

UNIVERSITY OF OXFORD

**Siamese-GNNs for Change-Point Detection and Analysis of
Economic Shocks within the International Trade Network**

Thesis submitted in partial fulfilment of the requirement for the degree of MSc in Social Data Science at the Oxford Internet Institute at the University of Oxford

Trinity Term 2024

12,291 words

Contents

1	Introduction	3
2	Literature and Hypotheses Development	6
2.1	Manifestation of Economic Shocks in Trade Patterns	6
2.2	Change-Point Detection of Financial Crises	7
2.3	Graph Machine Learning and Trade Networks	9
2.4	Resilience to Economic Shocks	10
3	Data	12
4	Methodology	13
4.1	Trade Network Construction	13
4.2	Economic Shock Operationalization	14
4.3	Change-Point Detection	16
4.4	Application of s-GNN Change-Point Detection	24
5	Results	26
5.1	Uncovering of Ground-Truth Economic Shocks	26
5.2	Performance of Traditional Distance Metrics	28
5.3	Performance of s-GNN Models	31
5.4	Synthetic Experiments	39
5.5	Analysis of Detected Change-Points	40
6	Discussion	47
6.1	Success of Graph Machine Learning	47
6.2	s-GNN-Based Change-Points for Economic Analysis	48
6.3	Contributions	49
6.4	Limitations	50
7	Conclusion	51
A	Appendix	58
A.1	Network-Driven Identified Years	58
A.2	Feature Subsets	58
A.3	Hyperparameter-Tuning Results	61

Abstract

This thesis introduces a novel approach to analyzing economic shocks within the international trade network (ITN) using graph machine learning. While traditional economic research often examines trade responses to shocks within specific crises, regions, or market sectors, the method proposed here aims to overcome these limitations of conventional econometric models by enabling a broader scope of research. By implementing graph machine learning for change-point detection, I demonstrate significant performance improvements over traditional network-based distance metrics in identifying all economic shocks present in the ITN. When further applied to specific subnetworks, the model successfully detects both well-known economic disruptions and potentially overlooked events, expanding the range of change-point detection beyond financial crises. The investigation of change-points by region additionally reveals evidence supporting an association between globalization and increased risk of economic shocks. However, when exploring change-points by product, I find no clear distinction in shock risk between manufactured and raw-material products. By bridging advanced machine learning techniques with economic analysis, this work can reveal where economic shocks actually manifest themselves, offering new avenues for understanding the intricate dynamics of global trade and informing more targeted and effective policy responses.

Key words: change-point detection; graph machine learning; International Trade Network; economic shocks; network analysis

1 Introduction

Understanding the differential impact of economic shocks on varied countries and market sectors is crucial for analyzing the onset of crises and formulating appropriate fiscal and monetary policies. Traditional statistical and econometric models, however, have been limited in the extent of such analysis due to their constraints in handling complex, interconnected data. Because of this, much existing research consists of case studies focused on specific regions, products, or types of economic shocks, such as disruptions in food production (Malesios et al., 2020) or the 2008 Global Financial Crisis (Allen et al., 2018). However, to gain a thorough understanding of the mechanisms of economic shocks, it is essential to examine their effects across all such factors. This thesis aims to overcome these limitations by employing advanced machine learning, thereby providing a more expansive perspective on the broader landscape of economic shocks. By offering a method for assessing shocks' diverse impacts across multiple sectors and geographic regions, this research emphasizes the importance of a comprehensive, global view while addressing the challenges posed by the heterogeneous nature of these shocks.

To enable this comprehensive approach, the association between the international trade network (ITN) and economic shocks can be leveraged. The nature of this relationship can be either direct or indirect, depending on the context of a given economic shock. For instance, when an economic shock occurs due to a specific trade agreement, it directly impacts future trade partners. Conversely, shocks stemming from geopolitical events may have more indirect effects on trade, leading to unforeseen changes in the importance of certain goods or country interactions (Z. Ma & Cheng, 2005). Given this intricate connection between economic shocks and trade, leveraging the ITN in a computational context could significantly enhance the ability to predict and analyze economic shocks.

Change-point detection, a statistical technique used to identify significant shifts in time series data, can be effectively applied to social networks such as the ITN. This method identifies points where statistical properties, like mean or variance, change substantially, indicating a shift in an underlying process (Enikeeva & Klopp, 2021). In economic analysis, change-point detection can reveal shifts in macroeconomic indicators, which has been previously used to identify financial crises, but not broader economic shocks (D. Ma & Mankad, 2020; Malesios et al., 2020; Pepelyshev & Polunchenko, 2015). Additionally, these prior works typically use stock market data, and although this type of data is effective, when uncovering broader economic shifts, trade network data provides more nuanced insights regarding country relationships and rippling effects.

With networks specifically, change-point detection can be applied to comparisons of

snapshots of a graph at different years. Common methods incorporate unsupervised network distance metrics measuring the differences in edges or a node’s neighborhood from year to year (Barnett & Onnela, 2016; Enikeeva & Klopp, 2021). Yet, these methods lack the awareness of ground-truth change-points and therefore potentially overlook well-known events. They also do not incorporate node-specific information, and purely rely on the changes in network connections. To address these limitations, a supervised graph machine learning approach can be employed.

Graph Neural Networks (GNNs) have emerged as a powerful tool for working with complex data contained in a graph structure, especially in the context of trade networks (Casellato, 2022). By leveraging the relational information in graph data, GNNs can capture intricate patterns and dependencies that traditional network distances may overlook. For change-point detection specifically, siamese-GNNs offer a promising solution (Sulem et al., 2022). These models are trained on pairs of graphs labeled by whether they are separated by a change-point, thus incorporating explicit knowledge of economic shocks into the detection process. This GNN-based approach combines the strength of networks in conveying evolving relationships between entities with the ability to tailor the desired outcome through supervised learning, enabling more accurate identification of significant economic shifts present in the ITN.

When applied to subgraphs of the ITN, this trained model can identify to what extent the trade within a certain context was truly impacted by a given event. As two examples of potential applications, this thesis uses the trained s-GNN to determine whether impacts of economic shocks on the ITN vary by connectedness of regions or type of good. Previous research on the relationship between increasing globalization and the prevalence of economic shocks has yielded conflicting results, often due to a narrow focus on specific regions or shock types (Elbadawi & Hegre, 2008; Lungová, 2013; Marginean & Orastean, 2011; Nieman, 2011). Meanwhile, product-level analyses have suggested that trade in consumer goods tends to recover faster than that of raw materials following economic disturbances (Bartik et al., 2020; Jaarsma et al., 2017; Szczygielski et al., 2022). However, these findings again have primarily emerged from isolated case studies. This work therefore aims to provide a more expansive analysis that is less susceptible to region- or product-specific factors.

I therefore address the limitations of these prior works by focusing on the following three research questions:

RQ1: Does graph machine learning-based change-point detection outperform traditional network methods in identifying economic shocks present in the international trade network?

RQ2: Is an increase in globalization associated with more countries commonly experiencing an economic shock?

RQ3: Do trade networks of raw materials display fewer identified economic shocks compared to trade networks of manufactured goods?

In addressing these research questions, the results demonstrate the effectiveness of graph machine learning in offline change-point detection, providing a remarkable 0.524 percentage point increase in F1 score compared to traditional network-based distance metrics. The s-GNN model's strength lies in its discriminative ability, enabling the identification of specific years associated with economic shocks. When applied to regional and product-specific sub-networks, the trained s-GNN successfully detects both well-known and previously unidentified economic shocks. This application yields two key findings: first, it provides evidence supporting the relationship between increased globalization and a higher frequency of economic shocks across regions. Second, it reveals unexpected patterns in the occurrence of economic shocks across manufactured and raw goods categories, challenging conventional assumptions about their relative stability.

Through these findings, this work introduces an innovative approach to analyzing international trade networks using change-point detection and graph machine learning techniques. Although these methods have individually been applied to the international trade network, they have never previously been combined. The proposed methodology not only contributes to our understanding of global trade dynamics but also provides practical tools for policymakers and economists to navigate the complexities of international commerce. By offering insights into economic shocks across various regions and product categories, this work aims to enhance our ability to predict, understand, and respond to significant economic events.

This thesis is structured as follows: first, it reviews relevant literature motivating the hypotheses associated with each research question (Section 2). Next, it outlines the construction of the trade network and details the s-GNN architecture, explaining how change-point detection is incorporated into this framework (Sections 3 and 4). The results section then compares the performance of traditional network distance metrics with the s-GNN method and demonstrates applications of the trained s-GNN model to specific subsets of the trade network (Section 5). Finally, the thesis concludes by discussing its contributions, limitations, and future research (Section 7).

2 Literature and Hypotheses Development

This work draws from four key research areas: the relationship between economic shocks and trade, change-point detection in economic applications, graph neural networks utilizing trade data, and relevant background for two example applications of siamese Graph Neural Networks (s-GNNs). This foundation establishes the reasoning for why the international trade network should be used in this work, identifies limitations in existing change-point detection methods, and demonstrates how s-GNNs can address these shortcomings.

2.1 Manifestation of Economic Shocks in Trade Patterns

To justify the use of trade data, this work first examines literature on economic shocks and how they manifest themselves in export and import values. An economic shock is typically defined as an unexpected event that affects an economy, either positively or negatively ([Besser et al., 2008](#)). These shocks can be broadly categorized into three main types: financial crises, industry-specific downturns, and external events such as natural disasters or geopolitical incidents ([Hill et al., 2012](#)). Each category uniquely influences international trade dynamics and affects countries differently depending on their economic context.

Previous research has examined these categories individually, often utilizing the gravity model of trade and employing empirical case studies of specific countries or regions. In financial crises, internal shocks leading to currency depreciation have varying effects on imports and exports, with the overall impact depending on the magnitude of these changes ([Tambunan et al., 2012](#)). While cheaper domestic production typically boosts export quantities, the increased relative cost of foreign goods tends to reduce imports. For instance, the 1997 Asian Financial Crisis led to significant currency depreciations in several Southeast Asian countries. While this boosted exports in some sectors due to increased price competitiveness, it also led to a sharp decline in imports and overall economic contraction in the affected countries ([Corsetti et al., 1999](#)).

External shocks, however, often counteract export benefits as they simultaneously affect other economies, preventing the increased demand observed in some financial crises ([Z. Ma & Cheng, 2005](#)). These shocks can trigger a cyclical decline in both exports and imports: reduced export values initially lead to decreased production, forcing lower employment and diminished household incomes, thereby causing weakened demand for goods.

Significant trade agreements can also function as economic shocks. While these agreements are designed to boost trade among participating nations, their impact is still un-

predictable, with nearly half of all economic integration agreements failing to produce a measurable effect on trade volumes (Kohl, 2014). Rose (2004) used the gravity model with bilateral trade data to understand three major trade agreements, including the World Trade Organization (WTO), the General Agreement on Tariffs and Trade (GATT), and the Generalized System of Preferences (GSP). This study similarly found little difference in trade patterns stemming from the WTO and GATT but found significant impacts from GSP, further confirming the differential impacts of trade agreements on trade networks.

External events encompass a wide array of shocks, including natural disasters, public health crises, and geopolitical events, each with varying impacts on trade patterns. The impact of natural disasters on trade, while somewhat predictable, largely depends on the affected country's economic status (F. Liu et al., 2023). Developing nations typically experience increased imports and decreased exports following climate disasters, whereas developed countries generally see no significant changes in overall trade volumes.

Pandemics present a more complex scenario, generating both supply and demand shocks with unclear trade implications (Baldwin & Di Mauro, 2020). While health concerns force many workers to stay home, reducing production capacity, the heightened demand for essential medical and household goods often boosts imports. The Covid-19 pandemic, for instance, led to a net decrease in imports from China overall, yet some countries increased their Chinese imports when facing supply chain disruptions elsewhere (X. Liu et al., 2022). Furthermore, pandemic-induced travel restrictions triggered recessions in numerous countries, adding another layer of complexity to trade impacts (Vo & Tran, 2021).

Geopolitical events can have indirect impacts on trade patterns. For example, the Brexit vote in 2016 led to immediate currency fluctuations and long-term changes in trade relationships between the UK and EU countries (Graziano et al., 2021). These events often create uncertainty in international markets, affecting trade flows and investment decisions.

The ambiguous impacts of economic shocks on trade, both between and within different types of shocks, underscore the complexity of these events. While trade data is applicable in investigating these events, the nuanced dynamics necessitate more sophisticated modeling approaches than traditional econometric methods allow. This work therefore focuses on developing a model that can capture these non-linear relationships between economic shocks, country-level factors, and trade patterns.

2.2 Change-Point Detection of Financial Crises

Change-point detection is a statistical technique used to identify abrupt shifts in the underlying patterns or distributions of time series data (Pepelyshev & Polunchenko, 2015).

It aims to pinpoint moments when the statistical properties of a sequence change significantly. These methods are broadly categorized into two types: offline and online change-point detection. Offline detection analyzes the entire dataset retrospectively, identifying all change points after all data has been collected. In contrast, online change-point detection processes data sequentially as it becomes available, making decisions about potential change points in real-time. While online methods are suited for prediction tasks, they may be less accurate than offline approaches due to limited information at each decision point. The choice between offline and online methods depends on the specific requirements of the analysis, such as the need for real-time detection versus historical analysis.

Macroeconomic forecasting has been the typical use-case for change-point detection within economic analysis but is still a work in progress due to the unpredictable nature of many economic events. Prior works in this area often use GDP, stock market, and other macroeconomic indicators. Some approaches involve nonparametric estimations of both the number of change points and when they occur ([Allen et al., 2018](#); [Zou et al., 2014](#)). Other studies utilize parametric methods to detect sudden changes in the time series based on varying distributions, such as Shiryaev-Roberts ([Pepelyshev & Polunchenko, 2015](#)) and CUSUM ([Ergashev, 2004](#)), but the number of change-points in these methods is unknown. Because of this, although these methods can provide insights into the years that were most impacted, their use is difficult to justify when there are known ground-truth change-point years. Bayesian multi-change point analysis was therefore developed where the number of change-points can be specified, better substantiating the correctness of the method in detection of future change-points ([Heßler et al., 2023](#)). Yet, because these methods are parametric, they therefore rely on the choice of an appropriate distribution which can hinder the applicability of these methods to different contexts.

A network (graph) is a method of representing interactions between entities (nodes), typically individuals, places, or communities connected by links (edges). Because of networks' ability to capture sudden shifts in dynamics between these nodes, prior works have started researching them as a new domain within change-point detection ([Peel & Clauset, 2015](#)). Network-specific change-point detection methods aim to quantify the difference between two graph snapshots. Popular metrics include Jaccard similarity, Frobenius distance, Laplacian spectral distance, and graph kernels ([Sulem et al., 2022](#)). In the context of detecting financial crises, stock market data is often used but in a network format where stocks are connected based on a certain level of similarity. Prior works have utilized distance metrics such as the Gaussian kernel function ([Banerjee & Guhathakurta, 2020](#)), the Laplacian energy measure ([C. Huang et al., 2023](#)) and non-negative matrix factorization ([D. Ma & Mankad, 2020](#)). Similar to the traditional change-point detection methods without networks, these network-based methods are unsupervised, meaning they

do not have any knowledge of the specific differences between graphs they should be looking for. The detected change-points are therefore very sensitive towards the metric used. Because of this, in-depth domain knowledge is required to ensure the correct choice of metric.

With the use case of detecting crises using typical finance data, stock market indices and GDP indicators are just one-dimensional metrics that primarily capture the financial evolution of high-income countries. Using trade data, however, encapsulates most countries and also shows indirect effects of other countries' actions. Overall, these prior works support the use of change-point detection within the context of detecting shocks while providing room for improvement through the novel use of the ITN.

2.3 Graph Machine Learning and Trade Networks

Graph machine learning, a subset of machine learning that focuses on data represented as graphs or networks, offers a promising approach to combine the benefits of supervised learning with the utilization of network structures. This approach is particularly relevant for analyzing complex relational data such as the ITN. Within this field, graph neural networks (GNNs) have gained significant prominence. GNNs are deep learning models designed to process and learn from graph-structured data, capturing both node features and the relationships between nodes ([Zhou et al., 2020](#)).

Recent years have seen a growing interest in applying GNNs to international trade network analysis. Many studies have focused on using GNNs to predict node-level features or future connections between countries. For instance, [Panford-Quainoo et al. \(2019\)](#) trained a GNN on trade flow data between countries, weighted by GDP and distance. They then applied these models to downstream tasks including predicting future trade partners, achieving 98% accuracy. This application demonstrates the potential of GNNs to capture and leverage the complex interconnections within global trade networks.

Expanding on this work, [Monken et al. \(2021\)](#) specifically examined soybean trade and trained a GNN to predict trade unit value. Particularly relevant to the current work, they further applied their model to two economic shocks relevant to soybean trade - the China-US trade war and the Covid-19 pandemic. They created a counterfactual model of soybean trade to simulate if trade had been cut off entirely between the US and China and looked at the impacts on the trade unit value of China, the US, India, and Brazil from the pandemic.

The temporal evolution of trade networks has also been explored using GNNs. [Casel-lato \(2022\)](#) developed a recurrent graph neural network to predict these evolutions for different categories of products. This approach allows for a dynamic understanding of how trade relationships change over time.

While these applications of GNNs to trade networks are promising, the use of GNNs for change-point detection in trade networks remains largely unexplored. Change-detection using GNNs has previously been limited primarily to computer vision contexts. Siamese GNNs (s-GNN) have been used to understand changes between two images, such as in environmental and medical contexts (Shuai et al., 2022; Song et al., 2023; You et al., 2023). However, the potential of this approach within network contexts has not been yet realized.

A significant step towards this application was made by Sulem et al. (2022), who developed an s-GNN for change-point detection and applied the model to detecting financial market crashes using networks of S&P 500 stocks. Their method more accurately identified market events compared to traditional change-point detection methods, demonstrating the power of this approach in financial contexts. This study serves as a model for the methodology of the current work in detecting broader economic shocks.

The primary benefit of the s-GNN approach is its lack of metric sensitivity issues, which can inhibit traditional network-based methods. It also enables the incorporation of node-based features, whereas traditional metrics purely rely on structural differences between graph snapshots. Additionally, s-GNNs' supervised nature provides the ability to specify the nature of which differences between years to detect. Based on the success of the s-GNN approach in detecting financial crises using stock market data, the following hypothesis is proposed regarding RQ1:

H1: The s-GNN method will achieve higher performance than traditional network-distance based metrics in identifying change-points in the international trade network.

2.4 Resilience to Economic Shocks

The resilience of countries to economic shocks has been a subject of extensive study, driven by the need to anticipate effective policies. However, much of the existing research on resilience consists of case studies focusing on specific regions or shock types, limiting their broader applicability. Some researchers have attempted to analyze multiple countries and their resilience to various economic shocks using traditional regression approaches with macroeconomic indicator features (Duval et al., 2007; Soufi et al., 2022). These studies have identified factors such as export-to-GDP ratio and flexible product markets as crucial determinants of economic resilience during crises. Yet, these analyses are often constrained to upper-middle and high-income countries, leaving a significant gap in our understanding of resilience in lower-income nations. When examining external shocks specifically, research has shown that countries implementing resilience-building policies

tend to fare better, regardless of their economic size (Briguglio, 2016). This suggests that proactive policy measures can play a crucial role in mitigating the impact of economic shocks. Therefore, a more inclusive analysis of resilience, encompassing a broader range of countries and a longer timeline of economic shocks, would provide valuable insights into global economic dynamics.

2.4.1 Region-Level Analysis

The evolution of globalization, particularly the Third and Fourth Waves marked by the World Trade Organization's formation in 1995 and technological advancements in the early 2000s, has led to unprecedented global interconnectedness (Targowski, 2014). This increased connectivity offers countries expanded access to products, services, and expertise, potentially fostering economic growth. However, it also heightens vulnerability to economic downturns in individual nations.

Previous research on globalization's impact on economic shocks has produced conflicting results, often due to the focused nature of studies on specific regions or shock types. Some studies have examined the effects on developed or emerging European countries (Lungová, 2013; Marginean & Orastean, 2011), while others have concentrated on specific economic shocks such as armed conflict (Elbadawi & Hegre, 2008) and civil war (Nieman, 2011).

The theory of economic interdependence provides a potential framework for understanding these dynamics. As countries become more interconnected through trade and financial flows, they may also become more vulnerable to external shocks (Keohane & Nye Jr, 1973). Supporting this theory, Kose et al. (2003) found that increased financial integration led to higher volatility in consumption growth for a sample of 76 countries from 1960 to 1999. This increased volatility could be interpreted as a higher susceptibility to economic shocks, suggesting that the intensified globalization of the early 2000s might have amplified countries' exposure to global economic fluctuations.

The inconsistent findings from these studies, coupled with the theoretical insights from economic interdependence, underscore the need for a more comprehensive analysis. Given this context, the following hypothesis is proposed for RQ2:

H2: Years following the significant increase in globalization in the early 2000s will exhibit a higher probability of economic shock.

2.4.2 Product-Level Analysis

Just as the impact of economic shocks is not uniform across regions, it is also not uniform across products. The concept of product-level macroeconomic resilience has gained

significant attention, particularly in the wake of the Covid-19 pandemic. Multiple studies have found that goods deemed most essential and those used as substitutes demonstrated the highest resilience, with food and telecommunications sectors being prime examples ([Bartik et al., 2020](#); [Szczygielski et al., 2022](#)). This aligns with the basic economic principle that demand for necessities tends to be more inelastic, making these sectors more resilient to economic fluctuations.

A study by [Jaarsma et al. \(2017\)](#), focusing on Dutch industries and their resilience to the Global Financial Crisis, revealed that trade in products closer to the final consumer recovered faster than that of raw materials. This finding suggests a potential hierarchy of resilience within supply chains, with downstream products benefiting from more stable demand patterns. These studies on the correlation between specific types of goods and economic resilience inform the following hypothesis regarding RQ3:

H3: Networks of crude and raw materials will exhibit more change-points compared to networks for manufactured products.

3 Data

To construct the trade networks for both the traditional and s-GNN change-point methods, I use the following data sources:

1. **Bilateral trade values data:** The trade data utilized in this research is derived from the Harvard Growth Lab’s Atlas of Economic Complexity dataset, which offers a cleaned version of international trade data sourced from the UN Comtrade database. This dataset provides yearly import and export values spanning from 1962 to 2021, covering a wide range of countries and products, where each row in the dataset specifies an exporter, importer, the amount being traded, and the product type. The products are categorized according to the Standard International Trade Classification (SITC) codes. These codes include varying levels of granularity, but the S4 codes, used in this work, are the broadest levels and group the products into 10 different categories. Table 1 displays the descriptive statistics for the trade data.
2. **Country characteristics data:** For the s-GNN modelling, the networks can contain information specific to the countries. This country-level information is derived from the World Development Indicators from the World Bank. For years 1962 to 2018, the set of all indicators was previously scraped for all available countries. This set of years therefore limits the overall scope of analysis to this timeline. The

	Mean	Std	25%	50%	75%
Export Value	16,777.53	209,185.81	4.07	43.68	702.60
Import Value	14,652.84	156,953.45	4.37	46.41	757.50
Trade Balance	2,124.69	113,387.38	-83.54	-0.69	73.02
Total Trade Value	31,430.38	352,036.19	17.11	138.85	1,846.29

Table 1: Descriptive statistics for export value, import value, total trade value, and trade balance across pairs of bilateral trade. These values are reported for all trade summed across products and averaged by year for each unique pair of countries. Total trade value is the combined export and import value and trade balance is the difference between export and import values. Values are in USD and reported in 10,000’s.

primary categories within this dataset include macroeconomic, demographic, institutional, and industry variables. Examples of each category are listed.

- (a) **Macroeconomic:** GDP, inflation rates, employment figures, and foreign direct investment inflows
- (b) **Demographic:** population size, growth rates, and age distribution
- (c) **Institutional:** governance quality, ease of doing business, and political stability
- (d) **Industry:** contributions of different industries (e.g. agriculture, manufacturing, construction) to national GDP

These variables can provide context alongside the raw trade values as to why certain countries might be more likely to form trade relationships.

4 Methodology

To test the research questions, two overarching methods were conducted. For RQ1, the traditional and s-GNN change-point detection methods were created, tested, and compared. RQ2 and RQ3 focus on the initial application of the trained s-GNN model to specific segments of the trade network and will rely on more qualitative analysis of the produced change-points.

4.1 Trade Network Construction

To model the ITN, directed networks were created from the trade data where countries are represented by nodes, and edges connect countries that have a trade relationship in the export direction. An edge starting at country A and linked to country B indicates

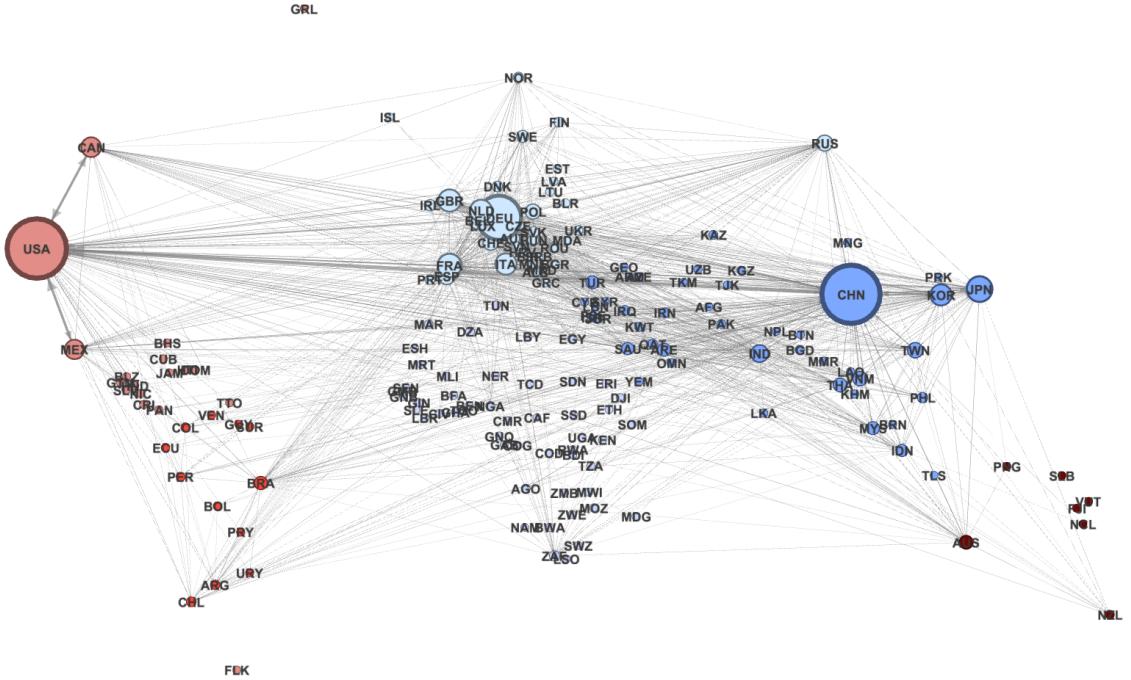


Figure 1: Network of 2018 trade where the edge weights are the total value of exports. Only edges with a value greater than 1.4×10^9 are shown. The edge width is proportional to the edge weight, where wider edges indicate higher export amounts. The node sizes are proportional to the in-degree of the node. The positions are based on the latitude and longitude of the countries.

that A exports goods to B. The edges are then weighted by export value across all products, where the edge weight indicates the proportion of all of country A’s exports that go to country B. The purpose of this was to ensure that smaller economies were not as overshadowed by the larger players, such as the United States and China. Fig. 1 displays an example of the network setup for 2018 trade. Node attributes were also added in the s-GNN method to give greater context to the differences between countries, and the selection of these node features is discussed in section 4.3.3.

This process of constructing the trade network was completed for each of the years between 1962 and 2018, leading to a final set of 57 graphs. These networks were then used in the change-point detection methods.

4.2 Economic Shock Operationalization

To train a model and determine its performance in predicting economic shocks, ground truth values of when economic shocks occurred between 1962 to 2018 are required. Yet, the definition of economic shocks, as described in Section 2, is expansive. Because of their unclear nature and inclusion of not only financial crises but also geopolitical events and natural disasters, there could be an economic shock of some magnitude contained in

every year between 1962 to 2018. This work will therefore only focus on the economic shocks reflected in the trade networks.

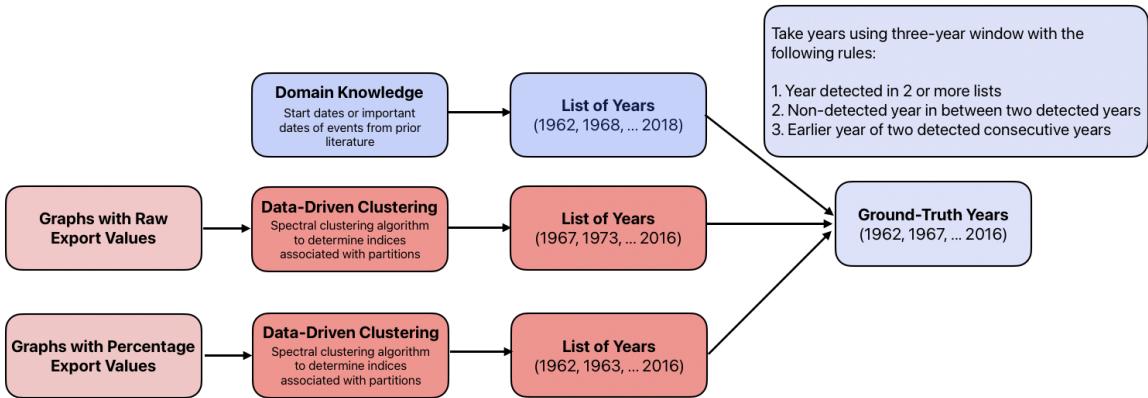


Figure 2: Overview of ensemble method to determine ground-truth economic shock years to use as supervised change-point labels. Two of the created lists of years depend on a data-driven spectral clustering method while the third relies on domain knowledge of large-scale economic shocks. The final list of ground-truth years takes a combination of the three lists.

To do this, an ensemble of domain-knowledge and network-driven methodologies was developed, as shown in Fig. 2, where each method outputs a unique list of years to possibly use as ground-truth values. The first method initially defined economic shocks through a culmination of events from domain knowledge and those studied in related prior literature. The crises were labeled using their associated starting year – for example, 1973 for the Oil Crisis and 2007 for the Global Financial Crisis.

To better identify the years when these shocks actually manifested themselves in trade networks, a data-driven clustering methodology from Sulem et al. (2022) was incorporated that utilized the networks developed in section 4.1. The method first clustered the networks into groups of countries via spectral clustering with the network’s normalized graph Laplacian matrix. Spectral clustering acts as a method of dimensionality reduction that is then able to identify communities within the network. At each year, the silhouette score of the clusters was computed, which computes the distance between clusters where a higher value indicates more distinct groups of nodes. The number of clusters tested ranged from 2 through 20, and the number associated with the highest average silhouette score was used.

The network at each year was then clustered, and the clustering labels of each possible pair of years were then compared. This was done using the Adjusted Rand Index (ARI), which measures the similarity between two sets of clustering labels, resulting in a 57x57 matrix of ARI scores. To determine which years were most similar, the ARI scores were then clustered using spectral clustering, where cluster values 10 through 25 were

tested. The silhouette score was again computed to determine the best number of clusters, corresponding to the number of economic shocks.

To determine the years of these economic shocks, centroids were computed for each of the clustered groups. This involved first obtaining the integer closest to the mean position of each of the labels. The year corresponding to that position in the array was then labeled as a possible change-point.

This clustering algorithm was implemented twice with different versions of the trade networks. The first used the export percentages per target country for edge weights as described in section 4.1. The second used the raw export values between countries as the edge weights. These were both used since the first allows for economic shocks of smaller economies to not be as overshadowed. The second, however, could emphasize larger economic shocks if the export values of stronger economies change drastically.

To get the final crisis year labels, an ensemble method of the three label lists (domain, export percentages, and raw exports) was used. When looking at windows of three years, the following rules were used in determining the final list of ground-truth economic shock years. If a year was detected in two or more of the lists, then it was considered ground-truth. If a year not present in any of the lists was in between two years detected by any of the lists, then it was considered ground-truth. Lastly, if there were two consecutive years identified by any of the lists, the earlier year was used as a ground-truth year.

4.3 Change-Point Detection

As described in section 2.2, change-point detection in the network context has typically been an unsupervised or semi-supervised process using different distance metrics. In the following section, these traditional distance metrics are first outlined and used as a baseline for the performance of the s-GNN. The architecture and training process of this model is then described as well as a synthetic s-GNN experiment used as a robustness check.

4.3.1 Traditional Network Change-Point Detection

Network distance metrics for traditional change-point detection were used as a baseline comparison for the s-GNN method. The choice of distance metrics within this work were guided by those used in Sulem et al. (2022). They include:

1. **Frobenius distance** takes the Frobenius norm of the difference between the two graphs' adjacency matrices. This norm takes the square root of the sum of the

absolute square of each element ([Barnett & Onnela, 2016](#)):

$$d_F(A, B) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (a_{ij} - b_{ij})^2}, \quad (1)$$

2. **DeltaCon similarity** uses the affinity matrices to understand how information propagates differently between the two networks ([Koutra et al., 2013](#)). The affinity matrix contains the pairwise similarities between graphs, and these similarities are calculated using the Fast Belief Propagation method with the following equation:

$$S = [I_n + \epsilon^2 D - \epsilon A]^{-1}, \quad (2)$$

where I is the identity matrix, D is the diagonal degree matrix, A is the adjacency matrix, and ϵ is a positive constant.

3. **Weisfeiler-Lehman (WL) kernel** is a graph similarity measure that iteratively refines node labels based on the labels of their neighbors. Starting with initial node labels, the algorithm updates these labels over several iterations, capturing the structural information of the graph. After each iteration, feature vectors are created from the frequency of node labels. The WL kernel value between two graphs is computed by summing the similarities of these feature vectors across all iterations, effectively capturing both local and global graph structures ([Shervashidze et al., 2011](#)).
4. **Network cumulative sums statistic (CUSUM)** takes the dot product of the matrices resulting from the following equation being applied to two independent samples of graphs (those from even and odd indices) ([Yu et al., 2021](#)):

$$C_t = \frac{1}{\sqrt{2L'}} \left(\sum_{s=t-L'+1}^t A_s - \sum_{s=t+1}^{t+L'} A_s \right), \quad L' \leq t \leq T - L', \quad (3)$$

where A_s is a graph's adjacency matrix and L' is $L/2$.

5. **Operator norm of network CUSUM (CUSUM 2)** is the same as CUSUM but only uses one sample that is instead partitioned into two groups corresponding to past and future observations ([Enikeeva & Klopp, 2021](#)).
6. **NCPD with spectral clustering** uses the graph Laplacian with k -means clustering to partition the networks ([Cribben & Yu, 2017](#)).

7. **Laplacian Anomaly Detection (LAD)** uses the equation

$$Z_t = 1 - |\tilde{\sigma}_t \sigma_t|, \quad (4)$$

where σ is a vector of ℓ largest values from the unnormalized graph Laplacian and $\tilde{\sigma}$ is the dot product of all the ℓ largest values from the graph Laplacian from each window of length L ([S. Huang et al., 2020](#)).

Using one of these metrics, the average similarity statistic is first calculated with the following equation ([Sulem et al., 2022](#)):

$$Z_t(L) = \frac{1}{L} \sum_{i=0}^{L-1} s(G_{t-i}, G_{t-i-1}), \quad t \geq L, \quad (5)$$

where s is a similarity function with two graphs as input coming from a window of length L . For most of the metrics, $L = 4$ was used while $L' = 2$ was used for CUSUM and CUSUM 2 due to these metrics' use of partitions. For each pair of consecutive graphs within the given window, the similarity scores are calculated and then averaged. The purpose of averaging is to limit the effects of noise within the data. Then if $Z_t \leq 0.5$, and $Z_{t'} \geq 0.5$ where t' is a value less than t within the window of size L , then t is a change-point detected by the algorithm. The threshold of 0.5 is arbitrary and can be fine-tuned to best suit a given similarity metric. This process was completed for each of the seven distance metrics listed above.

4.3.2 Siamese-GNN Change-Point Detection

As a supervised, deep learning approach to change-point detection, a siamese graph neural network (s-GNN) was constructed and tested. The output of this model can be used in place of s in the similarity equation (5). The structure and how the inputs are manipulated throughout the model are depicted in Fig. 3. Each input into the model is a pair of networks which come from two different years. The associated ground-truth label of this pair is 0 if there is an economic shock between the two years or 1 if there is not. For example in Fig. 3, the pair of networks for year 1972 and 1980 is labeled as 0 since the 1973 Oil Crisis occurred between these years. From each of the networks within the pair, a feature matrix and an adjacency matrix can be derived. The feature matrix contains the World Bank country-level characteristics, described further in section 4.3.3, and the adjacency matrix contains the edge weights corresponding to export percentages between each pair of countries.

The feature matrix and adjacency matrix for each graph are then provided to the graph encoder. This encoder is responsible for capturing the relationships between nodes

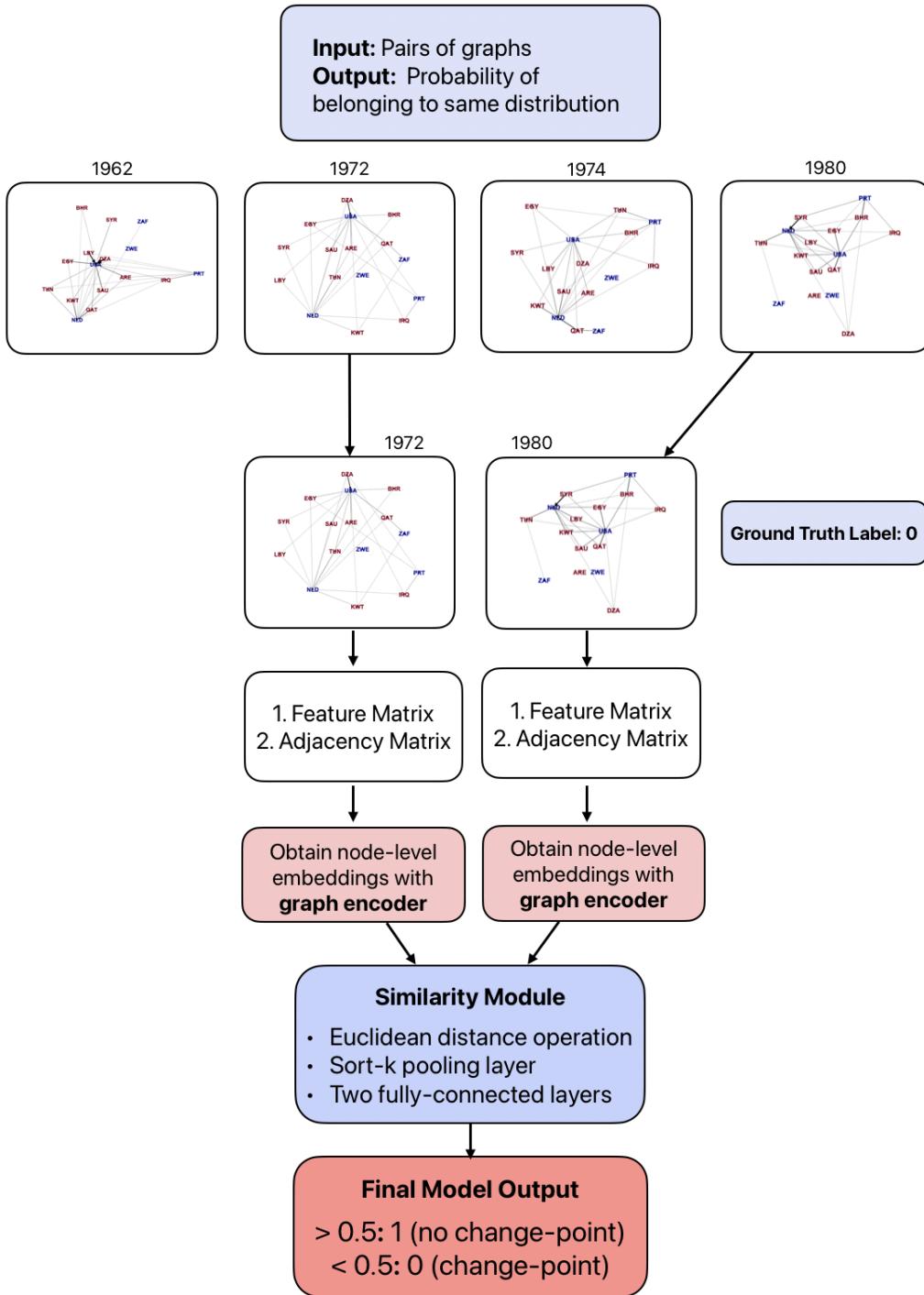


Figure 3: Diagram of the training process and layers included in the siamese graph neural network. Four networks are used as an example where a pair of them are trained on that are separated by an economic shock.

for an individual graph, but contains the same weights when applied to each graph within the pair. The following four encoding layers were tested:

1. **Graph Convolutional Network (GCN)** is an encoder derived from computer vision tasks where the input is now a graph instead of an image. Broadly, this encoder works by grouping similar nodes into clusters and taking the weighted average of the cluster where lower-degree nodes get weighted more highly ([Zhang et al., 2019](#)). One drawback of this approach is that it can be very local in scope where the encoder only receives information from a node's immediate neighbors and not the overall graph structure. A single layer within the GCN uses the following equation:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)} \right), \quad (6)$$

where σ is the activation function, \tilde{A} is the adjacency matrix of the network with added self-loops, \tilde{D} is the degree matrix of \tilde{A} , H is the matrix of activations in the l -th layer, and W is the weight matrix of the l -th layer.

2. **GraphSAGE** works best for large-scale networks since it samples a fixed number of neighbors for each node rather than looking at all neighbors like the GCN ([Hamilton et al., 2017](#)). It aggregates the sampled nodes' features through aggregation functions like mean, LSTM, and pooling. Although this sampling procedure reduces complexity, it can introduce noise due to the specifics of a given sample rather than looking at the entire neighborhood. The layer-wise propagation uses the following equation:

$$h_v^{(l+1)} = \sigma \left(W^{(l)} \cdot \text{AGGREGATE}^{(l)} \left(\{h_v^{(l)}\} \cup \{h_u^{(l)}, \forall u \in \mathcal{N}(v)\} \right) \right), \quad (7)$$

where $N(v)$ represents the neighbors of node v , $h(v)$ is the feature vector of node v , and AGGREGATE is the aggregation function (such as mean or LSTM).

3. **Graph Attention Networks (GAT)** uses attention mechanisms to focus on the most relevant information by assigning different importance to different neighbors. In the case of graphs, this is especially advantageous because the model can better understand the relationship between a node and its neighbors ([Veličković et al., 2017](#)). This encoder uses the following equation:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(v)} \alpha_{vj}^{(l)} W^{(l)} h_j^{(l)} \right), \quad (8)$$

where α_{vj} represents the attention coefficients.

4. **Graph Isomorphism Network (GIN)** is more sensitive to differences in graph structure (Xu et al., 2018), and uses a Multi-Layer Perceptron (MLP) to aggregate information from a node’s neighbors. It is a derivative of the Weisfeiler-Lehman isomorphism test, and uses the following equation:

$$h_v^{(l+1)} = \text{MLP}^{(l)} \left((1 + \epsilon^{(l)}) h_v^{(l)} + \sum_{u \in \mathcal{N}(v)} h_u^{(l)} \right), \quad (9)$$

where MLP is a multi-layer perceptron, and ϵ is a fixed scalar.

The encoder converts the networks to a vector which encapsulates the node and edge information. The pair of node embeddings are then passed through the similarity module to calculate the distance between the two graphs. This module consists of a Euclidean distance operation which obtains the node-level pairwise distances, a sort- k pooling layer which obtains the largest k distances, followed by two fully-connected layers. These fully connected layers consist of a linear layer, a normalization layer, and a ReLU activation layer. The output of the second fully-connected layer is then sent through a sigmoid activation function which transforms the output to a value between 0 and 1, which can be interpreted as the probability that the two graphs belong to the same phase. In training the model, gradient descent was utilized with an Adam optimizer and binary cross entropy loss function.

4.3.3 Feature Selection

Three different sets of variables were tested to use as the feature matrix within the s-GNN. The baseline set of features included the centrality of the given country in each of the 10 SITC product categories, the prior year’s GDP growth rate, and the current year’s GDP growth rate. The second set of features used mutual information selection (MIS) to obtain 15 features with the highest MIS score from the set of all World Development Indicators from the World Bank. These scores are based on a regression using current GDP growth as the dependent variable. MIS determines what the most important features are in predicting GDP but any variables directly related to GDP such as GDP per capita or GDP growth were excluded. With the s-GNN, there is no direct variable that corresponds to change-points, so GDP was used instead due to its importance in predicting financial crises. This method of feature selection, however, could depend too heavily on GDP and not acknowledge other important features. Therefore, the final method used a random subset of 15 features to try and capture variables that might not have been recognized by the MIS method. This number of features was used based on exploratory analysis that indicated a balance between the fewest features possible while maintaining model

performance. The same sets of features were then tested without the previous or current GDP variables. This is to understand how much the model depends on this type of information which is typically used in standard time-series financial models. The variables contained in the MIS and random feature subsets are listed in Appendix A.2.

4.3.4 Model Training and Testing

The first step in creating the s-GNN was appropriately splitting the pairs of graphs into train, validation, and test subsets. This data split was established using a random sampling scheme which consists of creating a set of all possible unique pairs of graphs and then randomly selecting 60% for the training set, 20% for the validation set, and 20% for the testing set. Another method, the window sampling scheme, instead uses the first 60% of graphs as train, the next 20% as validation, and the last 20% as test and then derives the pairs of graphs from within these defined groups. Because this work aims to apply the model retroactively in the offline change-point detection context, the random sampling scheme is more pertinent since the model is exposed to a broader range of years in training, whereas the window sampling method is better suited for online change-point detection with a traditional financial time-series prediction task since it has not seen the most recent years' data. Both methods, however, were tested to see if there were large differences in performance.

Fig. 4 displays the overall process of feature, encoder, and hyperparameter selection. The three sets of features, both with and without the GDP growth rate variables, were each tested on the four different encoders. For each of these, they underwent 5-fold cross validation. Additionally, each set of features was then tested with three different normalization methods – no normalization, log transformation, and standardization. From the cross-validation scores, the top-three performing models were then chosen to undergo hyperparameter tuning. The following parameters were tested with grid-search hyperparameter tuning: learning rate of 0.001, 0.01, dropout rate of 0.01, 0.05, 0.1, sort- k values of 30, 50, 100, and hidden layer values of 16, 32, 64.

Once the hyperparameters were chosen, the final three models were determined by the use of four metrics including accuracy, F1 score, F2 score, and F0.5 score. The F1, F2, and F0.5 scores generally measure how well a binary classifier discriminates between the two classes (graph pairs containing vs. not containing a change-point) but each give their own weight to either precision or recall. F1 balances the two equally, F2 emphasizes recall over precision, and F0.5 emphasizes precision over recall. Precision in the case of identifying change-points would be the number of pairs not including a crisis year correctly identified out of all pairs identified as not including a crisis. Recall would be the number of pairs not including a crisis correctly identified out of all pairs not including

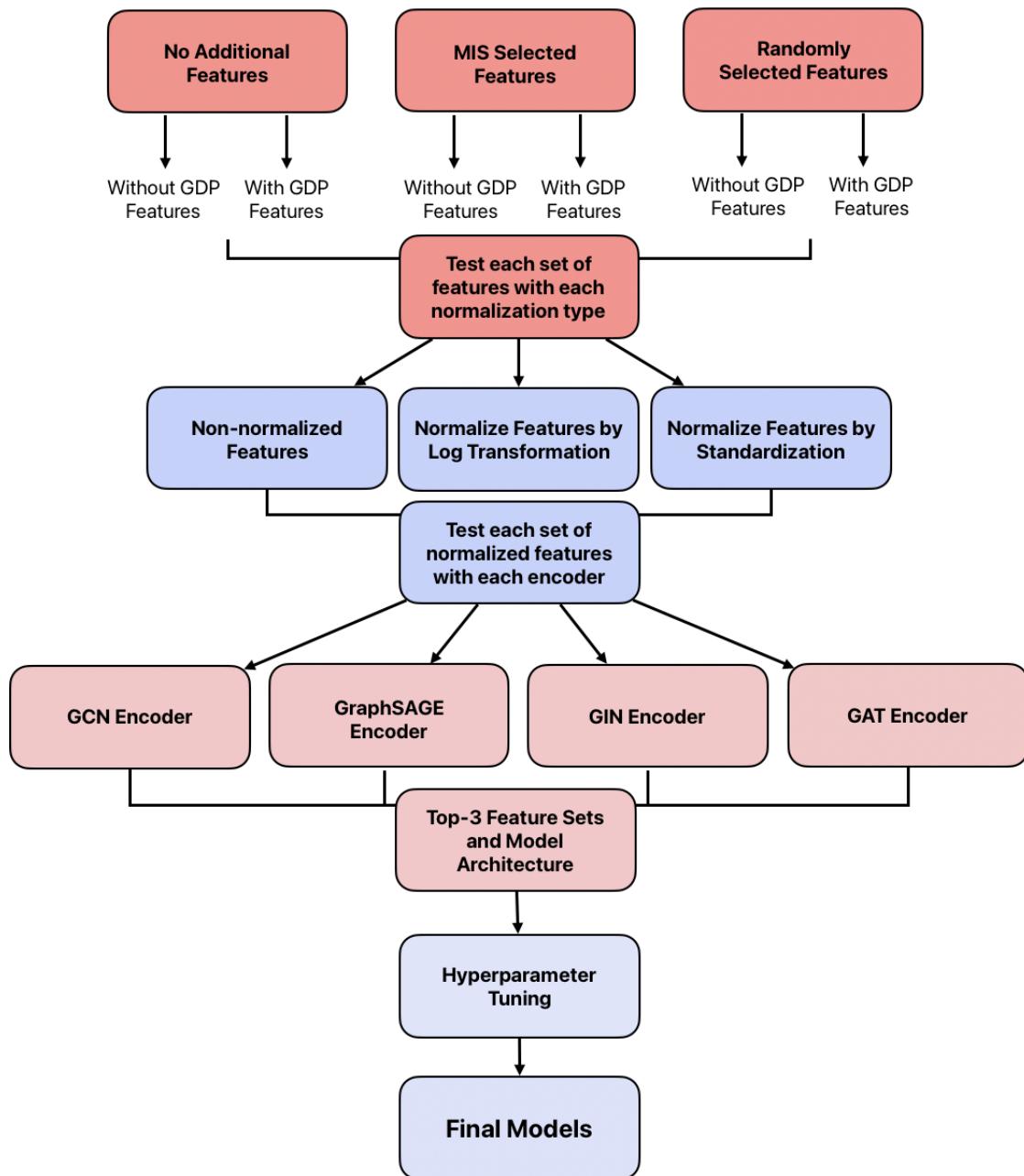


Figure 4: Diagram of the feature set, normalization type, and model selection process.

a crisis. These metrics were applied to each fold within the five-fold cross validation and then averaged to obtain the final score.

The top-three models were then applied to the test set of data. The same metrics were used in addition to AUC. These metrics were all applied to classifying pairs of networks as containing a change-point or not from their similarity scores. The next step was actually obtaining the change-points using the average similarity score equation (5), as was done with the traditional change-point detection methods. With each list of identified change-points from the three s-GNN models, the F1 and ARI scores were calculated to determine how well the lists of identified change-points aligned with the list of ground-truth change-points.

4.3.5 Synthetic Experiment Model Validation

Using synthetic data through generated stochastic block model networks, I validated the performance on real-world data for traditional vs. deep learning methods through synthetic experiments. This is to ensure that how the s-GNN performs on the real-world data is not unique to this trade-data scenario but it more widely applicable to other datasets. Four of the traditional network-distance change-point methods were implemented with the synthetic data - including NCPD, Frobenius distance, CUSUM, and CUSUM 2. The choice of these four methods was based on those used in [Sulem et al. \(2022\)](#). Since the synthetic networks do not have inherent node features, identity encodings were instead used.

The synthetic methods were implemented for the multiple change-point scenario since the single change-point scenario is not relevant to the trade data. A series of networks was created from a stochastic block model (SBM) and a Erdős–Rényi model where the switch from one to the other indicates a change-point. The cluster size (s) was tested for different values between 20 and 80, and this changes the number of communities present in the SBM networks. A time series of 2400 graphs was created and 9000 random pairs of graphs were sampled for training and validation.

4.4 Application of s-GNN Change-Point Detection

As an example of how the s-GNN model could be applied to analyze the ITN, specific subsets of the network were analyzed. These subsets include regions and products interpreted to understand how the change-points differ within these groups.

4.4.1 Analysis of Region Crises

To understand how shocks differ by region, the change-points were calculated for subsets of the ITN. These change-points were obtained using the trained, best-performing model from the test-set results of the performance metrics when applied to classifying pairs of networks. The s-GNN requires uniform nodes as those it was trained on, so to work with this, all nodes are still included even for countries not within a specified region, but only the edge values were kept for the countries being analyzed. The regions explored include six of the continents, including Asia, Africa, Europe, North America, Oceania, and South America. For each region, only the edges and their associated values are kept if countries are within a one-hop neighborhood of countries within the continent. This means that any countries one edge away or within the continent are kept in the network. The one-hop neighborhood is included to ensure that each continent has enough edges remaining within the network to show changes in their trade patterns.

The change-points produced were then analyzed qualitatively to see where countries do and do not overlap with the ground-truth change-points. Additionally, change-points not directly linked to the ground-truth economic shocks were analyzed to understand what events specific to the continent could produce that point.

To specifically answer RQ2 regarding how the probability of a shock corresponds with globalization, the level of globalization was measured for each region using the global clustering coefficient of the graph for a particular region at a particular year. The global clustering coefficient measures the ratio of actual triangles to possible triangles for a node, where a triangle is defined as three nodes being each connected to one another. By having a high clustering coefficient, this indicates the network is densely connected, aligning with the typical definition of globalization.

4.4.2 Analysis of Product Crises

To understand how the shocks differ by types of products for RQ3, the change-points were also calculated using the networks where the edge values only contain the proportions of exports for those within a given product category. The 10 categories explored include the 10 specified by the S4 SITC codes. Networks for each of the 10 product categories were created, and the final, trained s-GNN model was then applied to each network. To analyze the difference by product type, the product categories were classified as either raw goods or manufactured products. The categories, goods included, and classification are shown in table 2. Some categories are comprised of a combination of both raw materials and manufactured products, however, the classification was determined based on the majority of goods within a given S4 code. The average similarity at each year in each

group was then compared to understand the difference in detection of economic shocks between the two groups.

S4 Code	Product Category	Classification
0	Food And Live Animals	Manufactured
1	Beverages And Tobacco	Manufactured
2	Crude Materials, Inedible, Except Fuels	Raw Materials
3	Mineral Fuels, Lubric. And Related Mtrls	Raw Materials
4	Animal And Vegetable Oils, Fats And Waxes	Raw Materials
5	Chemicals And Related Products	Raw Materials
6	Manufactured Goods Classif. By Material	Manufactured
7	Machinery And Transport Equipment	Manufactured
8	Miscellaneous Manufactured Articles	Manufactured
9	Goods Not Classif. Elsewhere	Manufactured

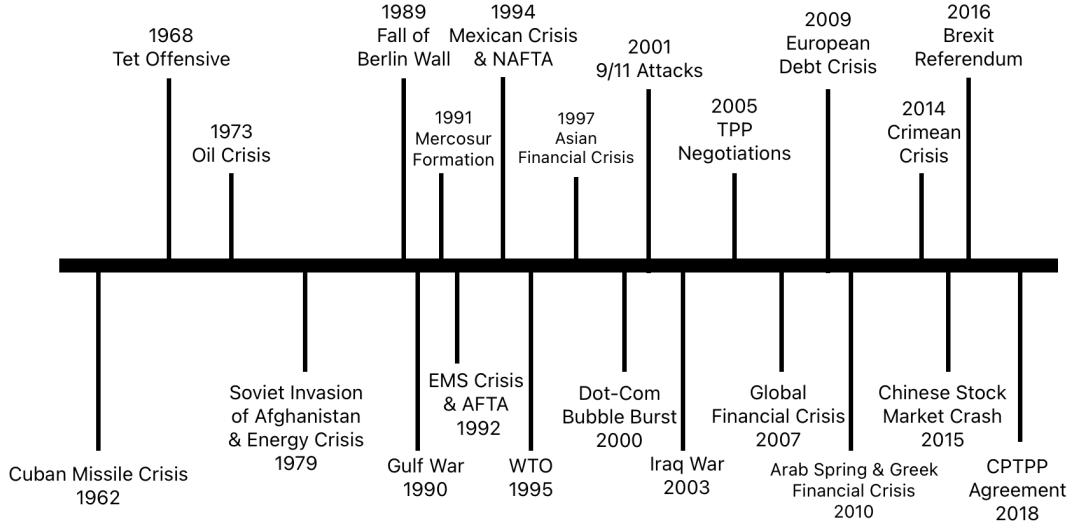
Table 2: SITC S4 categories with classifications as either majority manufactured products or raw materials.

5 Results

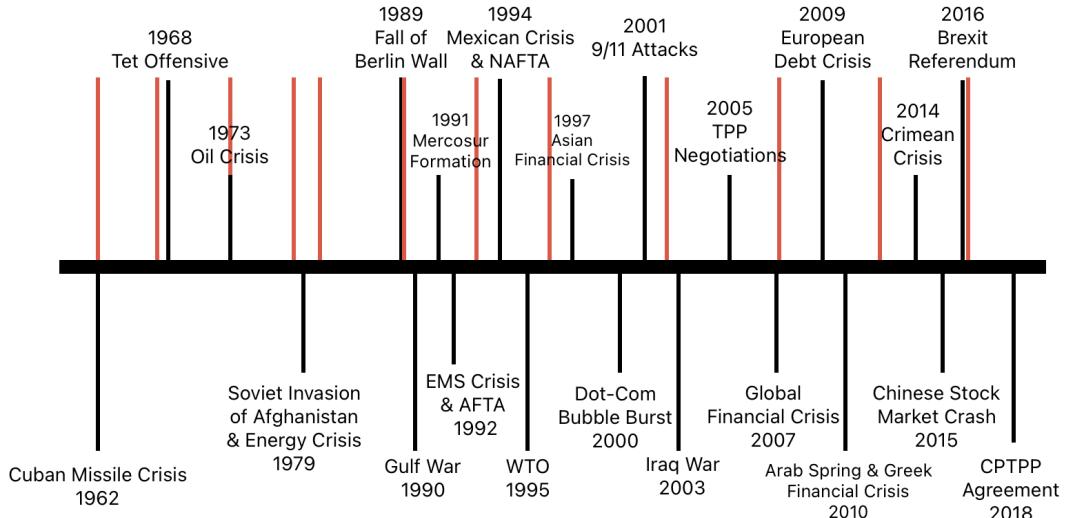
The results are broken into two main sections - change-point prediction and s-GNN application. Within the model results, the performance for both the traditional and s-GNN approaches are detailed and then compared. The analysis section then describes the results when the best change-point detection model is employed on the region and product subnetworks.

5.1 Uncovering of Ground-Truth Economic Shocks

Fig. 5a displays the timeline of events created using only domain knowledge and prior literature of financial crises, trade agreements, and major geopolitical events. 22 events are identified for the years 1962 through 2018. Fig. 5b displays the timeline of ground-truth change-point events as determined by the combination of the domain-knowledge years and years identified by the clustering algorithm (Fig. 2) for both the raw export values and the export percentage values. 12 change-points are identified, with four of the events aligning exactly with the year determined through domain-knowledge and the remaining eight being within one year of a known event, confirming the success of the clustering method in identifying major economic events. Fig. A.1 and A.2 display the individual timelines for the years using either the raw or percentage export values with the spectral clustering algorithm. The algorithm for the export percentages produced



(a) Domain knowledge



(b) Ground truth years from ensemble method

Figure 5: Comparison of years identified through domain knowledge (5a) and through the ensemble method (5b) which combined the years from domain knowledge and the clustering algorithm with raw and percentage export values. In 5b, the red lines display the years identified by the clustering algorithm and are used as the ground truth years for the remaining experiments.

four more change-points than the raw export values, perhaps pointing to the increased ability for the export percentages to pick up on changes in activity associated with smaller economies.

5.2 Performance of Traditional Distance Metrics

To first address whether graph machine learning methods perform better than traditional network methods in change-point detection, baseline performance metrics for the traditional methods must be determined. Fig. 6 displays the identified change-points from each of the traditional network similarity metrics with a similarity threshold of 0.5. Table 3 displays their associated F1 and ARI scores. The F1 score measures a combination of precision and recall to determine how well the two classes (change-point vs. non change-point) are predicted. The ARI score measures how similar the generated partition is to the ground-truth partition, where a value closer to 0 indicates the generated partition is closer to random. DeltaCon achieved the highest F1 score (0.39) while CUSUM achieved the best ARI score (0.53). Frobenius and Weisfeiler-Lehman both have an F1 score of 0.00 since neither of these methods were able to correctly identify any of the ground-truth change-points.

Method	F1 Score	ARI Score
NCPD	0.17	0.48
LAD	0.33	0.34
CUSUM	0.36	0.53
CUSUM 2	0.23	0.32
DeltaCon	0.39	0.18
Frobenius	0.00	0.08
WL	0.00	0.19

Table 3: F1 and ARI Scores for traditional network-based change-point detection methods using threshold of 0.5.

Method	F1 Score	ARI Score
NCPD	0.38	0.35
LAD	0.39	0.12
CUSUM	0.43	0.46
CUSUM 2	0.36	0.21
DeltaCon	0.45	0.31
Frobenius	0.39	0.17
WL	0.47	0.44

Table 4: F1 and ARI Scores for traditional network-based change-point detection methods using best-identified threshold value.

Fig. 7 displays the detected change-points from the traditional methods when using the best identified similarity score threshold. Table 4 displays the F1 and ARI scores for these metrics. Weisfeiler-Lehman (WL) achieved the highest F1 score (0.47) while

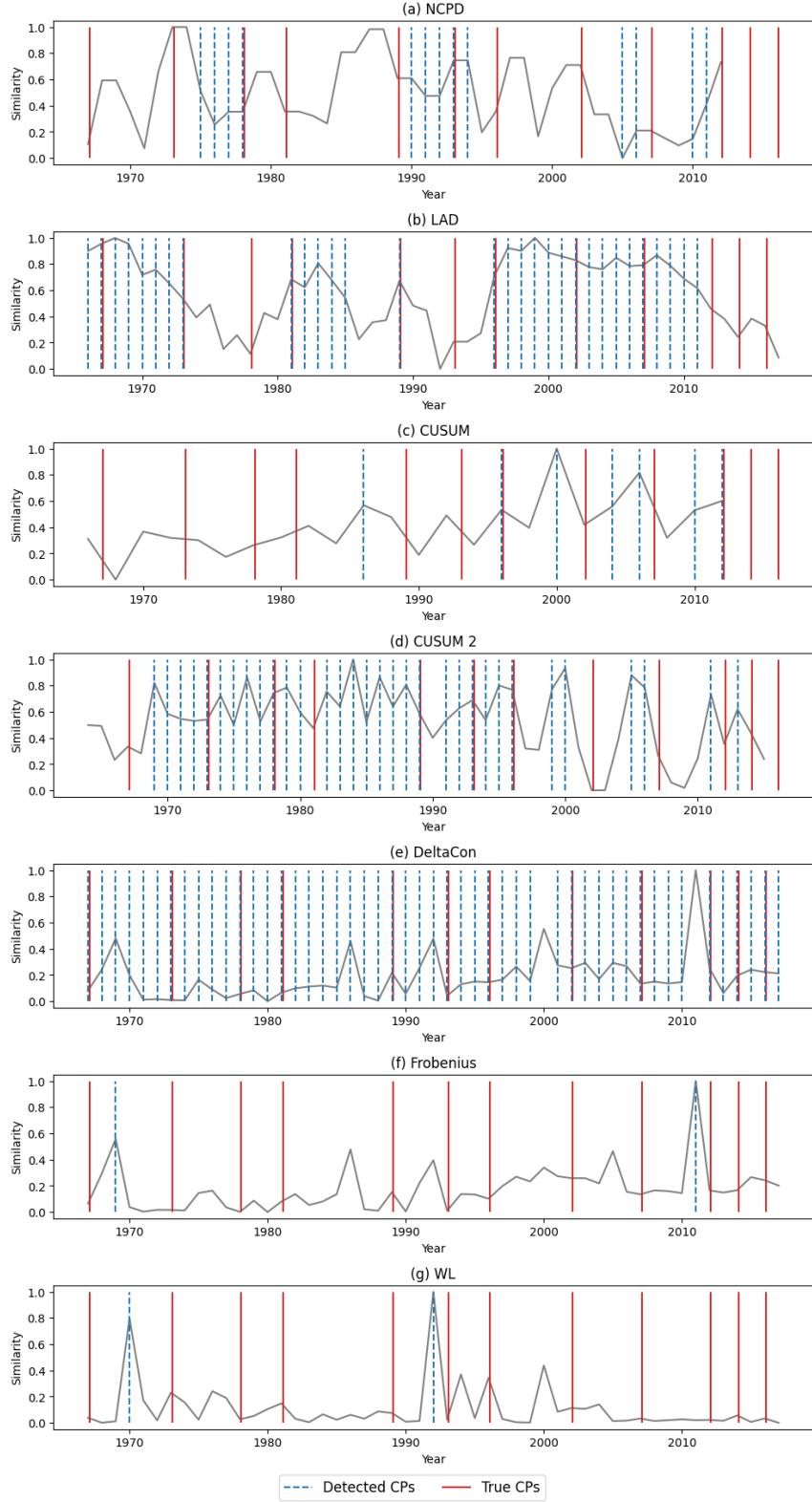


Figure 6: Identified change-points from 1966 to 2018 using the traditional network distance metrics with a threshold of 0.5. The red, solid bars indicate the ground-truth economic shocks, and the blue, dashed lines indicate those detected by the methods.

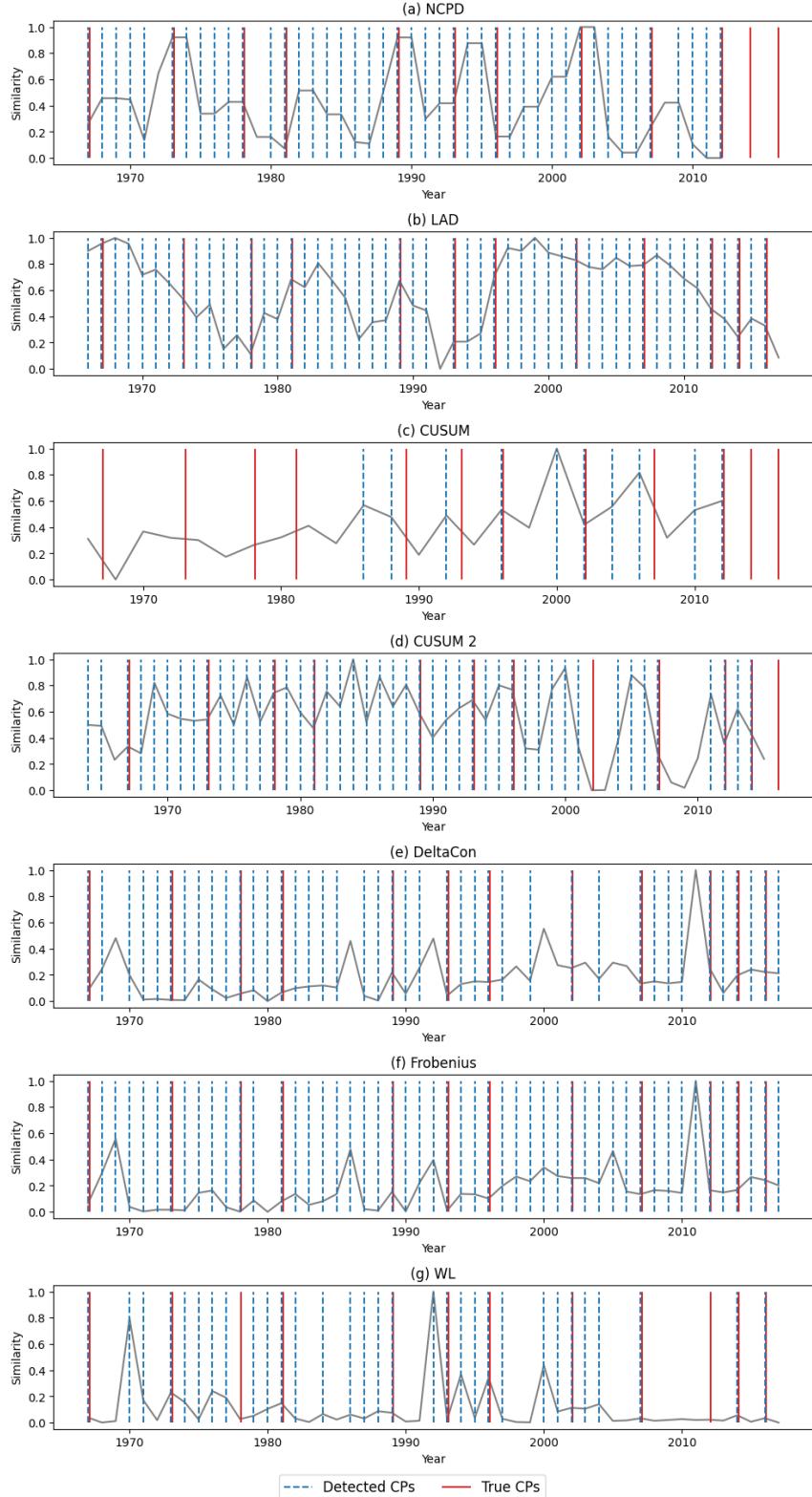


Figure 7: Identified change-points from 1966 to 2018 using the traditional network distance metrics with the best-identified threshold. The red, solid bars indicate the ground-truth economic shocks, and the blue, dashed lines indicate those detected by the methods.

Encoder	Features	Accuracy	F1	F2	F0.5
GCN	Exports	0.600	0.666	0.718	0.622
	MIS	0.544	0.666	0.770	0.589
	Random	0.558	0.677	0.777	0.601
GraphSAGE	Exports	0.786	0.799	0.800	0.800
	MIS	0.590	0.723	0.860	0.626
	Random	0.597	0.667	0.715	0.625
GIN	Exports	0.527	0.684	0.829	0.582
	MIS	0.530	0.673	0.801	0.581
	Random	0.536	0.697	0.851	0.591
GAT	Exports	0.751	0.787	0.810	0.774
	MIS	0.629	0.690	0.750	0.657
	Random	0.542	0.701	0.853	0.594

Table 5: Comparison of metrics for GCN, GraphSAGE, GIN, and GAT encoders with the inclusion of GDP growth variables with non-normalized validation data. The best score for each metric is bolded.

CUSUM achieved the highest ARI score (0.46). All of the F1 scores increased in value compared to original threshold of 0.5 (3). ARI scores, however, decreased for NCPD, LAD, CUSUM, and CUSUM 2 while they increased for DeltaCon, Frobenius, and WL. With the threshold of 0.5, there is some discretion in the change-points that most methods identify. Yet, with the adjusted threshold, the methods, with the exception of CUSUM, are far less specific in their identification. Methods like NCPD, LAD, and Frobenius identify almost every year as a change-point, which is not practical to the task even if they identify some of the years correctly.

5.3 Performance of s-GNN Models

To compare the traditional and graph machine learning approaches, the best s-GNN model must be identified. The results first describe the chosen model architecture based on validation dataset performance. The three best performing models are then compared through test set performance to identify the final model. Using this final model, change-point detection is implemented and compared to the traditional methods' results.

Table 5 displays the average metrics of 5-fold cross validation for the four different embedding methods with each of the three different feature segments with the inclusion of GDP features. These models were trained with the following parameters: learning rate of 0.01, dropout rate of 0.05, top-k of 50, and 16 hidden units. Across all metrics, the GraphSAGE encoder performs the best. Additionally, for three of the four metrics, the feature set only containing centrality information for the 10 product categories performed

Encoder	Features	Accuracy	F1	F2	F0.5
GCN	Exports	0.584	0.655	0.707	0.614
	MIS	0.558	0.676	0.782	0.599
	Random	0.569	0.670	0.752	0.607
GraphSAGE	Exports	0.587	0.698	0.805	0.618
	MIS	0.794	0.812	0.855	0.786
	Random	0.595	0.674	0.736	0.622
GIN	Exports	0.523	0.687	0.825	0.583
	MIS	0.531	0.677	0.799	0.587
	Random	0.530	0.696	0.855	0.590
GAT	Exports	0.558	0.630	0.677	0.593
	MIS	0.560	0.825	0.744	0.594
	Random	0.554	0.673	0.771	0.599

Table 6: Comparison of metrics for GCN, GraphSAGE, GIN, and GAT encoders with non-normalized validation data without the inclusion of GDP growth variables. The best score for each metric is bolded.

the best. The GAT encoder with no additional features also performed similarly to GraphSAGE with all metrics above 0.750.

When trained without prior and current GDP growth variables (Table 6), for three of four metrics, GraphSAGE again performs the best, however, the MIS feature set now achieves the highest metrics of the three feature sets. This is most likely due to this feature set still accounting for information related to GDP growth. The performance of the MIS feature set without GDP growth features and with the GraphSAGE encoder actually outperforms the feature set with GDP growth features with an accuracy now of 0.794 compared to 0.590. For the GCN and GIN encoders, there is not much difference in performance between the feature sets with and without GDP growth features. Performance decline is seen with the GAT encoder, with accuracy, F2, and F0.5 scores all decreasing.

Table 7 displays the results of feature sets with GDP growth features but after standardization of the features. For the MIS and random feature sets, the GraphSAGE, GCN, and GAT encoders see a performance increase. Across all encoders, the MIS feature subset now performs the best. The GraphSAGE encoder still achieves the best performance and now across all metrics. All of its values are now above 0.860 compared to 0.780 without standardization. After removing the GDP growth variables (Table 8), however, the performances mostly decline to similar values as in Table 6. GraphSAGE with MIS features, however, maintains its performance and actually sees a slight increase compared to Table 7. Its metrics are now above 0.890.

Table 9 displays the results of the model with GDP features and with normalization

Encoder	Features	Accuracy	F1	F2	F0.5
GCN	Exports	0.608	0.668	0.715	0.628
	MIS	0.810	0.858	0.870	0.847
	Random	0.795	0.818	0.845	0.793
GraphSAGE	Exports	0.748	0.773	0.792	0.759
	MIS	0.863	0.904	0.915	0.895
	Random	0.850	0.869	0.903	0.838
GIN	Exports	0.517	0.660	0.779	0.572
	MIS	0.535	0.675	0.799	0.585
	Random	0.516	0.671	0.804	0.576
GAT	Exports	0.774	0.792	0.791	0.801
	MIS	0.802	0.832	0.860	0.812
	Random	0.757	0.797	0.838	0.762

Table 7: Comparison of metrics for GCN, GraphSAGE, GIN, and GAT encoders with validation data normalized by standardization with the inclusion of GDP growth variables. The best score for each metric is bolded.

Encoder	Features	Accuracy	F1	F2	F0.5
GCN	Exports	0.567	0.656	0.725	0.603
	MIS	0.812	0.843	0.870	0.822
	Random	0.559	0.706	0.848	0.600
GraphSAGE	Exports	0.563	0.684	0.798	0.599
	MIS	0.895	0.915	0.942	0.893
	Random	0.630	0.726	0.811	0.665
GIN	Exports	0.528	0.683	0.825	0.582
	MIS	0.533	0.675	0.802	0.579
	Random	0.532	0.670	0.803	0.569
GAT	Exports	0.558	0.631	0.679	0.593
	MIS	0.694	0.723	0.736	0.714
	Random	0.620	0.702	0.760	0.658

Table 8: Comparison of metrics for GCN, GraphSAGE, GIN, and GAT encoders with validation data normalized by standardization without the inclusion of GDP growth variables. The best score for each metric is bolded.

Encoder	Features	Accuracy	F1	F2	F0.5
GCN	Exports	0.594	0.672	0.714	0.629
	MIS	0.762	0.787	0.809	0.770
	Random	0.519	0.628	0.717	0.562
GraphSAGE	Exports	0.781	0.803	0.796	0.805
	MIS	0.914	0.918	0.899	0.938
	Random	0.620	0.671	0.709	0.639
GIN	Exports	0.521	0.690	0.834	0.579
	MIS	0.533	0.669	0.795	0.586
	Random	0.539	0.705	0.849	0.593
GAT	Exports	0.748	0.791	0.806	0.770
	MIS	0.705	0.761	0.823	0.710
	Random	0.586	0.636	0.682	0.601

Table 9: Comparison of metrics for GCN, GraphSAGE, GIN, and GAT encoders with validation data normalized through logging with the inclusion of GDP growth variables. The best score for each metric is bolded.

Encoder	Features	Accuracy	F1	F2	F0.5
GCN	Exports	0.563	0.661	0.721	0.608
	MIS	0.589	0.705	0.820	0.620
	Random	0.526	0.684	0.829	0.582
GraphSAGE	Exports	0.568	0.679	0.793	0.604
	MIS	0.596	0.702	0.802	0.629
	Random	0.534	0.691	0.839	0.587
GIN	Exports	0.524	0.688	0.829	0.584
	MIS	0.537	0.672	0.798	0.575
	Random	0.529	0.665	0.807	0.565
GAT	Exports	0.553	0.636	0.674	0.598
	MIS	0.652	0.722	0.762	0.703
	Random	0.547	0.660	0.753	0.589

Table 10: Comparison of metrics for GCN, GraphSAGE, GIN, and GAT encoders with validation data normalized through logging without the inclusion of GDP growth variables. The best score for each metric is bolded.

Feature Set	Accuracy	F1	F2	F0.5	AUC
Standardized with GDP	0.843	0.875	0.854	0.897	0.850
Standardized without GDP	0.962	0.971	0.968	0.974	0.950
Logged with GDP	0.890	0.912	0.872	0.957	0.910

Table 11: Performance metrics on test data from random sampling scheme for top-three models.

Method	F1	ARI
sGNN-MIS-Norm-GDP	0.951	0.643
sGNN-MIS-Norm-No-GDP	0.994	0.820
sGNN-MIS-Logged-GDP	0.970	0.862

Table 12: Average F1 and ARI Scores over 10 iterations for s-GNN change-point detection methods using using equation 5 and threshold of 0.5 with data from random sampling scheme.

by logging. The MIS feature sets for the GCN and GAT encoders see a slight increase compared to the non-normalized features, but they do not achieve the same performance as the standardized features. The MIS features with the GraphSAGE encoder actually increases in performance compared to the normalized features, with all metrics above 0.899. When GDP growth variables are removed from this model (Table 10), unlike previously, all feature sets now perform similarly, and GraphSAGE no longer performs the best. GAT achieves the best performances for three out of four metrics, but the metrics are now only above 0.65. Out of all model iterations, these results show the lowest performance even compared to the non-normalized feature sets without GDP used in Table 6.

Overall, GraphSAGE and MIS features tend to provide the highest metrics, although GAT performs similarly with features normalized by standardization. With the validation set performance, the model with logged features achieved the highest metrics. This model was subsequently used for hyperparameter tuning – these results are shown in Fig. A.4. From this hyperparameter tuning, the following parameters were used for final testing of the models: lr of 0.001, sort-k of 50, and 16 hidden units.

The top-three best performing combinations of features and normalization methods are standardization with MIS and GDP features, standardization with MIS and no GDP features, and logging with MIS and GDP features. These three feature sets and the GraphSAGE encoder were then tested on the held-out test set (Table 11). The model using standardized features without GDP performs the best with all metrics above 0.95, and actually sees higher performances compared to the validation set. The results of using the same three models for change-point detection task with equation 5 are shown

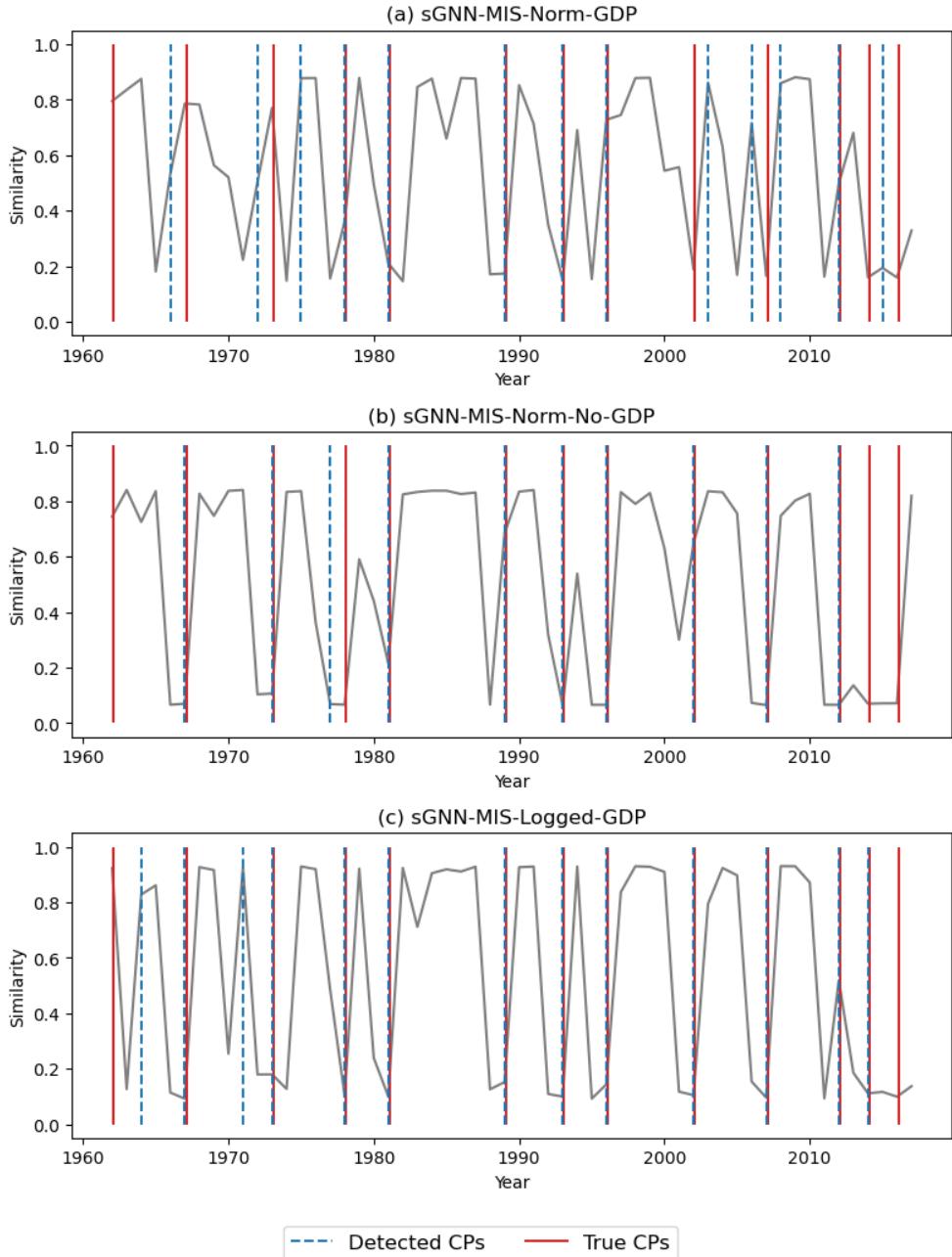


Figure 8: Identified change-points of the three best performing s-GNN models. The blue, dashed lines indicate the change-points detected by the specified model, and the red, solid lines indicate the ground-truth change-points.

in Table 12. Although the model with logged features achieves the highest ARI value of 0.862, the model with standardized features and no GDP features actually achieves the highest F1 score of 0.994. The associated change-points with each of the models are displayed in Fig. 8. Because the model using standardization and no GDP features performed the best on the test data and achieved the highest ARI value, this model is selected as the final model to be applied to RQs 2 and 3.

5.3.1 Window-Sampling Method

The results of the traditional metrics when applied to only the subset of years used in the test set for the window sampling scheme are shown in Table 13. These metrics now see improved performance compared to the random sampling setting given the much shorter time period and fewer change-points to detect, with CUSUM achieving maximum F1 and ARI scores of 1.00.

The test-set results of the top-three s-GNN models trained using the window sampling scheme are shown in Table 14. The model with logged features achieves the highest performance across four of the five metrics; however, the results are significantly lower than that of the random sampling method. The highest F1 score is now 0.658, which is 0.31 percentage points lower than the best F1 from the random sampling results. The highest AUC is also 0.36 percentage points lower.

Method	F1 Score	ARI Score
NCPD	0.00	0.66
LAD	0.33	0.24
CUSUM	1.00	1.00
CUSUM 2	0.40	0.75
DeltaCon	0.67	0.88
Frobenius	0.00	0.28
WL	0.80	0.62

Table 13: F1 and ARI Scores for traditional network-based change-point detection methods using a threshold of 0.5 and selection of years used in window sampling (2007-2018).

When applied to the change-point detection task, the model with standardized features and no GDP again achieves the highest F1 score of 0.861, but the model with logged features achieves the best ARI score of 0.607 (Table 15). These are again much lower than that of the random sampling scheme, however, because of the test set partition, they are only trying to predict four change-points rather than thirteen. The associated change-points detected from the window sampling models are shown in Fig. 9. The model with logged features exactly identifies two of the change-points, while the other models identify

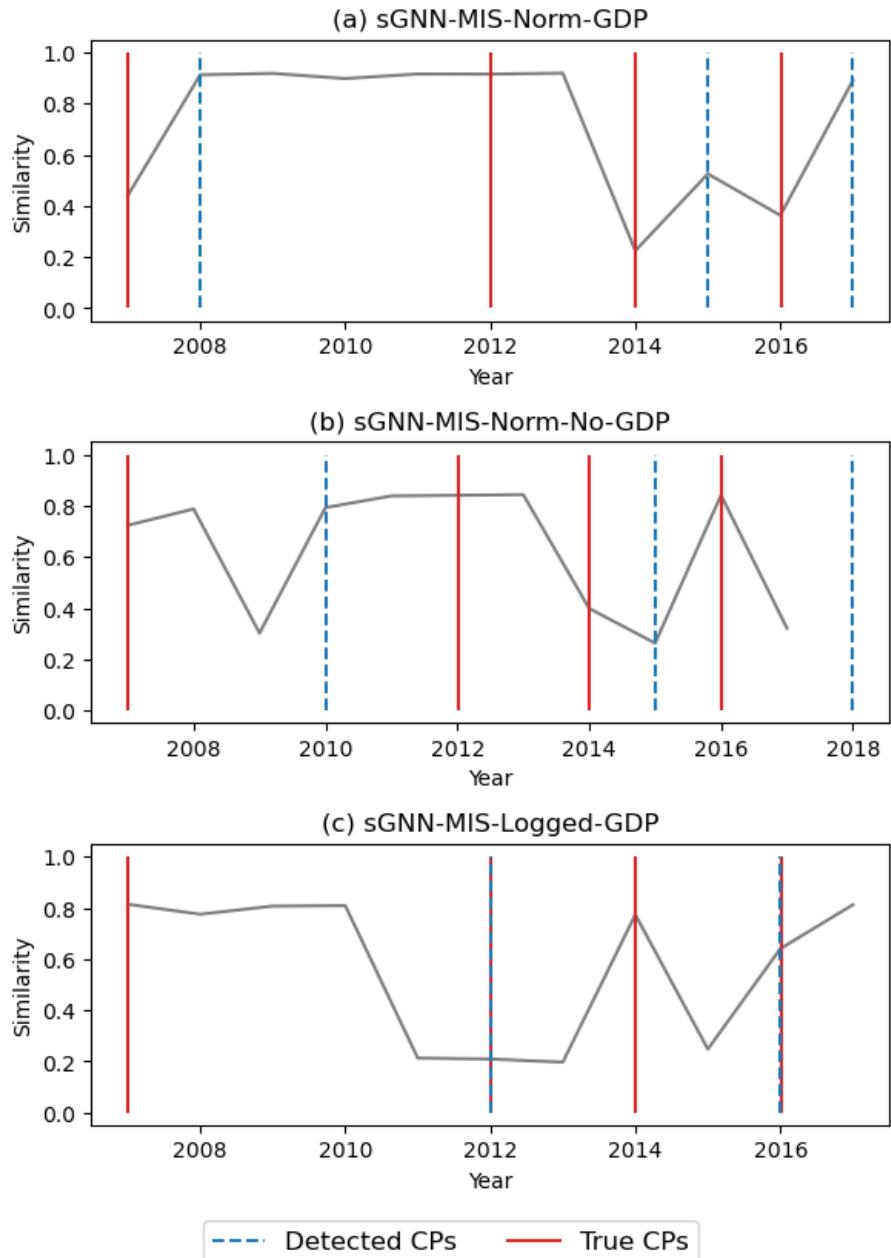


Figure 9: Identified change-points of the three best performing s-GNN models on test data using the window sampling method. The blue, dashed lines indicate the change-points detected by the specified model, and the red, solid lines indicate the ground-truth change-points.

Feature Set	Accuracy	F1	F2	F0.5	AUC
Standardized with GDP	0.545	0.651	0.745	0.579	0.540
Standardized without GDP	0.530	0.587	0.621	0.556	0.530
Logged with GDP	0.591	0.658	0.718	0.607	0.590

Table 14: Test set performance for top-3 models using window sampling with MIS-selected feature variants. Window samples use first 60% of years for train, next 20% for validation, and final 20% for testing.

Method	F1 Score	ARI Score
sGNN-MIS-Norm-GDP	0.860	0.248
sGNN-MIS-Norm-No-GDP	0.861	0.203
sGNN-MIS-Logged-GDP	0.842	0.607

Table 15: Average F1 and ARI Scores over 10 iterations for s-GNN change-point detection methods using equation 5 and threshold of 0.5 with test data from window sampling scheme.

years three years near the ground-truth years. Although this model performs relatively well, CUSUM achieves higher F1 scores, and the s-GNN’s ARI score only outperforms LAD and Frobenius.

5.4 Synthetic Experiments

Fig. 10 displays the results of testing the traditional and s-GNN change-point methods on the synthetic stochastic block model networks. At almost every point the s-GNN methods significantly outperform the traditional network-distance methods. The only exception is at clique size 80 where CUSUM 2 achieved a perfect F1 of 1.0, which is slightly higher than the 100-k s-GNN F1 score of 0.95. For most of the F1 scores, the traditional methods are below 0.10. This supports the results using the real-world data that the deep learning s-GNN methods perform better at change-point detection than network distances. Additionally, we generally see that increased clique size and increased k value are associated with higher F1 scores. Not only can these F1 scores be used as a robustness check to ensure that the model’s performance is not specific to the real-world data, but these values can also be used as benchmarks which the models should outperform with the real-world data since the SBM did not have any inherent node features to draw information from unlike the real-world data regarding country attributes.

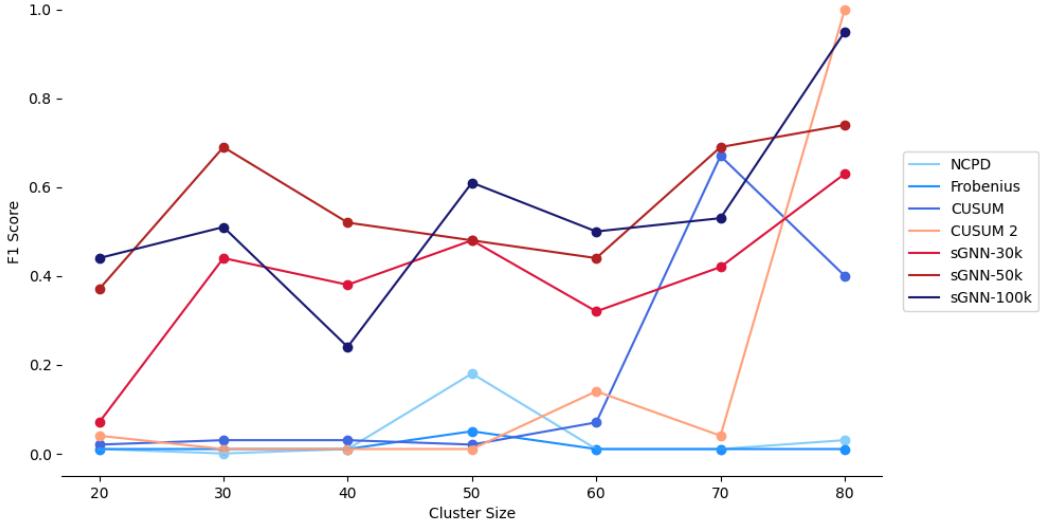


Figure 10: F1 scores averaged over 10 iterations for application of traditional network-distance and s-GNN change-point methods to synthetic clique data. Three top-k values for the s-GNN were tested, including 30, 50, and 100. For each method, they were tested with clique sizes 20, 30, 40, 50, 60, 70, and 80.

5.5 Analysis of Detected Change-Points

The next section will detail the identified change-points by region and product subnetworks when applying the trained s-GNN. This only represents an initial analysis of these results as examples of how the s-GNN change-point detection model could be applied.

5.5.1 Analysis by Region Network

Fig. 11 displays the detected change-points broken down by the six regions. No region has the same change-points, providing evidence that the method can be used for this level of analysis. North America’s change-points most closely align with the original, ground-truth economic shocks (ARI of 0.787), while Oceania’s change-points are the least aligned (ARI of 0.431). The model did not identify the year 2007 as a change-point, which is supposed to be associated with the Global Financial Crisis. Although the Great Recession did impact this region, the impact was less severe compared to North America and Europe, with countries like Australia experiencing an economic downturn but not at the magnitude of a recession (Edwards, 2010). Instead, Oceania indicates both 2004 and 2006 as economic shocks. Fiji did experience heightened geopolitical events in the early 2000s, including the 2006 Fijian coup d’état (Fraenkel & Firth, 2007). However, it is unclear exactly what the 2004 change-point would be an indication of.

Notably, the 1973 Oil Crisis is detected in every region’s timeline. This is plausible as this was a large-scale economic shock that impacted an essential product with very

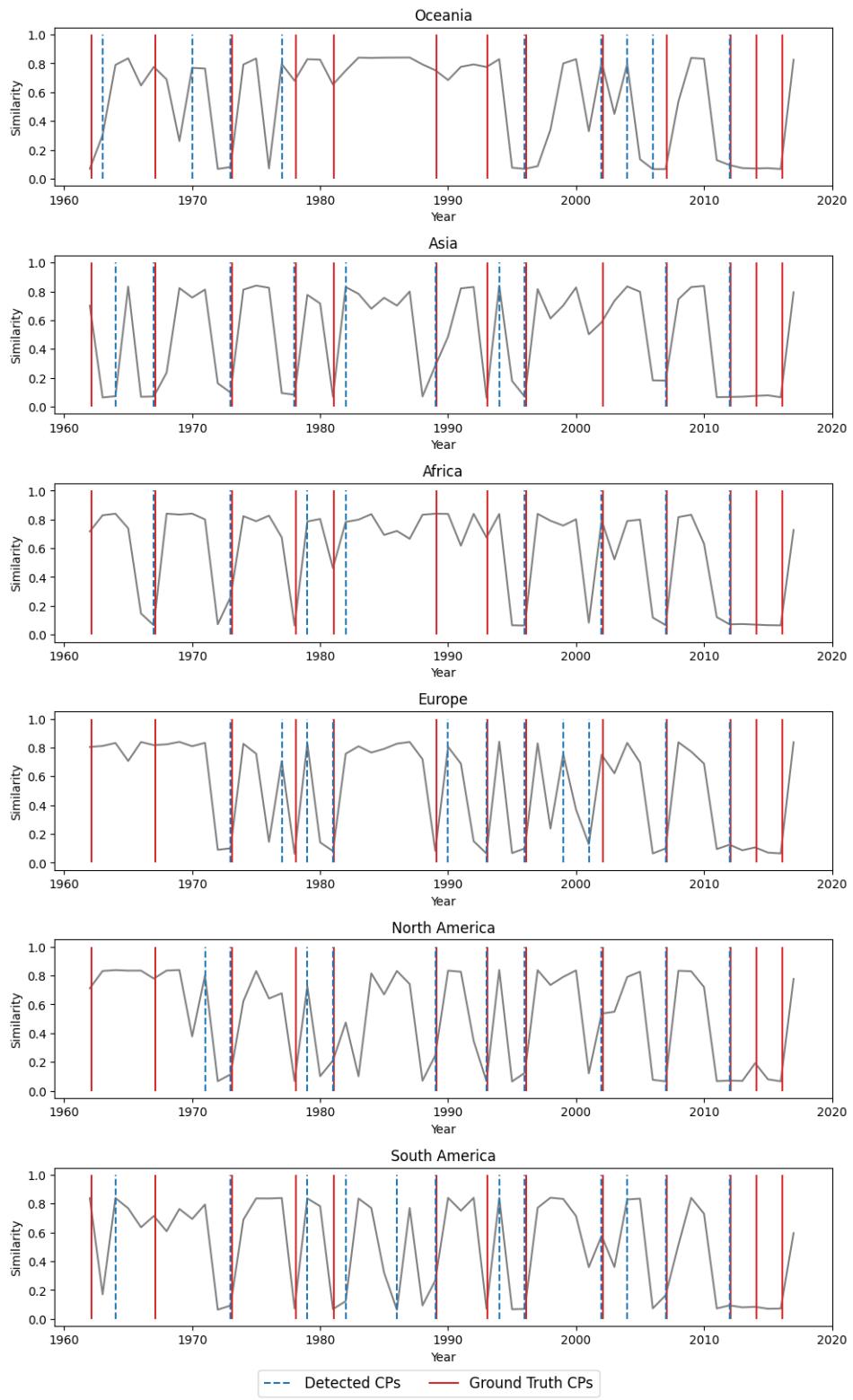


Figure 11: Detected change-points for each region using the MIS-Norm-No-GDP s-GNN model.

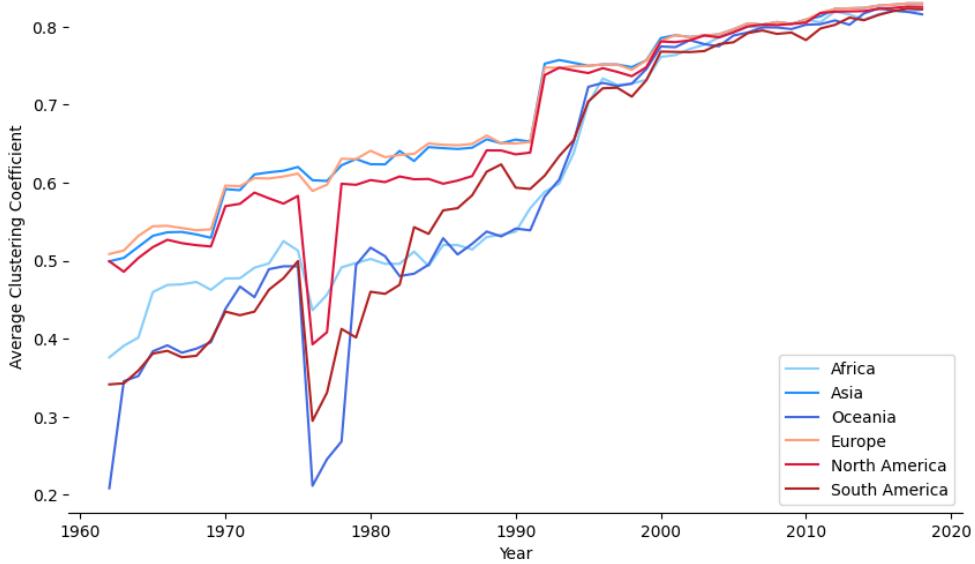


Figure 12: Average clustering coefficient for each region for each network from 1962 to 2018.

direct trade impacts. Even though the countries most directly impacted included the US, UK, Canada, Japan, and the Netherlands, the shock to oil had rippling effects on other regions' cost of shipping, trade balances, and trade policies ([Issawi, 1978](#)).

The years of Asia's economic shocks are relatively similar to North America's. The one exception is the lack of an economic shock in the early 2000's. North America and Europe felt a recession around 2000 and 2001, but Asia was still recovering from the 1997 Asian Financial Crisis, so the early 2000's recession did not have as much of a direct impact ([Das, 2012](#)).

Africa does not show many economic crises during the 1980's and 1990's, even though this is known to be a very turbulent time for the region especially regarding the end of apartheid. It would be expected for there to be a detected shock around the late 1980's when governments adopted trade sanctions against South Africa ([Levy, 1999](#)).

Europe, uniquely, has four identified shocks from 1990 to 2000. This could be associated with the formal collapse of the Soviet Union in 1991 and subsequent general economic upheaval in the following years especially in regard to the changes in trade among Eastern European countries ([Broadman, 2006](#)). Throughout the mid-1990's, Russia was also signing trade agreements, which could present themselves as economic shocks.

South America is shown to have three economic shocks between 1980 to 1990. This would align with the Latin American Debt Crisis of the 1980's, stemming from the 1973 Oil Crisis. There is a ground-truth label at 1981, but South America instead has a detected point at 1982, which could be associated with August of 1982 which was deemed the start of the crisis ([Sims & Romero, 2013](#)).

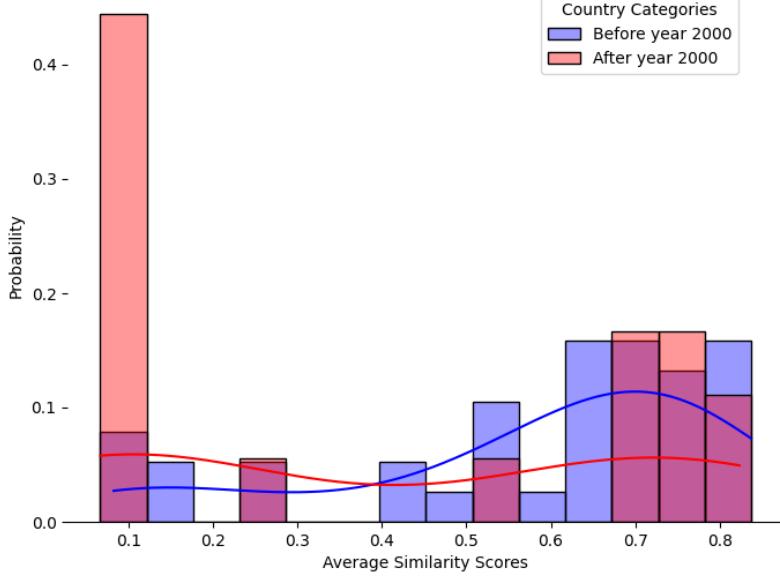


Figure 13: Distribution of average similarity scores across regions before and after the year 2000. The year 2000 was decided based on Fig. 12, where this year indicated a convergence of all regions towards high connectedness.

Fig. 12 displays the average clustering coefficient over the years 1962 to 2018. As supported by prior literature, the level of connectedness increases across most years but converges for all regions around the year of 2000. When looking at the average similarity scores separated by those before and after the year 2000 (Fig. 13), it is apparent that after 2000, regions seem to have a higher probability of crisis. For years before 2000, the values are more skewed towards higher similarity scores, indicating lower probability of an economic shock.

5.5.2 Analysis by Product Network

The following results are for the subgraphs separated by product category. Fig. 15 displays the detected change-points for the different categories of raw materials, while Fig. 16 displays those for the manufactured products. The change-points for product category 1 (beverages and tobacco) align the least with the ground-truth economic shocks (ARI of 0.504) while those for product category 6 (manufactured goods classified by material) align most closely (ARI of 0.761). Therefore the manufactured products are one of the categories with the fewest number of shocks, and those present align very closely with the known ground-truth change-points.

Beverages and tobacco show a large number of shocks during the 1970's and early 1980's. It is not clear what exactly this could be associated with since there were competing forces impacting the tobacco industry during this time. OECD countries began

implementing health awareness campaigns and advertising restrictions in 1973, directly negatively affecting tobacco consumption ([Laugesen & Meads, 1991](#)). Yet, Brazil, Zimbabwe, and other countries in Africa and Asia continued to expand their exporting of tobacco, creating ambiguous net effects on the product.

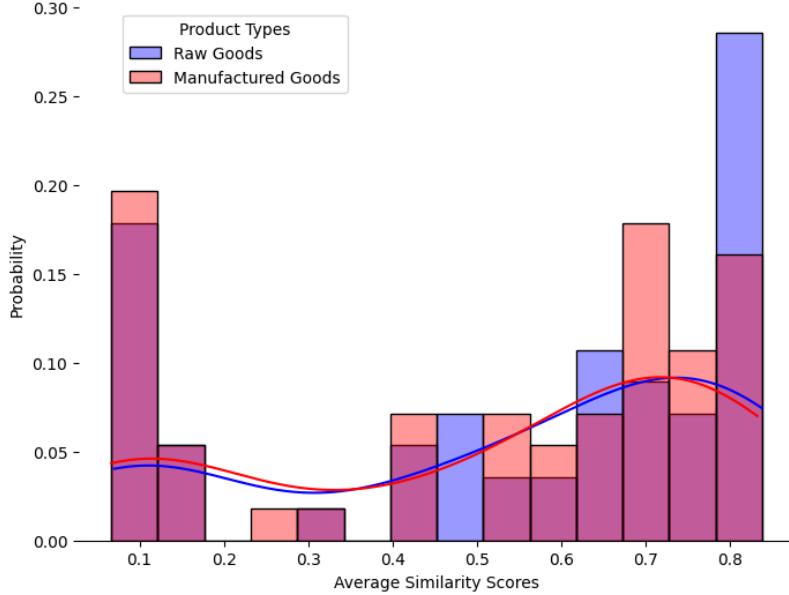


Figure 14: Distribution of similarity scores by product type. The solid lines are the kernel density estimators for each group.

Food and live animals shows 14 detected crises, the highest of all product categories. The small-world property of the food trade network is associated with highly connected countries. This property could be a conduit for economic shocks because a disruption by a central country can be felt throughout the rest of the network ([Ji et al., 2024](#)).

Oil is contained in the category 3 (mineral fuels, lubric. and related mtrls), and does show change-points at 1973 and 1975 which were the exact years of the 1973 Oil Crisis. This can serve as a robustness check to ensure that the model is correctly identifying crisis periods within these networks.

Fig. 14 displays the probability of average similarity scores broken down by raw vs. manufactured goods. The raw goods display almost twice the probability of having a pair of extremely similar networks, with the manufactured goods having a slightly higher probability of having a pair of two very distinct networks. When smoothed, however, the kernel density estimator shows that no significant difference in the distributions between the two groups of product categories.

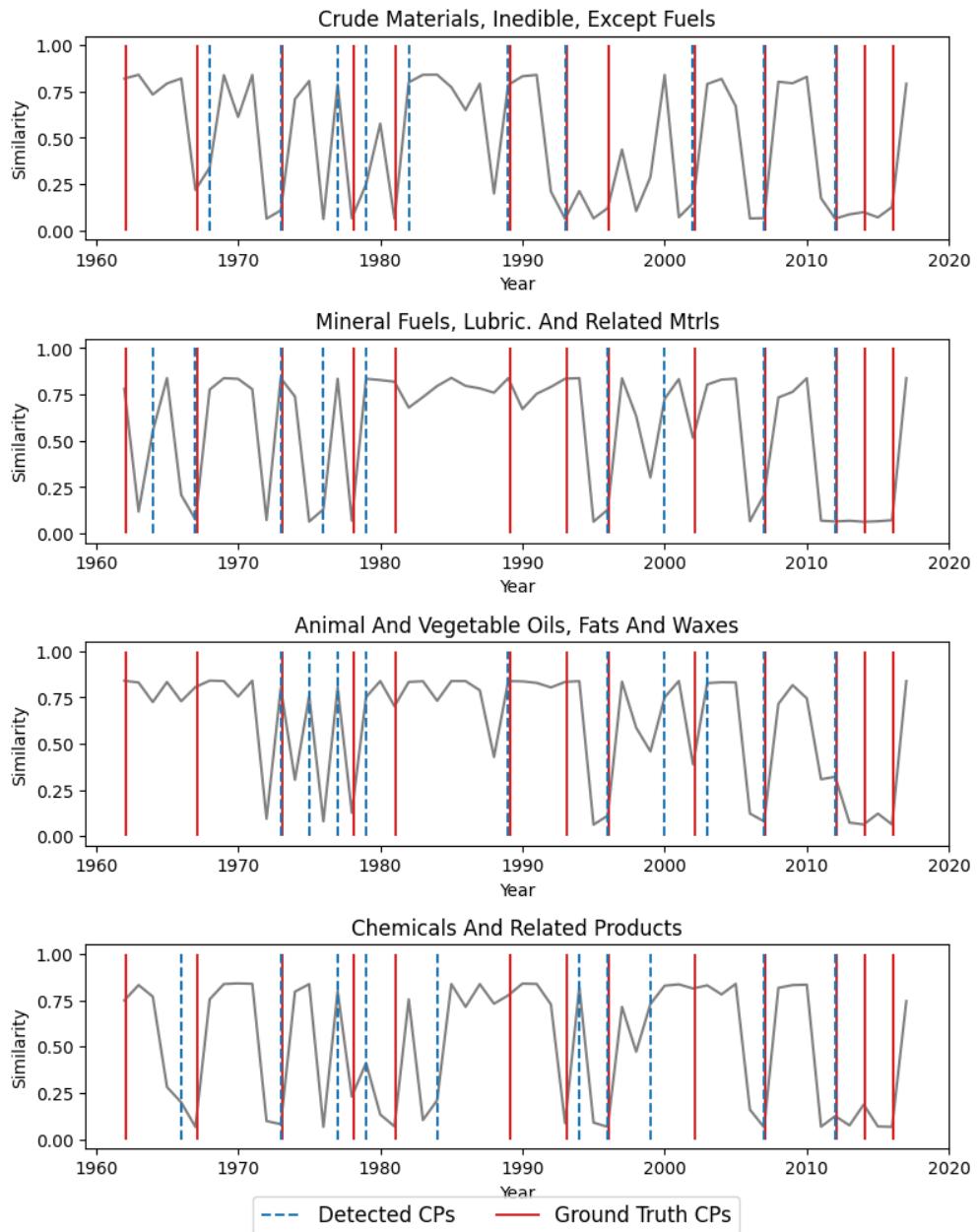


Figure 15: Detected change-points using the MIS-Norm-No-GDP s-GNN model for products classified as raw materials.

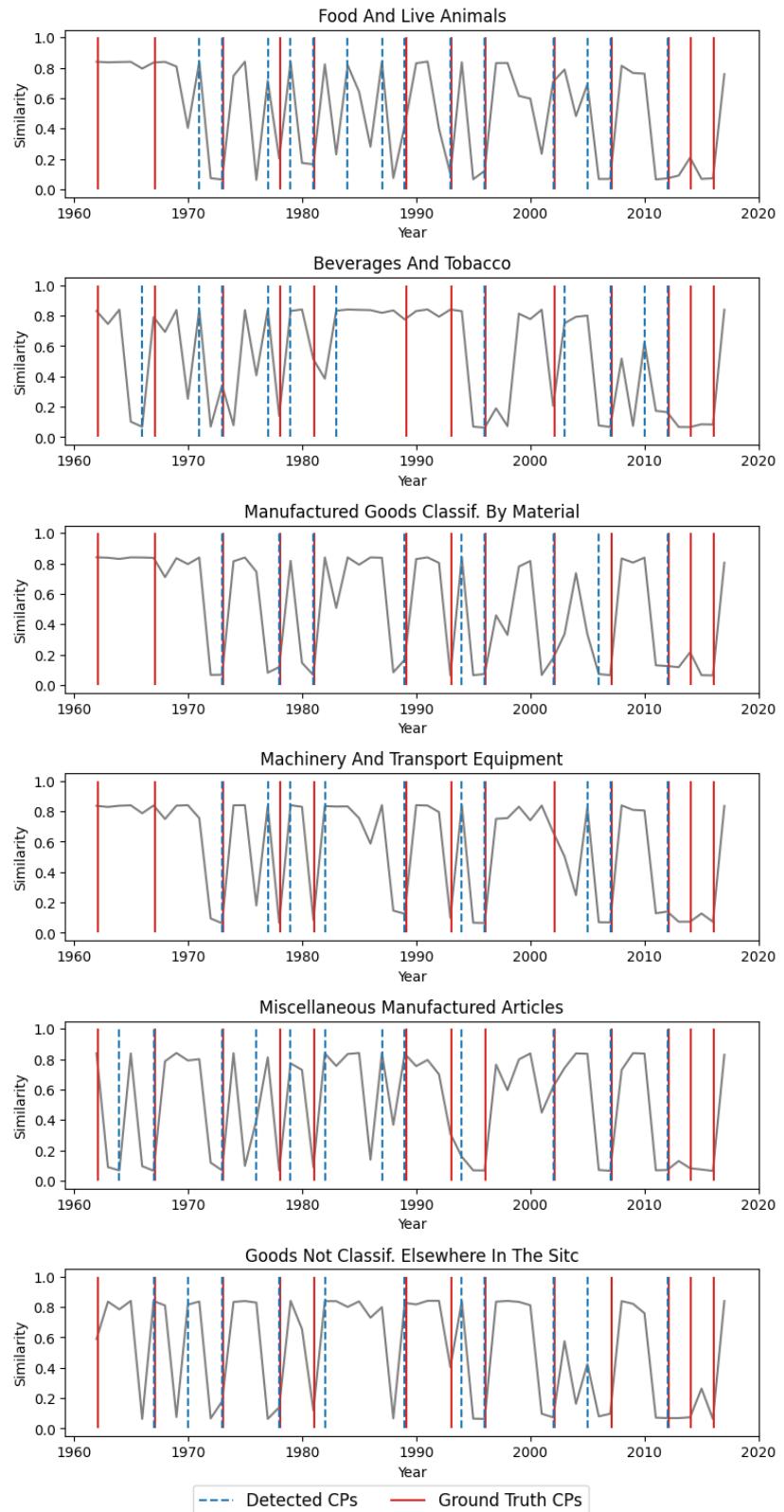


Figure 16: Detected change-points using the MIS-Norm-No-GDP s-GNN model for products classified as manufactured goods.

6 Discussion

This thesis has explored the feasibility of applying graph machine learning methods to analyze economic shocks' presence in trade networks. The following discussion will evaluate the success of the hypotheses, describe the novel contributions of this work, and suggest directions for future research in this area.

6.1 Success of Graph Machine Learning

The results presented above demonstrate the effectiveness of graph machine learning in offline change-point detection, particularly when compared to traditional network-distance based methods, providing a remarkable 0.524 percentage point increase in F1 score. This methodology's strength ultimately lies in its discriminative ability, enabling the identification of specific years associated with economic shocks, whereas traditional metrics tend to identify an excessive number of change points.

Additionally, specific architecture and feature subsets were proven to be more beneficial over others. The GraphSAGE encoder consistently provided the best performances across a range of metrics, while the GIN encoder consistently performed the worst. This points to the importance of the incorporation of node and edge feature values. While the GIN encoder can perform well for analyzing changing network structures, some networks, like the ITN, change edge values more than the presence of edges.

The model's performance was enhanced by features selected based on their correlation with GDP, highlighting the importance of GDP-related indicators. This held true regardless of whether GDP growth rates were directly included in the model. However, the literature presents conflicting views on the predictive power of GDP for economic shocks. For instance, [Hellwig \(2021\)](#) found limited value in variables such as real exchange rate and GDP growth rate when using random forest or gradient boosted tree models. In contrast, [L. Liu et al. \(2022\)](#) reported that numerous variables related to GDP growth demonstrated statistical significance. These divergent findings underscore the complexity of economic shock prediction and the potential value of my approach in leveraging GDP-related features within a graph-based framework.

While the primary focus of this research is the use of s-GNNs for offline change-point detection, online change-point detection was also briefly explored. The results suggest potential applications of s-GNNs in forecasting future economic shocks, an important area of study. This task was introduced through the window sampling method and the model with logged features was able to exactly identify two of the change-points; however, the model's performance decreased significantly compared to the offline task with random sampling. This decline is primarily due to the limited number of samples available when

using annual data. Additionally, the model was outperformed by most of the traditional change-point detection methods. Future research could focus more on the prediction task, potentially utilizing monthly or quarterly data for improved performance of the s-GNN in this setting. The applicability of this method is further supported by other studies that have successfully used GNNs for GDP and trade partner prediction ([Monken et al., 2021](#); [Panford-Quainoo et al., 2019](#)).

6.2 s-GNN-Based Change-Points for Economic Analysis

The hypothesis concerning the increased probability of economic shocks with heightened globalization found initial support through the analysis of detected change-points in regional subnetworks. The data revealed a distinct increase in the probability of extremely low similarity scores for regions post-2000 compared to the pre-2000 period. This finding provides compelling evidence that the growing interconnectedness among nations may indeed contribute to an increased vulnerability to economic spillover effects. The detected change-points in regional subnetworks suggest that local economies have become more susceptible to external influences, potentially leading to more severe or more frequent economic disruptions.

However, the hypothesis regarding the difference in the probability of economic shocks between raw and manufactured goods was not supported by the data. Initially, it was expected that manufactured goods would show fewer detected change-points, indicating greater stability. Contrary to this expectation, the analysis revealed inconsistent patterns across different categories of manufactured goods. Surprisingly, the category with the highest number of change-points belonged to the manufactured goods sector. This unexpected result challenges the assumption that processed goods inherently possess greater economic stability than raw materials.

Further evidence of this came from the analysis of overall probabilities of similarity scores. The data indicated that raw goods tended to exhibit similar year-to-year similarity as manufactured goods. This finding contradicts prior research, which has shown that products closer to consumers weather economic shocks better due to their perceived higher market value and more stable demand ([Bartik et al., 2020](#); [Jaarsma et al., 2017](#); [Szczygielski et al., 2022](#)).

Yet, when looking at individual categories, the patterns of identified change-points are supported by prior literature. A primary sector studied through network analysis includes food due to great importance these goods play to human well-being. [Fair et al. \(2017\)](#) specifically studied the global wheat trade network with respect to severe weather events. Through modeling the network as a preferential attachment model, they concluded that the wheat network is becoming less vulnerable to attacks. When looking at Fig. 16, the

food and live animals timeline clearly shows a reduction in the frequency of economic shocks over time, supporting the conclusions by [Fair et al. \(2017\)](#).

These two research questions were presented as applications of the s-GNN model, demonstrating its potential use cases in economic analysis. However, it is crucial to recognize that these questions and hypotheses only begin to scratch the surface of the complex economic dynamics at play. There are far more factors that should be explored while using s-GNNs as a tool for analysis.

Future research should better determine what events these change-points specifically point to. It is an assumption that the 2006 change-point could be related to the Fijian coup d'état, but this would have to be confirmed by isolating Fiji's subnetwork and analyzing the associated change-points. Additionally, a more granular analysis of different categories of manufactured goods could shed light on why certain sectors exhibit more change-points than others. The lack of detected change-points for Africa in the 80's is also a point of concern. This should be one of the main periods of change for the region due to the Anti-Apartheid Movement ([Levy, 1999](#)). More investigation into these specific subnetworks would be needed to understand why the model did not detect this period of change. Overall, while the detected change-points provide initial insights into the dynamics of economic shocks in a globalized world, they also underscore the need for continued research and refinement of economic models.

6.3 Contributions

This study adopts a network approach to understand trade interactions between countries and how these manifest impacts from economic shocks. The network representation of the International Trade Network has been previously utilized in economic analyses to study the effects of economic shocks on export and import relationships ([Gephart et al., 2016](#); [Tamea et al., 2016](#)). However, the use of these networks to predict shocks and identify potentially unknown economic disruptions represents a novel contribution to the field.

Previous research on network-based change-point detection in economics has primarily focused on financial crises, including stock market crashes ([Banerjee & Guhathakurta, 2020](#)) and banking crises ([D. Ma & Mankad, 2020](#)). However, the application of an s-GNN model trained on specific events enables more flexible change-point detection, allowing researchers to identify a broader range of economic shocks or more narrowly defined events as needed. Additionally, when these traditional network-based change-point detection methods are employed, they are limited to analyzing the shifts in graph structure over time. The s-GNN method enables the incorporation of node features, providing information regarding how changes in an individual country might impact its

position in the greater network.

The effectiveness of the s-GNN was further validated through its application to region and product-specific subnetworks. By accurately identifying many well-known, large-scale economic shocks, the model demonstrated its ability to detect relevant fluctuations in the trade network. Moreover, the identification of years beyond the established ground-truth values suggests potential for uncovering previously overlooked economic events. For example, this capacity could be seen by the model’s identification of 2004 as a significant year for Oceania. While not immediately associated with a known economic shock, this finding presents an opportunity for further investigation into the region’s economic dynamics during that period.

This study represents one of the first applications of the s-GNN architecture proposed by [Sulem et al. \(2022\)](#), extending its use beyond networks of S&P 500 index stock returns. This work not only validates the model’s versatility but also broadens its application scope beyond images, where s-GNNs have primarily been used for change-point detection previously ([Shuai et al., 2022](#); [Song et al., 2023](#)). By successfully applying this architecture to international trade networks, this thesis demonstrates this method’s potential for diverse economic analyses.

6.4 Limitations

The primary limitation of this study is how the ground-truth years are defined. The ensemble method that combines domain knowledge and data-driven approaches is meant to ensure that the years are those that are actually present in the trade networks. Yet, the final selection of years really depends on the way in which the three lists of years are combined. The rules outlined in section 4.2 are relatively subjective, and other methods of combining these lists could potentially have significant impacts especially on which years the model identifies within the region and product subnetworks.

During the training process using the random sampling scheme, there is potential that there is indirect data leakage between the train, validation, and test sets. If the pair of graphs for years 2002 and 2007 are present in the train set and the pair for 2003 and 2007 is present in the test set, the model already has some awareness that 2007 is associated with a change-point. This knowledge could be leaking through when evaluating on the test set, thereby artificially improving the performance metrics. Yet, if this occurring, performance across all models regardless of feature set or logging should see relatively high performance, but this is not the case, so it is unclear whether this issue is actually present.

In calculating the years associated with change-points, the s-GNN has trouble recognizing change-points at the ends of the distribution. Further work would need to be done

to determine whether this was an issue with the window within the similarity equation or because of issues in the pairs sampled for training.

In the implementation of the synthetic experiments from the paper by [Sulem et al. \(2022\)](#), the adjusted F1 scores for both the traditional metrics and the s-GNN models were far lower than the published results. It is unclear what this could be attributed to, so future work should further understand why this is and try to limit the difference between the results from the two studies.

To analyze regional change-points, I defined subnetworks using one-hop neighborhoods for each node. This approach balanced including connections beyond the immediate region while avoiding an excessive number of external edges. However, it potentially allowed economic shocks from countries outside the region to influence results. Additionally, I kept all country nodes in the network because of the model’s requirement for homogeneous networks. Future research should explore improved methods for refining network scope and develop s-GNNs capable of handling heterogeneous networks.

7 Conclusion

This thesis has introduced siamese graph neural networks to the task of change-point detection of economic shocks to understand whether this model outperforms traditional network-based distance metrics. The s-GNN model clearly provides a more nuanced and precise tool for economic analysis, capable of pinpointing specific years associated with economic shocks, including both well-known events and potentially overlooked economic disruptions.

This study provides novel contributions to the field by extending the application of change-point detection within economics beyond just financial crises and towards broader economic shocks. By uniquely applying s-GNNs to the ITN, this study demonstrates its versatility and opens up new avenues for economic research and analysis.

While this research has made significant contributions, it is not without limitations. The definition of ground-truth years and the model’s difficulty in recognizing change-points at the extremes of the distribution present challenges that future research must address. These limitations, along with some unexpected findings within the analysis of identified change-points, underscore the need for continued refinement.

Looking ahead, the potential for graph machine learning methods, particularly s-GNNs, in economic analysis is vast. Future research should focus on developing these methods further, especially in creating s-GNNs capable of handling heterogeneous networks. More detailed analyses of various economic sectors and regions could provide deeper insights into the intricate dynamics of global trade and economic shocks.

As our world becomes increasingly interconnected, the tools and methodologies developed in this thesis will become even more crucial for understanding and potentially predicting economic disruptions on both regional and global scales. By continuing to expand upon this work, we can enhance our ability to understand such a complex economic network, ultimately contributing to more resilient and adaptive economic systems worldwide.

References

- Allen, D. E., McAleer, M., Powell, R. J., & Singh, A. K. (2018). Non-parametric multiple change point analysis of the global financial crisis. *Annals of Financial Economics*, 13(02), 1850008.
- Baldwin, R., & Di Mauro, B. W. (2020). Economics in the time of covid-19: A new ebook. *Vox CEPR Policy Portal*, 2(3).
- Banerjee, S., & Guhathakurta, K. (2020). Change-point analysis in financial networks. , 9(1), e269. Retrieved 2024-06-05, from <https://onlinelibrary.wiley.com/doi/10.1002/sta4.269> doi: 10.1002/sta4.269
- Barnett, I., & Onnela, J.-P. (2016). Change point detection in correlation networks. *Scientific reports*, 6(1), 18893.
- Bartik, A. W., Bertrand, M., Cullen, Z., Glaeser, E. L., Luca, M., & Stanton, C. (2020). The impact of covid-19 on small business outcomes and expectations. *Proceedings of the national academy of sciences*, 117(30), 17656–17666.
- Besser, T. L., Recker, N., & Agnitsch, K. (2008). The impact of economic shocks on quality of life and social capital in small towns. *Rural Sociology*, 73(4), 580–604.
- Briguglio, L. P. (2016). Exposure to external shocks and economic resilience of countries: evidence from global indicators. *Journal of Economic Studies*, 43(6), 1057–1078.
- Broadman, H. G. (2006). *From disintegration to reintegration: Eastern europe and the former soviet union in international trade*. World Bank Publications.
- Casellato, C. (2022). International trade modelling with temporal relational graph neural networks. Retrieved 2024-06-05, from <http://dspace.unive.it/handle/10579/21414>
- Corsetti, G., Pesenti, P., & Roubini, N. (1999). What caused the asian currency and financial crisis? *Japan and the world economy*, 11(3), 305–373.
- Cribben, I., & Yu, Y. (2017). Estimating whole-brain dynamics by using spectral clustering. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66(3), 607–627.
- Das, D. K. (2012). How did the asian economy cope with the global financial crisis and recession? a revaluation and review. *Asia Pacific Business Review*, 18(1), 7–25.
- Duval, R., Elmeskov, J., & Vogel, L. (2007). Structural policies and economic resilience to shocks.
- Edwards, J. (2010). Australia after the global financial crisis. *Australian Journal of International Affairs*, 64(3), 359–371.
- Elbadawi, I., & Hegre, H. (2008). Globalization, economic shocks, and internal armed conflict. *Defence and Peace Economics*, 19(1), 37–60.

- Enikeeva, F., & Klopp, O. (2021). Change-point detection in dynamic networks with missing links. *arXiv preprint arXiv:2106.14470*.
- Ergashev, B. A. (2004). Sequential detection of us business cycle turning points: Performances of shiryayev-roberts, cusum and ewma procedures. *Econometrics from Econ-WPA*.
- Fair, K. R., Bauch, C. T., & Anand, M. (2017). Dynamics of the global wheat trade network and resilience to shocks. *Scientific reports*, 7(1), 7177.
- Fraenkel, J., & Firth, S. (2007). *From election to coup the 2006 campaign and its aftermath*. Anu Press.
- Gephart, J. A., Rovenskaya, E., Dieckmann, U., Pace, M. L., & Bränström, Å. (2016). Vulnerability to shocks in the global seafood trade network. , 11(3), 035008. Retrieved 2024-06-05, from <https://iopscience.iop.org/article/10.1088/1748-9326/11/3/035008/meta> doi: 10.1088/1748-9326/11/3/035008
- Graziano, A. G., Handley, K., & Limão, N. (2021). Brexit uncertainty and trade disintegration. *The Economic Journal*, 131(635), 1150–1185.
- Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.
- Hellwig, K.-P. (2021). *Predicting fiscal crises: A machine learning approach*. International Monetary Fund.
- Heßler, M., Wand, T., & Kamps, O. (2023). Efficient multi-change point analysis to decode economic crisis information from the s&p500 mean market correlation. , 25(9), 1265. Retrieved 2024-05-14, from <https://www.mdpi.com/1099-4300/25/9/1265> doi: 10.3390/e25091265
- Hill, E., Clair, T. S., Wial, H., Wolman, H., Atkins, P., Blumenthal, P., ... Friedhoff, A. (2012). Economic shocks and regional economic resilience. In *Urban and regional policy and its effects: Building resilient regions* (pp. 193–274). Brookings Institution Press.
- Huang, C., Deng, Y., Yang, X., Yang, X., & Cao, J. (2023). Can financial crisis be detected? laplacian energy measure. *The European Journal of Finance*, 29(9), 949–976.
- Huang, S., Hitti, Y., Rabusseau, G., & Rabbany, R. (2020). Laplacian change point detection for dynamic graphs. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 349–358).
- Issawi, C. (1978). The 1973 oil crisis and after. *Journal of Post Keynesian Economics*, 1(2), 3–26.
- Jaarsma, M., et al. (2017). What drives heterogeneity in the resilience of trade: Firm-specific versus regional characteristics. *Papers in Regional Science*, 96(1), 13–33.

- Ji, G., Zhong, H., Nzudie, H. L. F., Wang, P., & Tian, P. (2024). The structure, dynamics, and vulnerability of the global food trade network. *Journal of Cleaner Production*, 434, 140439.
- Keohane, R. O., & Nye Jr, J. S. (1973). Power and interdependence. *Survival*, 15(4), 158–165.
- Kohl, T. (2014). Do we really know that trade agreements increase trade? *Review of World Economics*, 150, 443–469.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2003). Financial integration and macroeconomic volatility. *IMF Staff papers*, 50(Suppl 1), 119–142.
- Koutra, D., Vogelstein, J. T., & Faloutsos, C. (2013). *DELTACON: A principled massive-graph similarity function*. Retrieved 2024-05-14, from <https://arxiv.org/abs/1304.4657v1>
- Laugesen, M., & Meads, C. (1991). Tobacco advertising restrictions, price, income and tobacco consumption in oecd countries, 1960–1986. *British journal of addiction*, 86(10), 1343–1354.
- Levy, P. I. (1999). Sanctions on south africa: What did they do? *American Economic Review*, 89(2), 415–420.
- Liu, F., Tawiah, V., Zakari, A., & Alessa, N. (2023). The impact of climate disaster on international trade: evidence from developed and developing countries. *Journal of Environmental Management*, 342, 118308.
- Liu, L., Chen, C., & Wang, B. (2022). Predicting financial crises with machine learning methods. *Journal of Forecasting*, 41(5), 871–910.
- Liu, X., Ornelas, E., & Shi, H. (2022). The trade impact of the covid-19 pandemic. *The World Economy*, 45(12), 3751–3779.
- Lungová, M. (2013). The resilience of regions to economic shocks. In *Proceedings of the 11th international conference liberec economic forum* (pp. 363–371).
- Ma, D., & Mankad, S. (2020). Change detection in core-periphery networks: A case study on detecting financial crises in the interbank market. Available at SSRN 3742790.
- Ma, Z., & Cheng, L. (2005). The effects of financial crises on international trade. In *International trade in east asia* (pp. 253–286). University of Chicago Press.
- Malesios, C., Jones, N., & Jones, A. (2020). A change-point analysis of food price shocks. *Climate risk management*, 27, 100208.
- Marginean, S., & Orastean, R. (2011). Globalization and economic crisis in european countries. *International Journal of Economics and Finance Studies*, 3(1), 209–218.

- Monken, A., Haberkorn, F., Gopinath, M., Freeman, L., & Batarseh, F. A. (2021). Graph neural networks for modeling causality in international trade. , 34. Retrieved 2024-06-05, from <https://journals.flvc.org/FLAIRS/article/view/128485> doi: 10.32473/flairs.v34i1.128485
- Nieman, M. D. (2011). Shocks and turbulence: Globalization and the occurrence of civil war. *International Interactions*, 37(3), 263–292.
- Panford-Quainoo, K., Bose, A. J., & Defferrard, M. (2019). *Bilateral trade modeling with graph neural networks* (phdthesis).
- Peel, L., & Clauset, A. (2015). Detecting change points in the large-scale structure of evolving networks. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 29).
- Pepelyshev, A., & Polunchenko, A. S. (2015). Real-time financial surveillance via quickest change-point detection methods. *arXiv preprint arXiv:1509.01570*.
- Rose, A. K. (2004). Do we really know that the wto increases trade? *American economic review*, 94(1), 98–114.
- Shervashidze, N., Schweitzer, P., van Leeuwen, E. J., Mehlhorn, K., & Borgwardt, K. M. (2011). Weisfeiler-lehman graph kernels. , 12, 2539–2561.
- Shuai, W., Jiang, F., Zheng, H., & Li, J. (2022). Msgatn: A superpixel-based multi-scale siamese graph attention network for change detection in remote sensing images. *Applied Sciences*, 12(10), 5158.
- Sims, J., & Romero, J. (2013). Latin american debt crisis of the 1980s. *Federal Reserve History*, 22.
- Song, X., Hua, Z., & Li, J. (2023). Gmtns: Gnn-based multi-scale transformer siamese network for remote sensing building change detection. *International Journal of Digital Earth*, 16(1), 1685–1706.
- Soufi, H. R., Esfahanipour, A., & Shirazi, M. A. (2022). A quantitative approach for analysis of macroeconomic resilience due to socio-economic shocks. *Socio-Economic Planning Sciences*, 79, 101101.
- Sulem, D., Kenlay, H., Cucuringu, M., & Dong, X. (2022). *Graph similarity learning for change-point detection in dynamic networks* (No. arXiv:2203.15470). arXiv. Retrieved 2024-05-14, from <http://arxiv.org/abs/2203.15470> doi: 10.48550/arXiv.2203.15470
- Szczygielski, J. J., Charteris, A., Bwanya, P. R., & Brzeszczyński, J. (2022). The impact and role of covid-19 uncertainty: A global industry analysis. *International Review of Financial Analysis*, 80, 101837.
- Tambunan, T. T., et al. (2012). Trade response to economic shocks in indonesia. *Journal of Business Management and Economics*, 3(8), 283–300.

- Tamea, S., Laio, F., & Ridolfi, L. (2016). Global effects of local food-production crises: a virtual water perspective. *Scientific reports*, 6(1), 18803.
- Targowski, A. (2014). From globalization waves to global civilization. *Comparative Civilizations Review*, 70(70), 7.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Vo, T. D., & Tran, M. D. (2021). The impact of covid-19 pandemic on the global trade. *International Journal of Social Science and Economics Invention*, 7(1), 1–7.
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.
- You, Z.-H., Wang, J.-X., Chen, S.-B., Ding, C. H., Wang, G.-Z., Tang, J., & Luo, B. (2023). Crossed siamese vision graph neural network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*.
- Yu, Y., Padilla, O. H. M., Wang, D., & Rinaldo, A. (2021). Optimal network online change point localisation. *arXiv preprint arXiv:2101.05477*.
- Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1), 1–23.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, 1, 57–81.
- Zou, C., Yin, G., Feng, L., & Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems.

A Appendix

A.1 Network-Driven Identified Years

Figures A.1 and A.2 display the two lists of years identified for the ensemble method using the raw export values and the export percentages, respectively.

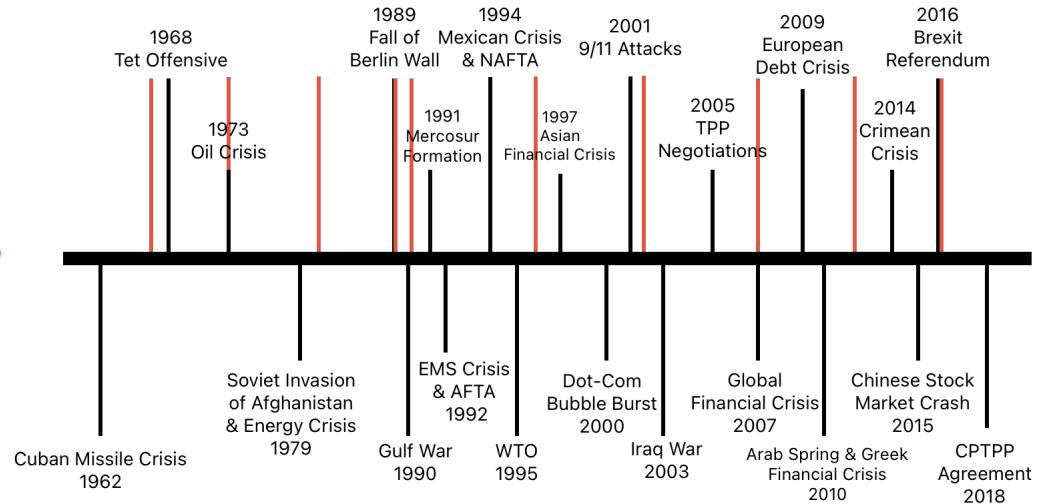


Figure A.1: Economic shocks identified through the clustering algorithm using raw export values.

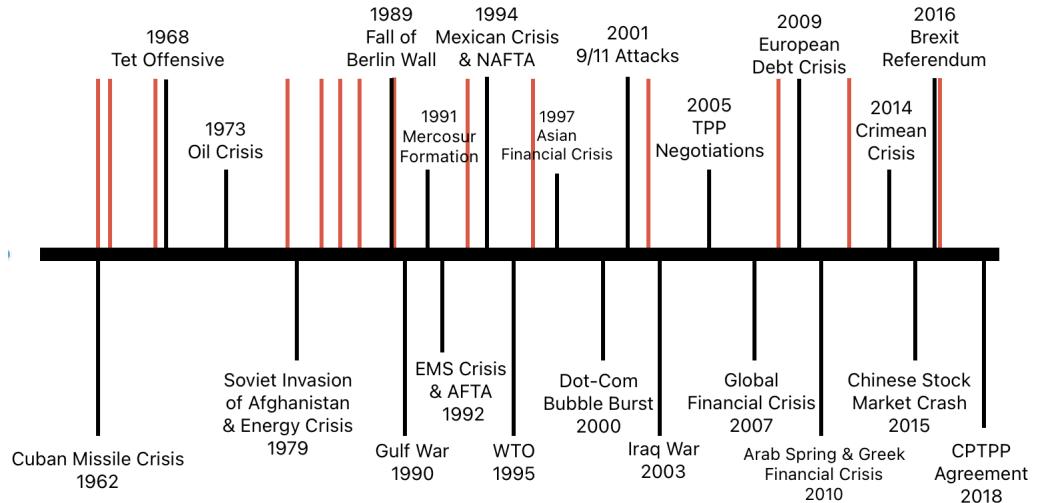


Figure A.2: Economic shocks identified through the clustering algorithm using percentage export values.

A.2 Feature Subsets

Tables A.1 and A.2 display the two subsets of World Development Indicators identified by MIS and random methods, respectively.

Feature Code	Definition
GC.DOD.TOTL.GD.ZS	Central government debt, total (% of GDP): The total amount of central government debt as a percentage of GDP.
IT.MLT.MAIN.P2	Mobile cellular subscriptions (per 100 people): Number of mobile cellular subscriptions per 100 people.
NE.CON.PRVT.KD.ZG	Final consumption expenditure, private sector (annual % growth): Annual percentage growth of private consumption expenditure in constant local currency.
NE.CON.TOTL.KD.ZG	Final consumption expenditure (annual % growth): Annual percentage growth of total consumption expenditure in constant local currency.
NV.AGR.TOTL.ZS	Agriculture, forestry, and fishing, value added (% of GDP): The value added of agriculture, forestry, and fishing as a percentage of GDP.
NV.IND.TOTL.KD.ZG	Industry (including construction), value added (annual % growth): Annual percentage growth of value added in the industry sector in constant local currency.
NV.SRV.TOTL.KD.ZG	Services, value added (annual % growth): Annual percentage growth of value added in the services sector in constant local currency.
NY.GNP.MKTP.KD.ZG	GNI growth (annual %): Annual percentage growth rate of GNI at market prices based on constant local currency.
NY.GNP.PCAP.KD.ZG	GNI per capita growth (annual %): Annual percentage growth rate of GNI per capita based on constant local currency.
SP.ADO.TFRT	Adolescent fertility rate (births per 1,000 women ages 15-19): Number of births per 1,000 women ages 15-19.
SP.POP.2024.FE.5Y	Population ages 20-24, female (% of female population): Percentage of the female population aged 20-24.
SP.POP.2024.MA.5Y	Population ages 20-24, male (% of male population): Percentage of the male population aged 20-24.
SP.POP.6569.FE.5Y	Population ages 65-69, female (% of female population): Percentage of the female population aged 65-69.
SP.POP.6569.MA.5Y	Population ages 65-69, male (% of male population): Percentage of the male population aged 65-69.
SP.POP.65UP.MA.ZS	Population ages 65 and above, male (% of male population): Percentage of the male population aged 65 and above.

Table A.1: Definitions of MIS selected features coming from World Bank data.

Feature Code	Definition
GC.XPN.TOTL.CN	Expense (current LCU): Expense is cash payments for operating activities of the government in providing goods and services.
DC.DAC.DNKL.CD	Net ODA received from DAC countries, Denmark (current US\$): Net official development assistance received from Denmark.
ST.INT.DPRT	International tourism, number of departures: Number of departures of residents from the country for international travel.
SP.POP.6064.FE.5Y	Population ages 60-64, female (% of female population): Percentage of the female population aged 60-64.
SL.IND.EMPL.MA.ZS	Employment in industry, male (% of male employment): Male employment in industry as a percentage of total male employment.
DT.NFL.PRVT.CD	Net financial flows, private (NFL, current US\$): Net financial flows to the private sector.
SE.XPD.CSEC.ZS	Expenditure on secondary education (% of government expenditure on education): Percentage of total government expenditure on secondary education.
SP.URB.TOTL.IN.ZS	Urban population (% of total population): Percentage of the total population living in urban areas.
NY.GNP.MKTP.PP.KD	GNI, PPP (constant 2017 international \$): Gross national income converted to international dollars using purchasing power parity rates.
DC.DAC.POLL.CD	Net ODA received from DAC countries, Poland (current US\$): Net official development assistance received from Poland.
TX.VAL.MRCH.R5.ZS	Merchandise exports to high-income economies (% of total merchandise exports): Percentage of total merchandise exports sent to high-income economies.
SE.PRM.OENR.FE.ZS	Adjusted net enrollment rate, primary, female (% of primary school-age children): Net enrollment rate in primary education for female children.
NE.CON.GOVT.CN	Final consumption expenditure, general government (current LCU): Government final consumption expenditure in current local currency.
SL.TLF.BASC.ZS	Labor force with basic education (% of total working-age population with basic education): Percentage of the working-age population with basic education in the labor force.
EN.ATM.PM25.MC.M3	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter): Average annual exposure to PM2.5 concentrations.

Table A.2: Definitions of randomly selected features coming from World Bank data.

A.3 Hyperparameter-Tuning Results

Table A.4 displays the results of the hyperparameter-tuning experiments.

Table A.3: Hyperparameter Tuning Results

Learning Rate	Dropout Rate	Sort k	Hidden Units	Accuracy	F1 Score	F2 Score	F0.5 Score
0.001	0.05	50	16	0.913	0.937	0.915	0.960
0.001	0.05	50	64	0.853	0.890	0.848	0.936
0.001	0.05	100	16	0.838	0.877	0.830	0.930
0.001	0.05	100	64	0.875	0.908	0.874	0.944
0.001	0.1	50	16	0.819	0.874	0.872	0.876
0.001	0.1	50	64	0.853	0.890	0.848	0.936
0.001	0.1	100	16	0.838	0.877	0.830	0.930
0.001	0.1	100	64	0.875	0.908	0.874	0.944
0.0001	0.05	50	16	0.819	0.874	0.872	0.876
0.0001	0.05	50	64	0.853	0.890	0.848	0.936
0.0001	0.05	100	16	0.838	0.877	0.830	0.930
0.0001	0.05	100	64	0.875	0.908	0.874	0.944
0.0001	0.1	50	16	0.819	0.874	0.872	0.876
0.0001	0.1	50	64	0.853	0.890	0.848	0.936
0.0001	0.1	100	16	0.838	0.877	0.830	0.930
0.0001	0.1	100	64	0.875	0.908	0.874	0.944

Table A.4: Results of hyperparameter tuning the final selected model using GraphSAGE, MIS selected logged features including current and previous GDP values, and 15 epochs. Best scores for each metric are bolded.