and informing more effective policy and decision-making.

Bibliography

[1] T. Mayer and S. Zignago, "Notes on CEPII's distances measures: The GeoDist database," CEPII working paper, 2011.

[2] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[3] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[4] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.

[5] R. Wen, K. Torkkola, B. Narayanaswamy, and D. Madeka, "A multi-horizon quantile recurrent forecaster," *arXiv preprint arXiv:1711.11053*, 2017.

[6] N. Laptev, J. Yosinski, L. Li, and S. Smyl, "Time-series extreme event forecasting with neural networks at Uber," in *International Conference on Machine Learning*, 2017.

[7] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, pp. 3104–3112, 2014.

[8] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

[10] J. Grigsby, "Long-range transformers for dynamic spatio-temporal forecasting," *arXiv preprint arXiv:1911.10529*, 2019.

[11] H. Yao, X. Tang, H. Wei, Y. Zheng, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," *arXiv preprint arXiv:1802.08714*, 2018.

[12] W. Liu, Y. Sun, S. Wang, J. Zhou, W. Lin, and X. Wang, "Contextual attention for human pose estimation," *arXiv preprint arXiv:1801.10863*, 2018.

[13] A. Prado, A. Conde, J. Vega, E. Villar, and J. L. De La Rosa, "Spatio-temporal neural networks for space-time series forecasting and imputation," in *IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pp. 32–36, 2019.

[14] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, pp. 802–810, 2015.

[15] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 3634–3640, 2018.

[16] T. Mayer and S. Zignago, "Notes on CEPII's distances measures: The GeoDist database," CEPII working paper, 2011.

[17] K. Head and T. Mayer (2002), "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade", CEPII Working Paper 2002-01.

[18] The Growth Lab at Harvard University, "International Trade Data (SITC, Rev. 2)," Harvard Dataverse, UNF:6:ZiFCiXb32UHj+6Nk044T5A==, 2019, version V7, doi:10.7910/DVN/H8SFD2. Available: `https://doi.org/10.7910/DVN/H8SFD2`.

[19] Development code repository. Available: `https://github.com/your-repository`

[20] R. C. Feenstra, "Advanced International Trade: Theory and Evidence," Princeton University Press, 2014.

[21] J. Tinbergen, "Shaping the World Economy; Suggestions for an International Economic Policy," Twentieth Century Fund, 1962.

[22] J. E. Anderson, "A Theoretical Foundation for the Gravity Equation," The American Economic Review, vol. 69, no. 1, pp. 106-116, 1979.

[23] M. C. Cuadrado, R. G. Sanchís, and D. Zanardini, "Predicting International Trade with Graph Neural Networks," 2024.

[24] C. Chatfield, "The Analysis of Time Series: An Introduction," Chapman and Hall/CRC, 2004.

[25] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, "Time Series Analysis: Forecasting and Control," John Wiley Sons, 2015.

[26] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016.

[27] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The Graph Neural Network Model," IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 61-80, 2009.

[28] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural Message Passing for Quantum Chemistry," in Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 1263-1272.

[29] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation Learning on Graphs: Methods and Applications," IEEE Data Engineering Bulletin, vol. 40, no. 3, pp. 52-74, 2017.

[30] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, et al., "Re-

lational inductive biases, deep learning, and graph networks," arXiv preprint arXiv:1806.01261, 2018.

[31] B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," in Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018, pp. 3634-3640.

[32] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting," in International Conference on Learning Representations, 2018.

[33] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," arXiv preprint arXiv:1609.02907, 2016.

[34] C. A. Hidalgo and R. Hausmann, "The building blocks of economic complexity," Proceedings of the National Academy of Sciences, vol. 106, no. 26, pp. 10570-10575, 2009.

[35] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero, "A New Metrics for Countries' Fitness and Products' Complexity," Scientific Reports, vol. 2, 2012.

[36] R. Hausmann, C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, A. Simoes, and M. Yildirim, "The Atlas of Economic Complexity: Mapping Paths to Prosperity," MIT Press, 2011.

[37] J. A. Frankel and A. K. Rose, "Estimating the Effect of Currency Unions on Trade and Output," NBER Working Paper No. 7857, 1999.

[38] A.-C. Disdier and K. Head, "The Puzzling Persistence of the Distance Effect on Bilateral Trade," The Review of Economics and Statistics, vol. 90, no. 1, pp. 37-48, 2008.

[39] T. Mayer and S. Zignago, "Notes on CEPII's distances measures: The GeoDist database," CEPII working paper, 2011.