Predicting Global Trade Hubs

A Scalable Hybrid Framework Combining Dynamic Network Analysis and Explainable AI

Muhammad Zeeshan Asghar

June 25, 2025

Data Science Master's Program, HSE University

The Challenge: Forecasting a Dynamic Global Economy

Forecasting global trade is critical for economic policy, but it's a major challenge because trade networks are:

- **Complex & Interdependent:** A single policy change can cascade through the entire system.
- **Constantly Evolving:** New trade blocs emerge while others fade.
- **High-Dimensional:** Thousands of relationships change simultaneously.

The Core Problem

How can we accurately forecast future trade and identify the next major economic hubs in this complex system?

A Three-Part Methodology

To address this, I designed a methodology that directly incorporated feedback from the project proposal.

1. Feature Engineering

Create a rich dataset combining time-series, network, and learned graph features.

2. Predictive Modeling

Train and evaluate advanced models (like XGBoost) against powerful baselines.

3. Explainable AI (XAI)

Use SHAP to interpret the model and understand the key drivers of its predictions.

Data Pipeline and Exploration

The project began with a robust data pipeline using UN Comtrade data (1988-2024).

- Raw data of 795K+ records was filtered to 17,000 core observations.
- Trade values were highly skewed, requiring a log transformation to stabilize variance for modeling.

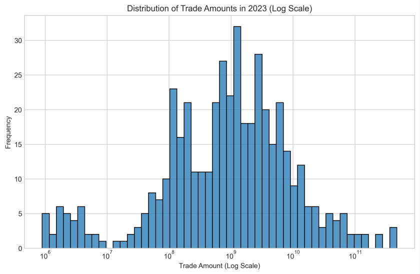


Figure 1: Distribution of trade values before and after log transformation.

Feature Engineering: Time-Series and Network Features

First, I built features based on history and network structure, as suggested in the professor's feedback.

- **Time-Series Features:** Using the `tsfresh` library, I automatically extracted over 777 statistical characteristics for each trade pair's history.
- **Dynamic Network Metrics:** I modeled the trade data as a series of yearly graphs and calculated evolving centrality scores like **PageRank**, **HITS**, and **Harmonic Centrality** for every country.

Feature Engineering: Learned Graph Embeddings

The most advanced step was to have the model learn its own features from the network structure.

Temporal Graph Network (TGN) Inspired Approach I implemented a GCN-LSTM model to process the sequence of 37 yearly trade graphs.

- This model learns a 32-dimensional **embedding** for each country for each year.
- These embeddings represent the model's own understanding of a country's evolving structural role.

This is a core innovation of the project, moving beyond just handcrafted features.

Analysis of the 2023 Trade Network

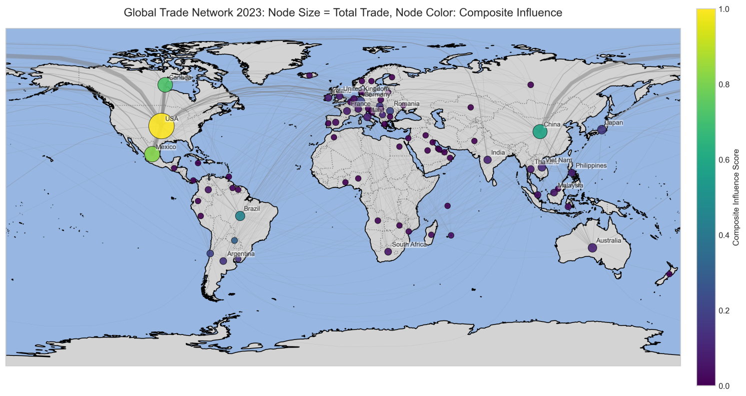


Figure 2: The 2023 global trade network. Node size is trade volume; color shows influence.

Key Insight: The network exhibits "small-world" properties (high local clustering, short paths), with the USA, China, and Germany as the most influential hubs.

Network Analysis: Identifying Key Players with Centrality

I used multiple centrality measures to identify important countries, as recommended.

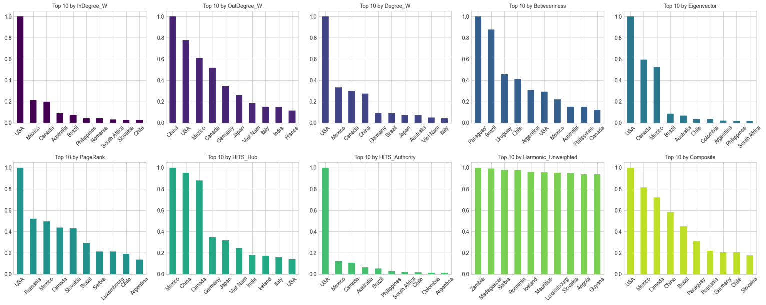


Figure 3: Top 10 countries by various centrality measures in 2023.

Finding: Major economies like the USA and China dominate most metrics, while countries like Paraguay and Uruguay emerge as critical "bridges" with high Betweenness Centrality.

Network Analysis: Detecting Trade Blocs

Using the Greedy Modularity algorithm, we can cluster countries into distinct trade communities.

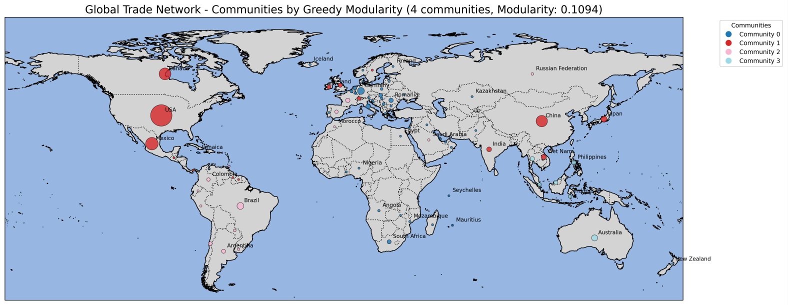


Figure 4: The 4 distinct trade communities detected in the 2023 network.

Finding: The analysis identified four major trade blocs, providing a structural baseline for tracking how these alliances shift over time.

Results: The Surprisingly Strong Baseline

Any useful model must outperform a simple baseline. The results were humbling.

Model	RMSE (Billions)	R-squared
Naive Forecast (predicts last year's value)	**5.66**	**0.9844**
Historical Average	15.67	0.8804

Primary Finding: The Naive Forecast's Superiority

Simply predicting that this year's value is the same as last year's proved more accurate than any complex model. This highlights the extreme persistence in trade data and sets a powerful, non-trivial baseline.

Results: The Value of Learning from Hybrid Features

While no model beat the Naive baseline, comparing the learning models reveals the value of our hybrid features.

Learning Model	RMSE (Billions)	R-squared
XGBoost + TGN Embeddings	**37.38**	**0.3197**
XGBoost (Standard Features)	38.36	0.2835
Random Forest	41.16	0.1751
LSTM	52.27	0.1443

The TGN-Augmented Model Wins The XGBoost model augmented with our TGN-inspired embeddings was the best-performing learning model, improving the R-squared by nearly 4 percentage points over the standard model.

Results: Qualitative Error Analysis

A qualitative error analysis reveals where the best model struggles.

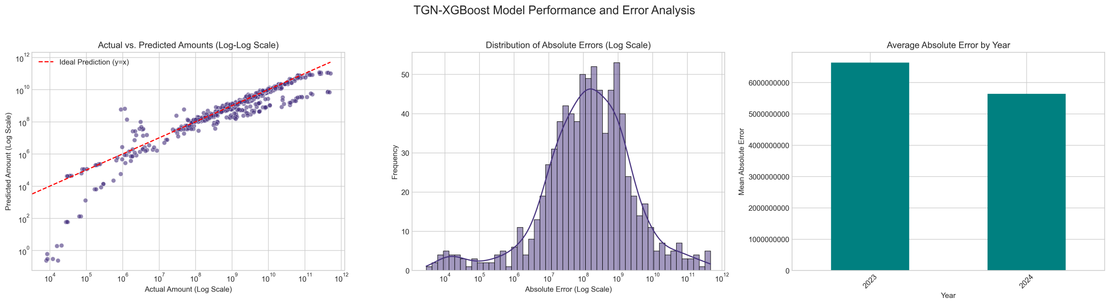


Figure 5: Error analysis of the TGN-XGBoost model.

Insight: The model is accurate for most predictions (left) but systematically under-predicts the highest-value trade flows (right), leading to a long-tailed error distribution.

Explainable AI: Understanding the "Why"

Moving beyond performance metrics, we use **SHAP (SHapley Additive exPlanations)** to understand our best model.

- SHAP is a game-theoretic approach to explain the output of any machine learning model.
- It allows us to see not just *what* the model predicts, but *why*.

We applied SHAP to answer three key questions:

1. What features are most important globally?
2. How do these features interact?
3. Why was a specific prediction made?

XAI: Global Drivers of Trade

SHAP confirms that our hybrid feature strategy was effective.

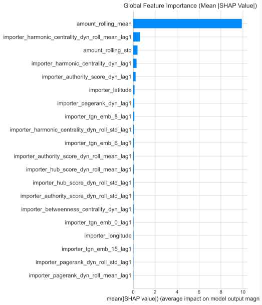


Figure 6: Mean absolute SHAP values (feature importance).

Key Drivers: History (`amount_rolling_mean`), network position (`harmonic_centrality`), and learned embeddings (`tgn_emb`) are all crucial.

XAI: Directional Effects of Top Features

The beeswarm plot shows not just importance, but also the directional impact of features.

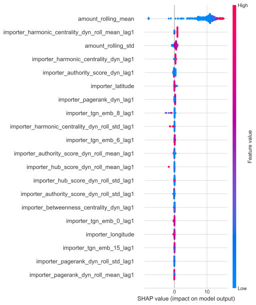


Figure 7: SHAP beeswarm plot showing feature values and their impact.

Insight: High values (red) of the top features, like rolling mean and centrality, consistently push the prediction higher (positive SHAP value).

XAI: Uncovering a Sophisticated Trade Pattern

This SHAP dependence plot shows how the top two features interact to influence predictions.

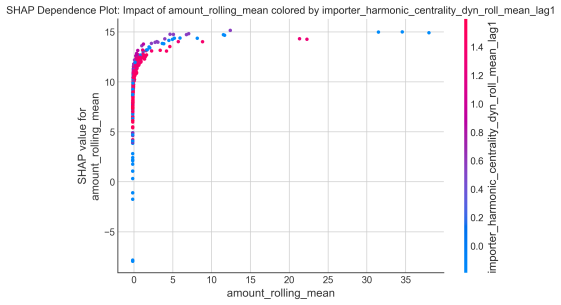


Figure 8: Impact of Rolling Mean, colored by Importer Harmonic Centrality.

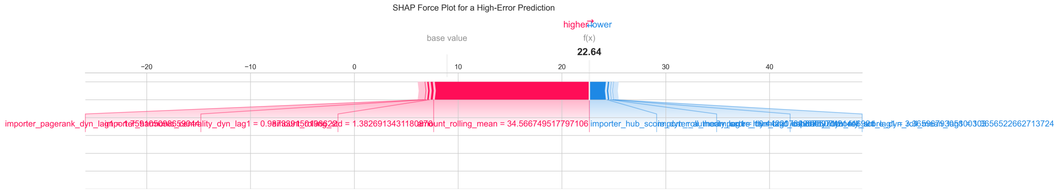
Interpretation: The model learned that for a given trade volume (x-axis), the prediction is pushed **even higher** if the importer has high network centrality (the red dots).

XAI: Explaining a High-Error Forecast

SHAP force plots can explain individual predictions. This one shows a large under-prediction for USA-Mexico trade.

High-Error Case: USA-Mexico 2023

- **Actual:** \$475 Billion
- **Predicted:** \$6.8 Billion



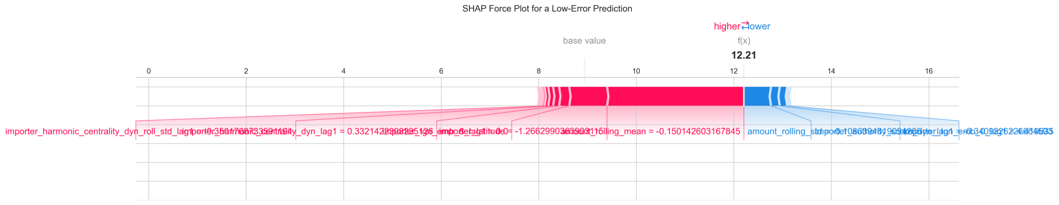
Reasoning: Many features, including a low rolling mean relative to the final value, pushed the prediction down, showing the model's struggle with massive, volatile trade relationships.

XAI: Explaining a Low-Error Forecast

In contrast, the force plot for a low-error prediction shows features in balance.

Low-Error Case: Mauritius-Niger 2024

- **Actual:** $195 Thousand
- **Predicted:** $200 Thousand



Reasoning: Here, the features correctly balance each other out, with the low rolling mean correctly pushing the prediction down toward the small final value.

Conclusion and Key Takeaways

This project successfully demonstrates the power of a hybrid approach to a complex forecasting problem.

- **Built a Scalable Pipeline:** Successfully processed and engineered features from a large, complex dataset.
- **Validated a Hybrid Feature Approach:** Proved that combining time-series, dynamic network, and learned graph embeddings improves model performance over simpler learning models.
- **Generated Interpretable Insights:** Used XAI to show *how* the model learned complex, non-linear relationships that drive global trade.
- **Established a Powerful Baseline:** Showed that the high persistence in trade data makes the Naive Forecast a formidable benchmark.

Future Work

This framework provides a strong foundation for several exciting future research directions:

- **Incorporate Exogenous Data:** Enhance the model by adding macroeconomic indicators like country-specific GDP, inflation rates, and trade tariffs.
- **Advanced GNN Architectures:** With more computational resources, explore more complex Temporal Graph Network architectures to capture deeper relational patterns.
- **Multi-Step Forecasting:** Extend the framework from single-step (one year ahead) to predict trade values several years into the future.

Thank You Questions?