

Predicting Global Migration with Multiplex Network Analysis

AUSTIN P. WRIGHT, Georgia Institute of Technology

CHITWAN KAUDAN, Georgia Institute of Technology

ARJUN GOYAL, Georgia Institute of Technology

LIPI SHAH, Georgia Institute of Technology

FRANK WHITESELL, Georgia Institute of Technology

AUSTIN HIMSCHOOT, Georgia Institute of Technology

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. 1, 1 (October 2019), 11 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

Authors' addresses: 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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Association for Computing Machinery.

XXXX-XXXX/2019/10-ART \$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

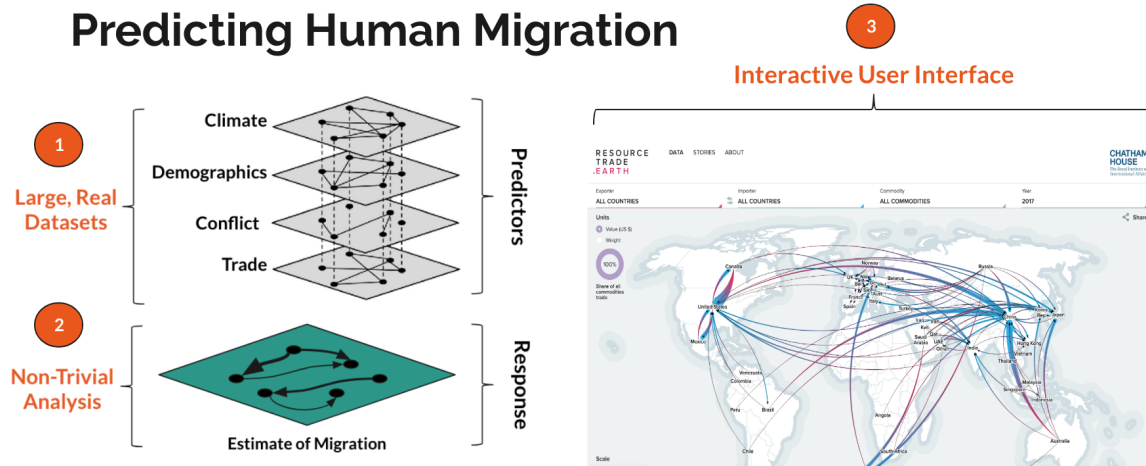


Fig. 1. Project Overview

1 INTRODUCTION

While understanding the causes of global migration is of great policy importance, there is disagreement on the degree to which the effects of climate change will influence global international migration when compared to other more traditional political or socioeconomic causes. Therefore understanding not only the individual effects of different causes of migration, but the relative importance and mutual interaction of those effects is of great interest for building future migration policies. This work will aim to build a multiplex network model to predict global migration flows and understand their causes. A further objective will be to develop interpretable interactive visualisations to display these results to non-technical policy makers. The choice of multiplex network analysis for international relations is the key concept of this work, based heavily on previous work in the field of nuclear nonproliferation[11]. This framework consists of considering many different data sources encoding individual countries as nodes, and relationships between countries as edges, where each different category of relationship is represented as an individual layer within the larger model. Combining these layers allows the analysis to consider multiple features and their structural relationships between nations, combining them for an final analysis goal. In our case, this analysis aims to understand the edge attribute of migration flows between countries. The rest of this proposal will further justify this multiplex network analysis framework, the data streams we are considering and their policy relevance, and the scope of work and plan for the proposal. Figure 1 is illustrative of the analysis and visualization components of this proposal, explicitly marking the key requirements of the project.

2 HEILMEIER QUESTIONS

2.1 What are you trying to do?

Our objective is to use network analysis to predict international migration patterns. We aim to integrate datasets pertaining to diverse factors that have been studied to contribute to global migration flows: political[10][15], socioeconomic[16], demographic, climate[17], and more [7][8]. Using these methods we can display our predictions in a world map interface that allows users to parse out the quantitative importance of individual factors and how they interact with each other in influencing migration patterns over time.[11]

2.2 How is it done today; what are the limits of current practice?

Currently, academic research relies on qualitative methods such as country or region-specific case studies to study the impact of factors on migration. [18] [15] [12] [9] [3] Also, quantitative research that examines international migration is frequently one-dimensional such that studies can only focus on a select number of correlated factors in their analysis.[5][2] Because the motivations behind human migration are diverse[7][15], it is important to consider these factors from multiple lenses so that an accurate understanding of this topic can be generated. Furthermore, individual datasets can be sparse and differing data collection methods can affect the quality of each study.[4]

2.3 What's new in your approach? Why will it be successful?

Our approach is novel in the use of a multiplex network to predict migration patterns across different theoretical triggers of migration. Using a multiplex network to model an international scenario with time-series data, with individual countries as nodes and relationships between them as edges, has been used to predict states of nuclear proliferation[11]. We will build on this by including quantitative node and edge attributes in our model. Another multiplex model [14] examining different metrics to use for edge weights will be useful in our prediction analysis. Coinciding with these different metrics is the idea of using regression between layers [6]. Combining the edge weight metrics and layer regression can be used for our model. Each layer will examine a unique factor using a different dataset, improving the robustness of our results. Some of the attributes we have considered are trade [16], conflict [10], climate change impact on conflict [5], environmental hazards [18], and socioeconomic data [15]. We believe including more attributes will improve the probability of success of our predictions. Moreover, there are available theories of migration that we can compare our results to, which will give us a baseline of plausibility.[17] [7][4][12][18]

2.4 Who cares?

As evidenced by the vast array of publications written on the topic, global migration is of interest to governments [8][13], international agencies, non-profits and businesses, academic

institutions, the media, and societies as whole [7]. It is important to note that migration always impacts two communities, the origin and the destination. The United Nations [15] and the World Bank [17] have expended considerable time and resources on the collection of data which will be used in our model.

2.5 If you're successful, what difference and impact will it make, and how do you measure them?

Modelling migration flows will help nations and organizations better understand the existing drivers for migration as well as plan for future migration. Because a government's immigration policy heavily influences migration [8] [13], informed policy-making decisions may lead to better outcomes, both for migrants and the citizens of the destination country. The model can be validated by future migration data on an annual basis and testing multiple methods for estimating bilateral migration flows to ensure our model's accuracy[1].

2.6 What are the risks and payoffs?

One important risk to account for is an accidental exclusion of a migration factor that may turn out to be important for prediction. Missing an important attribute, land-locked status for instance, could skew the results of our predicted migration patterns. Organizations or agents using our results in their studies or decisions may be impacted by this outcome. On the other hand, successful predictions could lead to a better understanding of migration patterns and relationships between countries.

2.7 How much will it cost?

Currently, there does not appear to be a substantial cost associated with this project, save for time and energy. The software and hardware used will either be open-source or free to use. The datasets we plan to use are publicly available with both [18] and [15] having links to potential datasets: environmental hazards and socioeconomic data, respectively. The International Organization of Migration aggregates country data to produce robust international datasets as well.

2.8 How long will it take?

The totality of the project should take from now until the due date on December 3rd. Each group member is expected to contribute a few hours each week.

2.9 How will progress be measured?

The midterm exam will culminate with the progress report due on November 8th. At this time, we plan to obtain and clean all datasets. Furthermore, specifications for the model parameters

and different node/edge attributes for the multi-layer network will be started. The final exam will consist of the completed model and visualization with the final paper.

3 PLAN OF ACTIVITIES

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

All team members will have contributed similar amount of effort.

REFERENCES

- [1] Guy J. Abel and Joel E. Cohen. 2019. Bilateral international migration flow estimates for 200 countries. *Scientific Data* 6, 1 (2019), 82. <https://doi.org/10.1038/s41597-019-0089-3>
- [2] 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.
- [3] 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>
- [4] Thomas Bauer and Klaus Zimmermann. 1995. *Modelling International Migration: Economic and Econometric Issues*. 95–115.
- [5] 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>
- [6] Giona Casiraghi. 2017. Multiplex Network Regression: How do relations drive interactions? arXiv:physics.soc-ph/1702.02048

- [7] 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.
- [8] 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.
- [9] 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>
- [10] 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> arXiv:<https://www.tandfonline.com/doi/pdf/10.1080/07388940490463861>
- [11] 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>
- [12] 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>
- [13] 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> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/imig.12136>
- [14] 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>
- [15] 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>
- [16] John R. Oneal and Bruce M. Russett. 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>
- [17] Clionadh Raleigh and Lisa Jordan. 2008. *Assessing the Impact of Climate Change on Migration and Conflict*. Technical Report. World Bank Social Development Department.
- [18] 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>

Appendices

4 LITERATURE REVIEW

4.1 Austin Wright Papers

- (1) Nuclear Network [11] The main idea of this paper is that we can understand how many different factors together affect the decisions of countries, in this case nuclear proliferation, by modeling international relations as a multi-layer network. We use this idea for predicting migration instead of proliferation, but the fundamental idea of multi-layer networking to model international relations is the key insight for this work. The work in this paper is however limited, in that it tries to only use edge level attributes (e.g. bilateral trade) and only tries to predict a node level attribute (proliferation), we will aim to expand the kinds of input data streams, as well as predict an edge attribute in migration.
- (2) Trade Dependence [16] One of the data layers we believe is important and will carry over is the trade dependence of nations. This dependence is based on the formulation of Oneal and Russett [16], which calculated the relative trade dependence between nations as a fraction of a nation's GDP. This aims to measure how dependent a specific nation is on another, and the level of economic cooperation between nations. This has been shown to reduce conflict, and we expect it will also be positively correlated with migration. A downside of the data set from these authors is that these specific authors do not have versions going up to the current year, however given that the formula is rather straightforward we can calculate it from more open and up to date economic sources.
- (3) Conflict [10] Another factor we feel will be strong connected with migration is armed conflict. The comprehensive Dyadic Militarized Interstate Dispute (MID) database [10] includes detailed information about the asymmetric use of force between nations in conflict over a long time interval. However it similarly to Oneal, it lacks extremely up to date data, and filling in the gaps for recent years may be more difficult in this case. This may introduce a limitation of the direct timeliness of the work, however the richness of the data still makes it work using for a strong methodology, and if this work is useful for policy applications, they likely will be able to acquire better updated data, and so we can maintain intellectual merits while the broader impacts of the work are not diminished.

4.2 Himschoot Papers

- (1) Assessment on Climate Change Impact [17] This is a survey of the social consequences of climate change with its main focus on environmental hazards and migration. Provides descriptive data, but little in terms of hard data. Migration is no longer considered a consequence of climate change, and the vulnerability of groups depends on socioeconomic characteristics. Will give an idea of how environmental hazards impact migration and give

a comparison to our model. We will improve upon this study by including socioeconomic information as well as environmental hazards.

- (2) Environmental Degradation [18] Has a link to environmental disasters database. Identifies expected links between hazards and migration for islands and drylands. Mentions that future migration is not quantified which could be useful for the results of our model. We may have to think about permanent vs non-permanent migration. Does a lot of the groundwork for expected results. Distinguish between forced and voluntary migration. Did a case study for three countries, which was not technical.
- (3) Climate Change/Migration/Conflict [5] Examines links between climate change and migration as well as migration and conflict. Conflict here can be on any level: person, group, nations. Identifies that it may be difficult to tell the direct effect of climate change on migration. Would climate variability be useful to look at? No actual study done in this paper; it is an analysis of the types of migration and what are potential causes. It is valuable in that it can provide us potential node and edge attributes. Adding data to this information would be interesting.

4.3 Chitwan Papers

- (1) Demographic Drivers [7] This is a survey of the macro (country level), meso (industry/system level), and micro(individual level) drivers of global migration. The author cites that demographics impacts migration on both the macro and micro levels. On the macro level, a country's demographics such as population density, age distribution, population growth have a large impact on whether a country has a large immigration trend. More micro or personal level factors such as each individual's education level, age, sex, gender, religion, marital status play a significant role in whether an individual decides to immigrate. The paper wants to dispel the idea that poor, illiterate rural immigrants since it find that most immigrants tend to be better educated, healthier and more financial stable than the average citizen in their native countries.
- (2) Link Prediction in Multiplex Networks based on Interlayer Similarity [14] This paper proposes an Link Prediction Interlayer Similarity (LPIS) method that outperforms state of the art Entropy Based Method (EBM) and Meta Path Based Method (MPBM) on the Twitter-Foresquare (TF), Twitter-Instagram (TI), and Offline and Online Relationships (OOR) networks. The main idea is to consider a weighted sum of intralayer (within layer) features and interlayer (between layers) similarities between nodes to predict a link between them. The weights for intralayer and interlayer measure need to optimized for each network. The paper states for the intralayer link prediction, the authors used a logistic regression classification with features Adamic-Adar (AA) measure, Jaccard coefficient, Preferential Attachment (PA) indexes. (FYI Jaccard is the number of common neighbors between 2 nodes divided by the total number of neighbors for both of them. AA inversely weights the number of common neighbors with the degrees of the nodes.

PA is simply the product of the degrees of 2 nodes.) The main focus of the paper is on experimenting with different interlayer similarities indexes such as Clustering Coefficient (CC) similarity which measure the extent to which 2 nodes in a network tend to cluster together, Average Similarity of Neighbors (ASN) which is the ratio of common links in 2 layers for node i . The paper proposes Asymmetric ASN (AASN) which normalizes the ASN with the total number of links in the layer to account for the fact that some layers may be denser than others. Their results find that their hybrid approach with the AASN index does a better job than state of the art link prediction algorithms.

- (3) Regression in Multiplex Models [6] This paper provides a formal statistical model for regression between layers in a multiplex network. It uses generalized hypergeometric ensembles (Hypergeometric distribution describes the probability of k successes in n trials without replacement; basically binomial distribution but without replacement. I cannot figure out the ensembles part :(). Now we can look at the influence of each layer and conduct statistical significance tests on the predictor layers. The paper runs the regression on a high school students' interactions network. The layers include students organized by shared classes, broader subject topics (i.e. math classes, English classes), gender, self-reported friendships, and Facebook friendships. They find that shared classes has the greatest influence on whether students will interact. They are able to quantify the degree of influence with an odds ratio (probability an interaction occurs given information about shared classes). They also report the significance of each explanatory layer. Pretty neat stuff!

4.4 Frank Whitesell Papers

- (1) Modelling International Migration: Economic and Econometric Issues [4] This paper surveys the general economic theories which attempt to explain the phenomenon of international migration, from the older neo-classical approach to the recently-developed network migration theory. It discusses the difficulties of analyzing migration data, especially lack of comparability of data between countries, issues separating the true effects from trends in time-series data, and the importance of microeconomic and individual factors on the migration decision, which is difficult to accurately sample and model. Being aware of these difficulties will help us as we collect data sets. Overall, there is strong macroeconomic evidence that migration is largely determined by the search of individuals for better economic conditions for their households. Understanding the individual decision-making process for migration requires better micro-economic data.
- (2) Conflict, disasters and no jobs: Reasons for international migration from Sub-Saharan Africa [15] This paper investigates the factors behind high growth in net international migration in Sub-Saharan Africa. This paper is useful because it collectively and empirically examines the impacts of a number of factors we wish to model in our multiplex network, including conflict, lack of job opportunities, and climate change and natural disasters. In

voluntary migration, the decision to migrate is usually an economic decision, but in forced migration, the need to survive or protect oneself is the highest priority. We will need to consider voluntary versus forced migration in our model. The source of the data used in the study is the UN Population Division, and the data is widely available. The paper found that armed conflict and GDP growth have the highest impacts on international migration. It also makes the point that it is not easy to separate individual effects on immigration, because they interact in very complex ways. The shortcoming of this paper as it relates to our project is that it focuses only on Sub-Saharan Africa.

- (3) Causality Chains in the International Migration Systems Approach [12] This article identifies four specific groups of factors (economic, social, political, and “linkages”) and incorporates causality into the systems approach of investigating international migration. It surveys the current theories of international migration, broken down into those which explain why immigration from one country to another is initiated, and those which explain why it continues. The paper explains the relationships between the groups, and includes a helpful figure defining these relationships. While the explanations of how individual factors interact with one another will be useful to us in the variable selection phase, the paper does not present any new study of these theories in terms of data.

4.5 Lipi Papers

- (1) Selective Migration Policy Models and Changing Realities of Implementation [13] This paper talks about various models of migration based on the type of country. Various countries tend to have selection migration policies. Migration policies can be generally modelled into three types namely human capital model, neo-corporatist model and market oriented demand-driven model. Such types of market economies can enforce migration and the implications can differ depending on the type of government the country has. Nations also tend to shift their inclination for the types of skills or points required for an immigrant to become a resident. This shift could also alter the migration patterns of the nation. This paper does not explore analysis or prediction of migration based on these models.
- (2) Modes of Immigration Politics in Liberal Democratic States [8] This paper goes over the the politics of immigration in liberal democracies. It is stated that official policies tend to be more liberal than public opinion and annual intakes larger than is politically optimal. If the state policy makers think that immigration policy could maximize their votes, they respond to this organized pressure to favour immigration ignoring the widespread but poorly articulated opposition of immigration by general public. Comparing different democracies, most of the nations are now falling within three models mentioned in [13]
- (3) Environmental Motivations for Migration: Population Pressure, Poverty, and Deforestation in the Philippines [3] This paper predicts migrant’s destination choice based on

various features like poverty, population pressure and forestation. They have used multinomial modelling on data obtained from Philippines's migration report and 1990 national census. Population pressure turns out to be an integral factor in determining the place of migration. They have also added political unrest and travel cost variables as extra parameters. The population shifts more towards availability of natural resources rather than availability of employment or upland income.

4.6 Arjun Papers

- (1) Network Analysis of International Migration [2] This paper covers the gravity model of international migration, a widely used convention for estimation migration flows between countries. It considers migrations largely as a result of two factors, the size of each country (origin and destination) and the distance between countries. We may use distance or population size differences between countries as an individual layer to evaluate in our analysis.
- (2) Measuring the Environmental Dimensions of Human Migration: The Demographer's Toolkit [9] The key takeaway from this paper is that data collection can influence describing explicit linkages between environmental factors and migration. It can be difficult to parse out whether climate related issues are the real or only reason for temporary or permanent migration. This study provides an overview of various regression techniques, when specifically analyzing population and household outcomes. This paper looks at 4 specific studies, all qualitative case studies, and the limitations and challenges associated with each.
- (3) Bilateral International Migration Flow Estimates for 200 Countries [1] This paper evaluates 6 different ways to estimate bilateral migration numbers, and includes R code to help determine these figures. Using the methodologies of this paper, we may be able to validate our predictions and ensure we are using historical migration data effectively when we create our multiplex model.