How Homophily and Consolidation Shape Cumulative Migration Dynamics from Mexico to the United States

Filiz Garip fgarip@wjh.harvard.edu
Maocan Guo mguo@fas.harvard.edu

Harvard University

Abstract

Migration flows from Mexico to the United States originated in the central-western Mexican states, a pattern prior research connected to the early development of railroads in the region. Over time, these areas continued to be major suppliers of migrants to the United States, but new areas (such as the border states) also emerged. The level of migration out of the emerging areas eventually came to surpass that from the historic regions. This pattern presents a puzzle for cumulative causation theory, which predicts ever-increasing migration rates due to the accumulation of social resources. This project seeks to solve this puzzle. Using data from the Mexican Migration Project, we will study how the social structure in sending communities, specifically the homophily and consolidation in the distribution of socio-economic characteristics, shapes the diffusion of migration behavior.

Extended abstract

Motivation

Blau and Schwartz (1984) were the first to discuss the importance of homophily (the formation of social ties to people who are proximate within a dimension of social life) and consolidation (the correlation of social positions across different dimensions) for understanding the link between the social structure and the observed pattern of social relations. Recently, using a computational modeling framework, Centola (2012) extended these ideas to study the diffusion of social behaviors. Our study will test these ideas with survey data from the Mexico-U.S. context, and thus, identify the structural mechanisms that shape the diffusion of migration behavior. The findings are likely to alter how we think about the cumulative causation of migration (Massey 1990).

Methodology

We use data from 124 communities surveyed by the Mexican Migration Project. We create a panel data set for all individuals (not just household heads) using the retrospective information on household members. Let's take a household surveyed in 1990, where the daughter has migrated to the United States for the first time in 1980. Her attributes, like age and education, were recorded in 1990, but could be projected linearly to 1980. The economic status of her household could be reconstructed using the data on the timing of asset purchases. The characteristics of her community could be traced back using the retrospective community history. All these plausible steps rely on one crucial assumption: that the daughter in question was living in the same household and community in 1980. (While this assumption is viable for most cases, our study cannot account for the cases for which it is not.)

Our key measures capture homophily and consolidation in each community in each year. We consider five social dimensions: (i) degree of education, (ii) type of occupation, (iii) land ownership, (iv) property ownership, and (v) business ownership. In Blau and Schwartz's (1984) words, each of these social dimensions provides opportunities for social interaction.

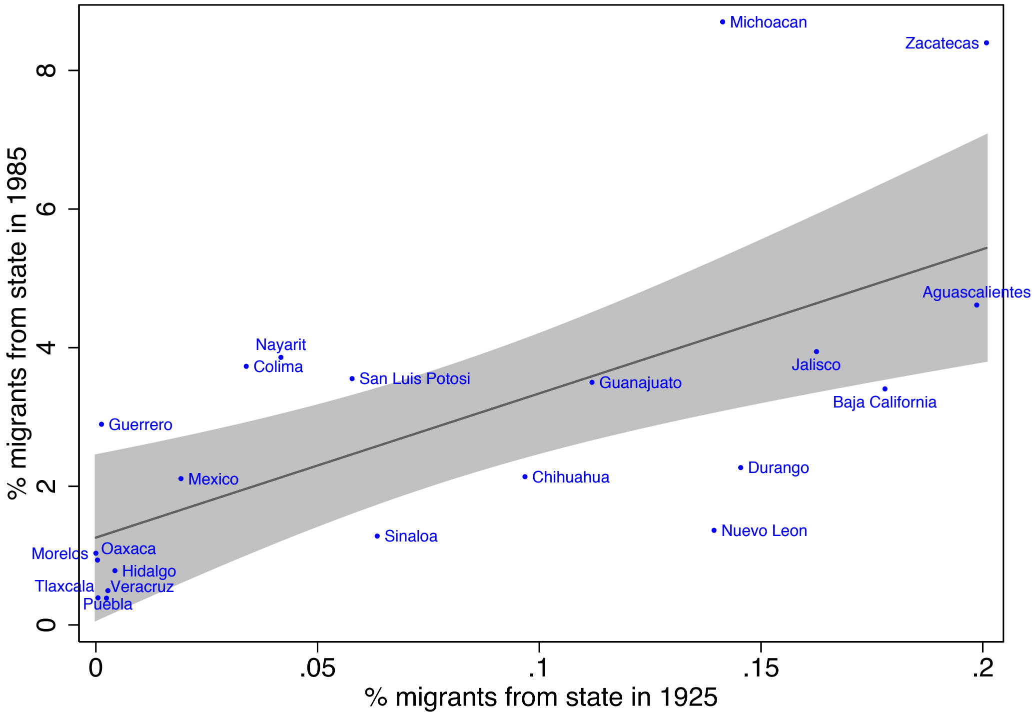
To measure homophily, we first compute a separate proxy for each dimension and run separate models. We also compute a composite measure, which equals the mean of all pairwise social distances in each community-year. (Social distance is based on the Euclidean distance based on the five dimensions, where each dimension is standardized to 0-1 range. Standardizing the dimensions to the same variance leads to similar results.)

To measure consolidation, following McPherson and Smith-Lovin (1987), we use the mean pairwise correlation between different dimensions. Because some dimensions (e.g., occupation) are measured on an ordinal scale, we use Spearman's rank correlation.

Preliminary results

Visualizing the 'puzzle'
Figure 1 compares the percentage of migrants in each of the 21 Mexican states in 1925 and 1985. (The former are from Foerster (1925), which reports the number of migrants by state of origin entering the United States through Texas and California in 1925. According to Borjas and Katz (2007), these states accounted for more than 90% of the Mexican-born population in the United States at the time.) We divide this number by the state population reported in the 1930 Mexican Census (INEGI), and report the percentages on the x-axis. The y-axis shows the percentage of migrants from each state in 1985 according to the MMP data. (MMP communities were surveyed in different years in the 1982-2008 period. Using retrospective life histories, we can trace communities in time, but the number of observations drops as we go further back. For most descriptives, we fix the time at 1985, when all communities are observed, and the number of observations is still abundant.)

Figure 1 State-level migration rates to the United States in 1925 and 1985

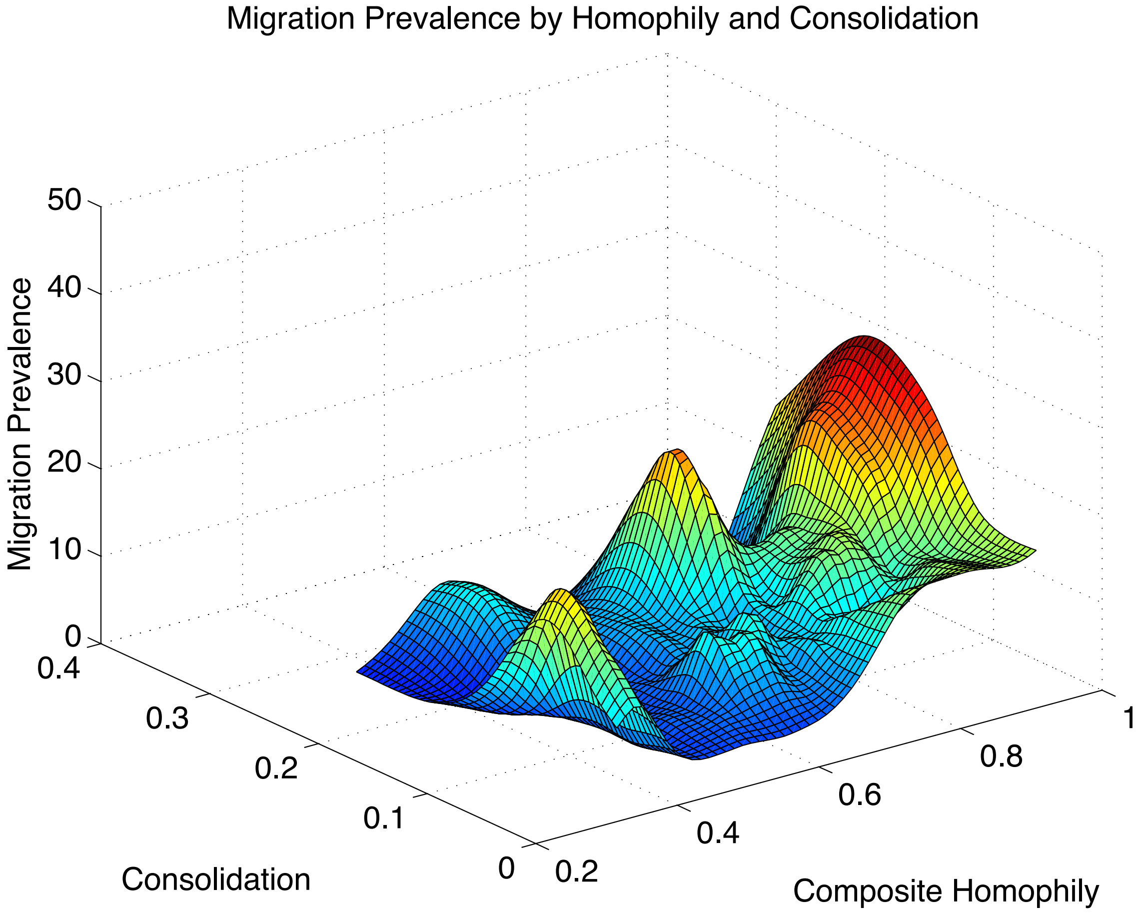


The points show both continuity and divergence over time. The points for 10 out of the 21 states remain within the 95% confidence band around the fitted line obtained by regressing the 1925 rates on the 1985 rates. (Quadratic term is dropped, as it is not statistically significant.) The estimated slope (20.8, $p<0.05$) suggests that states, on average, have experienced a 20-fold increase in their out-migration rates over 60 years. But the rate of increase has varied across states. Six states (Guerrero, Nayarit, Colima, San Luis Potosi, Michoacán and Zacatecas) enjoyed larger-than-expected increases in migration rates given the linear model, while six others (Sinaloa, Chihuahua, Nuevo Leon, Durango and Baja California) saw lower-than-expected increases. In this paper, we seek to explain the sources of this variation in migration outcomes.

Visualizing the relationship among homophily, consolidation and migration prevalence
Using a thin-plate smoothing spline (a spatial interpolation method), we generate a 3-dimensional surface graph depicting the relationship among homophily, consolidation and migration prevalence. Centola (2012) has provided a similar plot based on simulations from a computational model, which suggested a nonlinear (first increasing then decreasing) relationship between homophily and diffusion, as well as between consolidation and social diffusion. Specifically, his results showed that consolidation and homophily are interdependent, and each parameter – without any contribution from the other- cannot support successful diffusion.

Although based on real data (which is a lot more variable than data generated from a controlled experiment in a computational setting), our results are strikingly similar. <u>In the 124 Mexican communities, those with moderate to high levels of consolidation, and moderate to high levels of homophily, are the most likely to achieve high levels of migration prevalence.</u>

Figure 2



The output below shows the coefficient estimates from two models of migration prevalence: (i) baseline model (base), and (ii) model with state dummies (state_dum). Both models include linear and quadratic terms for composite homophily (chom) and consolidation (cons) as well as their interaction. The unit of analysis is the community-year (N=3,512). In the baseline model, both homophily and consolidation have a nonlinear relationship to prevalence. The interaction between the linear terms of the variables is positive, while that between the quadratic terms is negative.

```
----------------------------------------
    Variable |     base      state_dum
-------------+--------------------------
        chom | -133.40***   -114.83***
       chom2 |   71.75***     71.30***
        cons | -479.69***   -191.87*
       cons2 |  591.41**     -66.75
   chom_cons |  787.27***    411.99***
 chom2_cons2 | -1616.71***   -453.80
-------------+--------------------------
           N |    3512         3512
        r2_a |    0.12         0.34
----------------------------------------
   legend: * p<.05; ** p<.01; *** p<.001
```

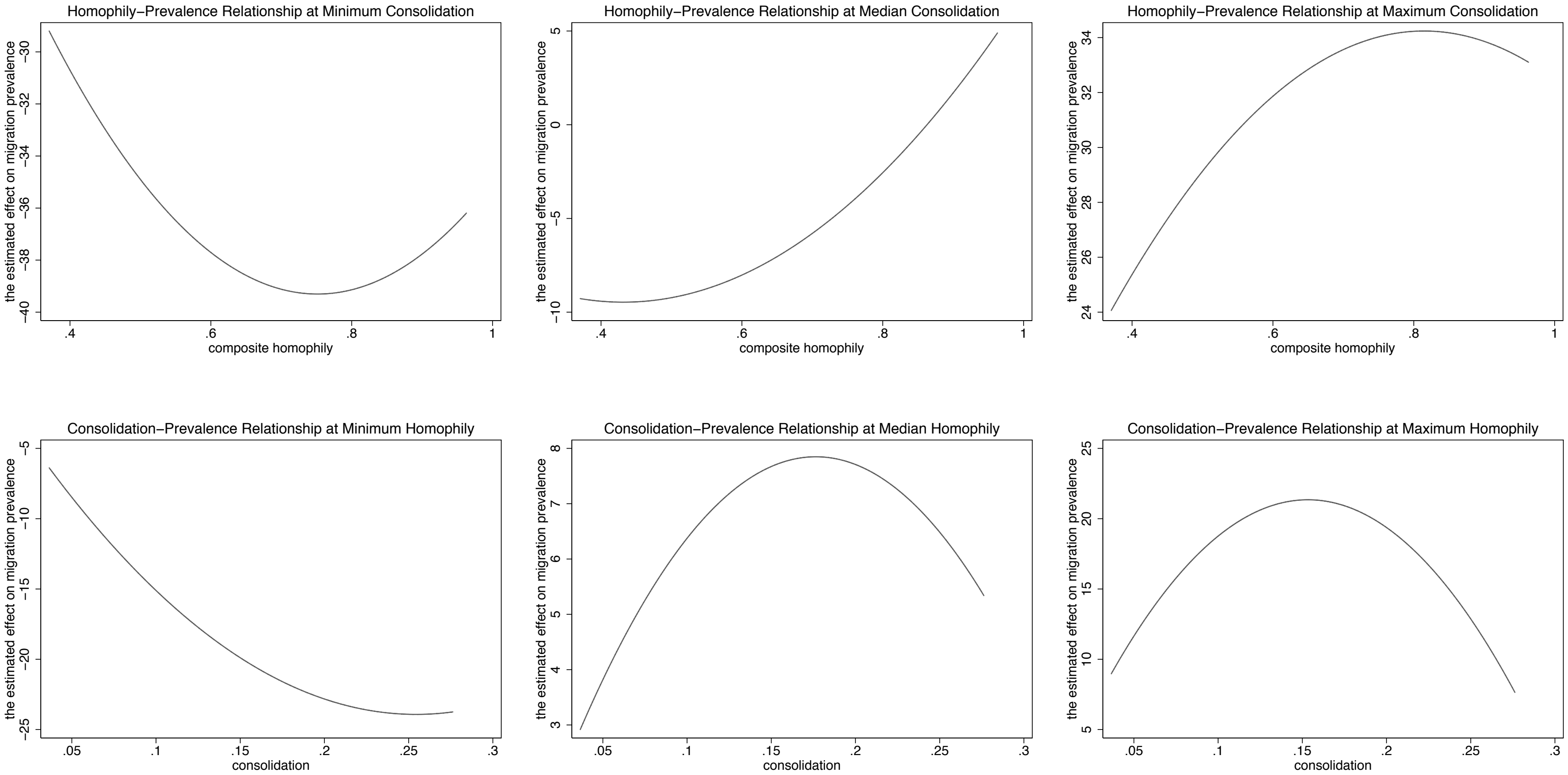
The figures below explore these relationships complicated by the presence of interaction terms. The upper-left figure shows the effect of homophily on prevalence, holding consolidation at its minimum value (0.03). The effect is negative and increasing in size, reaching its minimum value at 0.75. The upper-middle figure shows the relationship when consolidation at its median value (0.11). The effect is initially negative, increasing in absolute size until it reaches 0.43 and declining thereafter. The effect becomes positive once homophily reaches 0.86. The upper-right figure repeats the analysis with consolidation at its maximum value (0.28). The effect of homophily on prevalence is positive and increasing until it reaches 0.82, and declining thereafter. Overall, homophily is associated with lower prevalence at low levels of consolidation, and with higher prevalence at high levels of consolidation.

The lower-left figure shows the estimated relationship between consolidation and prevalence while homophily is held at its minimum (0.37). The effect of consolidation is negative and increasing in size, reaching its minimum point at 0.25. When homophily is fixed at its median value (0.72) in the lower-middle figure, the effect of consolidation becomes positive and increasing in size, albeit at a decreasing rate, peaking at 0.18. Similarly, when homophily is set at its maximum value (0.96) in the lower-right figure, the effect of consolidation on prevalence is positive, and increasing at a declining rate, with a peak at 0.15.

These findings largely confirm several insights from the Centola (2012) study. First, homophily and consolidation are interdependent – each amplifying the effect of the other on the diffusion of migration. Second, a combination of (i) high homophily and moderate consolidation (as seen in the upper-middle and lower-right figures), or (ii) moderate homophily and moderate to high consolidation (as seen in the upper-right and lower-middle figures) leads to the highest diffusion levels. Third, a combination of low homophily and low consolidation (upper-left and lower left figures) leads to lower (‘failed’ in Centola’s terms) diffusion outcomes.

Although real data covers a much narrower range of consolidation ([0.03, 0.28]) and homophily ([0.37, 0.96]) than the theoretical range of [0,1] explored in Centola (2012), the observed relationships among consolidation, homophily and migration rates are strikingly similar to those revealed in his simulations.

Figure 3 Homophily, Consolidation, Migration Prevalence Relationship


Homophily–Prevalence Relationship at Minimum Consolidation

Homophily–Prevalence Relationship at Median Consolidation

Homophily–Prevalence Relationship at Maximum Consolidation

Consolidation–Prevalence Relationship at Minimum Homophily

Consolidation–Prevalence Relationship at Median Homophily

Consolidation–Prevalence Relationship at Maximum Homophily

References

Blau, P. M., and J. E. Schwartz. 1984. *Crossing Social Circles.* Orlando: Academic Press

Borjas, George and Lawrence F. Katz. 2007. “The Evolution of the Mexican-Born Workforce in the U.S. Labor Market.” Pp. 13-55 in *Mexican Immigration to the United States*, edited by George J. Borjas. University of Chicago Press.

Centola, Damon. 2012. “Overlapping Groups and Wide Bridges: Structural and Agential Constraints in the Dynamics of Social Diffusion.” Working Paper.

Foerster, Robert F., 1925. The Racial Problems Involved in Immigration from Latin America and the West Indies to the United States. United States Department of Labor.

Massey, Douglas S. 1990. “Social Structure, Household Strategies, and the Cumulative Causation of Migration.” *Population Index* 56: 3-26.

McPherson, J. Miller and Lynn Smith-Lovin. 1987. “Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups.” *American Sociological Review* 52(3): 370-379.