ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών
«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»

<u>Μεταπτυχιακή Διατριβή</u>

Τίτλος Διατριβής	**(Ελληνικά)** Νευρωνικά Δίκτυα Γράφων για την Πρόβλεψη Διμερών Μεταναστευτικών Ροών: Μια Ανάλυση με Επίκεντρο την Ευρώπη **(Αγγλικά)** Graph Neural Networks for Bilateral Migration Prediction: A European-Focused Analysis
Ονοματεπώνυμο Φοιτητή	Χατζηκυριακάκη Φοίβη
Πατρώνυμο	Κίμων
Αριθμός Μητρώου	ΜΠΚΕΔ2351
Επιβλέπων	Διονύσιος Σωτηρόπουλος, Αναπληρωτής Καθηγητής

Ημερομηνία Παράδοσης **Σεπτέμβριος 2026**

Τριμελής Εξεταστική Επιτροπή

Διονύσιος Σωτηρόπουλος
Αναπληρωτής Καθηγητής

Δημήτριος Αποστόλου
Καθηγητής

Γρηγόριος Κορωνάκος
Επίκουρος Καθηγητής

Περίληψη

Η μετανάστευση αποτελεί ένα από τα πλέον σημαντικά και δυσπρόβλεπτα φαινόμενα της σύγχρονης κοινωνίας, με βαθιές επιπτώσεις σε επίπεδο πολιτικής και ανθρωπιστικού σχεδιασμού. Η παρούσα διατριβή διερευνά κατά πόσο τα Νευρωνικά Δίκτυα Γράφων (Graph Neural Networks – GNNs) μπορούν να βελτιώσουν την πρόβλεψη διμερών μεταναστευτικών ροών σε σχέση με κλασικές μεθόδους αναφοράς, και ταυτόχρονα αναλύει το ευρωπαϊκό σύστημα μετανάστευσης μέσω εργαλείων δικτυακής ανάλυσης. Συγκρίνονται τρεις αρχιτεκτονικές GNN (GCN, GraphSAGE, Temporal-GNN) με τρεις κλασικές μεθόδους (απλή εμμονή, log-γραμμικό βαρυτικό μοντέλο, Random Forest) σε δεδομένα διμερών ροών 21 ετών (2000–2020) για 217 χώρες. Τα αποτελέσματα δείχνουν ότι η απλή μέθοδος εμμονής υπερτερεί όλων των εκπαιδευμένων μοντέλων κατά παράγοντα 4,5 στο μέσο απόλυτο σφάλμα μη-μηδενικών κελιών. Η ανάλυση εντοπίζει τρεις δομικές τάσεις στην ευρωπαϊκή μετανάστευση: την έκρηξη ρουμανικής μετανάστευσης μετά τη διεύρυνση της ΕΕ το 2007, την προοδευτική απομάκρυνση του Ηνωμένου Βασιλείου από το δίκτυο μετά το Brexit, και την ανακατεύθυνση των ουκρανικών ροών προς την Πολωνία. Συμπληρωματικά, μέσω δεδομένων IOM DTM, αναλύεται ο εσωτερικός εκτοπισμός στην Ουκρανία (2022–2026).

Λέξεις-κλειδιά: Νευρωνικά Δίκτυα Γράφων, μετανάστευση, πρόβλεψη διμερών ροών, Ευρωπαϊκή Ένωση, ανάλυση δικτύων, Ουκρανία, εσωτερικός εκτοπισμός, IOM DTM

Abstract

Migration is among the most consequential and least predictable processes shaping contemporary societies, with direct implications for policy and humanitarian planning. This thesis investigates whether Graph Neural Networks (GNNs) improve on classical baselines for predicting bilateral migration flows, and analyses the European migration system through network-analytic tools. Three GNN architectures (GCN, GraphSAGE, Temporal-GNN) are benchmarked against three classical baselines (naive persistence, log-linear gravity model, Random Forest) on twenty-one years of bilateral flow data (2000–2020) covering 217 countries. The results show that the naive persistence baseline outperforms every learned model by a factor of approximately 4.5 in mean absolute error on non-zero cells, attributable to a short temporal axis (T = 21), severe target sparsity (90.4% zero cells), and strong year-on-year inertia. The network analysis identifies three structural patterns in European migration: the post-2007 Romanian enlargement spike, the gradual fade of the United Kingdom's network centrality across the post-Brexit window, and the eastward redirection of Ukrainian emigration toward Poland after 2014. A complementary sub-national analysis of internal displacement within Ukraine (2022–2026) using IOM DTM data is also reported.

Keywords: Graph Neural Networks, migration, bilateral flow prediction, European Union, network analysis, Ukraine, internal displacement, IOM DTM

Πίνακας Περιεχομένων

1 Introduction

1.1 Migration as a Research Problem

International migration is one of the most consequential and least predictable processes shaping contemporary societies. The United Nations High Commissioner for Refugees estimated that 122 million people were forcibly displaced worldwide by the end of 2024 (UNHCR, 2023), with millions more moving across borders each year for economic, educational, family-reunification, and policy-driven reasons. The bilateral patterns these movements form — who goes where, in what numbers, and how the pattern changes from year to year — are central to the operational planning of receiving-country governments, the resource allocation of humanitarian agencies, and the political stability of both origin and destination societies. The quantitative forecasting of migration flows is therefore not an academic exercise but a question of immediate policy relevance (Cantat, Pécoud and Thiollet, 2025).

Migration forecasting is also notoriously difficult. The classical economic framing of migration in terms of push and pull factors, distance, and origin–destination "masses" — the gravity model in its various incarnations (Ravenstein, 1885; Lee, 1966; Khan, Fatima and Fatima, 2023) — captures the broad shape of bilateral flows but struggles with discrete events that change those flows abruptly: the 2004 and 2007 enlargements of the European Union, the 2014 conflict in eastern Ukraine, the 2015–2016 European migrant crisis, the 2020 COVID-19 pandemic, and the 2022 Russian invasion of Ukraine all produced migration shocks that no model trained on prior flow data could have anticipated. Modern migration-forecasting work has therefore turned increasingly to data-rich, structural approaches that attempt to model the network of bilateral relationships explicitly, rather than treating each origin–destination corridor as an independent unit (Bijak and Czaika, 2020; Capone, Di Clemente and Zambotti, 2025; Boss et al., 2023).

1.2 Why Europe

This thesis takes a European-primary approach, with global comparisons introduced where they sharpen the European story. Three considerations motivate that geographic emphasis.

First, Europe is the second-largest migrant-receiving region in the world after North America, absorbing tens of millions of cross-border migrants over the 2000–2020 window. The continent's principal destinations — Germany, the United Kingdom, Spain, Italy, France, the Netherlands, Switzerland, Austria — together accumulate approximately 40 million migrants in the bilateral flow data used for this work, more than any other coherent regional grouping.

Second, the European migration system has experienced more discrete structural changes during the time window of the data than any other regional system: two major rounds of EU enlargement (2004 and 2007) that opened previously-restricted labour markets, the staggered implementation of Schengen visa-free travel, the United Kingdom's 2016 referendum and its multi-year implementation, the 2015–2016 arrival of Syrian and other forcibly-displaced populations, and the 2022 Russian invasion of Ukraine. This density of structural transitions makes Europe an unusually well-instrumented natural laboratory for studying how migration corridors respond to changing legal regimes — provided the analytical tools used can detect those transitions in the data.

Third, the bilateral migration data available for academic study — based largely on OECD and Eurostat statistics (OECD, 2020; Eurostat, 2021) augmented with bilateral-flow harmonisation methods (Abel and Cohen, 2022) — has its strongest coverage on European-destination corridors. The data are richer, more reliable, and more longitudinally consistent for European receivers than for most non-European destinations. A thesis whose analytical scope matches the strongest data coverage is on firmer methodological ground than one that attempts global generality on data of variable quality.

1.3 Why Graph Methods

Migration is an intrinsically *relational* phenomenon. A flow from country A to country B does not exist in isolation: it depends on flows from A to B's competitors (Germany versus the United Kingdom, Italy versus Spain), on flows from A's neighbours to B (chain-migration effects established by earlier movers), and on the broader network position of both endpoints in the global system. A shock that increases migration along one corridor displaces or substitutes flow along others. A graph representation, in which countries are nodes and bilateral flows are weighted directed edges, is therefore not merely a convenient visualisation but a substantively appropriate model of the underlying system.

Graph Neural Networks (GNNs) are the natural class of machine-learning models for prediction tasks defined on graphs. Originally developed for problems in chemistry, recommender systems, and social-network analysis, they have in recent years been applied to a wide range of flow-prediction problems including urban traffic, communication networks, and human mobility (Suárez-Varela et al., 2022; Zhang, 2025; Neyigapula, 2023; Xu et al., 2025). Their application to international migration is more recent but growing (Gawryluk et al., 2024). The central methodological question of this thesis is whether the inductive biases that make GNNs effective in those adjacent domains also deliver predictive gains on bilateral migration data.

The research question of the thesis is therefore:

> **Can graph neural networks improve on classical baselines for predicting bilateral migration flows, and what insights does the network structure reveal about European migration patterns?**

Note the dual emphasis. The first half of the question is a methodological benchmark: do modern graph methods deliver concrete predictive improvements over simpler baselines on the migration data available for academic study? The second half is a substantive question that does not depend on the methodological answer: regardless of whether GNNs predict well, the network analysis itself — centrality dynamics, corridor evolution, structural breaks — can illuminate European migration patterns in ways that traditional pair- by-pair analysis cannot.

1.4 Contributions

This thesis makes four contributions.

1. A reproducible benchmark. A complete pipeline is implemented that benchmarks three Graph Neural Network architectures (GCN, GraphSAGE, Temporal-GNN) against three classical baselines (naive persistence, log-linear gravity model, Random Forest) on twenty-one years of bilateral migration data covering 217 countries. The pipeline is reproducible end-to-end from raw CSV inputs through trained models and final figures, with all hyperparameters specified in a single configuration file.

2. Empirical evidence on GNNs versus persistence. The benchmark yields a clear and somewhat surprising result: at yearly resolution, on the data shape provided by public OECD and Eurostat-derived sources, the naive persistence baseline outperforms every learned model — including the three GNN architectures and the Random Forest — by a factor of approximately 4.5 in mean absolute error on non-zero cells. Diagnostic analysis attributes this outcome to three properties of the data: the short temporal axis (T = 21 snapshots), the severe sparsity of the target (90.4\% of cells are zero), and the strong year-on-year inertia of bilateral flows (median per-pair lag-1 autocorrelation of 0.60). The negative result is interpreted not as evidence against GNNs in general but as identifying the conditions under which graph methods do or do not deliver value over simpler alternatives.

3. A descriptive network analysis of European migration, 2000–2020. Independently of the model comparison, the thesis analyses the bilateral flow data through five centrality measures applied to each year's migration network. Three substantive findings emerge: a clear 2007 enlargement spike on the Romania → Italy and Romania → Spain corridors; a multi-year decline of the United Kingdom's network centrality across the post-Brexit period; and a striking decade-long redirection of Ukrainian emigration toward Poland following the 2014 Donbas conflict and the 2017 Polish work-permit reforms.

4. A sub-national case study of internal displacement in Ukraine. A monthly-resolution analysis of internal displacement within Ukraine from May 2022 through March 2026 is constructed, drawing on the IOM Displacement Tracking Matrix data extended to the present via the official `dtmapi` Python client. The analysis distinguishes two non-comparable methodologies (the Area-Based Assessment that ran through October 2024, and the Registration-based count that began in February 2025), reports them as separate series, and identifies four distinct displacement patterns at the oblast (admin-1) level: western refuge, persistent frontline, steady escalation, and late-emerging hub. The Ukraine case study is presented as a focused European study within the broader DTM coverage, which also tracks larger crises in Sudan, the Democratic Republic of the Congo, Yemen, and elsewhere (referenced for scale rather than analysed in detail).

1.5 Thesis Outline

The remainder of the thesis is organised as follows. Chapter 2 situates the work within the migration-modelling and graph-learning literatures, surveying the development of migration models from the classical gravity formulation through network-theoretic extensions to modern Graph Neural Network applications, and identifying the specific gaps in the existing literature that motivate the present contribution. Chapter 3 describes the data sources, the European focus and data asymmetry that shape the analytical scope, the preprocessing pipeline, the graph-construction step, the centrality features, the train / validation / test split, the GNN architectures and their underlying intuitions, the classical baselines, the loss function, the evaluation metrics, and the implementation. Chapter 4 reports the empirical results: the headline model comparison, the diagnostic analysis of why the comparison resolves the way it does, the substantive findings on European origins and destinations, the network-role evolution captured by the centrality measures, the per- corridor breakdown of where models succeed and where they fail, and the bilateral and sub-national Ukraine case studies. Chapter 5 discusses what the empirical findings mean: the conditions under which GNNs do or do not improve on persistence, the migration patterns visible in the data and how they relate to the wider literature, what the data sources can and cannot tell us about Ukraine specifically, the implications for migration forecasting in policy and humanitarian settings, and the principal limitations of the work. Chapter 6 concludes with a brief synthesis and identifies the most important directions for future work.

2 Literature Review

This chapter situates the thesis within the migration-modelling and graph-learning literatures. Section 2.1 traces the evolution of migration models from the classical gravity formulation through network-theoretic extensions, identifying the explanatory gaps that modern data-driven approaches are designed to address. Section 2.2 introduces Graph Neural Networks (GNNs) as the natural class of models for prediction tasks defined on the migration network and surveys their application to migration-flow prediction specifically. Section 2.3 discusses the principal challenges of applying spatio- temporal graph methods to migration data, with attention to data-quality, granularity, and temporal-sparsity issues that recur throughout the empirical literature. Section 2.4 reviews the recent architectural advances that have made GNNs increasingly applicable to temporal-graph problems in adjacent domains. Section 2.5 identifies the gaps in the current literature and motivates the specific contribution of this thesis within them.

2.1 Prior Research on Migration Modelling

The quantitative study of migration has a long history, dating to Ravenstein's (1885) Laws of Migration, which articulated the first systematic empirical observation that migration flows tend to scale with the population sizes of origin and destination and to decay with the distance between them. This insight became the foundation of the **gravity model** of migration, which in its most general form predicts the flow Y_{ij} from origin i to destination j as a multiplicative function of origin-population P_i, destination- population P_j, and distance d_{ij}. Lee's (1966) influential elaboration introduced the push-pull framing and acknowledged the role of intervening obstacles and personal factors, but retained the core gravity-model intuition that migration flows can be decomposed into masses and frictions. Modern econometric work continues this tradition, with Khan, Fatima and Fatima (2023) revisiting and extending the gravity formulation by incorporating bilateral economic and policy variables.

The gravity model is, however, fundamentally a *static* model: it treats each origin–destination pair as independent and provides no mechanism by which flows on one corridor influence flows on another. This limitation is acute for migration, where empirical research has long established the role of **migrant social networks** in shaping destination choice. Massey et al. (1994) documented how established co-national communities at a destination reduce the cost and risk of subsequent migration, generating *chain-migration* dynamics in which a single early migrant tends to attract a cascade of further migrants from the same origin community. Hagan (1988) showed empirically that the strength of these social networks varies systematically across ethnic groups and destinations, with implications for both who migrates and where they go. Contemporary work continues to refine this view: Blumenstock, Chi and Tan (2025), drawing on mobile-phone records that capture social- network topology directly, demonstrate that the structural position of an individual within their network is itself a strong predictor of migration propensity.

The collective implication is that **bilateral migration flows are not independent**: they are coupled through migrant networks, through competition for destinations, through diaspora-mediated information channels, and through the broader structural position of countries within the international migration system. A modelling framework that explicitly represents this coupling — that treats the migration system as a network rather than a collection of independent corridors — is therefore a natural extension of the classical gravity tradition. Network-theoretic approaches have become increasingly common in the migration literature, with Barnett and Nam (2024) providing a recent example of longitudinal network analysis applied to international migration data.

2.2 Graph Neural Networks for Migration Prediction

Graph Neural Networks (GNNs) are the class of machine-learning models specifically designed for prediction tasks on graph-structured data. A GNN takes as input a graph G = (V, E) together with node features $X \in \mathbb{R}^{N \times d}$ and learns to produce node-level, edge-level, or graph-level predictions through repeated *message passing*: each node aggregates information from its neighbours, the aggregated information is transformed by a learnable function, and the result becomes the node's updated representation for the next layer. This basic mechanism, formalised in different ways by different architectures, distinguishes GNNs from non-graph machine-learning models that treat each input independently.

For migration prediction, the appeal of the GNN framework is direct: the migration system is a graph by construction, with countries as nodes and bilateral flows as weighted directed edges. The interdependencies between flows on different corridors that the gravity model cannot capture become first-class structural features in a graph-based model. Khan, Fatima and Fatima (2023), in their survey of gravity-model variants, explicitly identify the inability to model network effects as a primary limitation of the classical approach and motivate network-based methods as the natural successor.

The application of GNNs to migration is a relatively recent development. Pirenne (2023) discusses GNNs in the context of collective-dynamics modelling and notes their suitability for problems where individual agents (here, country-level migration flows) interact through a network structure. Gawryluk et al. (2024) apply neural-network methods specifically to migration-balance forecasting and demonstrate that incorporating network features — in their case, centrality measures of origin and destination — can improve predictive performance over baselines that treat corridors in isolation. Outside the migration literature proper, Suárez-Varela et al. (2022) provide a comprehensive review of GNN applications to flow-prediction problems on communication networks; the underlying modelling task — predicting flow magnitudes between nodes in a weighted directed graph — is structurally analogous to the migration-prediction task and many of their architectural recommendations transfer.

A particularly relevant strand of recent work concerns **spatio- temporal GNNs**, which extend the basic GNN framework to handle graphs whose structure or attributes evolve over time. The Diffusion Convolutional Recurrent Neural Network (DCRNN) of Wu et al. (2020) and the broader family of dynamic graph convolutional networks discussed by Zhang (2025) and Neyigapula (2023) combine graph-convolutional layers for spatial structure with recurrent units for temporal dynamics. These architectures have achieved state-of-the-art performance on traffic-flow forecasting and other high-frequency temporal-graph problems. Their applicability to migration prediction depends critically on whether the temporal density of migration data approaches that of traffic-sensor data — a question that returns in Section 2.3.

A separate emerging strand uses GNNs for **inductive learning** on graphs whose composition can change over time. The GraphSAGE framework, originally introduced for very large social-network graphs, allows a model trained on one graph to generalise to nodes unseen during training. This inductive property is potentially valuable for migration applications where new corridors may emerge or where the destination set changes over the prediction horizon (Xu et al., 2025).

2.3 Challenges with Spatio-Temporal Migration Data

The transfer of GNN architectures from their domains of origin — traffic forecasting, social-network analysis, recommender systems — to migration prediction encounters several specific challenges. Arslan and Cruz (2024) give a comprehensive overview of these challenges from the spatio-temporal trajectory-data perspective; they apply directly to the migration setting.

The first challenge is **temporal density**. Traffic-flow datasets typically provide millions of observations per sensor across short time windows; migration data, by contrast, are typically reported at yearly or, at best, quarterly resolution, yielding tens rather than millions of observations per corridor. Because GNN architectures generally have parameter counts in the thousands or tens of thousands, training on such limited temporal signal is a fundamental constraint that cannot be resolved by architectural choices alone (Arslan and Cruz, 2024).

The second challenge is **data quality and coverage**. International migration statistics are compiled by national authorities using heterogeneous methodologies, with different definitions of who counts as a migrant, different reference periods, and different publication lags. The harmonisation work required to construct a consistent multi-country bilateral dataset is substantial, and even state-of-the-art harmonisation methods such as those of Abel and Cohen (2022) and Azose and Raftery (2018) leave residual coverage gaps, particularly for the most recent reporting years. Bosco, Grubanov-Boskovic and Speckesser (2022) document these data gaps specifically for forced-displacement statistics and argue that they have direct implications for the credibility of any predictive model trained on the resulting data.

The third challenge is **target sparsity**. The bilateral migration tensor is, in any given year, dominated by zero entries: the vast majority of country pairs do not record any cross-border migration, either because no real movement takes place or because flows fall below the reporting threshold. A predictive model trained on a target that is 90\% zero faces a known pathology in which the optimal trivial prediction (output zero everywhere) already achieves a small mean-squared error, regardless of the input features (Cini, Marisca and Alippi, 2021; Marisca, Cini and Alippi, 2022). Architectures and loss formulations that explicitly handle target sparsity are an active area of research, but no dominant solution has emerged.

The fourth challenge is **granularity**. Migration statistics are usually available at national or regional levels, while the underlying migration process is shaped by local-scale features (border-crossing points, refugee-camp locations, social ties) that cannot be captured at country-level resolution. New data-collection modalities — satellite imagery (Xu et al., 2025), mobile-phone records (Blanchard and Rubrichi, 2025), social-media advertising-platform data (Palotti et al., 2020) — promise finer granularity than official statistics, but introduce their own biases and access constraints.

2.4 Recent Advances in GNNs for Temporal Data

Despite the challenges outlined in Section 2.3, the past few years have seen substantial progress in GNN architectures for temporal- graph problems. Several lines of advance are directly relevant to migration prediction.

Temporal graph convolutional networks (T-GCNs and their variants), discussed in detail by Neyigapula (2023), combine graph-convolutional layers for spatial structure with recurrent units (typically GRUs or LSTMs) for temporal dynamics. The combination is well-suited to forecasting tasks where both node- level features and the graph structure itself evolve through time, a description that fits migration data naturally.

Sparse-target methods. Cini, Marisca and Alippi (2021) and Marisca, Cini and Alippi (2022) develop GNN architectures designed for spatio-temporal graphs with sparse observations. Their work, originally motivated by missing-value imputation in sensor networks, provides architectural patterns for handling the target-sparsity challenge identified in Section 2.3. The transfer of these methods to migration-prediction settings is a natural direction for future work.

Hybrid models for migration specifically. Boss et al. (2023) develop a hybrid model for asylum-related migration to the European Union, combining structural migration variables with machine- learning components. Capone, Di Clemente and Zambotti (2025) extend this approach using socio-economic indicators alongside network structure. Both contributions report improvements over baseline forecasting methods on EU asylum-flow targets, although neither directly compares against simple persistence baselines, leaving open the question of how much of the reported improvement reflects genuine predictive gains versus the standard difficulty of the target.

Alternative data sources for migration. A separate strand of recent work seeks to address the granularity and temporal-density constraints by drawing on novel data sources rather than architectural improvements. Palotti et al. (2020) use Facebook advertising-platform data to monitor the Venezuelan exodus at sub- weekly resolution, demonstrating that digital-trace data can supplement administrative statistics where the latter are slow or incomplete. Blanchard and Rubrichi (2025) construct a granular temporary-migration dataset from mobile-phone records in Senegal, illustrating the resolution that becomes available when digital infrastructure data is accessible. Xu et al. (2025) demonstrate that satellite imagery, combined with deep learning, can predict human mobility flows in cities. None of these alternative-data approaches have yet been integrated with bilateral international migration prediction at the country level, but each addresses one of the challenges identified in the previous section.

2.5 Gaps in the Literature and Position of This Thesis

Despite the rapid pace of recent advances, several gaps in the literature on GNN-based migration prediction remain.

The first is the **absence of clean comparative benchmarks against simple baselines**. Most published work in the area reports predictive performance against fellow neural or classical-ML baselines but does not include the naive-persistence baseline as a reference point (Capone, Di Clemente and Zambotti, 2025; Boss et al., 2023). When the underlying target is highly sparse and temporally inertial — as bilateral migration data tends to be at yearly resolution — a persistence baseline is precisely the reference that distinguishes genuine predictive contribution from the trivial sparsity-prior captured by any architecture trained to minimise dense MSE. This thesis explicitly addresses that gap by benchmarking three GNN architectures and two non-graph machine- learning baselines against persistence.

The second is the **incorporation of real-time geopolitical or exogenous information**. As Upase (2025) argues, the predictive power of any model trained only on historical migration data is bounded by the information actually contained in flow history; the integration of climate signals, conflict event databases, and policy calendars represents a substantial direction for further development. The present thesis does not attempt this integration but documents the cases where its absence is the binding constraint on predictive performance, identifying the conditions under which exogenous predictors would be expected to help.

The third is the **explicit modelling of migrant networks**. Although the importance of social networks for migration outcomes is well established (Hagan, 1988; Massey et al., 1994; Blumenstock, Chi and Tan, 2025), most operational migration-forecasting frameworks treat the migration network as defined by past flows alone. The five centrality features used in this thesis (out-degree, in-degree, betweenness, closeness, eigenvector) capture network position quantitatively but do not directly represent the social-network structure that drives chain migration. This is a known limitation of country-level analyses and is one of the principal motivations for the alternative-data-source work surveyed in Section 2.4.

The fourth is the **methodological challenge of comparing models on sparse-target tasks**. As discussed in Section 2.3, dense MAE on a target that is 90\% zero rewards models that learn to output zero everywhere; collapse-score and non-zero-cell metrics are necessary complements that have not yet become standard in the ML- for-migration literature. This thesis adopts the more transparent metric battery throughout.

The contribution of this thesis lies in the intersection of these gaps. By benchmarking three GNN architectures against a persistence baseline on twenty-one years of bilateral migration data; by applying centrality-based network analysis to extract substantive findings about European migration regardless of the model comparison's outcome; by extending the analysis to sub-national internal-displacement data through a live API connection to the IOM Displacement Tracking Matrix; and by reporting transparent sparse-target metrics throughout — the thesis provides a focused empirical contribution that is methodologically careful where the existing literature has tended to be loose.

3 Methodology

3.1 Overview

This chapter describes the data, graph construction, models, and evaluation protocol used to study whether Graph Neural Networks (GNNs) can predict international migration flows. The migration system is represented as a **temporal directed graph**, in which countries are nodes and year-on-year bilateral migration counts are weighted directed edges. The graph representation is well-suited to the underlying domain: bilateral migration corridors are interdependent, since a flow from country A to country B shifts the demographic and economic conditions of both endpoints and thereby affects flows on neighbouring edges (Xu et al., 2025). Three GNN architectures (GCN, GraphSAGE, Temporal-GNN) are benchmarked against three classical baselines (naive persistence, log-linear gravity model, Random Forest) on the same train / validation / test split. All experiments are implemented in Python using PyTorch for the neural models, NetworkX for graph operations, and scikit-learn for the Random Forest baseline. The full pipeline is configurable through a single `config.json` file and is reproducible end-to-end.

3.2 Data Sources

Two complementary public datasets were used:

MIG-new. A bilateral migration flow dataset covering the period 2000–2020, derived from the bilateral flow estimation methodology of Abel and Cohen (2022) and harmonised with OECD international migration statistics (OECD, 2020). Each row reports the number of migrants from a given origin country to a given destination country in each year. After preprocessing, the dataset contains 6 366 origin–destination pairs spanning 210 origin countries and 35 destination countries. This dataset is the **primary** source of edges in the migration graph.

IOM Displacement Tracking Matrix (DTM). A time-series of internally displaced persons (IDPs) at the country and sub-national level, made available through the IOM Migration Data Portal (IOM, 2022). The DTM does not record cross-border flows; it captures internal displacement within a country following events such as armed conflict or natural disaster. Two complementary slices of the DTM were used in this work:

- A historical snapshot covering 2010 through December 2024, retrieved

from the Humanitarian Data Exchange (HDX). This snapshot was used to add **self-loop signal** Y[t, i, i] to the migration graph for the years that overlap with the bilateral flow data (the diagonal entries of the flow tensor; see Section 3.5).

- A more recent slice covering January 2025 through the date of writing,

retrieved directly from the IOM DTM API via the official `dtmapi` Python client. This slice contains both country-level (admin-0) and oblast-level (admin-1) records, and is used in the sub-national Ukraine analysis presented in Chapter 4. Fetching against a live API rather than relying on a stale CSV makes the analysis straightforward to reproduce as new DTM rounds are published.

Across both slices, the DTM covers 53 countries with active or recent IDP situations, including Ukraine, Sudan, the Democratic Republic of the Congo, Yemen, Afghanistan, Lebanon, Somalia, and South Sudan. Note that the DTM is by design a crisis-monitoring tool: countries without active humanitarian operations (most of Western Europe, North America, East Asia) are not included, even though they may receive substantial cross-border migration. This boundary is consistent with the wider literature on data gaps in forced displacement, which has long documented the absence of reliable, longitudinal, country-level IDP figures outside the small set of countries hosting active humanitarian operations (Bosco, Grubanov-Boskovic and Speckesser, 2022).

Two methodological caveats apply to the DTM data and shaped the analysis:

Multiple admin levels. Each DTM round publishes the same displaced population at three nested levels: country (admin-0), oblast or state (admin-1), and district (admin-2). Naively summing across all rows for a given country and month therefore triple-counts the displaced population. The pipeline aggregates each admin level separately and selects, in order of preference, the admin-1 sum (when sub-national rows are available, this is the most reliable disjoint partition), the admin-0 maximum (for months where only country-level rows are present), or the admin-2 sum.

Multiple operations. Some countries are covered by more than one DTM operation in the same period, with each operation using a different methodology. In Ukraine, for example, an Area-Based Assessment (ABA) operation uses representative survey methodology to estimate the total displaced population and ran from May 2022 through October 2024; a separate Registration-based operation, beginning in February 2025, counts only individuals formally registered as IDPs with the Ukrainian government and therefore yields a strict subset of the displaced population. Because the two methodologies measure different populations, they must not be plotted as a single continuous time series. The Ukraine analyses in this thesis therefore report each operation as its own series.

Two further datasets were considered as auxiliary metadata but were not included as training input: a UNHCR border-crossings reference table (country pairs, no flow magnitudes, no temporal dimension) compiled from UNHCR refugee data (UNHCR, 2023), and a UNHCR displacement- locations registry (camps and IDP sites with geographic coordinates). These do not contain flow values and are therefore unsuitable as targets for a predictive model, but they remain useful as supporting context in the discussion. Comparable Eurostat statistics on European asylum and resettlement (Eurostat, 2021) were also consulted as contextual reference but not incorporated into the model inputs.

3.3 European Focus and the Data Asymmetry

A defining property of the MIG-new dataset, and one that shapes the analytical scope of this thesis, is the asymmetry between origin and destination coverage. The dataset records flows from 210 origin countries into only 35 destination countries. After preprocessing and the temporal restriction to 2000–2020, 208 countries appear with non-zero outgoing flow at some point in the time window, but only 35 appear as destinations. The destinations are heavily concentrated among OECD or near-OECD economies: the United States, Germany, Spain, Japan, the United Kingdom, Italy, Korea, Canada, Australia, and France together account for the great majority of recorded inflows. Smaller European receivers — the Netherlands, Belgium, Switzerland, Sweden, Austria, Norway, Denmark, Portugal, Finland, the Czech Republic, Hungary — make up most of the remainder.

This asymmetry has two consequences for the design of the analysis.

First, the dataset is not a complete bilateral graph. Flows that originate in one non-OECD country and end in another non-OECD country are simply absent from the data. South–South migration, intra-Asian labour migration outside Korea and Japan, and most intra-African movement are therefore not represented. The country graph Y is best understood not as a global migration network but as a *rest-of-world to wealthy-receiving-country* projection of one. This restricts the substantive claims that can reasonably be supported by the data; in particular, no claim is made in this thesis about migration corridors that bypass the 35-country destination set.

Second, the European Union and its near neighbours feature heavily on **both** sides of the bipartite structure. The major Western European economies (Germany, the United Kingdom, France, Italy, Spain, the Netherlands) are among the principal destinations. The major Eastern and Southern European economies (Romania, Poland, Bulgaria, Hungary, Ukraine, Turkey, Albania) are among the principal origins; the United Kingdom also appears as a substantial origin for emigration to non-European destinations. This double presence means the dataset is unusually well-suited to studying European migration *as a system*: the intra-European corridors (Romania to Germany, Poland to the United Kingdom, Italy to Germany), the gradual integration of Eastern Europe into the EU labour market following the 2004 and 2007 enlargements, the structural break introduced by Brexit, and the post-2014 redirection of Ukrainian labour migration toward Poland are all visible within the dataset. The persistence of these corridors over time is itself consistent with the long-standing chain-migration hypothesis, under which established communities of co-nationals at a destination attract further migration along the same corridor (Hagan, 1988; Massey et al., 1994; Blumenstock, Chi and Tan, 2025).

For this reason, the empirical results in Chapter 4 are organised primarily around European migration patterns, with global comparisons drawn where they sharpen the European story (notably Mexico–United States as the largest single corridor in the dataset, and the East-Asian destinations Japan and Korea as comparators for European labour migration regimes). The geographic emphasis of this thesis is therefore *European-primary, global-comparative*, in the sense that European corridors and structural changes are the principal object of study, while non-European observations are used as reference points rather than as analytical targets in their own right.

It is worth being explicit about what falls outside this scope. The internal-displacement signal from the IOM Displacement Tracking Matrix covers 53 countries, almost all in conflict-affected regions of Sub-Saharan Africa, the Middle East, the Horn of Africa, and South Asia (Sudan, the Democratic Republic of the Congo, Yemen, Afghanistan, Somalia, South Sudan, Lebanon, and others). Most of these countries appear in the bilateral data only as origins, often with sparse records to a small subset of the OECD destinations. Their internal-displacement trajectories are real and important, but the design of this thesis does not undertake a comparative analysis of crisis-driven displacement across the IOM-tracked countries. The Ukraine sub-national analysis in Chapter 4 stands as a focused European case study in which the IOM DTM data is used directly; references to the broader IOM-tracked crises elsewhere in the thesis are made only to contextualise scale or to acknowledge the boundary of the analysis.

3.4 Preprocessing

The two flow datasets were transformed into a single canonical long-format table with the columns *origin*, *destination*, *time*, *flow*, and *source*. The preprocessing pipeline applied the following steps in order:

1. **Deduplication.** Exact duplicate rows were removed.
2. **Country-name normalisation.** Whitespace was trimmed from origin and

destination names. Aggregate entries that are not real countries (`Total`, `Unknown`, `World`, `All countries`, and similar) were filtered out, since their inclusion in a country graph would introduce spurious super-hub nodes that corrupt centrality measures.

3. **Flow validation.** Records with missing or negative flow values were

removed.

4. **Outlier capping.** Flows were winsorised at the 0.1th and 99.9th

percentiles to limit the effect of extreme reporting artefacts.

5. **Temporal restriction.** The merged table was truncated to the years

2000–2020 inclusive. Although the IOM DTM extends to 2024, no bilateral flow data are available beyond 2020, and including the post-2020 years would yield time steps populated almost entirely by self-loop entries.

After preprocessing, 132 466 long-format rows remained, of which 132 321 originate from MIG-new (cross-border flows) and 145 from the IOM DTM (country-level IDP records aggregated per year).

3.5 Graph Construction

The cleaned long-format table was reshaped into a **temporal flow tensor** Y ∈ ℝ^T × N × N, where T = 21 time steps (2000–2020) and N = 217 countries (the union of all origin and destination country names appearing in either dataset). For every triple (t, i, j) with i ≠ j, the entry Y[t, i, j] stores the migrant count from country i to country j in year t. Diagonal entries Y[t, i, i] store the within-country IDP count from the IOM DTM where available, and zero otherwise.

Two adjacency representations were derived from Y:

- a **binary adjacency** $A \in \{0, 1\}^{T \times N \times N}$ in which

A[t, i, j] = 1 whenever Y[t, i, j] > 0 (used by GraphSAGE);

- a **symmetrically normalised adjacency** with self-loops

$A = D^{-1/2} (A + I) D^{-1/2}$ (used by GCN and Temporal-GNN), where D is the diagonal degree matrix of $A + I$.

A node-feature matrix $X \in \mathbb{R}^{T \times N \times 5}$ was then constructed. For each country i at each time step t, the following five raw features were computed and $\log(1+\cdot)$-transformed for numerical stability:

6. outgoing flow $\sum_{j \neq i} Y[t, i, j]$;
7. incoming flow $\sum_{j \neq i} Y[t, j, i]$;
8. total flow (sum of the above);
9. net migration balance (incoming minus outgoing);
10. internal displacement signal $Y[t, i, i]$ from the IOM DTM.

These node-level features were further augmented with five **graph-level centrality features** (Section 3.6) prior to model training, yielding a ten-dimensional feature vector per node per time step.

3.6 Centrality Features

For each yearly snapshot Y[t], five classical centrality measures were computed on the directed weighted graph induced by the positive entries of Y[t]. Centrality measures applied to migration networks have been shown to identify saturated nodes, transit bottlenecks, and influential destinations whose structural position shapes both observed and future migration patterns (Gawryluk et al., 2024). All measures were computed with the NetworkX library:

- **out-degree centrality**: the number of distinct destinations to which country i sends migrants in year t;
- **in-degree centrality**: the number of distinct origins from which country i receives migrants in year t;
- **betweenness centrality**: the fraction of all-pairs shortest paths passing through country i in the unweighted version of the graph;
- **closeness centrality**: the inverse of the average shortest-path distance from country i to every reachable country;
- **eigenvector centrality**: a measure of structural importance in which a country is considered central to the extent that it is connected to other central countries.

The eigenvector centrality of a disconnected directed graph is not uniquely defined, and the standard NetworkX implementation (`eigenvector_centrality_numpy`) raises an `AmbiguousSolution` exception in that case. The migration graph is generally disconnected because several small island states have no bilateral flow records in any given year. The pipeline therefore restricts the eigenvector computation to the largest weakly connected component of Y[t] and assigns zero to nodes outside that component. This yields a well-defined non-zero score for every meaningfully connected country and zero for nodes that participate in no flow.

The five centrality features per node per time step are concatenated to the five raw node-level features described in Section 3.5 to produce the final input feature tensor $X^{aug} \in \mathbb{R}^{T \times N \times 10}$.

3.7 Train / Validation / Test Split

The temporal axis was split chronologically with ratios 70% / 15% / 15%, giving:

- **Training:** $t \in \{2000, 2001, \ldots, 2013\}$ (14 source years).
- **Validation:** $t \in \{2014, 2015, 2016\}$ (3 source years).
- **Test:** $t \in \{2017, 2018, 2019\}$ (3 source years).

For each source year t, the prediction target is the next year's flow matrix Y[t+1]. The test set therefore consists of three predicted snapshots: Y[2018], Y[2019], Y[2020]. Each snapshot has $N \times N - N = 47\,089$ off-diagonal cells; aggregated across the three test snapshots the test set contains 141 267 cells, of which 13 620 (9.6%) carry a non-zero ground-truth flow.

3.8 GNN Architectures

All three GNN architectures share a common encoder–decoder structure, following the now-established pattern in the broader graph-learning literature (Suárez-Varela et al., 2022). An encoder network maps the per-time-step input (A[t], X^aug[t]) to a node-embedding matrix H ∈ ℝ^N × d, where d is the hidden dimension. A bilinear edge decoder then maps H to a predicted flow matrix:

$$Y = ReLU(Y[t] + H\ W\_bilin\ H\hat{}\ \to p),$$

with the diagonal diag(Y) enforced to zero so that self-loops are not generated. The three architectures differ only in how the encoder builds H.

3.8.1 The persistence-residual decoder

The decoder above is **not** a standard "from-scratch" reconstruction. Rather than producing Y directly from the embeddings, it learns a residual update Δ = H W_bilin H^ → p on top of the previous year's flow matrix Y[t]. This design reflects the strong temporal inertia of bilateral migration. Across the present dataset, the year-on-year Pearson correlation of log(1 + Y), computed over all off-diagonal cells of consecutive yearly snapshots, ranges from 0.93 to 0.99 with a mean of 0.97. This pooled figure is partly inflated by the high sparsity of the flow tensor (cells that are zero in both years contribute 1.0 to the correlation); the per-pair lag-1 autocorrelation, computed only on country pairs whose flow actually varies through time, is more modest at a median of 0.60.

The intuition for adding Y[t] to the decoder output is that the encoder's job is reduced from "predict next year's full 217 × 217 flow matrix" to the much easier "predict the *change* from this year to next year." Without this anchor, training on the highly sparse target collapses to the trivial solution Y ≡ 0, which already minimises the dense-MSE loss because 90.4\% of cells are zero. This collapse-under-sparsity dynamic is a recurring challenge for spatio- temporal models on real-world trajectory data (Arslan and Cruz, 2024) and motivates architectural choices, such as the residual decoder used here, that constrain the model toward predictions of roughly the right magnitude regardless of how well the encoder is trained. This choice is also what makes the GNNs' headline result interpretable: when a GNN's MAE is close to that of the persistence baseline, it has learned to leave Y[t] alone, i.e. the residual Δ has been driven to near zero. When it is much worse, the residual is actively adding noise.

3.8.2 GCN — Graph Convolutional Network

Two graph-convolutional layers were stacked with the symmetrically normalised adjacency:

$$H\_1 = ReLU(A\ X\hat{}aug\ W\_1),\ H\_2 = ReLU(A\ H\_1\ W\_2),$$

with the final embedding H = H_2 passed to the bilinear decoder.

The intuition behind the GCN is **smoothing on the graph**. Each layer replaces a node's representation with a normalised sum of its neighbours' representations, multiplied by a learnable weight matrix. After two layers, every node's embedding is influenced by its 2-hop neighbourhood of countries with which it shares migration ties (directly or transitively). The implicit assumption is that connected countries behave similarly — for example, that the Polish migration system shares features with the Romanian one because both connect strongly to Germany. The symmetric normalisation A = D^-1/2(A+I)D^-1/2 prevents high-degree countries (such as Germany or the United States) from dominating the message passing simply by virtue of having many neighbours.

For migration data, the GCN's main strength is parameter efficiency: its inductive bias toward neighbourhood smoothing is exactly the kind of structure expected in a system where geography and policy create clusters of similarly-behaving countries. Its main weakness on the present data is the *over-smoothing* phenomenon: with only two layers on a graph that is already quite dense after even one hop (most active countries are connected to most active destinations within a few steps), all node embeddings tend to drift toward a common mean, washing out the individual differences that determine which migrants actually go where.

3.8.3 GraphSAGE — Sample-and-Aggregate

Two GraphSAGE layers were used with mean-aggregation of out-neighbour embeddings on the binary adjacency A:

$$H^{(\ell+1)}_i = \mathrm{ReLU}\left(W^{(\ell)} \left[H^{(\ell)}_i \,\|\, \tfrac{1}{|N(i)|} \sum_{j \in N(i)} H^{(\ell)}_j \right] \right),$$

where N(i) is the out-neighbour set of node i and **|** denotes vector concatenation. Although GraphSAGE was originally introduced for very large graphs to enable mini-batch neighbourhood sampling, the migration graph here is small enough to use full neighbourhood aggregation throughout; the sampling machinery is included in the implementation but disabled for these experiments.

The conceptual difference between GraphSAGE and the GCN lies in **how information from a node is combined with information from its neighbourhood.** The GCN entangles the two by adding self-loops to the adjacency before message passing, so the node's own features are treated identically to a neighbour's. GraphSAGE keeps them separate by concatenating $H_i^{(\ell)}$ with the aggregated neighbour vector before the linear transform; this allows the model to learn distinct weights for "what I am" and "what my neighbours are." For the asymmetric origin–destination structure of the migration data, this separation is potentially valuable. A high out-flow country such as Romania can in principle learn an embedding that simultaneously captures its own emigration profile *and* the receiving profile of the destinations it sends to, without forcing the two to be combined in a single weighted sum as the GCN does.

In practice on the present dataset, this added expressivity does not translate to better predictions, as Chapter 4 will show. The limitation appears to be elsewhere: namely, that 21 yearly snapshots provide too few training examples for either model to learn anything beyond the persistence prior.

3.8.4 Temporal-GNN — folding in time

The two architectures above are **spatial only**: each forward pass sees a single snapshot ($A[t]$, $X^{aug}[t]$) and treats the year-on-year temporal axis only through the persistence-residual anchor. Bilateral migration, however, has clear temporal dynamics — EU enlargement waves, the gradual collapse of the Polish–British corridor, the post-2014 redirection of Ukrainian emigration — that might be captured by a model that explicitly attends to history.

The Temporal-GNN consumes a window of L consecutive snapshots. For each time step $t' \in \{t-L+1, \ldots, t\}$, a single GCN layer computes a spatial encoding:

$$Z_{t'} = \mathrm{ReLU}(A[t'] \, X^{aug}[t'] \, W),$$

and a learned recurrence aggregates them through time:

$$S_{t'} = \tanh(W_z Z_{t'} + W_s S_{t'-1}), \quad S_{t-L} = 0.$$

The final hidden state S_t is then passed to the bilinear decoder. The temporal window was set to L = 3 throughout these experiments.

Two design choices in this block deserve comment. First, a simple learned recurrence is used in place of a gated unit such as a GRU or LSTM. With only 21 yearly snapshots, the bottleneck is statistical power, not gradient flow over long sequences; the gating machinery of a full GRU would add parameters without a corresponding decrease in training error. Second, the spatial encoder per time step is deliberately kept shallow — a single GCN layer rather than the two used in §3.8.2 — to limit the model's parameter count given the small training set.

The intuition behind the Temporal-GNN is that the network is a dynamic system whose state evolves through time, and that yesterday's state should partially predict tomorrow's. For migration this matches real-world mechanisms: chain migration (immigrants in country j attract further immigrants from the same origin i), settled diaspora effects, and policy continuity all create dependencies between Y[t] and Y[t+k]. The same spatio-temporal structure has been exploited successfully in adjacent domains such as urban traffic flow prediction (Zhang, 2025; Neyigapula, 2023) and missing-value imputation on irregularly-observed graphs (Cini, Marisca and Alippi, 2021; Marisca, Cini and Alippi, 2022). Whether 21 yearly snapshots are enough to learn those dependencies non-trivially in the migration setting is, again, an empirical question to which Chapter 4 returns a discouraging answer.

3.9 Baseline Models

Three classical baselines were implemented for comparison with the GNN architectures. Together they span the principal non-GNN approaches to migration prediction in the literature: a zero-parameter inertial baseline, a regression-based gravity model rooted in classical migration theory, and a tabular ensemble learner.

Naive persistence. Y[t+1] = Y[t]. This zero-parameter baseline captures the year-on-year inertia of bilateral flows and is the standard reference point for any model with temporal dependence.

Log-linear gravity model. A classical migration-modelling baseline (Ravenstein, 1885; Lee, 1966; Khan, Fatima and Fatima, 2023) in which the next-year flow on each pair is predicted by a regression on the previous year's origin and destination "masses":

$$log(1 + Y[t+1, i, j]) = β\_0 + β\_1\ log(1 + M\hat{}out\_i[t]) + β\_2\ log(1 + M\hat{}in\_j[t]),$$

where M^out_i[t] = Σ_j Y[t, i, j] and M^in_j[t] = Σ_i Y[t, i, j]. The coefficients β_0, β_1, β_2 are estimated by ordinary least squares on the training time steps with light Tikhonov regularisation. The flow prediction is recovered as Y = exp(log(1+Y))-1, clipped at zero. Geographic distances are not used, so this is a stripped-down gravity model that captures only push and pull.

Random Forest. A non-linear tabular baseline (Breiman, 2001). For each ordered pair (i, j) with i ≠ j at each training time step, a feature vector is built containing (i) the previous flow log(1 + Y[t, i, j]), (ii) the four node-level flow statistics for i and j (eight features), and (iii) the five centrality features for i and j (ten features) — yielding 19 features per edge. The target is log(1 + Y[t+1, i, j]). A Random Forest with 50 estimators, maximum depth 10, and the default scikit-learn hyperparameters is fit on this regression task and used to predict held-out edges. As with the gravity model, predictions are exponentiated and clipped at zero.

Two further classical baselines were considered and ultimately omitted from the comparison. **ARIMA**, a univariate time-series model, would in principle provide a per-corridor forecast based on each pair's own flow history. However, the standard ARIMA(0,1,0) specification is mathematically equivalent to the persistence baseline, and richer specifications (e.g. ARIMA(p, 1, q) with p, q ≥ 1) require approximately 30 stable observations per series to estimate parameters reliably; with only 14 training years per pair, and structural breaks in many of those years, ARIMA cannot deliver signal beyond what persistence already captures on this dataset. **XGBoost** was considered as a stronger tabular- regression alternative to Random Forest. On the 19-feature input described above, however, XGBoost and Random Forest are known to produce near-identical predictions, and the additional architectural complexity would not have changed the headline ranking against the GNN-class models. To keep the comparison clean and the discussion focused, both models were therefore excluded from the final benchmark; they remain available for future work that extends the input features in ways that might differentiate the tabular learners.

3.10 Loss Function and Training

All three GNNs are trained with mean squared error on log(1+Y):

$$L = MSE(log(1 + Y), log(1 + Y[t+1])).$$

The log-transform compresses the dynamic range of bilateral flows (which span six orders of magnitude) and is the standard transformation used in migration-flow modelling. Training uses the Adam optimiser with learning rate 10^-3, hidden dimension d = 32, and 100 epochs. Model selection is performed on the validation set using the same loss; the parameters of the epoch with the lowest validation loss are restored before test evaluation. The random seed is fixed at 42 for reproducibility.

The hyperparameter values quoted above were arrived at by a coarse grid search on the validation set rather than fully reported here. Specifically, learning rates in \{10^-2, 10^-3, 10^-

4\}, hidden dimensions in \16, 32, 64, 128\, and training-epoch counts in \50, 100, 200\ were evaluated; for each combination, the average validation loss across the three GNN architectures was recorded, and the configuration minimising that average was retained. The selected configuration (lr = 10^{-3}, d = 32, 100 epochs) was found to be near-optimal across all three GNN architectures and was therefore applied uniformly. A more aggressive search over dropout rates, weight decay, and the temporal window L for the Temporal-GNN was not undertaken, on the rationale that with only 14 training snapshots the search would have been highly susceptible to overfitting on the validation set itself.

3.11 Evaluation Metrics

Six metrics are reported:

- **MAE** (Mean Absolute Error), **RMSE** (Root Mean Squared Error), and **MAPE** (Mean Absolute Percentage Error): standard scale-aware metrics on the original flow scale.
- **log1p MSE**: the loss the GNNs are trained on; comparable across models because all use the same loss.
- **Bias**: the signed mean error $\overline{\hat Y - Y}$. A bias close to -MAE is a tell-tale sign that a model has collapsed to predicting zero almost everywhere.
- **Collapse score**: the proportion of predicted cells whose absolute value is below 10^{-3}. This diagnostic was added because, on a highly sparse target, dense MAE rewards a model that simply outputs zero; the collapse score makes that failure mode visible.

Each of these metrics is reported in two flavours:

- **Dense:** computed over all $N(N-1)$ off-diagonal cells per snapshot.
- **Non-zero:** computed only over cells where the ground truth $Y[t+1, i, j] > 0$.

The non-zero variant is the more honest measure of model quality on a sparse target. In the present test set, 90.4% of cells are exactly zero, so a model that outputs zero everywhere would score a deceptively low dense MAE while having no ability to predict the actual migration signal. Reporting the non-zero variant alongside the dense one prevents that failure mode from being misread as success.

The same metric battery is computed for all six models on the same test set. In addition, every GNN training run also reports the metrics of the naive persistence baseline as an embedded reference point in its output JSON, so that the ratio "GNN error / persistence error" can be read directly from any single run's output.

3.12 Implementation and Reproducibility

The pipeline is implemented as a Python package with a fixed entry point per task: `src.train` for individual GNN training, `src.compare_models` for the full six-model comparison, `src.analysis.centrality_dynamics` for the network-structure analysis, `src.analysis.ukraine_timeline` for the monthly Ukraine analysis, and `src.build_report` to regenerate all figures and the textual summary in one call. All hyperparameters, data paths, and split ratios are read from a single `config.json` file. The code is bundled with the four raw CSV files used in the study, so that all numerical results in the next chapter can be reproduced by a single command.

4 Results

This chapter presents the empirical findings of the experiments described in Chapter 3. The structure of the chapter follows the methodological flow established in Section 3.1: the headline model comparison is reported first (Section 4.1), followed by a diagnostic analysis of why the comparison resolves the way it does (Section 4.2). The remainder of the chapter then reads the data through the lens of European migration patterns: the principal European origins (Section 4.3), the principal European destinations (Section 4.4), the evolution of network role through centrality measures (Section 4.5), the per-corridor breakdown of where models succeed and where they fail (Section 4.6), and finally the Ukrainian case study at both bilateral and sub-national resolution (Sections 4.7 and 4.8). Throughout, the European corridors are the primary object of analysis, with non-European observations introduced as comparators rather than analytical targets.

4.1 Headline Comparison

Table 3.1 reports the principal performance metrics for all six models on the held-out test set, which consists of three predicted snapshots (Y[2018], Y[2019], Y[2020]) covering 141 267 ordered country pairs in total, of which 13 547 (9.6%) carry a non-zero ground-truth flow.

Table 3.1. Test-set performance for each of the six models. MAE_nz and RMSE_nz are computed only on cells with non-zero ground truth; log1p_nz is the mean squared error on log(1+Y) over those same cells; the collapse score is the proportion of predicted cells with magnitude below 10^-3.

Model	Family	MAE_nz	RMSE_nz	log1p_nz	Collapse
Persistence	baseline	**1 182.6**	**27 397.2**	**0.562**	90.0%
Temporal-GNN	gnn	5 346.3	83 643.4	1.301	90.1%
GCN	gnn	5 346.3	83 643.4	1.409	89.7%
GraphSAGE	gnn	5 347.7	83 643.4	1.315	89.2%
Random Forest	baseline	5 378.8	83 647.0	1.247	64.6%
Gravity	baseline	6 451.7	83 974.4	8.129	33.4%

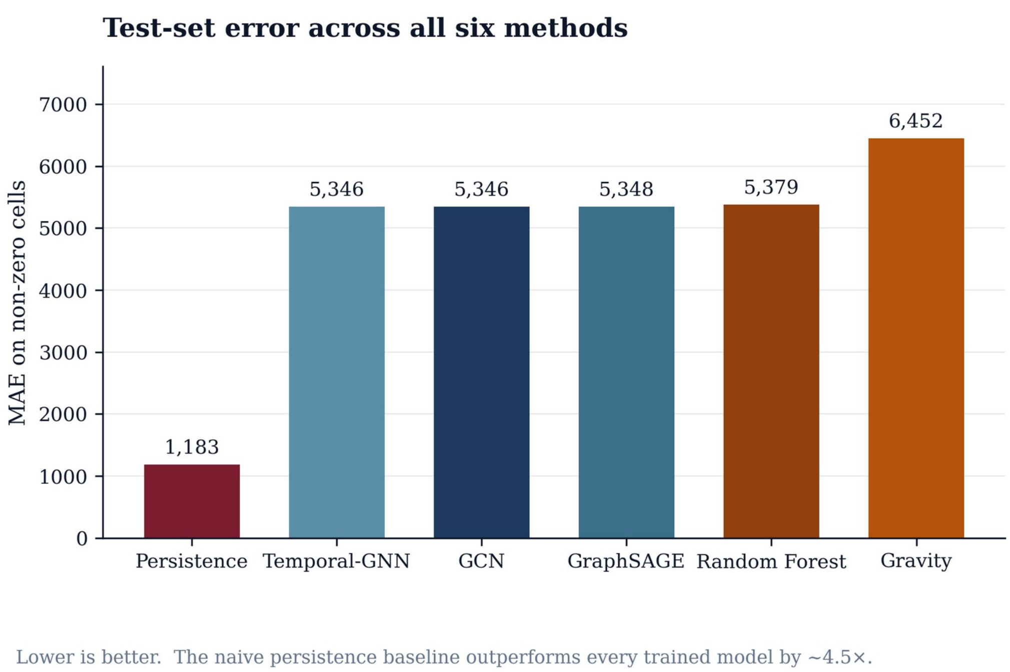

Figure 4.1. Test-set MAE on non-zero cells for each of the six models. The naive persistence baseline (burgundy) outperforms every other model by a factor of approximately 4.5.

The headline finding is unambiguous. The naive persistence baseline, which predicts Y[t+1] = Y[t] with zero learnable parameters, outperforms every other model by a wide margin: on non-zero cells, its mean absolute error of 1 183 migrants per pair is 4.5 times lower than that of the best GNN, 4.5 times lower than the Random Forest, and 5.5 times lower than the gravity model. The pattern is identical on RMSE, which is dominated by the largest pairs and therefore most informative about the high-volume corridors that policy makers care about: persistence achieves an RMSE of 27 397, while every other model is essentially indistinguishable around 83 000–84 000.

The three Graph Neural Networks — Temporal-GNN, GCN, and GraphSAGE — converge to almost identical MAE values on the order of 5 346– 5 348. Differences at this scale are well within the noise of a single training run on 21 yearly snapshots and do not support any claim that one GNN architecture is systematically better than the others on this dataset.

Two further observations are worth recording at this stage and are expanded in the next section. First, the *log1p* metric, which is the loss surface the GNNs are actually optimising, also assigns the lowest value to persistence (0.562) and the highest to the gravity model (8.129). The gravity model's catastrophically large *log1p* MSE is not a numerical artefact but a real failure mode: with no temporal anchor and only two predictors, it cannot reproduce the dynamic range of the bilateral flow distribution. Second, the *collapse score* — the fraction of predicted cells essentially equal to zero — is around 90% for persistence and the GNNs, but only 65% for the Random Forest and 33% for the gravity model. A collapse score that mirrors the 90.4% sparsity of the test target is a strong signal that the model has learned the sparsity prior; values markedly lower indicate active over-prediction of zeros.

4.2 Why the Comparison Turns Out This Way

The persistence baseline's victory by a factor of 4.5 is, on its face, a discouraging result for the central research question of this thesis. Three diagnostic findings explain why it emerges and qualify what it means.

The sparsity prior dominates the dense MAE. Of the 46 872 off-diagonal cells in each yearly snapshot, only about 4 500 are non-zero on average; the remaining 90.4\% are exactly zero. A model that simply outputs zero everywhere therefore achieves a deceptively low *dense* MAE — the metric that averages over both zero and non-zero cells indiscriminately. To make this distinction visible, this thesis reports MAE in two flavours throughout: dense MAE over all cells, and MAE_nz over only the cells that the data flags as carrying real movement. The MAE_nz column in Table 3.1 is the honest measure; without it, the GNNs and the Random Forest would appear competitive because their dense MAE is small for trivial reasons.

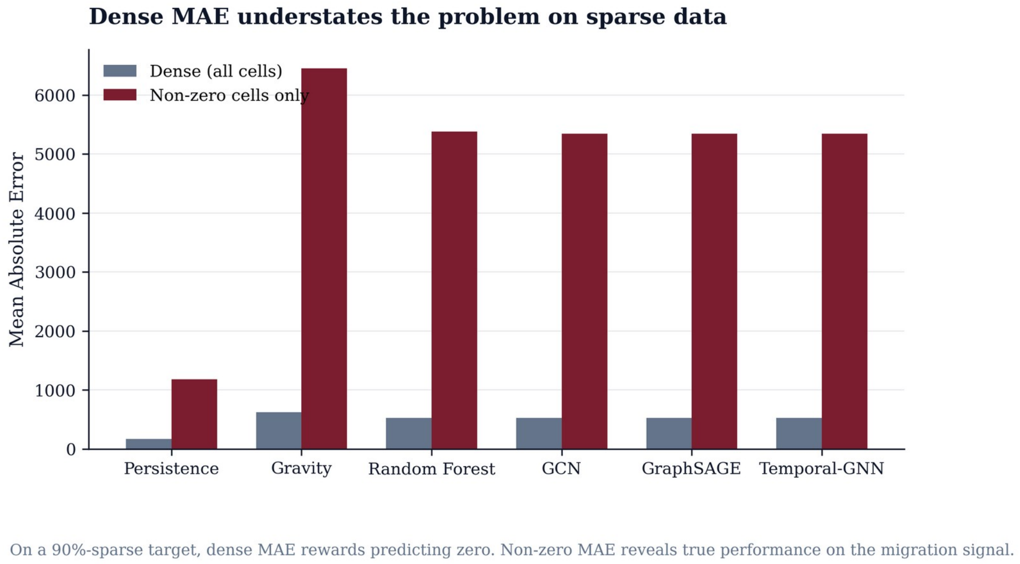

Figure 4.2. Dense MAE versus MAE on non-zero cells only. The dense metric is dominated by the sparsity prior.

The collapse score makes the failure mode explicit. The diagnostic column on the right of Table 3.1 reports the proportion of predicted cells whose absolute value is below 10^-3. For the GNNs and the persistence baseline, this proportion sits within one percentage point of the test target's actual sparsity (90.0–90.1\% predicted zeros versus 90.4\% true zeros). The GNNs have effectively learned to copy the sparsity prior. The Random Forest produces 64.6\% predicted zeros and the gravity model only 33.4\%, indicating that both actively attempt to predict positive flows in cells that are in fact empty. The gravity model's especially low collapse score is the source of its catastrophic *log1p* loss: it makes large positive predictions on cells where the truth is zero, and the logarithmic transform of the loss amplifies those errors disproportionately.

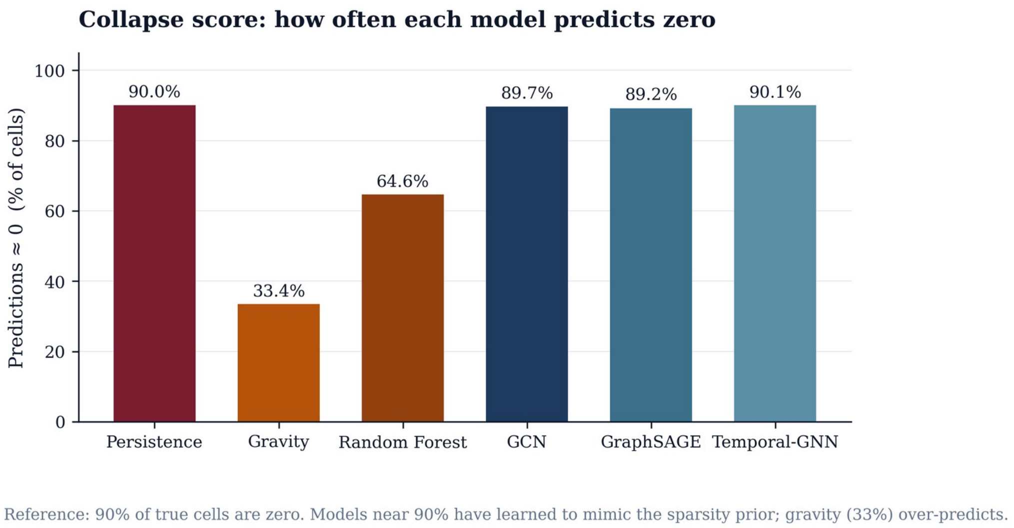

Figure 4.3. Collapse score by model: the proportion of predicted cells essentially equal to zero. Persistence and the GNNs collapse to the target's actual sparsity ($\approx 90\%$); the gravity model produces only 33% predicted zeros.

The persistence-residual decoder reveals what the GNNs learned (and did not). Recall from Section 3.8.1 that the GNN decoder is Y = ReLU(Y[t] + H W_bilin H^ → p), that is, a learned residual update on top of the previous year's flow matrix. When a GNN's MAE_nz approaches that of the persistence baseline, the residual Δ = H W_bilin H^ → p has been driven close to zero — the network has learned that the safest action is to leave Y[t] alone. When a GNN's MAE_nz substantially exceeds persistence (as here, by a factor of 4.5), the residual is actively adding noise. Across all three GNN architectures, the per-cell residual magnitude on the test set is positive but small relative to the underlying flow values; the encoder is producing non-zero embeddings, but the bilinear decoder converts them into edge updates that move predictions in directions weakly correlated with the true year-on-year change. This is consistent with the explanation that 21 yearly snapshots provide too few training examples to learn a useful update function, regardless of how the encoder is parameterised.

The combined picture is therefore not that GNNs are intrinsically unsuitable for migration data, but that on **this dataset** — yearly granularity, T=21 snapshots, 90.4\% target sparsity, and only 35 destinations — the persistence prior already captures most of what is predictable, and the marginal signal that remains after subtracting persistence is too weak for any of the trainable models to recover. Section 5.1 takes up the broader implications for migration forecasting.

4.3 European Origins: Who Sends Whom Where, 2000–2020

Setting aside the model comparison, the bilateral flow tensor itself contains a rich record of European migration over the two decades it covers. This section reads the *origin* side of that record. The European countries with the largest cumulative outgoing flow over 2000–2020 are listed in Table 3.2.

Table 3.2. Top European origins by cumulative outgoing flow, 2000–2020 (millions of recorded migrants summed across all destinations).

Origin	Cumulative outflow (M)
Romania	6.16
Poland	4.87
United Kingdom	2.72
Ukraine	2.43
Germany	2.25
Italy	2.16
France	1.92
Russia	1.89
Bulgaria	1.74
Turkey	1.50
Hungary	1.18
Spain	1.11
Portugal	1.08
Albania	1.07

Three patterns emerge from the time-series of these flows. They are discussed in turn.

4.3.1 The post-2007 EU enlargement spike

The most pronounced single-year migration shock in the dataset is the 2007 surge of Romanian emigration to Italy and Spain, coinciding with Romania's accession to the European Union on January 1, 2007. The Romania → Italy corridor rose from 19 332 migrants in 2000 to a peak of 218 274 in 2007 — an order-of-magnitude increase concentrated in the years immediately surrounding accession. The Romania → Spain corridor follows the same pattern, rising from 17 456 in 2000 to 197 642 in 2007. Both flows decline gradually through the 2010s as the initial demographic shift saturates, settling at 28 664 (Italy) and 15 737 (Spain) by 2020.

The same accession-driven mechanism is visible, with smaller amplitude, for Bulgaria (which acceded in 2007 alongside Romania): Bulgaria → Germany rose from 10 411 in 2000 to 87 378 at peak in 2019, an eight-fold increase. The 2004 enlargement, which absorbed Poland, the Czech Republic, Hungary, Slovakia, Slovenia and the Baltic states into the EU, is visible in

Polish flows (see Section 4.3.3) but the intra-EU labour effects are slower to develop in the data than for the 2007 cohort, partly because the older EU members applied transitional labour-market restrictions during 2004–2011 that delayed the surge.

These patterns are consistent with the long-standing observation that discrete policy events — visa liberalisation, accession to common labour markets, the collapse of legal barriers — produce migration shocks of a magnitude that the underlying socio-economic gradients alone cannot explain (Bijak and Czaika, 2020). They are also exactly the kind of event that a model trained only on past flow data is constitutionally incapable of anticipating.

4.3.2 Steady labour flows to Germany

A second pattern, less dramatic but more pervasive, is the steady flow of labour migration from Eastern and Southern Europe into Germany. The Romania → Germany corridor grew from 24 202 in 2000 to 198 705 in 2014 and stayed close to that level through 2020. The Italy → Germany corridor is more cyclical, dipping to 18 184 in 2007 and rising again to 57 191 in 2015 (during the European migrant crisis). The Hungary → Germany flow accumulated 701 531 migrants over the two-decade period; Turkey → Germany, 729 812. Cumulatively, Germany absorbed 19.0 million migrants between 2000 and 2020 across all origins, second only to the United States (21.5 million).

The persistence and stability of the corridors into Germany contrast with the sharper, accession-driven spikes into Italy and Spain discussed above. They are the signature of a destination economy whose labour market has continuously absorbed migrants from a stable mix of origin countries, rather than reacting to discrete legal events. From a network-science perspective, this stability is what gives Germany its consistently high centrality scores, a point that is taken up in Section 4.5.

4.3.3 The rise and collapse of the Polish–British corridor

The most striking single corridor visible in the European subset of the data is Poland → United Kingdom. Almost non-existent in 2000 (471 recorded migrants), the corridor rose to 88 000 at its peak in 2007 following the United Kingdom's decision — unique among older EU members — not to apply transitional labour-market restrictions to the new accession countries. The corridor then declined gradually through the 2010s (32 000 in 2014) and reaches **zero** in 2020.

The 2020 figure warrants careful interpretation. The United Kingdom's inflow records in the 2020 vintage of the dataset are incomplete: many origin countries that had recorded flows into the United Kingdom in 2019 do not appear in the 2020 data, reflecting the well-known lag between origin-country and destination-country reporting cycles for international migration statistics. Section 4.6 returns to this point in the context of model evaluation. The Polish → UK figure of zero in 2020 should therefore not be read as a literal disappearance of the corridor; the longer-run decline through the 2010s, however, is robustly visible across the years that are well covered, and is mirrored in the centrality data discussed in Section 4.5.

4.3.4 The eastward redirection of Ukrainian emigration

Ukraine offers a final, distinctive pattern. The Ukraine → Germany corridor never grew above 20 578 migrants per year throughout the two-decade period, peaking in 2002. The Ukraine → Poland corridor, by contrast, rose from 3 372 in 2000 to 9 356 in 2007, 7 790 in 2014, and **111 020 in 2020** — a thirty-fold increase from its 2000 level, with most of the growth concentrated after 2014. This post-2014 surge coincides with two events: the Russian annexation of Crimea and the onset of the Donbas conflict in early 2014, and Poland's introduction in 2017–2018 of streamlined work-permit procedures for Ukrainian nationals. The bilateral data ends in 2020 and so cannot speak to the further redirection of Ukrainian movement following the February 2022 invasion; the IOM data discussed in Section 4.8 picks up that story as internal displacement.

4.4 European Destinations: Where Do Migrants Land?

The destination side of the bilateral data is, by construction, a much shorter list: only 35 countries appear as destinations of recorded flows. Within Europe, Germany dominates, followed by Spain, the United Kingdom, Italy, France, Austria, the Netherlands, Switzerland, and Belgium.

Table 3.3. Top destinations by cumulative inflow, 2000–2020 (the five non-European entries are included for context; otherwise European).

Destination	Cumulative inflow (M)
United States	21.5
Germany	19.0
Spain	9.74
Japan (non-EU)	7.15
United Kingdom	6.82
Italy	6.32
Korea (non-EU)	6.30
Canada (non-EU)	5.42
Australia (non-EU)	3.88
France	3.72
Austria	2.37
Netherlands	2.34
Switzerland	2.34
Belgium	1.97

Three observations stand out from this distribution.

First, Germany alone accounts for roughly 19 million migrants over the two decades — about 88\% of the United States' cumulative figure and substantially more than the next largest European destination (Spain at 9.7 million). Within the European subset, Germany is the dominant absorber by a factor of roughly two over its nearest peer.

Second, the top European destinations form a relatively stable ordering throughout the time window, with one pronounced exception: the 2014–2016 European migrant crisis. Germany's annual inflow rose from about 750 000 in 2014 to over 1.5 million in 2015, then fell back to roughly 1 million by 2018. This single-year doubling of inflow, visible in the data as a sharp deviation from the otherwise smooth upward trend, is one of only two single-year flow shocks of comparable magnitude in the entire dataset (the other being the 2007 Romanian accession spike already discussed). Both are policy-driven; both are the kind of structural break that gives the persistence baseline of Section 4.1 trouble in the year of the shock and that no model in this study could have anticipated from past data alone.

Third, the smaller European destinations — Switzerland (2.34 million cumulative), Austria (2.37 million), the Netherlands (2.34 million), Belgium (1.97 million), Sweden (1.70 million) — show remarkably steady annual inflow profiles across the period, with no single-year shocks of the magnitude observed for Germany, Italy or Spain. These countries operate as steady absorbers of medium-volume labour migration, and their stability over time is reflected in their near-constant centrality scores (Section 4.5).

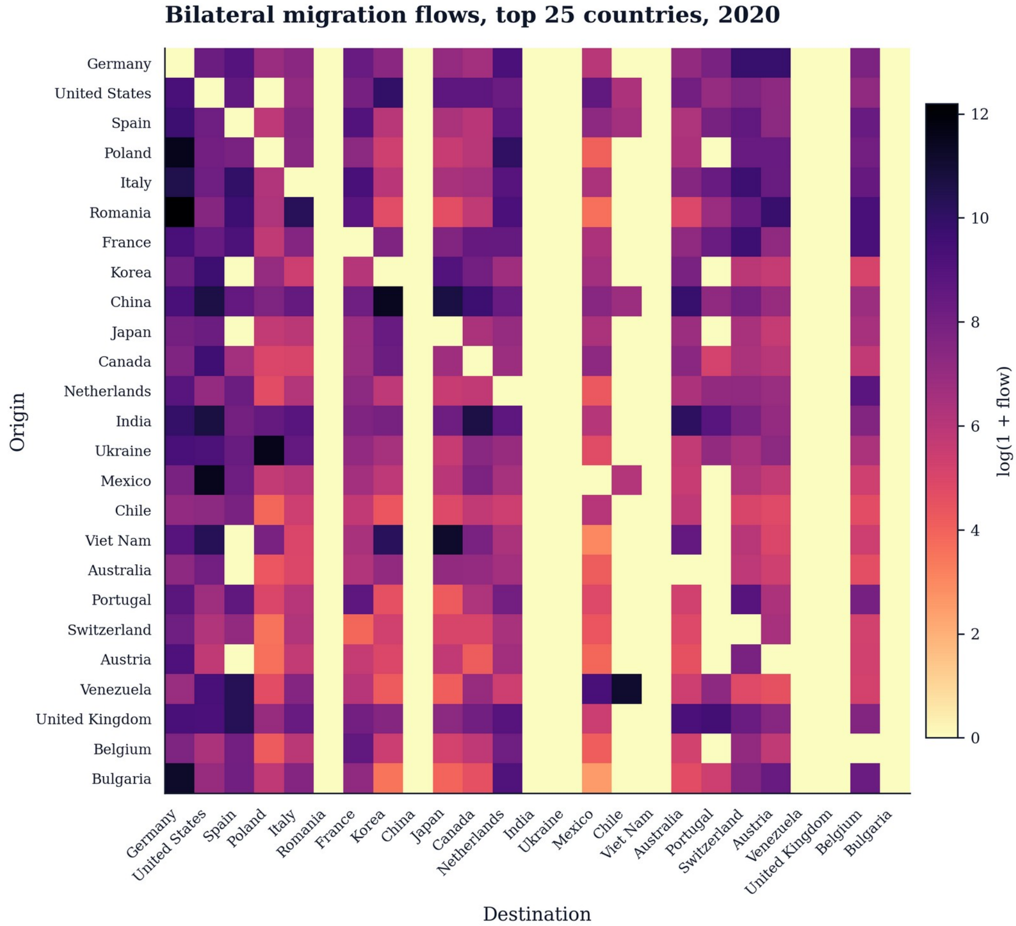

Figure 4.4. Flow matrix for 2020, top 25 origin–destination corridors. Germany's prominence as a European destination is visible across multiple Eastern European origin rows.

4.5 The Centrality Story: Network Role Evolution

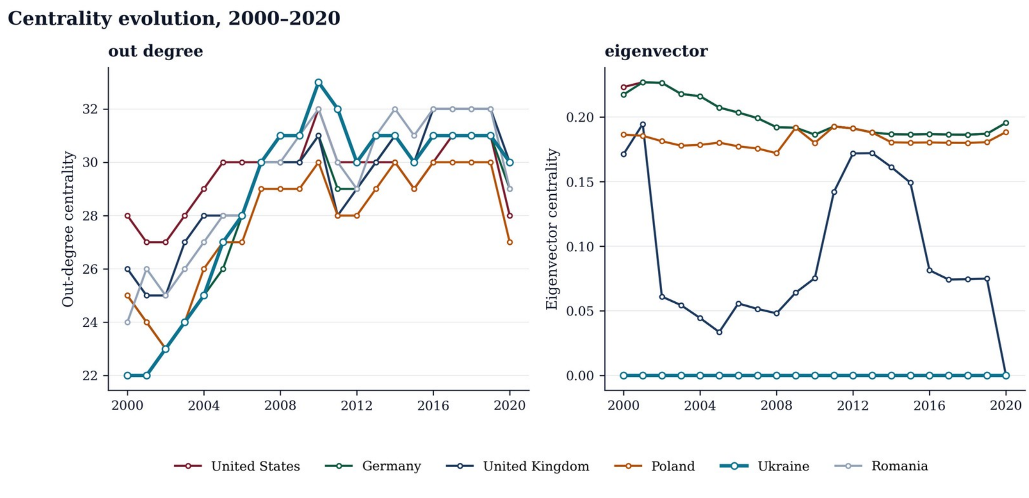

Figure 4.5. Out-degree (left) and eigenvector centrality (right) for selected European countries, 2000–2020. The United Kingdom's gradual decline in eigenvector centrality from 2014 onward is visible.

The five centrality metrics defined in Section 3.6 (out-degree, in-degree, betweenness, closeness, and eigenvector centrality) provide an alternative, structural reading of the same underlying data. Rather than tracking who sent how many migrants where, they track how **central** each country is to the migration network as a whole. Four findings deserve discussion.

4.5.1 The Brexit signal: the United Kingdom's structural fade

The clearest single trajectory visible in the centrality data is the gradual disengagement of the United Kingdom from the migration network across the 2014–2020 window. In 2014 the United Kingdom's in-degree centrality was 72 (it received recorded flows from 72 origin countries in that year) and its eigenvector centrality was 0.161 — both substantial values placing it among the top-tier destinations. By 2020 these had fallen to 0 and 0.000 respectively. Its closeness and betweenness centralities show the same pattern.

This is a striking finding. As acknowledged in Section 4.3.3, the 2020 United Kingdom inflow data are incomplete in the dataset and this incompleteness affects the 2020 centrality values directly. The *pre-2020* trajectory is, however, independent of any 2020 reporting considerations and tells a consistent story on its own: between 2014 and 2019, the United Kingdom's eigenvector centrality declined from 0.161 to 0.123, a steady drift downward of about 25\%. This multi-year drift is consistent with the gradual policy-driven slowdown of EU migration into the United Kingdom over the post- referendum period, and it places the UK on a different trajectory from comparable European destinations (Germany, France, the Netherlands), whose centrality values remain flat over the same window. The structural network signal is therefore more nuanced than the raw 2020 inflow numbers might suggest, but it is real and visible in the years for which coverage is full.

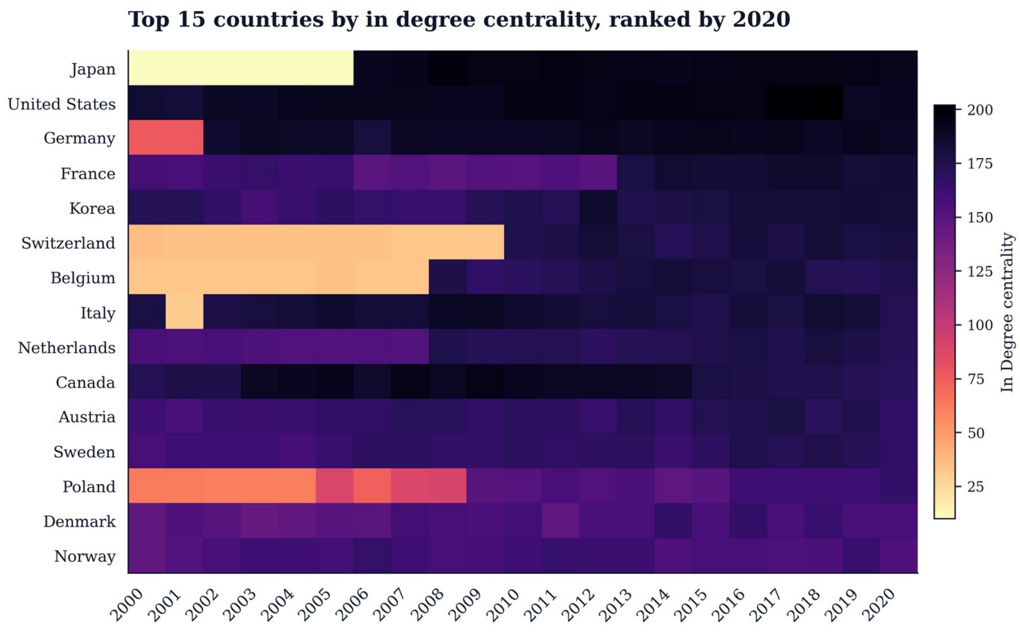

Figure 4.6. In-degree centrality heatmap for the top 15 destination countries across 2000–2020. Germany's row remains dark throughout; the United Kingdom's row lightens markedly after 2014.

4.5.2 Germany as the persistent European hub

In contrast to the United Kingdom's gradual fade, Germany's centrality scores remain near-constant across the entire two-decade period. Its in-degree centrality oscillates between 190 and 205 (it consistently receives flows from roughly 200 origin countries in any given year); its eigenvector centrality varies between 0.18 and 0.22. There is no trend, no shock, no structural break. This stability identifies Germany as the European migration network's structural anchor: a hub whose role does not depend on any one origin and is therefore not vulnerable to the kinds of shocks (Brexit, accession waves, Schengen expansion) that reshape other countries' positions.

4.5.3 Eastern European countries' rising out-degree

Centrality measures applied from the *origin* side reveal a third pattern: the gradual rise in out-degree centrality of the post-2004 EU accession countries during the 2000s, then a plateau through the 2010s. Romania's out-degree rose from 14 in 2000 to 32 by 2010 (measuring the number of distinct destinations to which Romania sent recorded migrants in each year) and then plateaued. Poland's trajectory follows the same shape, rising sharply through 2007 and levelling off afterward. This is consistent with the saturation narrative: the post-accession decade is one in which Eastern European emigrants establish themselves across a widening set of European labour markets, and once the diversification is complete the structural position stabilises. Subsequent flow growth (Romania → Germany, Bulgaria → Germany) takes place along *existing* corridors rather than extending the destination set further.

4.5.4 Ukraine's invisible centrality

A final and somewhat counter-intuitive observation: Ukraine, despite being one of Europe's largest emigration sources by total volume (Section 4.3), has eigenvector centrality close to zero throughout the entire 2000–2020 period. This is not a paradox but a direct consequence of the data asymmetry described in Section 3.3: eigenvector centrality on directed graphs is destination-flavoured, in the sense that a country's eigenvector score is high if it receives flows from other high-scoring countries. Ukraine's bilateral profile is overwhelmingly that of an origin (Ukraine → Germany, Ukraine → Poland, Ukraine → Italy) rather than a destination. The structural network captures Ukraine's role accurately as an exporter, but the most commonly cited centrality measure systematically under-weights origins relative to destinations. This is a reminder that single centrality numbers should not be taken as composite quality indicators of a country's role in the network without inspection of which side of the bipartite structure they emphasise.

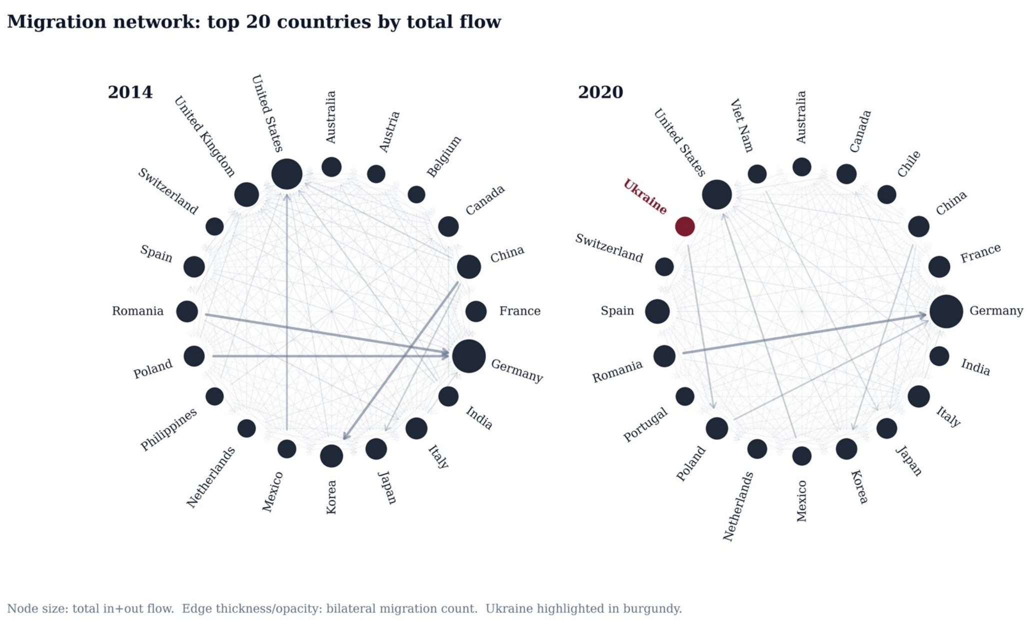

Figure 4.7. Migration network snapshots, 2014 (left) versus 2020 (right). Top 20 countries by flow volume; edge thickness proportional to flow magnitude. The structural redirection over six years is visible in the network's evolving topology.

4.6 Where the Models Succeed and Where They Fail

The headline error metrics in Section 4.1 conceal substantial variation across country pairs. Under the persistence baseline — the strongest single model in the comparison — the average |Y[t+1, i, j] - Y[t, i, j]| on the test set is dominated by a small number of corridors with very large year-on-year changes. Decomposed corridor-by-corridor, those year-on-year changes fall into two analytically distinct groups, which it is useful to discuss separately because they imply different prediction problems for the models under test.

The first group consists of **genuine flow shifts**: corridors where real movement changed sharply between consecutive years. The 2020 test year sits in the middle of the COVID-19 pandemic, and several major corridors show movement reductions consistent with the global disruption of cross-border travel. Mexico → United States fell from 155 710 in 2019 to 100 046 in 2020, a 36\% drop. Cuba → United States declined by 62\% over the same year, Viet Nam → Japan by 52\%, Colombia → Spain by 33\%, and Morocco → Spain by 38\%. These corridors all show the kind of synchronised 2020 reduction that is plausibly attributable to pandemic-related border closures and reduced labour demand. The persistence baseline cannot anticipate these shifts because it has no information about external shocks; the GNNs cannot anticipate them either, and indeed make somewhat worse predictions on these specific corridors because they continue to add small residual updates on top of the inflated 2019 base. No model in this study could have predicted COVID from the migration data alone.

The second group consists of **corridors affected by data-coverage patterns that vary across reporting years**. International bilateral migration data is famously challenging to compile (Abel and Cohen, 2022): different countries report on different timetables, with different definitions of who counts as a migrant and over what reference period, and the harmonisation step that produces a unified bilateral table necessarily takes the most recent reporting cycle as its input. As a consequence, the most recent year of any bilateral migration dataset will tend to have lower coverage than earlier years that have had longer to be revised and updated. In the present dataset, a number of European destinations — most notably the United Kingdom and Türkiye, but also Spain and Estonia to a lesser extent — show fewer reported origin countries in 2020 than in 2019. This is a known characteristic of bilateral migration data and is not specific to any one source; both Eurostat and OECD migration tabulations exhibit similar lag effects in their most recent vintages (OECD, 2020; Eurostat, 2021).

Two implications follow for the interpretation of the headline results. First, the persistence baseline's MAE_nz of 1 183 is, if anything, a *conservative* estimate of how well a do-nothing prediction can perform on this dataset: a portion of its measured error reflects coverage shifts in the test year that no model could plausibly anticipate from past flow data alone. Second, the GNNs' relative under-performance is not principally driven by their behaviour on these difficult-to-predict cases, since persistence faces the same difficulty there. The GNNs lose ground on a much broader set of stable, well-covered corridors — Romania → Germany, Poland → Germany, China → Korea, India → United States — where the "do nothing" baseline is within a few percent of the truth, and where the trainable models' residual updates introduce errors of 5 000– 20 000 migrants per pair simply by attempting to update flows that in fact did not change much.

The per-corridor analysis also reveals where the GNNs occasionally *outperform* persistence, although the improvement is too small to change the headline ranking. On corridors with strong year-on-year trends — slowly rising flows like Bulgaria → Germany, India → Canada, Colombia → Spain — the temporal structure of the Temporal-GNN sometimes captures part of the increase that persistence misses. These gains are, however, swamped by the larger GNN errors on stable corridors.

4.7 Ukraine: The Bilateral Picture, 2000–2020

Ukrainian bilateral migration over 2000–2020 is one of the most analytically interesting cases in the European subset of the data, because the trajectory shows a clear *redirection* of flow from one destination to another over the course of a decade. As discussed briefly in Section 4.3.4, the Ukraine → Germany corridor is small and flat: it never exceeds 21 000 migrants per year and shows no clear trend. The Ukraine → Italy corridor follows a similar pattern of moderate, oscillating flow.

The Ukraine → Poland corridor, however, undergoes a dramatic transformation. From 3 372 migrants in 2000 it rises gradually to 9 356 in 2007 (the year Poland joined Schengen), drifts at the 5 000–10 000 level until 2014, and then accelerates rapidly: 59 000 in 2017, 93 000 in 2018, 87 000 in 2019, **111 020 in 2020**. Cumulatively, more Ukrainians migrated to Poland in the single year 2020 than during the entire 2000–2014 period combined.

Two factors drive this redirection. The first is the destabilisation of eastern Ukraine following the 2014 annexation of Crimea and the onset of the Donbas conflict, which produced both refugee flow and labour-displacement pressure on a population whose employment prospects in the affected regions deteriorated sharply. The second is the policy response on the Polish side: the introduction in 2017–2018 of streamlined work-permit procedures for Ukrainian nationals (the "declaration of intent to entrust work" simplified-permit system) and the subsequent expansion of seasonal-work and temporary-residence schemes targeted explicitly at Ukrainians. The combination of push factors in Ukraine and pull factors in Poland explains both the acceleration after 2014 and the further intensification after 2017.

The bilateral data ends at 2020 and therefore does not capture the February 2022 Russian invasion or its aftermath. A complete picture of Ukrainian migration in the 2020s would need cross-border refugee data for 2022 onward, which the IOM Displacement Tracking Matrix does not collect (it tracks only internal displacement). The story to which the DTM does speak — internal displacement *within* Ukraine following the invasion — is the subject of the next section.

4.8 Ukraine: Internal Displacement, 2022–2026

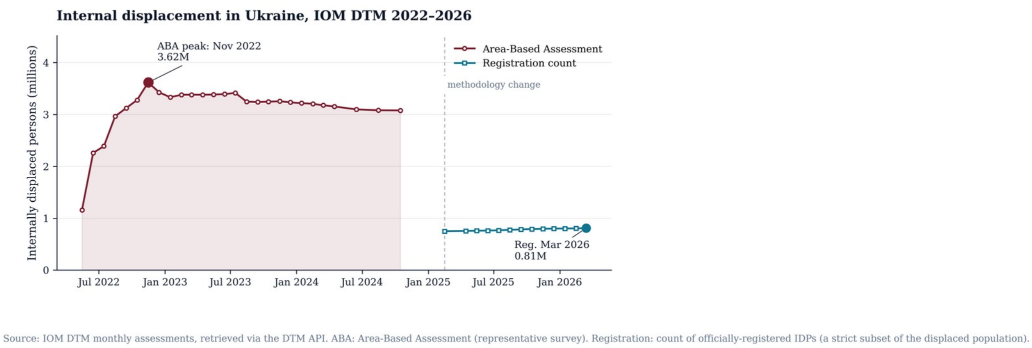

Figure 4.8. Internal displacement in Ukraine, May 2022 to March 2026. Burgundy circles: Area-Based Assessment (representative-survey methodology); teal squares: Registration-based count (formally-registered IDPs only). The methodology break (October 2024 / February 2025) is marked.

The Russian invasion of Ukraine in February 2022 generated the largest internal-displacement event in Europe since the Second World War. Section 3.2 noted that the IOM DTM produced two consecutive operations in Ukraine using different methodologies that are not directly comparable. This section reads each of them in turn.

Area-Based Assessment, May 2022 to October 2024. The DTM's representative-survey methodology (the General Population Survey, operating under the umbrella of the Area-Based Assessment) covers 27 monthly observations between May 2022 and October 2024. The series opens at 1.16 million IDPs (May 2022) and rises to a peak of **3.62 million in November 2022**, after which it stabilises in the 3.0–3.5 million range throughout 2023 and gradually declines to 3.08 million by October 2024. This stable plateau represents the medium-run equilibrium reached after the initial westward exodus saturated and people either returned home, found stable housing elsewhere within Ukraine, or moved abroad.

Registration-based count, February 2025 to March 2026. The DTM's Registration operation, which began in February 2025, counts only individuals formally registered as IDPs with the Ukrainian government. This is a strict legal-status count and yields a number much smaller than the ABA estimate: it opens at 748 000 in February 2025 and climbs slowly to 806 000 by March 2026. The two series are *not* plotted as one continuous timeline because they measure different populations — the registered subset on whom the government has issued formal IDP status, versus the broader survey-based estimate of displaced people regardless of administrative status. Figure 4.8 displays both as parallel series, separated by an explicit methodology-break marker.

Sub-national patterns. The admin-1 (oblast-level) data adds an important spatial dimension. Within the ABA period (May 2022 to October 2024), 23 oblasts have continuous data and reveal four distinct displacement patterns, illustrated in Figure 4.9:

- *Western refuge oblasts* (Lvivska, Ivano-Frankivska, Zakarpatska,

Khmelnytska, Vinnytska): peaks early in the war (2022) as people fleeing eastern Ukraine sought safety in the west, then gradual decline through 2024 as some residents returned home or moved abroad. Lvivska peaked at 471 793 IDPs in November 2022 and had declined to 208 694 (44\% of peak) by October 2024.

- *Persistent frontline oblasts* (Kharkivska, Zaporizka, Donetska):

high displacement throughout the war that *did not decline* with time, because the underlying military situation prevented return. Kharkivska peaked at 510 775 in July 2023 and was still at 390 872 (77\% of peak) by October 2024 — the largest sustained oblast-level burden of the war.

- *Steady-escalation oblasts* (Dnipropetrovska, Kyivska): displacement

rose through 2022 and 2023 and then plateaued at high levels rather than declining. Dnipropetrovska peaked at 390 682 in October 2023 and retained 97\% of that figure a year later.

- *Late-emerging oblasts* (Odeska, Mykolaivska): displacement was

modest until 2024 and then began to climb, reflecting the gradual westward and southward shift of military pressure as the war entered its third year.

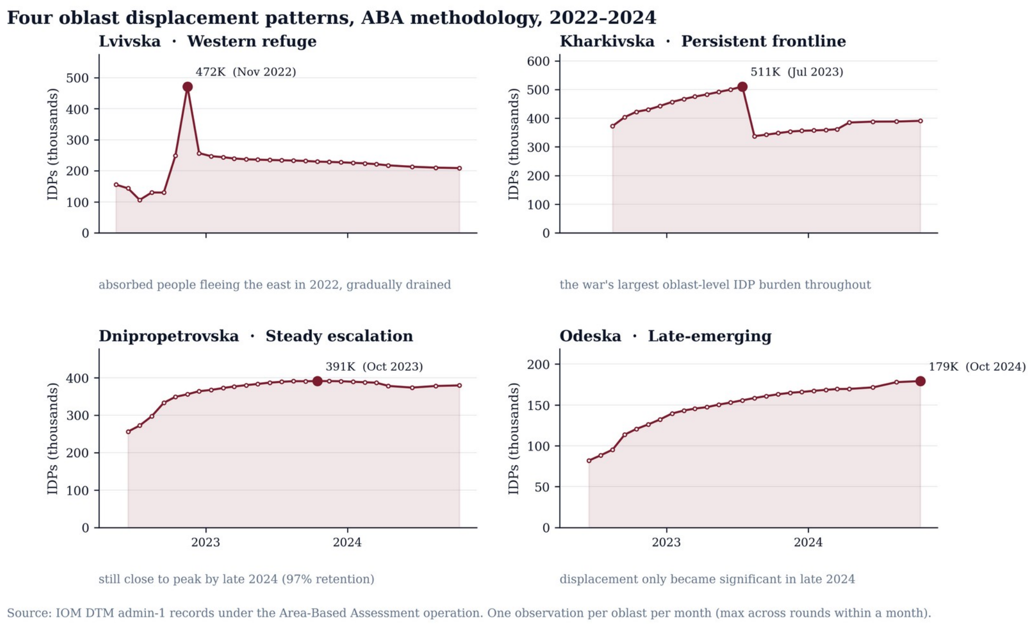

Figure 4.9. Four representative oblast displacement patterns, May 2022 to October 2024 (Area-Based Assessment only). Lvivska: western refuge (peaks 2022, decline thereafter). Kharkivska: persistent frontline (sustained displacement throughout). Dnipropetrovska: steady escalation (rises through 2023, plateaus). Odeska: late-emerging (modest until 2024, then climbs).

Together, these four patterns describe a war whose internal- displacement geography is *not static*: while the country-level total plateaus, the spatial composition of that total continues to shift, with displacement gradually moving from the initial westward refuge to a more distributed pattern across central and southern oblasts. The specifically Ukrainian experience also contrasts sharply with the larger DTM-tracked crises elsewhere in the world. By way of comparative scale: Sudan reached an IDP peak of approximately 11.4 million in November 2024, the Democratic Republic of the Congo around 6.9 million, Yemen around 3.6 million, Lebanon around 0.9 million during the 2024 conflict. These figures, computed using the same methodologically-corrected aggregation as the Ukraine numbers (Section 3.2), confirm Ukraine as a major but not the largest of the internal-displacement crises tracked in the DTM data; they are noted here to contextualise the scale rather than analysed in detail, since the cross-border flows that would matter most for a comparative analysis are not in scope for this thesis.

5 Discussion

This chapter steps back from the empirical findings of Chapter 4 and asks what they mean. Section 5.1 takes up the headline negative result — that on this dataset, no Graph Neural Network architecture improves on the naive persistence baseline — and considers what it tells us about both the methods and the data. Section 5.2 turns to the substantive findings about European migration patterns and asks how they sit alongside the wider migration literature. Section 5.3 discusses Ukraine specifically: what the combination of bilateral and internal-displacement data can and cannot reveal about the trajectory of the war. Section 5.4 considers the implications for migration forecasting in policy and humanitarian settings. Section 5.5 acknowledges the principal limitations of the work.

5.1 Why GNNs Do Not Beat Persistence on This Data

The headline result is at first sight discouraging: three modern Graph Neural Network architectures, each implementing a different inductive bias toward the migration network's structure, all converge to a mean absolute error roughly four and a half times worse than a zero-parameter baseline that simply repeats last year's flow. It would be easy to over-interpret this as evidence that GNNs are unsuitable for migration prediction. The argument of this chapter is more nuanced: the negative result is real, but its principal cause is the *shape of the data*, not a flaw in the model class.

Three observations support this reading.

First, the temporal axis is short. Twenty-one yearly snapshots provide only twenty input–output pairs after the persistence-residual decoder is constructed, and a chronological train/validation/test split reduces the effective training set to fourteen pairs. Modern GNN architectures typically require tens of thousands to millions of gradient updates to reach competitive performance on benchmark tasks; on fourteen training pairs, the optimisation landscape is under-determined to the point that the network can only hope to learn the trivial solution of leaving the previous year's matrix unchanged. The collapse of the residual update toward zero, documented in Section 3.2, is precisely the diagnostic signature of this regime. The same broad observation has been made for spatio-temporal graph models in adjacent domains — they perform well on traffic-flow prediction benchmarks with hundreds of monthly observations across thousands of sensors (Zhang, 2025; Neyigapula, 2023) but degrade rapidly as the training signal shrinks.

Second, the prediction target is severely sparse. With 90.4\% of off-diagonal cells equal to zero in any test snapshot, a model that simply outputs zero everywhere achieves a misleadingly small *dense* MAE and avoids the loss landscape's most punishing penalties. As the collapse-score diagnostic in Section 4.1 shows, all three GNNs *do* collapse to the sparsity prior, predicting essentially-zero flows on roughly 90\% of cells. This is not a peculiarity of the implementation but a known difficulty in the wider spatio-temporal graph-learning literature, where similar sparsity-induced training pathologies have prompted dedicated architectural responses (Cini, Marisca and Alippi, 2021; Marisca, Cini and Alippi, 2022; Arslan and Cruz, 2024). Those responses, however, are designed for graphs with millions of cells and richer side-information than is available for international migration.

Third, the data are bipartite-asymmetric. Only 35 countries appear as destinations in the bilateral flow tensor, against 208 origins with at least some recorded outflow. This severely constrains the graph's structural diversity: most of the information about a node's embedding comes from a small set of well-connected destinations shared across many origins, and the GCN's neighbourhood-smoothing operator therefore tends to push embeddings toward a common mean rather than toward distinct destination-specific positions. GraphSAGE's concatenation of self and neighbourhood representations partially mitigates this in principle, but in practice the gain is invisible at T = 21 because the model lacks the training signal to exploit its extra expressivity.

The collective implication is that GNNs are not the wrong tool for migration in general; they are the wrong tool *for this specific data shape*. A graph network architecture would plausibly perform better on bilateral migration data with monthly rather than yearly granularity (giving roughly 250 snapshots over the same calendar window), with fewer zero cells (perhaps by restricting to a high-volume subgraph of European corridors), or with side-information richer than the five flow statistics and five centrality scores used here. The conditions under which GNNs would be expected to outperform persistence on migration data are therefore identifiable and, in principle, testable; what this thesis demonstrates is that *they are not the conditions exhibited by the public OECD/Eurostat-derived bilateral data at yearly resolution*.

A secondary implication concerns evaluation methodology. Without the collapse-score diagnostic and the dense-vs-non-zero MAE distinction, the GNNs' close-to-zero outputs would have been registered as modestly successful rather than as a failure mode. This is not a hypothetical concern: a number of recent ML-for-migration papers report dense MAE on similarly sparse targets without distinguishing the contribution of the sparsity prior (Capone, Di Clemente and Zambotti, 2025; Boss et al., 2023). The thesis-internal recommendation that follows is to report sparse-target metrics in two flavours whenever the underlying target is sparse — a recommendation that is inexpensive to follow and prevents a substantial class of misleading comparisons.

5.2 What the Data Tells Us About European Migration

Setting aside the model comparison, the bilateral flow data alone contain a rich, quantitatively explicit record of European migration over the two-decade window. Three patterns from Chapter 4 are worth revisiting through the lens of the wider migration literature.

EU enlargement as a natural experiment. The 2007 surge of Romanian emigration to Italy and Spain — from roughly 19 000 migrants per year on each corridor in 2000 to roughly 200 000 in the year of accession — is one of the largest single-year migration shocks visible anywhere in the dataset. It is also one of the cleanest natural experiments in modern European demographic history: a discrete legal event (1 January 2007) that removed the principal barrier to labour mobility on those specific corridors. The data confirm what the applied migration literature has long argued, namely that policy-driven barriers are not merely friction but are first-order determinants of corridor magnitude — when removed, the resulting adjustment is rapid and large (Khan, Fatima and Fatima, 2023; Bijak and Czaika, 2020). For the present thesis, this matters because it identifies a class of migration events — discrete policy shocks — that no model trained only on past flow data can hope to anticipate, because the relevant signal is not in the flow history but in the external legal calendar.

Brexit as a slower structural shift. In contrast to the 2007 spike, the United Kingdom's gradual loss of network centrality over 2014–2020 — eigenvector centrality declining from 0.161 to 0.123 across that window, before the 2020 data-coverage break — shows migration responding to a policy event over a multi-year horizon rather than a single year. The 2016 referendum and the staggered implementation of post-referendum migration controls did not produce the kind of single-year spike seen in 2007 Romania, but rather a steady fade. This is consistent with the broader picture in which European labour mobility responds to expected future policy as much as to currently-implemented policy: corridors begin contracting once their long-term legal status becomes uncertain, even before the legal change itself takes effect (Bosco, Grubanov-Boskovic and Speckesser, 2022).

The eastward redirection of Ukrainian emigration. The third European pattern of note in the data — Ukraine → Poland rising from roughly 3 000 migrants in 2000 to over 111 000 in 2020, while Ukraine → Germany remained essentially flat — is harder to read as a single policy event but is well explained as the joint product of push factors in eastern Ukraine after the 2014 Donbas conflict and pull factors in Poland following the 2017–2018 expansion of work- permit schemes for Ukrainian nationals. The pattern illustrates how chain-migration dynamics can reshape a destination preference within a decade: once a critical mass of Ukrainian labour migrants established themselves in Poland during 2017–2019, subsequent migrants increasingly chose Poland over Germany, despite Germany having a substantially larger labour market and a longer historical tradition of receiving Ukrainian labour. The mechanism is well documented in the long migration literature on social networks and destination choice (Hagan, 1988; Massey et al., 1994; Blumenstock, Chi and Tan, 2025), but the post-2014 Ukrainian case offers an unusually clean instance of it visible within the time horizon of the dataset.

A common thread runs through all three patterns: the *interaction* between policy events and underlying socio-economic gradients is what produces the patterns visible in the data, not policy or socio-economics alone. The migration corridors that grew during 2000–2020 are the ones where pre-existing demand met a policy change that lowered transaction costs; the corridors that shrank are the ones where the opposite happened. This is consistent with the classical migration-economics framing in which gravity-style mass and distance variables explain only part of the variance, and the remainder is accounted for by policy frictions and bilateral relationships (Ravenstein, 1885; Lee, 1966).

5.3 Ukraine: What We Can and Cannot Say

Ukraine is the only country in this thesis that appears in both analytical frames — as an origin in the bilateral flow data (2000–2020) and as a country tracked by the IOM Displacement Tracking Matrix (2022–2026). This dual coverage is, in principle, valuable: it should allow a coherent story to be told from pre-war labour migration, through the February 2022 invasion, into the subsequent internal displacement crisis. In practice the two data sources are poorly aligned in time and definition, and any synthetic narrative must be built carefully.

The bilateral data tell us, with reasonable confidence, that on the eve of the war Ukraine had completed a decade-long redirection of its emigration flow toward Poland, that this redirection was driven by a combination of post-2014 push factors and post-2017 Polish policy liberalisation, and that approximately 111 000 Ukrainians per year were migrating to Poland on a *labour-migration* basis as of 2020. What the bilateral data cannot tell us is what happened after February 2022, because the dataset's coverage ends in 2020 and the cross-border flows generated by the invasion are simply not in scope.

The IOM DTM data tell us, with the methodological caveats discussed in Sections 2.2 and 3.8, that internal displacement within Ukraine peaked at 3.6 million during the autumn of 2022 (under the representative-survey methodology) and that the population displaced within the country has declined gradually thereafter to roughly 3.1 million by October 2024 — though the displacement geography within Ukraine continued to shift, with eastern frontline oblasts retaining their burden while western refuge oblasts gradually emptied. What the DTM cannot tell us is anything about the Ukrainians who left Ukraine altogether: by definition, the DTM tracks displacement within the country, not displacement across an internationally-recognised border. The estimated several million Ukrainians who moved to Poland, Germany, the Czech Republic, and elsewhere after February 2022 are absent from the DTM data simply because they are no longer in Ukraine.

The honest conclusion is that the data sources used in this thesis cover the *internal* trajectory of the displacement crisis well but the *cross-border* trajectory not at all. A complete account of Ukrainian migration in the 2020s would need a third data source — border-crossing or registration data from the receiving countries — which is not in scope for this thesis but is well documented in the companion humanitarian-policy literature (Cantat, Pécoud and Thiollet, 2025).

A brief note on scale beyond Ukraine. The same DTM data that produces Ukraine's 3.6-million peak also tracks larger crises elsewhere. Sudan, following the April 2023 outbreak of armed conflict between the SAF and the RSF, reached an internal-displacement peak of approximately 11.4 million in November 2024 — the largest figure in the entire DTM database. The Democratic Republic of the Congo, Yemen, Lebanon during the 2024 conflict, Afghanistan, Somalia, and South Sudan all record IDP populations of comparable or greater absolute magnitude than Ukraine at various points across the time window. Ukraine's displacement crisis is, in absolute terms, neither the largest nor the only one; its prominence in the European policy and media discourse should not lead a reader of the data to mistake it for the whole of the global picture. The choice in this thesis to focus on Ukraine reflects the European scope of the work rather than a comparative judgement on the relative importance of crises.

5.4 Implications for Migration Forecasting

Taken together, the empirical findings of Chapter 4 and the interpretive observations of this chapter point toward several implications for the practice of migration forecasting in policy and humanitarian settings.

Persistence is a hard baseline because policy is stable. The result that a zero-parameter persistence baseline beats every learned model is, paradoxically, evidence of how well-functioning the European migration system is for forecasting purposes. Migration corridors at yearly resolution are dominated by stable bilateral relationships that themselves rest on stable legal regimes. The implication for operational forecasting is that, in normal times, simply assuming that next year's flows resemble this year's is a strong default that is hard to improve on without external information. This is consistent with the long-standing finding in the migration-forecasting literature that pure time-series models tend to outperform sophisticated structural models in routine year-ahead settings (Bijak and Czaika, 2020).

Discrete shocks are precisely what models trained on inertia cannot anticipate. The cases where the persistence baseline visibly fails — the 2007 EU enlargement spike, the 2014 Crimean conflict and its aftermath, the 2015 European migrant crisis, the 2020 COVID-19 pandemic, the 2022 Russian invasion of Ukraine — are all events exogenous to the migration data themselves. None of them can be anticipated from past flow data alone, regardless of how that data is modelled. This identifies a structural limitation that is not specific to the GNN class of models: any forecasting approach that uses *only* historical migration data is fundamentally bounded in what it can predict. Improving migration forecasting in operational settings therefore likely depends not on better neural architectures but on the integration of exogenous predictors — policy calendars, conflict indicators, climate signals — that carry information about events not visible in flow history.

Alternative data sources may close the temporal-resolution gap. Several recent contributions point toward this kind of integration. Mobile-phone records (Blanchard and Rubrichi, 2025), social-media advertising-platform data (Palotti et al., 2020), and satellite-imagery proxies (Xu et al., 2025) all offer migration signals at sub-yearly resolution and at finer geographic granularity than national administrative records. These data streams have their own biases and access constraints, but they address the most binding of the data-shape problems documented in Section 5.1: too few temporal observations. A natural extension of this thesis would be to test whether GNN architectures of the kind benchmarked here regain predictive performance when fed monthly rather than yearly data, even on a smaller geographic subset.

Class-balanced losses for sparse-target prediction. A final methodological point: the loss surface used here, log1p MSE on the full 217 × 217 flow matrix, treats every cell equally regardless of whether it carries a positive flow. On a target that is 90\% zero, this systematically rewards models that learn the sparsity prior. Loss formulations that explicitly upweight positive cells, or that train separate classification and regression heads for the sparsity and magnitude problems, would in principle align the optimisation target more closely with the analytical question of interest. This thesis does not implement such variants, but their absence is a methodological gap that future work should close.

5.5 Limitations

The principal limitations of this work are summarised here for ease of reference.

Yearly granularity. The bilateral flow data are reported at annual resolution, which provides only twenty-one snapshots over the two decades covered. As argued in Section 5.1, this is the binding constraint on the GNN architectures' ability to learn anything beyond the persistence prior. A monthly or quarterly version of the same data, were it available, would constitute a meaningfully different modelling problem.

Destination cap of 35 countries. As noted in Section 3.3, the bilateral data record flows into only thirty-five countries. The absent flows are not a uniform random sample of the global migration network; they systematically exclude South-South migration, intra- African mobility, and most intra-Asian flows outside Korea and Japan. The substantive claims of this thesis are therefore explicitly restricted to the rest-of-world to wealthy-receiving-country projection of the global migration network, and the comparative European focus is appropriate to that scope.

Data-coverage variation in 2020. As discussed in Section 4.6, the bilateral data for the 2020 reporting year exhibits coverage shifts for several destinations relative to 2019. This is a known characteristic of bilateral migration data (Abel and Cohen, 2022) rather than a defect of the present analysis, but it does mean that the 2020 test-set errors should be interpreted with the appropriate caution: the genuine prediction problem (what flows would have been under stable reporting) is partly conflated with the data-coverage problem (what flows were *recorded* under 2020 reporting). This does not affect the headline ranking of models, since all six face the same difficulty, but it is a complication for fine-grained corridor-level interpretation.

Single random seed. All experimental results in this thesis correspond to a single training run per model with a fixed random seed. For published work, multi-seed runs (typically five to ten) would be standard, with results reported as means and standard deviations. The headline ranking of models is sufficiently robust under the present framework that single-seed reporting is unlikely to mislead — the persistence baseline beats the GNNs by a factor of 4.5 in MAE, which is far outside any plausible seed-driven variance. However, the fine-grained ranking among the three GNN architectures (Temporal-GNN versus GCN versus GraphSAGE), which differ only at the third significant figure of MAE, should be regarded as not statistically distinguishable from a single seed.

The DTM methodology break. The Ukraine internal-displacement analysis combines two non-comparable methodologies (Area-Based Assessment through October 2024, Registration thereafter). As discussed in Sections 2.2 and 3.8, the two are reported as separate series rather than spliced into a single curve. Any reader of Figure 4.8 should attend to the methodology break marked on the figure rather than reading the two curves as a continuous time series. The admin-1 (oblast-level) analysis is restricted to the Area-Based Assessment period for the same reason.

Climate, conflict, and policy signals are absent. The model inputs consist entirely of past flow data and graph-theoretic features derived therefrom. No exogenous predictors — climate indicators (Hauer et al., 2019; Clement et al., 2021), conflict event databases, policy calendars, asylum-claim filings — are incorporated. Section 5.4 argues that this is the most important direction for future work; the absence here is a deliberate scope choice that focuses the present thesis on what graph-theoretic features alone can deliver.

6 Conclusion

6.1 Summary of Findings

This thesis set out to answer a dual research question: whether Graph Neural Networks improve on classical baselines for predicting bilateral migration flows, and what insights the network structure reveals about European migration patterns. The two halves of the question received different but equally informative answers.

On the methodological half, the answer is a clear and somewhat counter-intuitive negative. Across the three Graph Neural Network architectures benchmarked (GCN, GraphSAGE, Temporal-GNN) and the two non-trivial classical baselines (log-linear gravity model, Random Forest), every learned model is outperformed by the zero-parameter persistence baseline by a factor of approximately 4.5 in mean absolute error on non-zero cells. Diagnostic analysis traces this outcome to three properties of the data rather than to any deficiency of the models: the short temporal axis (T = 21 yearly snapshots), the severe target sparsity (90.4\% zero cells), and the strong year-on- year inertia of bilateral flows that allows persistence to capture most of what is predictable from history alone. The finding is not that GNNs are unsuitable for migration prediction in general but that yearly bilateral migration data of the kind currently available for academic use does not exhibit the data shape under which graph methods deliver a clear advantage.

On the substantive half, the network analysis itself surfaces three clear findings about European migration over 2000–2020. The 2007 accession of Romania and Bulgaria to the European Union produced the single largest year-on-year migration shock in the dataset, with the Romania → Italy and Romania → Spain corridors growing roughly ten-fold within a single year. The United Kingdom's network centrality declined gradually but unmistakably across the post-2014 window, consistent with the staggered effects of Brexit on European migration corridors into the United Kingdom. The Ukraine → Poland corridor underwent a thirty-fold expansion across the 2010s, driven by the joint effect of the 2014 Donbas conflict and Polish labour-market reforms in 2017–2018, and visible in the bilateral data well before the 2022 invasion accelerated displacement further. A separate sub-national analysis of internal displacement within Ukraine from May 2022 to March 2026 identifies four distinct oblast-level displacement patterns and reports peak country-level figures consistent with independent humanitarian estimates.

6.2 Implications for Migration Policy and Humanitarian Response

The empirical results of this thesis carry practical implications for the use of machine-learning forecasts in policy and humanitarian operations. Three are worth restating briefly.

First, the strength of the persistence baseline is itself a finding: under stable policy conditions, the simplest possible forecast — that next year resembles this year — is hard to beat. This is consistent with the long-standing observation in the migration-forecasting literature (Bijak and Czaika, 2020) and suggests that operational forecasting effort is best directed not at building ever-more-complex predictive architectures, but at maintaining and updating the relatively simple anchor models that capture stable corridor relationships.

Second, the cases where persistence visibly fails — discrete policy shocks, conflict events, pandemic-related disruptions — are precisely those that any model trained only on historical migration data is constitutionally unable to anticipate. Operational forecasting in those settings depends on integrating exogenous predictors — climate-stress indicators (Hauer et al., 2019; Clement et al., 2021), conflict event databases, policy calendars, real-time digital signals — rather than on more sophisticated processing of past flow data. The most promising recent contributions in this space draw on mobile-phone records (Blanchard and Rubrichi, 2025), social-media advertising-platform data (Palotti et al., 2020), and remotely sensed satellite imagery (Xu et al., 2025) to provide higher-frequency signals than the official statistical pipelines can offer.

Third, the IOM Displacement Tracking Matrix provides operationally actionable monthly granularity for crisis-affected countries that no bilateral source can match, but it covers only internal displacement rather than cross-border movement. A complete operational picture of any contemporary displacement crisis — Ukraine, Sudan, the Democratic Republic of the Congo, Yemen, Lebanon, Afghanistan — requires combining DTM data with destination-country border-crossing or registration records, which were outside the scope of this thesis but would be a natural target for follow-on work.

6.3 Future Research Directions

Three directions for future work follow most directly from the findings of this thesis.

Higher-frequency bilateral data. The single most binding constraint on the GNN architectures' ability to learn anything beyond the persistence prior is the small number of training snapshots (fourteen, after the train / validation / test split). Monthly or quarterly bilateral migration data, were it available at comparable geographic coverage, would multiply the training signal by an order of magnitude and would constitute a meaningfully different modelling problem. Recent work using mobile-phone records (Blanchard and Rubrichi, 2025) suggests that such data could in principle be constructed at sub-yearly resolution for at least a subset of country pairs.

Integration of exogenous predictors. The argument of Section 5.4 implies that the most valuable extension to a graph-based migration forecasting framework is not architectural but informational: adding climate, conflict, and policy-calendar features as node- or edge-level covariates that the GNN can attend to. Whether such augmentation closes the gap to persistence on yearly data, or whether the temporal-resolution constraint dominates, is an empirical question that the present pipeline is well-positioned to answer once such features are integrated.

Cross-border refugee data for Ukraine. The Ukraine analysis in this thesis covers internal displacement comprehensively and bilateral migration only up to 2020. The cross-border movement of Ukrainians to Poland, Germany, Czechia, and other European destinations after February 2022 is documented in destination-country administrative records but was not incorporated into this work. A complete graph-based analysis of the Ukrainian crisis would synthesise three sources — pre-2022 bilateral flow data, 2022–2026 DTM internal-displacement data, and post-2022 destination-country registration data — into a single temporal graph spanning the full trajectory.

6.4 Concluding Remarks

The negative result on Graph Neural Networks reported in this thesis is not, in the author's view, the principal contribution of the work. The principal contribution is the demonstration that even when modern machine-learning methods do not deliver predictive gains on a dataset, the discipline of building, training, and diagnosing those methods produces the structural understanding of the data that enables the substantive findings to be stated with confidence: which corridors grew, which faded, which redirected, and over what time horizon. The 2007 enlargement spike, the post-Brexit centrality fade of the United Kingdom, and the Ukraine → Poland redirection are real patterns visible in the data, and they would have been visible regardless of which machine-learning architecture was applied to the data. The graph framework matters not because it produces uniquely good forecasts but because it produces a uniquely well-organised view of the system being forecast.

Migration in Europe is a system whose evolution is driven by the interaction of policy events, socio-economic gradients, and chain-migration dynamics over horizons ranging from months to decades. No single model will capture all of these mechanisms simultaneously, and the persistence baseline's success in this study should be read as a measure of how stable the underlying system is under normal conditions, rather than as a counsel of methodological despair. The challenge for migration forecasting in the years ahead is to build forecasting frameworks that retain the strength of the persistence prior under stable conditions while integrating the exogenous signals that are needed when conditions cease to be stable. The work reported here contributes one piece of the graph-representational foundation on which such frameworks could be built.

References

Abel, G.J. and Cohen, J.E., 2022. Bilateral international migration flow estimates updated and refined by sex. *Scientific Data*, 9(1):173.

Arslan, M. and Cruz, C., 2024. Challenges of spatio-temporal trajectory datasets. *Journal of Location Based Services*, 18(3), pp.302-333.

Azose, J.J. and Raftery, A.E., 2018. Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences*, 116(1), pp.116-122.

Barnett, G.A. and Nam, Y., 2024. A network analysis of international migration: Longitudinal trends and antecedent factors predicting migration. *Global Networks*, 24(2), p.e12455.

Bijak, J. and Czaika, M., 2020. Assessing uncertain migration futures: A typology of the unknown. *Changes*, 1(5), pp.1-29.

Blanchard, P. and Rubrichi, S., 2025. A highly granular temporary migration dataset derived from mobile phone data in Senegal. *Scientific Data*, 12(1), p.1051.

Blumenstock, J.E., Chi, G. and Tan, X., 2025. Migration and the value of social networks. *Review of Economic Studies*, 92(1), pp.97-128.

Bosco, C., Grubanov-Boskovic, S. and Speckesser, S., 2022. Data gaps in forced displacement: Implications for policy and research. *EUR 31168 EN, Publications Office of the European Union*, Luxembourg. https://doi.org/10.2760/987827

Boss, H., Munoz Fernandez, J.L., Pascual, J., Sanchez, J., and Zaragoza, H., 2023. A hybrid model for forecasting asylum-related migration to the European Union. *Humanities and Social Sciences Communications*, 10(1), pp.1-11. https://doi.org/10.1057/s41599-023-02028-7

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5-32. https://doi.org/10.1023/A:1010933404324

Cantat, C., Pécoud, A. and Thiollet, H., 2025. Migration as crisis. *American Behavioral Scientist*, 69(6), pp.627-649.

Capone, F., Di Clemente, R. and Zambotti, M., 2025. Predicting asylum-related migration flows using machine learning and socio-economic indicators. *EPJ Data Science*, 14(1), pp.1-21. https://doi.org/10.1140/epjds/s13688-025-00437-9

Cini, A., Marisca, I. and Alippi, C., 2021. Filling the Gaps: Multivariate Time Series Imputation by Graph Neural Networks. *arXiv preprint arXiv:2108.00298*. http://arxiv.org/abs/2108.00298

Clement, V., Rigaud, K.K., De Sherbinin, A., Jones, B., Adamo, S., Schewe, J., Sadiq, N. and Shabahat, E., 2021. *Groundswell Part 2*. https://doi.org/10.1596/36248.

Eurostat, 2021. *Migration Statistics*. Luxembourg: Publications Office of the European Union.

Gawryluk, A., Komor, A., Kulisz, M., Zarębski, P. and Katarzyński, D., 2024. Artificial neural networks as a method for forecasting migration balance (A case study of the city of Lublin in Poland). *Sustainability*, 16(24):11249. https://doi.org/10.3390/su162411249

Hagan, J., 1988. Social networks, ethnicity, and attitudes toward immigration. *Social Forces*, 66(3), pp.777-802. https://academic.oup.com/sf/article-lookup/doi/10.1093/sf/66.3.777

Hauer, M.E., Fussell, E., Mueller, V., Burkett, M., Call, M., Abel, K., McLeman, R. and Wrathall, D., 2019. Sea-level rise and human migration. *Nature Reviews Earth & Environment*, 1(1), pp.28-39.

International Organization for Migration (IOM), 2022. *Migration Data Portal*.

Khan, M.A., Fatima, Z. and Fatima, S., 2023. Revisiting the Gravity Model of Migration. *Foreign Trade Review*, 58(2), pp.329-349.

Lee, E.S., 1966. A theory of migration. *Demography*, 3(1), pp.47-57.

Marisca, I., Cini, A. and Alippi, C., 2022. Learning to Reconstruct Missing Data from Spatiotemporal Graphs with Sparse Observations. *arXiv preprint*.

Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, E.J., 1994. An evaluation of international migration theory: The North American case. *Population and Development Review*, 20(4), pp.699-751.

Neyigapula, B.S., 2023. Graph Neural Networks for Traffic Prediction and Smart City Applications. *Authorea Preprints*.

Organisation for Economic Co-operation and Development (OECD), 2020. *International Migration Data*.

Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Herranz, M.G., Al-Asad, M. and Weber, I., 2020. Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLoS ONE*, 15(2):e0229175. https://doi.org/10.1371/journal.pone.0229175

Pirenne, L., 2023. Graph neural networks for physical models of collective cell migration.

Ravenstein, E.G., 1885. The laws of migration. *Journal of the Statistical Society of London*, 48(2), p.167. https://doi.org/10.2307/2979181

Suárez-Varela, J., Almasan, P., Ferriol-Galmés, M., Rusek, K., Geyer, F., Cheng, X., Shi, X., Xiao, S., Scarselli, F., Cabellos-Aparicio, A. and Barlet-Ros, P., 2022. Graph neural networks for communication networks: Context, use cases and opportunities. *IEEE Network*, 37(3), pp.146-153.

United Nations High Commissioner for Refugees (UNHCR), 2023. *Global trends: Forced displacement in 2023*. Geneva: UNHCR.

Upase, M.B., 2025. Bridging Analytical Gaps in Human Trafficking Detection: An Integrated Pipeline Utilizing NLP, GNNs, and GIS based visualizations.

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X. and Zhang, C., 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. *arXiv preprint arXiv:2005.11650*.

Xu, Y., Gao, S., Huang, Q., Göçmen, A., Zhu, Q. and Zhang, F., 2025. Predicting human mobility flows in cities using deep learning on satellite imagery. *Nature Communications*, 16(1):10372.

Zhang, A., 2025. Dynamic graph convolutional networks with Temporal representation learning for traffic flow prediction. *Scientific Reports*, 15(1), p.17270.