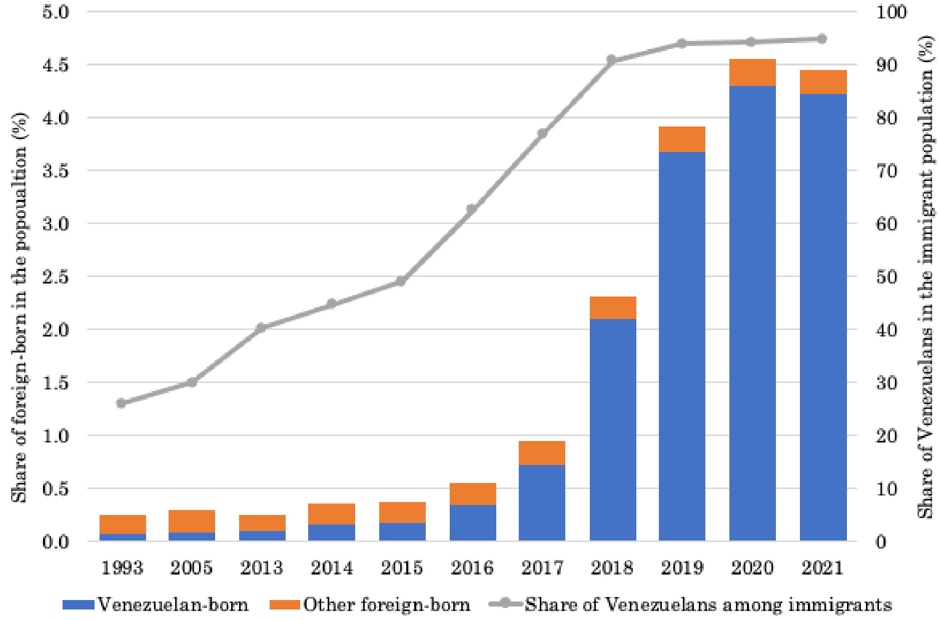
Online Appendix
Immigrant Networks in the Labor Market

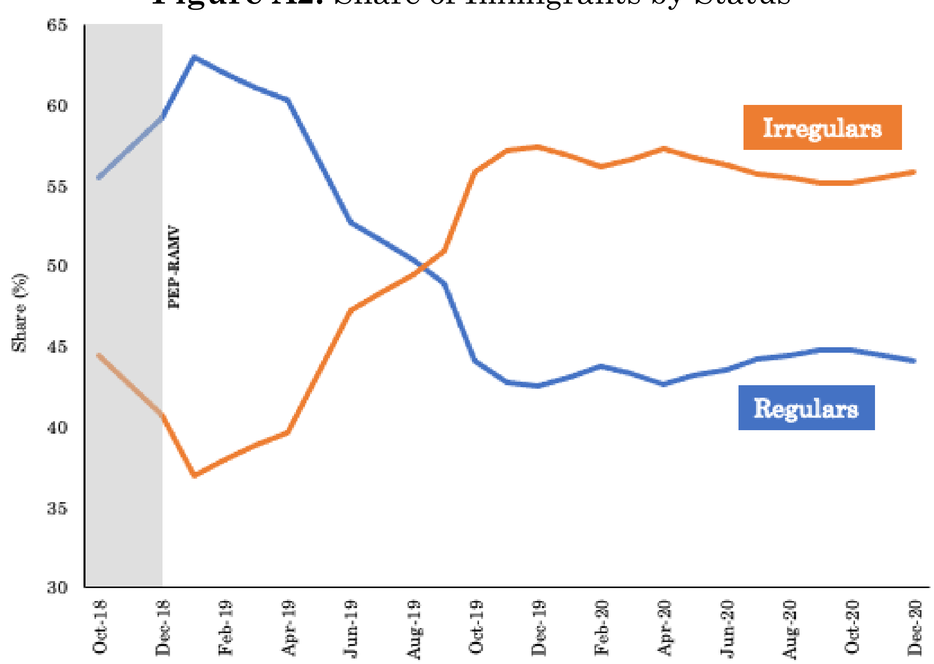
A Additional Figures and Tables

Figure A1: Immigration Shock



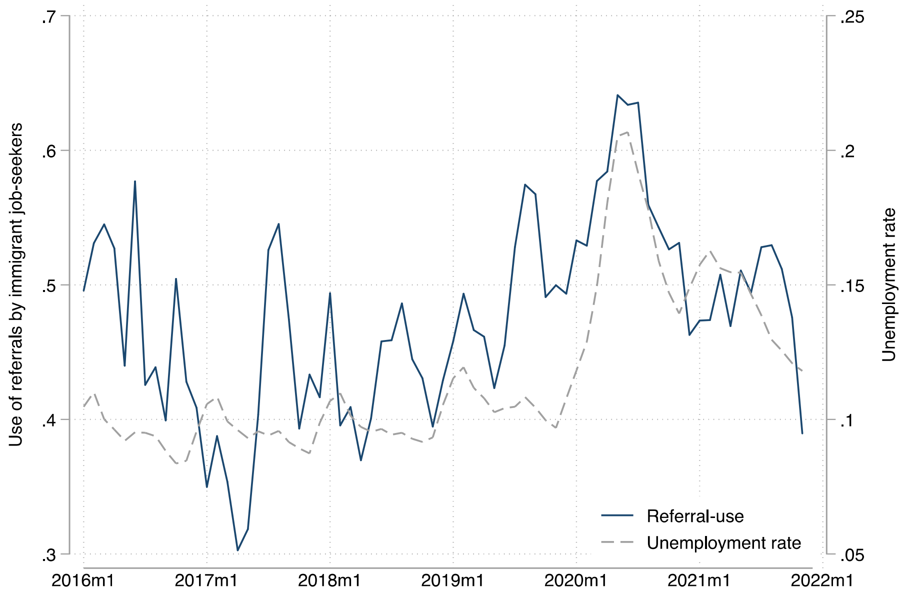
Notes: The Figure presents the foreign-born population as a percentage of the total population in Colombia (left axis) and the share of Venezuelan-born immigrants in the total foreign-born population (right axis) between 1993 and 2021. Shares are estimated using the population aged 15 to 64 years. Sample weights are based on the 2018 Population Census projections. *Source:* GEIH (2013-2020), Population Census (1993, 2005).

Figure A2: Share of Immigrants by Status



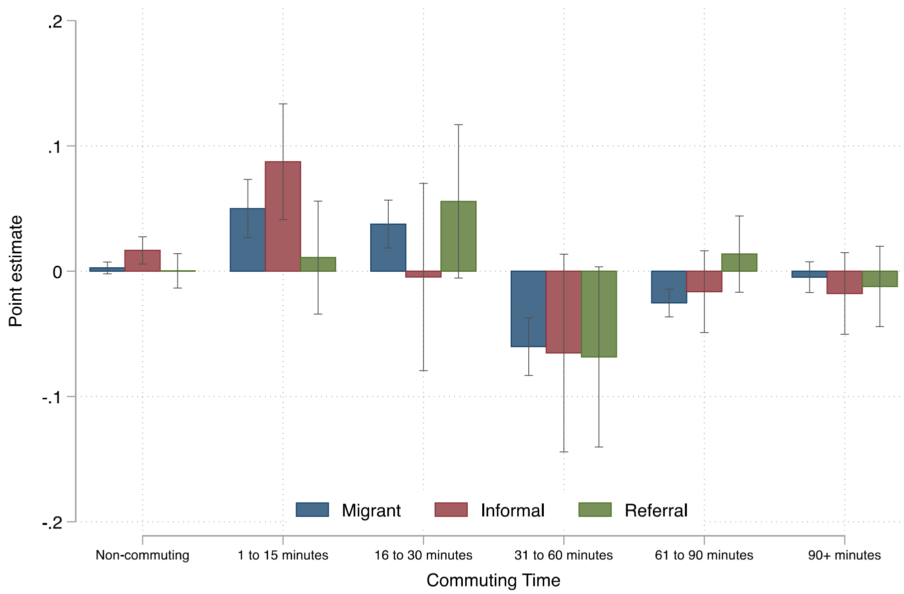
Notes: The Figure shows the share of immigrants with a regular and irregular status between October 2018 and December 2020. The shaded area indicates the first regularization period of undocumented immigrants, known as PEP-RAMV. *Source:* Migración Colombia; R4V.

Figure A3: Unemployment and the use of Job Referrals



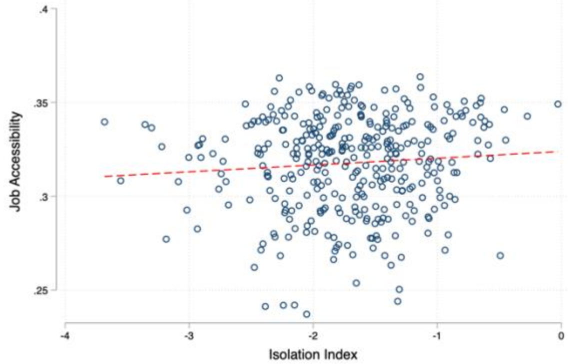
Notes: The Figure displays the evolution between 2016 and 2021 of the use of job referrals among immigrant job seekers and total unemployment rate in Colombia. Each point in time corresponds to a 3-month moving average. Sample is restricted to workers aged 15 to 64 years. *Source:* 2016-2021 GEIH.

Figure A4: Differences in Commuting Time among Immigrants

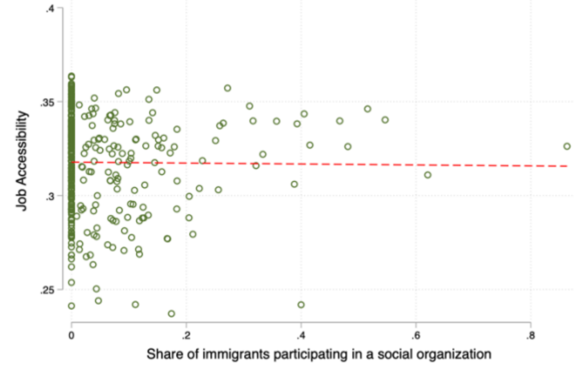


Notes: The Figure plots the point estimates and 95-percent confidence intervals of multiple regressions relating the probability of commuting within some window of time on a dummy variable for group membership (immigrant, informal worker, or finding a job through referrals). Sample is restricted to Venezuelan-born workers aged 15 to 64 years living in Bogotá. All regressions control for age and number of household members in the labor force, and include dummies for sex, marital status, head of household, educational attainment, work permit, residential neighborhood, and mode of transportation. Standard errors are clustered at the neighborhood level. *Source:* 2021 EMB.

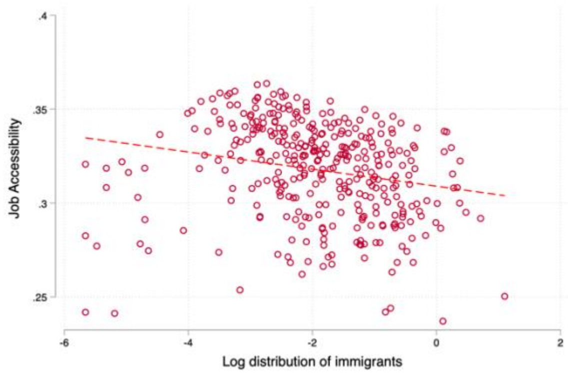
Figure A5: Spatial vs. Social Mismatch



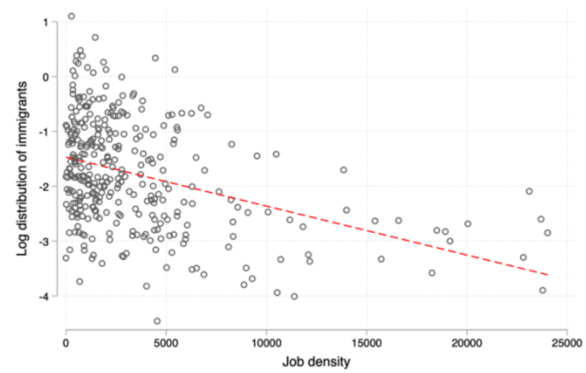
(a)



(b)



(c)



(d)

Notes: Panel (a) plots the relationship between the measure of job accessibility for each neighborhood (sector) and the (log) isolation index. Panel (b) plots the measure of job accessibility against the share of immigrants aged 15 to 64 years participating in a social, cultural, political, religious, productive, or union organization. Panel (c) plots the measure of job accessibility against the (log) distribution of immigrants across neighborhoods. Panel (d) plots the (log) distribution of immigrants across neighborhoods against the job density for high-wage jobs. High-wage jobs are defined as jobs paying above two-thirds of the median hourly wage (including self-employment labor income) for full-time, male workers in the city. All plots exclude neighborhoods with fewer than five sampled immigrant workers. *Source:* 2021 EMB.

Table A1: Extent of Sorting within Neighborhoods

Variable	R^2 Method Unconditional (1)	R^2 Method Conditional (2)	Residual Method Unconditional (3)	Residual Method Conditional (4)
Age	.006	.000	.080	.012
Male	.001	.001	.024	.028
Married	.002	.000	.043	.011
With children aged 0-14	.003	.000	.053	.016
High school graduate or lower	.163	.024	.403	.154
College graduate	.265	.046	.515	.214
Immigrant	.070	.039	.264	.197

Notes: The Table reports estimates of the extent of sorting within census blocks by comparing a series of individual characteristics for a randomly selected worker with the corresponding average characteristics in the block (not including the individual or someone in the same household). Only blocks with five or more workers are kept in the sample. Columns 2 and 4 condition on block group fixed effects. The first two columns report the R^2 from a regression of the individuals' characteristics on the block-level average. The last two columns report the pairwise correlation of the residuals of a regression of the individuals' characteristics and the block-level average on the block group fixed effect. *Source:* EMB–RELAB.

Table A2: Wages and Commuting Time

	(1)	(2)	(3)	(4)	(5)
A. All Venezuelan-born immigrant workers					
	(*log*) Hourly wage				
Commuting time (in minutes)	.0021	.0017	.0014	.0015	–.0006
	(.0006)	(.0005)	(.0005)	(.0006)	(.0020)
Demographic controls		✓	✓	✓	✓
Allowed to work dummy			✓	✓	✓
Sector of residence FEs				✓	✓
Sector of employment FEs					✓
Observations	2,023	2,023	2,023	2,023	1,118
B. Immigrant workers who lost their job one year before					
	(*log*) Hourly wage				
Commuting time (in minutes)	.0030	.0028	.0027	–.0006	–.0059
	(.0014)	(.0013)	(.0012)	(.0016)	(.0554)
Demographic controls		✓	✓	✓	✓
Allowed to work dummy			✓	✓	✓
Sector of residence FEs				✓	✓
Sector of employment FEs					✓
Observations	502	502	502	502	261

Notes: The Table reports results after regressing (log) wages on commuting time at the individual level. Panel A estimates results using all Venezuelan-born workers aged 15 to 64 years living in Bogotá. Panel B restrict the sample to immigrants that lost their job due to the Covid-19 restrictions and were living in the same neighborhood the year before. Wages include earnings of wage and salary workers and independent contractors. Demographic controls include age groups, educational attainment, number of household members in the labor force, and dummies for male, marital status, and head of household. Reported trips longer than 3 hours are excluded. *Source:* 2021 EMB.

B Empirical Facts Appendix

B.1 Fact 5: Effect of Network Strength on Labor Market Outcomes

I examine whether very local interactions (match strength) have an impact on labor market outcomes. Following Bayer *et al.* (2008), I construct a proxy of network strength at the individual level that intends to capture how likely are other workers in the block in helping an individual find a job.

I start by constructing a sample of all possible pairings of individual i with other individuals who reside in the same block $b(i)$ and do not belong to the same household, using all working-age individuals (aged 15 to 64 years).[1] For each pair (i, j), I compute a linear combination of the pair's covariates using the estimated parameters from the interaction of these variables with R_{ij} in Eq. (1) in the paper: $M_{ij} = \hat{\alpha}_1' X_{ij}$.[2] I then average M_{ij} over all matches for individual i, where $N_{b(i)}$ is i's number of neighbors, to get our network strength proxy, Q_i:

$$Q_i = \frac{1}{N_{b(i)}} \sum_{j \in N_{b(i)}} M_{ij} \,. \tag{B.1}$$

Since the network strength measure does a better job of characterizing the referral effect for workers who are less attached to the labor market, I focus on the sample of immigrant workers that lost their job due to the COVID-19 pandemic or that were living in another country 12 months before. Taking as the unit of observation an individual rather than a pair, I estimate the following equation:

$$y_i = \theta_{b(i)} + \delta_1 Q_i + \delta_2' X_i + \mu_i \,, \tag{B.2}$$

where y_i is a labor market outcome; $\theta_{b(i)}$ is a block-level fixed effect; X_i is a vector of individual characteristics (see Table 3 in the paper); and μ_i is an individual error term. I standardize Q_i to express results as a one-standard-deviation increase in network strength on the corresponding labor market outcome. By including block-level fixed effects, δ_1 identifies the additional effect of network strength once we account for average outcomes and attributes of workers in the block. For all employment outcomes and the probability that a worker's commuting time is less than 30 minutes, I estimate a linear probability model. For hours worked and hourly wage, I estimate a linear regression.

All results are presented in Table B.1. For the specifications using all origin-country groups (including natives), match strength has a positive and statistically significant effect on informality, hours worked, and commuting short distances. For

[1] Before constructing the pairs, I drop blocks with fewer than five observations and sectors with fewer than two blocks.

[2] In the computation, I only include parameters that are statistically significant at a minimum at the 10% percent level.

instance, a one-standard-deviation increase in match strength rises the probability of finding an informal job by about 5.5 percentage points, average hours worked per week by about 0.7 hours, and the probability of commuting within 30 minutes by 1.6 percentage points.

Table B1: Effect of Network Strength on Immigrant's Labor Market Outcomes

	Origin-country group					
	All groups			*Venezuelan-born*		
Dependent variable	Obs.	Coefficient	S.E.	Obs.	Coefficient	S.E.
Employment	25,019	.001	.005	2,191	.000	.011
Employment: wage and salary workers	17,098	–.022	.007	1,757	.004	.013
Employment: informal	17,098	.055	.008	1,757	–.006	.012
Hours worked per week	17,098	.704	.232	1,757	.175	.515
(*log*) Hourly wage	9,285	–.120	.020	811	–.002	.030
Commuting time	12,546	–.376	.389	1,456	.193	.594
Pr(commuting ≤ 30 minutes)	12,546	.015	.006	1,456	.006	.013
Standard deviation of network strength (%)		.691			1.703	

Notes: The Table reports results of a single regression for each of the six labor market outcomes on a proxy for network strength (Q_i) and the full set of individual characteristics reported in Table 3. Block fixed effects are included in all regressions. The regression for commuting time includes, in addition, origin and mode of transportation fixed effects. Results are for a sample of workers aged 15 to 64 years that lost their job due to the Covid-19 pandemic or that were living in another country 12 months before. The coefficients reported in the table characterize the effect of a one-standard-deviation increase in match quality on the corresponding labor market outcome. Standard errors are clustered at the block level. *Source:* EMB–RELAB.

B.2 Fact 6: Estimation of Clustering at Industries and Occupations

To study whether workers employed in the same industry or occupation are likely to live in the same neighborhoods, I follow broadly Hellerstein *et al.* (2011) and compare the observed clustering of immigrants versus what a random clustering would yield. Using the 2021 EMB sample, I match recently arrived immigrants (those arriving in the las 12 months to the country) to immigrants arriving in earlier waves. The sample is restricted to Venezuelan-born immigrants living in Bogotá who are between 15 and 64 years of age.

Let i (recent arrival) and j (earlier cohort) be a pair of immigrant workers; $I^R(i,j)$ is a dummy variable equal to one if i and j live in the same neighborhood (sector); and $I^W(i,j)$ is a dummy variable equal to one if i and j work in the same 4-digit industry or occupation, respectively.[3] Using the sample of pairs, I compute for each recent arrival the percentage of immigrant workers from earlier cohorts working in the same industry (respectively occupation) who live in the same neighborhood (sector)—excluding the individual worker. I average this share across all N recently arrived immigrants to create the *network isolation index*, NI^O:

[3] I restrict the sample to industries and occupations with at least two observed immigrant workers and drop pairs where both workers belong to the same household.

$$NI^O = \frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{\sum_i I^R(i,j) \times I^W(i,j)}{\sum_i I^W(i,j)} \times 100\,. \tag{B.3}$$

Note that the sums in the numerator and denominator are taken over all pairs for worker i. Their ratio is the fraction of previous immigrants in the same industry or occupation that live in the same neighborhood as worker i. To do inference, I bootstrap the entire sample of pairs with replacement 100 times and compute NI^O with the corresponding standard deviation and sample size; then, I estimate the mean standard error and report it along the network isolation index.

Since some neighbors are likely to work in the same industries or occupations, even if workers are assigned randomly to industries or occupations, I compare the network isolation measure to the extent of clustering that occurs *randomly* and denote this measure as NI^R. I randomly assign immigrant workers to industries and occupations, ensuring that I generate the same size distribution of industries and occupations (in terms of matched workers) in the city as I have in the sample. This is basically assigning workers to industries or occupations holding constant every time the number of workers that end up employed in a given industry or occupation. For each simulation, I compute NI^O. I repeat this 100 times and compute NI^R as the mean over these simulations.

All results are presented in Table 4 in the paper.

B.3 Fact 7: Urban Mismatch

To analyze if distance to jobs (spatial mismatch) and limited social connections (social mismatch) affect immigrants' labor market outcomes, I construct the following measures:

(i) *Social mismatch*. I start with a measure of residential segregation: isolation index. This measures the extent to which immigrants are exposed only to one another. Let M_n be the number of immigrants aged 15 to 64 years in block n, M_S the number of immigrants aged 15 to 64 years in neighborhood (sector) S, and L_n the total population aged 15 to 64 years in block n, then the isolation index $I_{S(n)}$ at the sector-level is constructed using the following formula:

$$I_{S(n)} = \sum_{n\in S}\left(\frac{M_n}{M_S}\right)\left(\frac{M_n}{L_n}\right).$$

As a second proxy of social mismatch, I estimate the share of immigrants in the neighborhood (sector) aged 15 to 64 years participating in a social, cultural, political, religious, productive, or union organization. This measures the membership to institutions that provide social capital, providing information about the degree of interactions with weak ties (*e.g.*, natives).

(ii) *Spatial mismatch*. I measure job-access using a gravity-based accessibility measure following Shen (1998). Let A_n be the accessibility to employment from residential neighborhood n; N represents the total number of residential and employment location; J_m is the number of jobs in neighborhood m (workplace location); T_{nm} is the average commuting time from residential neighborhood n to each workplace location m (one-way distance); C_m is the competition or potential demand for jobs in neighborhood m; and W_n is the number of workers (employed and unemployed) living in n. The job-access measure that incorporates the location of competing workers is estimated as follow:

$$A_n = \sum_{m=1}^{N} f(T_{nm}) \frac{J_m}{C_m} \quad \text{where} \quad C_m = \sum_{n=1}^{N} f(T_{nm}) W_n \, .$$

The term $f(T_{nm})$, also known as the “distance decay” effect, increases the spatial variation in the competition for jobs that is being driven by variation in population density across neighborhoods. I model the distance decay function an iceberg commuting cost such that $f(T_{nm}) = (e^{\nu T_{nm}})^{-1}$. I take $\nu = 0.012$, the rate of spatial decay or disutility from commuting, from Tsivanidis (2019) who estimated it for Bogotá using the 2015 Mobility Survey.

For a pair of residential (n) and workplace (m) locations where I do not observe in the data commuting flows, and therefore I cannot compute the average commuting time, I impute the average commuting time from residence n to workplace m using the STATA command *osrmtime* (Huber & Rust, 2016). The command uses the Open-Source Routing Machine (OSRM) and OpenStreetMap to find the optimal route by car.

B.4 Fact 8: Informal to Formal Employment Transitions

To estimate the effect of residence-based networks on immigrants' informal-to-formal job transitions, I leverage the expansion in 2018 of a two-year special permit (known as PEP-RAMV) that allowed irregular or undocumented immigrants to stay and work in Colombia. I use information in the 2021 EMB on Venezuelan-born immigrants aged 15 to 64 years living in Bogotá with a PEP. Because the information in the EMB does not distinguish between PEP (first wave) and PEP-RAMV (second wave), I rely on both the timing of when each policy was introduced and eligibility requirements to restrict the sample to those most likely to be holding a PEP-RAMV instead of the traditional PEP. The sample is constructed by excluding the following workers:

(i) Those living in Colombia for more than 5 years or less than 12 months. The PEP-RAMV targeted workers who arrived between 2017 and 2018.
(ii) Those holding only a Colombian ID and those with valid work visa. Because the PEP-RAMV targeted undocumented migrants, those with a Colombian ID or a work visa are less likely to have been part of the cohort of interest.

(iii) Those that show up in the RELAB before August 2, 2018. Immigrants employed in formal jobs before the introduction of the PEP-RAMV would not have had the irregular status.

(iv) Those workers that changed neighborhood in the last year or moved from a different municipality. I'm interested in looking at the effect for workers who did not changed neighborhood from the time the policy was introduced. The assumption made here is that the residential location of workers who report not moving in the past year has remained the same since 2018. Some evidence indicates that the fraction of movers on a yearly basis is small.

I measure the quality of social contacts in each neighborhood based on the extent to which information about formal (high-wage) jobs could potentially be diffused through the network, weighted by the size of the initial network. Using information on a previous wave of the EMB for 2017, I construct an index that ranks neighborhoods based on the unemployment rate, the share employed in the formal sector, and the share employed in low-income jobs for Venezuelan immigrants.[4] I begin by sorting neighborhoods by each measure. The share of formal employment is sorted from low to high, while the unemployment rate and share of low-wage jobs are sorted from high to low. I then create a cumulative percentile distribution of the total number of immigrant workers in each neighborhood based on the ranking for each measure. I average the three cumulative percentage distributions. Scores can range from 0 to 100.

References

Huber, S., & Rust, C. (2016). Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM). *Stata Journal, 16*(2), 416–423.

Shen, Q. (1998). Location Characteristics of Inner-City Neighborhoods and Employment Accessibility of Low-Wage Workers. *Environment and Planning B: Planning and Design, 25*(3), 345–365.

Tsivanidis, N. (2019). Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio. Mimeo, University of California, Berkeley.

[4] Low-income jobs are defined as immigrants earning lower than two-thirds of the median hourly income for full-time, male workers in the city. In the data, the threshold is slightly above the legal minimum wage which is the wage floor in the formal sector.