Predicting Global Migration with Multiplex Network Analysis

Austin P. Wright
apwright@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Chitwan Kaudan
ckaudan3@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Arjun Goyal
agoyal305@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Lipi Shah
lshah3@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Frank Whitesell
fwhitesell3@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Austin Himschoot
ahimschoot3@gatech.edu
Georgia Institute of Technology
Atlanta, GA

ABSTRACT

Understanding the causes of global migration has become increasingly important as the effects of climate change as well as global conflicts and economic development begin to take shape. However most current work only analyzes causes in isolation, finding many individually important factors but not quantifying how effects interact with each other. This work aims to build an interpretable and interactive analysis tool using a multiplex network modeling framework, in order to understand and visualize the relationships of these causes and effects on international migration.

ACM Reference Format:
Austin P. Wright, Chitwan Kaudan, Arjun Goyal, Lipi Shah, Frank Whitesell, and Austin Himschoot. 2019. Predicting Global Migration with Multiplex Network Analysis. In *Proceedings of Data and Visual Analytics (DVA'19).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

DVA'19, November 2019, Atlanta, GA, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Understanding the relative importance of and the mutual interaction between the drivers of global migration is of interest for building future migration policies. This work introduces a multiplex network model to predict global migration flows. The choice of multiplex network analysis for international relations is the key concept of this work, based on previous work in the field of nuclear nonproliferation[19]. This framework consists of combining unique data sources and encoding individual countries as nodes and relationships between countries (such as migration) as edges, where different properties represent individual layers within a larger model. Combining these layers allows analysis of multiple features and structural relationships between nations.

2 LITERATURE REVIEW

Motivation

As evidenced by publications on this topic, global migration is of interest to governments [15] [21], international agencies, non-profits and businesses, academic institutions, the media, and societies as whole [12]. It is important to note that migration always impacts two communities, the origin and the destination.

Migration Analysis

Currently, academic research relies on qualitative methods such as country or region-specific case studies to analyze factor impact on migration. [26] [23] [20] [17] [7] Also, quantitative research that examines international migration is frequently one-dimensional: studies focus on a small number of correlated factors in their analysis.[10][6] Because migration motivations are diverse[12][23], it is important to consider these factors from multiple lenses for a more accurate understanding. Furthermore, individual datasets can be sparse and differing data collection methods can affect study quality.[9]

Multiplex Networks

Using a multiplex network to model an international scenario with time-series data, with countries as nodes and relationships as edges, has been used to predict states of nuclear proliferation[19]. We will build on this by including quantitative node and edge attributes in our model. Another multiplex model [22] examining different metrics for edge weights will be useful in our prediction analysis. Coinciding with these different metrics is the idea of regression between layers [11]. Combining edge weight metrics and layer regression can be used for our model. Each layer will examine a unique factor with new data to improve the robustness. Some of the attributes we have considered are trade [24], conflict [18], climate change impact on conflict [10], environmental hazards [26], and socioeconomic data [23]. We believe including more attributes will improve the quality of our predictions. Moreover, current migration theories can be used to assess results.[25] [12][9][20][26]

3 METHOD

Causes of Migration

Past research suggests that migration is driven by individuals searching for better economic conditions for their households[9]. Trade dependencies play a role in determining the destination country of a person who decides to migrate.

A country's aggregate demographic and socioeconomic metrics could help determine its bilateral migration flow.[15]. We included a layer that represents population growth and another for the level of education of the general public for each country.

Communities' ability to mitigate environmental risk and chronic abundance of this risk has been hypothesized to affect migration [25]. Threats to land and fresh water resources from climate change also put burdens on individuals to move [26]. We included data on average temperature, natural hazards, water scarcity, and landlocked status to incorporate some of the natural effects of migration.

Political systems also have an impact on a person's ability and desire to emigrate from, or migrate to, a specific country. We explore these aspects by including datasets of autocracy scores and democracy scores of countries around the world.

Data Collection

We considered and cleaned the datasets included in Table 1 to use as our predictors for human migration.

Selecting Countries and Time-frames. Since the regression analysis is dependent on bilateral migration flow, selecting a standard list of countries across datasets was necessary. The initial dataset[5] used the ISO-3166-1 standard for source and target country abbreviations and numeric codes. We applied this standard across all node and edge attribute datasets for consistency. Furthermore, this dataset constructed 5 year estimates of migration flow from 1960-2015. As a result, we restricted our node and edge attribute datasets to this time-frame.

Table 1: Metrics and Datasets

Metric	Description
Democracy Score[1]	extent of democracy on a scale of 0-10
Autocracy Score[1]	extent of autocracy on a scale of 0-10
Birth Rate per 1000[8]	live births per 1000 people
Avg Years in School[2]	average years of total schooling for population age 25+
Landlocked Status[4]	binary variable
Water Scarcity Status[13]	ordinal variable ranging from low to high
Natural Hazards[3]	binary variables for types of hazards affecting countries
Average Temperature[14]	average temperature for countries by year in degrees Celsius
Trade Dependence [8][16]	total trade between two states as a fraction of source state's GDP
GDP per Capita [8]	annual GDP per capita

Node Attributes

Democracy Score: The Democracy score measures the country's extent of democracy on a scale of 0 through 10.

Autocracy Score: The Autocracy score also measures the country's extent of autocracy on a scale of 0 through 10.

GDP per Capita: The gross domestic product (GDP) per capita is the sum of gross economic value produced by all residents in the country plus any taxes and not including any subsidies, divided by population.

Average Temperature: The average temperature data for NOAA worldwide weather stations. The data was grouped on country and year to calculate average temperatures.

Average Years in School: Average Years in School is the mean years of schooling completed by the adult population (25 years or older) at a given time.

Birth rate per 1000: Birth rate per 1000 is the live births occurring during the year, per 1000 population estimated at midyear. Live births data is collected from national birth registration systems and population size is estimated mid-year from census data.

Landlocked Status: The landlocked status data identifies countries bordered only by land. A 1 represents landlocked and 0 otherwise.

Water Scarcity Category: This data categorically represents each country by their fresh water scarcity. There are five different levels from extremely-high to low. Aqueduct combines quantitative and qualitative physical risks and regulatory risks to categorize countries by these levels.

Natural Disasters Risk: A table of worldwide locations and natural hazards that affect them. Key natural disaster words were collected and compared to the CIA description. Words were partitioned based on a common theme. A binary score was given if a country did or did not experience these types of disasters.

Edge Attributes

Trade Dependence: Trade dependence is defined between country i and country j in [19] as:

$$D_{ij} = \frac{Ex_{ij} + Im_{ji}}{GDP_i}$$

where Ex_{ij} is the total value of exports from country i to country j, Im_{ji} is the total value of imports from j to i, and GDP_i is the GDP of i.

Bilateral Migration Flow: Estimates of bilateral (country to country) migration flow for 194 countries from 1960-2015 were generated using a combination of demographic and migrant stock datasets.[5]

Model

Learning Task. After collecting the data in the networked framework, the goal of the model was to predict migration flows between countries as a dependent variable of the independent data sources and graph structures. We can define this modeling task as finding a function, f, as a function of the node and edge variables for each migration flow pair. Here $\vec{n}$ is vector of node attributes of all layers for a given country in a given year, and $\vec{e}$ is the set of edge attributes of all layers for that year.

$$f\left(\vec{n}_{source}, \vec{n}_{target}, \vec{e}_{source\rightarrow target}, \vec{e}_{target\rightarrow source}\right)$$

We can represent this function as a linear combination of input features from the network, and find the optimal solution using linear regression using standard OLS.

Training. Using the migration edge attribute as the target variable, the model will learn coefficients for each input feature representing their contribution to the prediction of migration. In order to make sure features with very different scales as seen in table 2 are learned properly, each feature is normalized to have zero mean and unit standard deviation before training the model, this way features with high input values (like GDP) do not overpower potentially important input features with smaller values.

4 EXPERIMENTS/EVALUATION

Current Evaluation of Regression Coefficients

After training the model, the coefficients learned in the regression task are shown in figure 1. These each correspond to how much a feature contributes, either positively or negatively, to a prediction of migration from a source (src) country to a target

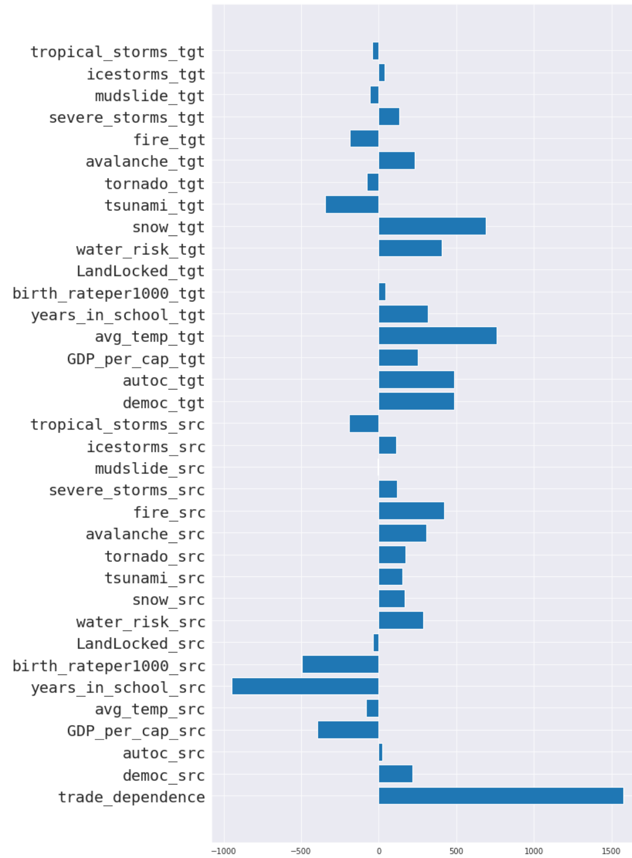

Figure 1: Regression Coefficients

(tgt) country in a specific year. The most significant variable is the edge attribute of trade dependence, strongly correlating with high migration, showing that economic ties are seemingly the most influential in migration patterns. Other factors of note are the relative value of temperature node variables, which is relevant given the effects of climate change.

Planned Future Evaluation

Quantitative Evaluation. After finalizing the edge and node attribute datasets we will include for our final analysis, we plan to evaluate our model quantitatively by a few methods. First, we intend to use a cross-validation approach to reduce the likelihood of over-fitting our model. We will also use prediction accuracy measures like root mean

Table 2: Input Features and Encoding

Feature	Encoding	Data Type	Mean	Standard Deviation
Democracy Score	Ordinal $\in [0, 10]$	Integer	5.45	3.90
Autocracy Score	Ordinal $\in [0, 10]$	Integer	2.08	3.02
GDP per Capita	USD	Float	$8911.72	$14965.96
Average Temperature	Degrees Celsius	Float	19.08	7.47
Average Years in School	Years	Float	6.75	3.36
Birth rate per 1000	Unitless Rate	Float	25.11	12.69
Land Locked Status	Categorical $\in \{0, 1\}$	Integer	0.23	0.42
Water Disaster Risk Score	Ordinal $\in [1, 5]$	Integer	2.54	1.40
Natural Disasters Risk	Categorical $\in \{0, 1\}^6$	Integer		

squared error to measure the predictive power of our model. We plan to perform significance tests for our regression coefficients to determine which predictor variables are integral for our model. Finally, we could utilize variable selection tools like LASSO regression to further reduce the complexity of our model and improve interpretability.

Visualization. Beyond quantitative analysis we would like to create an interactive visualization that showcases historical migration predictions. The user will be able to interactively select source and target countries, and filter on different time frames, node and edge level attributes. We want the visualization to convey not only the final migration prediction but the underlying explanatory variables that led to that prediction.

Qualitative Evaluation. As a final step, we would like to highlight a feature of our model that differentiates it from previous analyses of migration - its universality. Because our model includes data from all countries of the world, rather than one country or region, we have the ability to easily investigate historical migration patterns anywhere in the world in the past sixty years. We hope to identify interesting phenomena, explore the context of the trends, and to use the model to suggest possible causes for these outcomes.

Contributions and Plan of Activities

[1]

5 CONCLUSION AND FUTURE WORK

Under Construction

6 LIST OF INNOVATIONS

Migration studies normally consider case studies or focus on only a few of the factors influencing migration. In this sense, the data is normally country specific and/or have few dimensions. Our project combines longitudinal data over a variety of factors. With data on more countries and factors, our analysis and predictions should be stronger.

Another innovation is the technical approach we are using. Most migration analyses employ descriptive statistics for case studies or use simple single variable tests. Our multiplex network approach combines regression with predictions for each individual graph. The idea is to leverage more factors and more countries to create better results.

REFERENCES

[1] [n. d.]. INSCR Data Page. https://www.systemicpeace.org/inscrdata.html.

[1] All team members have contributed similar amounts of effort. Individual contributions and due dates are outlined in Table 3 and Table 4. Arjun Goyal (AG), Austin Himschoot (AH), Austin Wright (AW), Chitwan Kaudan (CK), Frank Whitesell (FW), Lipi Shah (LS)

Due Date	Activity	Actors
10/02	Generate ideas and select most viable	All
10/08	Written Proposal Proposal Presentation	All AW, CK
10/18	Phase 1: Data collection/preparation	All
11/04	Phase 2: Training and prediction Multiplex Regression Link Prediction	 AG, FW, CK AW, AH, LS
11/08	Progress Report	All
11/22	Phase 3: Interactive Visualization	All
12/03	Poster Presentation Poster Design Demo Presentation	 AH, LS, CK AW, AH, LS All
12/03	Final Report	All

Table 3: Old Plan of Activities.

Due Date	Activity	Actors	Status
10/02	Generate ideas and select most viable	All	Complete
10/08	Written Proposal Proposal Presentation	All AW, CK	Complete Complete
10/18	Phase 1: Data collection/preparation	All	Complete
11/08	Progress Report	All	Complete
11/15	Phase 2: Training and prediction Regression Method/Demo Full Dataset Training Predicting Edge Weights	 AW AH, LS, CK AG, FW	 Complete In Progress In Progress
11/22	Phase 3: Interactive Visualization	All	To Do
12/03	Poster Presentation Poster Design Demo Presentation	 AH, LS, CK AW, AH, LS All	To Do
12/03	Final Report	All	To Do

Table 4: Current Plan of Activities.

[2] 2018. Mean years of schooling. https://ourworldindata.org/grapher/mean-years-of-schooling-1

[3] 2018. The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/fields/292.html

[4] 2019. Landlocked Countries 2019. http://worldpopulationreview.com/countries/landlocked-countries/

[5] Guy J. Abel. 2018. Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015. *International Migration Review* 52, 3

(2018), 809–852. https://doi.org/10.1111/imre.12327 arXiv:https://doi.org/10.1111/imre.12327
[6] Fuad Aleskerov, Natalia Meshcheryakova, Anna Rezyapova, and Sergey Shvydun. 2017. Network Analysis of International Migration. In *Models, Algorithms, and Technologies for Network Analysis*, Valery A. Kalyagin, Alexey I. Nikolaev, Panos M. Pardalos, and Oleg A. Prokopyev (Eds.). Springer International Publishing, Cham, 177–185.
[7] Gregory S. Amacher, Wilfrido Cruz, Donald Grebner, and William F. Hyde. 1998. Environmental Motivations for Migration: Population Pressure, Poverty, and Deforestation in the Philippines. *Land Economics* 74, 1 (1998), 92–101. http://www.jstor.org/stable/3147215
[8] The World Bank. [n. d.]. World Bank Open Data. https://data.worldbank.org/
[9] Thomas Bauer and Klaus Zimmermann. 1995. *Modelling International Migration: Economic and Econometric Issues.* 95–115.
[10] Kate Burrows and Patrick Kinney. 2016. Exploring the Climate Change, Migration and Conflict Nexus. *International Journal of Environmental Research and Public Health* 13, 4 (Apr 2016), 443. https://doi.org/10.3390/ijerph13040443
[11] Giona Casiraghi. 2017. Multiplex Network Regression: How do relations drive interactions? arXiv:physics.soc-ph/1702.02048
[12] Francesco Castelli. 2018. Drivers of migration: why do people move? *Journal of Travel Medicine* 25, 1 (07 2018). https://doi.org/10.1093/jtm/tay040 arXiv:http://oup.prod.sis.lan/jtm/article-pdf/25/1/tay040/25811725/tay040.pdf tay040.
[13] Hannah Dormido. 2019. These Countries are the Most at Risk from a Water Crisis. https://www.bloomberg.com/graphics/2019-countries-facing-water-crisis/
[14] National Centers for Environmental Information and Ncei. [n. d.]. Climate Data Online. https://www.ncdc.noaa.gov/cdo-web/
[15] Gary P. Freeman. 1995. Modes of Immigration Politics in Liberal Democratic States. *International Migration Review* 29, 4 (1995), 881–902. https://doi.org/10.1177/019791839502900401 arXiv:https://doi.org/10.1177/019791839502900401 PMID: 12291223.
[16] International Monetary Fund. 2019. Direction of Trade Statistics (DOTS). https://data.imf.org/?sk=9D6028D4-F14A-464C-A2F2-59B2CD424B85
[17] Elizabeth Fussell, Lori M. Hunter, and Clark L. Gray. 2014. Measuring the environmental dimensions of human migration: The demographer's toolkit. *Global Environmental Change* 28 (2014), 182 – 191. https://doi.org/10.1016/j.gloenvcha.2014.07.001
[18] Faten Ghosn, Glenn Palmer, and Stuart A. Bremer. 2004. The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description. *Conflict Management and Peace Science* 21, 2 (2004), 133–154. https://doi.org/10.1080/07388940490463861
[19] Bethany L. Goldblum, Andrew W. Reddie, Thomas C. Hickey, James E. Bevins, Sarah Laderman, Nathaniel Mahowald, Austin P. Wright, Elie Katzenson, and Yara Mubarak. 2019. The nuclear network: multiplex network analysis for interconnected systems. *Applied Network Science* 4, 1 (2019), 36. https://doi.org/10.1007/s41109-019-0141-4
[20] Roel Jennissen. 2007. Causality Chains in the International Migration Systems Approach. *Population Research and Policy Review* 26 (08 2007), 411–436. https://doi.org/10.1007/s11113-007-9039-4
[21] Rey Koslowski. 2014. Selective Migration Policy Models and Changing Realities of Implementation. *International Migration* 52, 3 (2014), 26–39. https://doi.org/10.1111/imig.12136
[22] Shaghayegh Najari, Mostafa Salehi, Vahid Ranjbar, and Mahdi Jalili. 2019. Link Prediction in Multiplex Networks based on Interlayer Similarity. *CoRR* abs/1904.10169 (2019). arXiv:1904.10169 http://arxiv.org/abs/1904.10169
[23] Wim Naudé. 2008. *Conflict, disasters and no jobs: Reasons for international migration from Sub-Saharan Africa.* WIDER Research Paper 2008/85. Helsinki. http://hdl.handle.net/10419/45125
[24] John R. Oneal and Bruce M. Russet. 2002. The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985. *International Studies Quarterly* 41, 2 (12 2002), 267–293. https://doi.org/10.1111/1468-2478.00042 arXiv:http://oup.prod.sis.lan/isq/article-pdf/41/2/267/5154908/41-2-267.pdf
[25] Clionadh Raleigh and Lisa Jordan. 2008. *Assessing the Impact of Climate Change on Migration and Conflict.* Technical Report. World Bank Social Development Department.
[26] K. Warner, M. Hamza, A. Oliver-Smith, F. Renaud, and A. Julca. 2010. Climate change, environmental degradation and migration. *Natural Hazards* 55, 3 (01 Dec 2010), 689–715. https://doi.org/10.1007/s11069-009-9419-7