

Forecasting International Trade Flows Using Graph Attention Networks

Lorenzo Giamartino

Data Scientist, Publicis

MA Quant Research Methods (QMSS), Columbia University

October 2025

Abstract

I develop a graph neural network (GNN) framework to forecast bilateral trade flows across products and countries. The model represents the global trade system as a heterogeneous graph linking countries and products. Message passing via Graph Attention Networks (GATv2) enables countries and products to learn embeddings that reflect structural similarity in their productive capabilities. Combined with lagged trade and seasonality features, the model predicts both the probability and expected magnitude of monthly trade flows, achieving structure-aware forecasting motivated by the economic complexity literature.

1 1. Model Overview

The model jointly predicts (i) the probability that a positive trade flow will occur between two countries for a given product, and (ii) the conditional trade value if positive. This “hurdle” structure captures both the *extensive* and *intensive* margins of trade.

2 2. Graph Construction

I build a bipartite graph between countries and products using the standard Revealed Comparative Advantage (RCA) measure. For country c and product p , let

$$\text{RCA}_{c,p} = \frac{\frac{X_{c,p}}{\sum_p X_{c,p}}}{\frac{\sum_c X_{c,p}}{\sum_{c,p} X_{c,p}}},$$

where $X_{c,p}$ is the export (or import) value of product p by country c . An RCA greater than 1 indicates a comparative advantage.

- Each **country node** has binary features indicating $\text{RCA} > 1$ across all HS4 products, for both exports and imports.
- Each **product node** connects to all countries with $\text{RCA} > 1$ in that product.

- Two relation types are created:

$$(\text{country}, \text{exports_rca}, \text{product}), \quad (\text{country}, \text{imports_rca}, \text{product}).$$

This produces a heterogeneous graph $\mathcal{G} = (\mathcal{V}_C, \mathcal{V}_P, \mathcal{E})$ on which message passing allows information exchange between structurally similar countries and related products.

3 3. Graph Attention Encoding

Each node embedding is updated using multi-head graph attention (GATv2):

$$\mathbf{h}_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}, \quad \alpha_{ij}^{(l)} = \text{softmax}_j(a(\mathbf{W}^{(l)} \mathbf{h}_i^{(l)}, \mathbf{W}^{(l)} \mathbf{h}_j^{(l)})),$$

where α_{ij} is an attention coefficient over neighboring nodes. Separate heads handle export and import relations, enabling asymmetric propagation of trade capabilities. Through L layers, the encoder learns country and product embeddings \mathbf{h}_c and \mathbf{h}_p representing structural similarity in production and trade patterns.

4 4. Temporal and Exogenous Features

Each country–pair–product triple (c_i, c_j, p, t) at month t includes:

- Lagged trade features $(\text{lag}_1, \dots, \text{lag}_K)$ of past K months.
- Seasonal encodings $(\sin(2\pi m_t/12), \cos(2\pi m_t/12))$ of the month m_t .

These augment the graph embeddings with short-term temporal dynamics.

5 5. Decoder and Prediction Heads

The decoder concatenates embeddings and temporal features:

$$\mathbf{z} = [\mathbf{h}_{\text{exp}}, \mathbf{h}_{\text{imp}}, \mathbf{h}_{\text{prod}}, \mathbf{lags}, \mathbf{month}],$$

and feeds \mathbf{z} through two parallel MLPs:

$$\text{Head}_A : z \mapsto \hat{p} = \sigma(f_A(\mathbf{z})), \quad \text{Head}_B : z \mapsto \hat{y} = f_B(\mathbf{z}).$$

The final forecast combines occurrence and magnitude:

$$\widehat{X} = \hat{p} \times (\exp(\hat{y}) - 1).$$

6 6. Training and Backtesting

The model minimizes a joint loss:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}}(\hat{p}, y_{\text{pos}}) + \lambda \mathcal{L}_{\text{Huber}}(\hat{y}, \log(1 + y_{\text{amt}})),$$

where y_{pos} is a binary indicator for positive flows. Training uses the AdamW optimizer with $\lambda = 1.0$ and class-weighted BCE for imbalance.

A sliding-window backtest with $W = 18$ months and $K = 3$ lags evaluates each configuration on one-step-ahead forecasting. The best W is selected by minimum validation sMAPE:

$$\text{sMAPE} = \frac{2}{N} \sum_i \frac{|\hat{y}_i - y_i|}{|\hat{y}_i| + |y_i| + \epsilon}.$$

The final model retrains on the last W months and forecasts the next unseen month.

7 7. Interpretation

This framework builds on the concept of the *product space* [1]: countries with similar productive capabilities tend to develop similar trade structures. The GAT encoder thus learns these country latent capabilities that generalizes across products and trading partners. Temporal lags allow adaptation to recent trends while grounded in long-run comparative advantage.

References

- [1] Hidalgo, C.A., Klinger, B., Barabási, A.L., Hausmann, R. (2007). *The product space conditions the development of nations*. Science, 317(5837), 482–487.