IEB

Institut d'Economia de Barcelona

Document de treball de l'IEB 2013/20

DO LABOR MARKET NETWORKS HAVE AN IMPORTANT SPATIAL DIMENSION?

Judith K. Hellerstein, Mark J. Kutzbach, David Neumark

Cities and Innovation

DO LABOR MARKET NETWORKS HAVE AN IMPORTANT SPATIAL DIMENSION?

Judith K. Hellerstein, Mark J. Kutzbach, David Neumark

The **IEB** research program in **Cities and Innovation** aims at promoting research in the Economics of Cities and Regions. The main objective of this program is to contribute to a better understanding of agglomeration economies and 'knowledge spillovers'. The effects of agglomeration economies and 'knowledge spillovers' on the Location of economic Activities, Innovation, the Labor Market and the Role of Universities in the transfer of Knowledge and Human Capital are particularly relevant to the program. The effects of Public Policy on the Economics of Cities are also considered to be of interest. This program puts special emphasis on applied research and on work that sheds light on policy-design issues. Research that is particularly policy-relevant from a Spanish perspective is given special consideration. Disseminating research findings to a broader audience is also an aim of the program. The program enjoys the support from the **IEB-Foundation.**

The **Barcelona Institute of Economics (IEB)** is a research centre at the University of Barcelona which specializes in the field of applied economics. Through the **IEB-Foundation**, several private institutions (Applus, Abertis, Ajuntament de Barcelona, Diputació de Barcelona, Gas Natural and La Caixa) support several research programs.

Postal Address:
Institut d'Economia de Barcelona
Facultat d'Economia i Empresa
Universitat de Barcelona
C/ Tinent Coronel Valenzuela, 1-11
(08034) Barcelona, Spain
Tel.: + 34 93 403 46 46
Fax: + 34 93 403 98 32
ieb@ub.edu
http://www.ieb.ub.edu

DO LABOR MARKET NETWORKS HAVE AN IMPORTANT SPATIAL DIMENSION? *

Judith K. Hellerstein, Mark J. Kutzbach, David Neumark

ABSTRACT: We test for evidence of spatial, residence-based labor market networks. Turnover is lower for workers more connected to their neighbors generally and more connected to neighbors of the same race or ethnic group. Both results are consistent with networks producing better job matches, while the latter could also reflect preferences for working with neighbors of the same race or ethnicity. For earnings, we find a robust positive effect of the overall residence-based network measure, whereas we usually find a negative effect of the same-group measure, suggesting that the overall network measure reflects productivity-enhancing positive network effects, while the same-group measure may capture a non-wage amenity.

JEL Codes: J15, J30, J63

Keywords: Networks, job matches, wages, turnover.

Judith K. Hellerstein
University of Maryland and NBER
Department of Economics
University of Maryland
College Park
MD 20742, USA
E-mail: hellerst@econ.umd.edu

Mark J. Kutzbach
Center for Economic Studies,
U.S. Bureau of the Census
4600 Silver Hill Rd
Washington, DC 20233, USA
E-mail: mark.j.kutzbach@census.gov

David Neumark
UCI, NBER, and IZA
Department of Economics
3151 Social Science Plaza
Irvine, CA 92697-5100, USA
E-mail: dneumark@uci.edu

* We are grateful to Gilles Duranton, Kristin McCue, Erika McEntarfer, Henry Overman, and Giorgio Topa for helpful comments. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

I. Introduction

Research in labor economics has explored the potential for network connections among workers along a number of dimensions – including common military service, attending the same school, or coming from the same family. There appears to be a relatively common finding that workers who are connected to each other in some way that could plausibly result in them sharing labor market information also seem to have similarities in labor market outcomes, consistent with a role for labor market networks.

In a series of papers, two of us have explored a particular dimension of labor market networks – spatial labor market networks that connect workers who live in the same neighborhood. Using matched employer-employee data, Hellerstein et al. (2011) show that neighbors are more likely to work in the same establishments than would be predicted simply by the fact that neighbors are likely to work near where they live,[1] and for minorities and especially Hispanic immigrants the clustering of neighbors in the same workplace is dramatic. In addition, the study finds evidence suggesting that these residence-based networks are racially stratified. In particular, blacks are much more clustered at work with their black neighbors than with their neighbors overall (i.e., without regard to race). These findings suggest that labor market connections among neighbors may be an important source of network connections in the labor market.

In this paper, we turn to different types of evidence on labor market networks to explore further the role of residence-based labor market networks. In particular, we study evidence of the productivity of these networks in terms of turnover and earnings. We draw on theoretical work (Dustmann et al., 2011; Brown et al., 2012) that derives implications of labor market networks that arise because network connections lead to better labor market matches. What is new, however, is that we test these predictions for residence-based networks.

This inquiry is useful for two reasons. First, if the evidence we have assembled in our prior work is really identifying labor market networks, then these theoretical implications for job matches should carry over to workers potentially connected via residence-based labor market networks. Second, although

[1] This is also true conditional on skill measures.

there is every reason to expect that there are various types of labor market networks, and workers may be connected to others through more than one type of network, it is nonetheless useful to try to identify the most important sources of labor market connections. Nothing in our evidence contrasts the importance of residence-based networks with networks along fundamentally different connections, like common military service. However, our evidence can help distinguish between networks based on place of residence – which therefore have an important spatial dimension – and networks based simply on common race or ethnicity.

Given racial and ethnic segregation of neighborhoods, residence-based labor market networks could give rise to evidence that looks simply like networks based on common race or ethnicity. For example, Dustmann et al. (2011) find that workers who work with a larger share of workers from their ethnic group have lower turnover and higher wages – consistent with the predictions of their model. However, it may be that this arises because workers from the same neighborhood work together and are likely to be of the same ethnic group. Thus, looking explicitly at the role of residence-based labor market networks, and contrasting this with the role of networks that may be based solely on common race or ethnicity, can help us pin down the spatial nature of labor market networks.

Identifying an important spatial dimension of labor market networks is potentially significant for a number of reasons. First, if policy is to try to leverage labor market networks to get multiplier effects,[2] then policymakers have to know which connections among workers are productive. Second, some evidence that is consistent with the existence of networks is alternatively consistent with the existence of discrimination, and distinguishing between the two is obviously important. For example, evidence like that in Dustmann et al. (2011) – showing lower turnover and higher wages when one works with more co-ethnics – could stem from labor market discrimination, if employers with a large share of an ethnic group treat workers from that ethnic group better. However, evidence of residence-based networks – which can give rise to the type of evidence Dustmann et al. generate – is harder to explain as stemming from discrimination, and in that sense can give us more solid evidence on labor market networks.

[2] An example is the Jobs-Plus program, discussed in more detail in Hellerstein and Neumark (2013).

And third, establishing that residence-based networks are important in determining labor market outcomes provides new perspectives on how to think about the interrelations between space – in particular, where people live – and the labor market. These issues are central to questions at the intersection of urban economics and labor economics. Residence-based labor networks can, for example, help explain how ethnic and racial residential segregation reinforces poorer labor market outcomes for minorities. But they can also potentially lead us to think about how to increase labor market connections among neighbors that might help offset some of these disadvantages – as may happen for Hispanic immigrants who often live in highly-segregated ethnic enclaves.

II. Related Previous Research and Our Approach

Research on the Presence and Nature of Labor Market Networks

A growing body of evidence in labor economics points to the importance of labor market networks. Earlier evidence consists largely of survey findings indicating widespread reliance on friends, relatives, and acquaintances to find jobs. (See Ioannides and Datcher Loury, 2004, and more recent evidence for European countries in Pellizzari, 2010.) More recent research has studied and typically documented the similarity of employment outcomes for Veterans who served together in World War I (Laschever, 2009), workers displaced from the same firm (Cingano and Rosolia, 2009), family relations (Kramarz and Nordström Skans, 2007), those who attended the same educational institution (Oyer and Schaefer, 2009), and, most commonly, those from similar racial or ethnic groups (Giuliano et al., 2009; Åslund et al., 2009; Dustmann et al., 2011). This research literature provides more specific evidence on particular kinds of network connections that can arise between workers and the possible effects of these network connections on labor market outcomes. A common theme of this evidence is that employment of an individual is likely to be boosted by employment of others in their network.

Other recent work has focused on the geographic or spatial dimension of networks. Bayer et al. (2008) look for evidence of network effects among neighbors using confidential Census data on Boston-area workers. They find that two individuals living on the same Census block are more likely to work on the same Census block than are two individuals living in nearby areas (the same block group) but not on the same block.[3] As long as networks are stronger within blocks than within block groups, but unobserved differences are similar within blocks and block groups, this evidence suggests that residence-based labor market networks affect hiring.

Building on this work, and using data for the United States as a whole, Hellerstein et al. (2011) assess evidence on the importance of labor market networks among neighbors. The evidence improves on Bayer et al. by looking explicitly at who works at which establishment. The study tests for the importance of residence-based labor market networks in determining the establishments at which people work, using matched employer-employee data at the establishment level, based on a large-scale data set covering most of the United States (the 2000 DEED, described in Hellerstein and Neumark, 2003). The measure of labor market networks captures the extent to which employees of a business establishment come disproportionately from the same sets of residential neighborhoods (defined as Census tracts), relative to the residential locations of other employees working in the same Census tract but in different establishments.

Overall, the evidence indicates that residence-based labor market networks play an important role in hiring. For whites, about 10 percent of the maximum amount to which residential networks *could* contribute to the sorting of workers by establishment *is* actually reflected in the sorting of workers into establishments. This figure doubles for blacks in comparable (smaller) establishments. Networks also appear more important for less-skilled workers, as would be expected for network connections among neighbors because of the more local nature of low-skill markets. Finally, residence-based networks are considerably more important for Hispanics. Overall, the grouping of Hispanic workers from the same neighborhoods in the same business establishments is about 22 percent of the maximum, but it is twice as high for Hispanic immigrants and those with poor English skills. These results suggest that informal labor market networks may be particularly important for workers who are not as well-integrated into the labor market, and for whom employers may have less reliable information.

A potential alternative explanation of this evidence pointing to a role for labor market networks is

[3] Block groups comprise a set of Census blocks, and have a target size of about 1,500 people.

that rather than residential neighborhood influencing where one works – via residence-based networks – place of work determines where one lives. If, for example, co-workers recommend neighborhoods or houses to which workers then move, then we would see clustering of neighbors in the same establishments, but this would not be due to the operation of residence-based networks. Hellerstein et al. (2011) are able to rule out this alternative hypothesis. The Census data they use indicate whether a person changed addresses in the past five years, and the establishment-level data has establishment age. Thus, they can restrict attention to residents who have *not* moved in the past five years and who work in establishments that are fewer than five years old, for whom the choice of residential location necessarily preceded the decision to work at a new establishment. This restriction resulted in *stronger* evidence of residence-based networks.

Research has also noted that labor market networks may be race- (or ethnic-) based so that, for example, reliance on informal referrals in a predominantly white labor market benefits whites at the expense of other groups (Kmec, 2007). The simple fact that networks are based on neighborhood of residence implies some racial stratification of networks, given pervasive racial residential segregation in the United States. However, Hellerstein et al. (2011) also present evidence on whether there is racial stratification of networks even *within* neighborhoods, with labor market information less likely to flow between black and white co-residents than between co-residents of the same race. If networks among co-residents are racially stratified, then the likelihood that a black works with a neighbor regardless of race should be smaller than the likelihood that a black works with a black neighbor. The evidence points to weaker network connections between black and white neighbors than among black neighbors; specifically, the empirical importance of networks disregarding the race of neighbors and co-workers falls by more than 40 percent.[4]

Dustmann et al. (2011) also present evidence consistent with ethnicity-based networks. Studying urban workers and firms in Germany, they find that firms are more likely to hire workers from a particular ethnic group if they already employ a high share of workers from that group. No information

[4] There is some other evidence consistent with racially- or ethnically-stratified networks in both the United States and Europe (Kasinitz and Rosenberg, 1996; Semyonov and Glikman, 2009).

on potential network connections among neighbors (or along dimensions other than ethnicity) is used; the network connections are presumed to simply connect members of the same ethnic group. Andersson et al. (2010) also present evidence on the ethnic stratification of workers using LEHD employer-employee matched data. They find that migrants in U.S. metropolitan areas have a high propensity to work with other migrants, and that about half of migrant stratification can be explained by observable employee and employer characteristics (such as industry, residential location, and English-language skills).[5] Residential segregation where a worker lives is found to be especially predictive of segregation at that worker's place of work, while the presence of commute flows from the place of residence to the workplace (both measured at the tract level) do not have additional predictive power.

Research on the Productivity of Networks

There is also research on the effects of networks on labor market outcomes, to a large extent focusing on whether networks enhance efficiency in the job market, presumably by reducing search frictions – as in the Montgomery (1991) model – and leading to better job matches. Dustmann et al. (2011) provide a brief survey of this work, pointing out that the findings on network referrals and wages are mixed and (as in Loury, 2006) may depend on the nature of the referral. However, they also point out a clear identification problem: who uses a network referral is not randomly chosen, nor are the firms that use these referrals. As a consequence, the effect of network referrals – or proxies for them – may be unidentified in cross-sectional estimates.

Dustmann et al. (2011) develop an explicit model of the mechanism linking referrals to labor market outcomes, whereby there is less uncertainty about the productivity of a specific match between a worker and an employer when the hire comes through a referral network, rather than the external labor market. However, this relationship declines with tenure in the firm, as the low-productivity matches are terminated. In this model, then, the "productivity" of networks comes from better matches, reflected in higher wages and lower turnover, especially for less-tenured workers.

Dustmann et al. (2011) test the predictions of their model using matched employer-employee

[5] Hellerstein and Neumark (2008) present some similar evidence on the role of English-language skills in ethnic segregation.

data. Naturally, they do not know how the worker obtained a job. However, their model assumes that referrals are more likely to come from someone in the same ethnic group, so if an employer has a higher share of employment in a worker's ethnic group, the worker was more likely to have been hired via a referral than via the external market. They therefore estimate the relationship between this share and both wages and turnover. Once they include fixed firm effects (and whether or not they include worker fixed effects), the effect of "own-group" share is positive for wages and negative for turnover.[6] Interestingly, the results are the opposite *without* the fixed firm effects, perhaps because there is more hiring of minorities into lower-quality firms that pay lower wages and have high turnover. They also find evidence that these effects diminish with worker tenure, especially in the specification with both worker and firm fixed effects.[7]

Finally, they also consider whether these effects could instead come from "productivity spillovers," whereby workers are more productive if they work with members of their own group. Clearly this could generate the same kinds of effects, including why they decline with tenure if, for example, the complementarities between workers from the same group have to do with language differences that diminish over time. Dustmann et al. test and reject this alternative explanation by looking at an ethnic group that speaks the same language – namely, Austrians. They find the same results on own-group share for Austrians, which is hard to reconcile with productivity spillovers, at least via language

[6] The model in Brown et al. (2012) is similar, and is tested using data from a single company that includes information on whether a person was referred by a current employee of the company. Brown et al. find a long-term effect of "network" hires on reducing turnover, while network hires earn higher wages initially but this wage advantage dissipates and reverses after about six years with the firm.

[7] We have some reservations about two aspects of this analysis. First, in estimating how the effects of the group share on turnover or wages vary with tenure, the sample likely changes with tenure in nonrandom ways, making it difficult to sort out true changes in the effects with tenure from changes in the sample composition. (This applies to the Brown et al. analysis as well.)

Second, if network connections between workers are intended to improve matches by providing employers with information on unobservable characteristics of workers, one may not want to condition on worker fixed effects that capture the unobservables. Equivalently, the interpretation of the group share variables as reflecting network effects is less apparent in the specifications with individual fixed effects. Nonetheless, there may be missing worker-level variables that are spuriously associated with the strength of network connections. For example, in the data we use education is not measured. If low-education workers are more likely to search for jobs in local labor markets, they will appear more networked with their co-workers for network measures based on place of residence.

complementarities.[8]

Schmutte (2010) takes up a similar question by trying to identify the role of local referral networks in the allocation, across workers, of pay differentials associated with specific employers. Following Bayer et al. (2008), a worker's network is assumed to be captured by the distribution of employer-specific wage premia of workers from the same residential block, and the identifying variation comes from the block-level variation relative to the larger neighborhood, which is assumed to capture residential sorting. Using the LEHD data we use in this paper, the basic result is that workers in local referral networks with other workers who earn higher employer-specific wage premia are more likely to change jobs, and when they do change jobs, are more likely to move to a job with a higher employer-specific wage premium. The idea that drives this approach is that workers do not have information about employer-specific wage premia, and networks facilitate obtaining this information. Consistent with the findings in Hellerstein et al. (2011), Schmutte also finds that there is a stronger effect of local referral networks on the pay changes for non-natives than for natives.

Finally, Damm (2012) also presents evidence consistent with labor market networks based on place of residence and ethnicity-based networks boosting both employment prospects and earnings. Specifically, taking advantage of a quasi-experiment involving the settlement of refugee immigrants in Denmark, she finds that those who were settled in areas with higher overall employment rates of non-Western immigrants and co-nationals had a greater probability of finding employment, and had higher annual earnings if employed.

The productivity of networks with respect to turnover has also been addressed in some other work. Simon and Warner (1992), in an earlier study of scientists and engineers, derived a direct prediction that workers hired through referrals should stay on the job longer, because they are more likely

[8] Consistent with this evidence, Hellerstein and Neumark (2003) fail to find evidence of language complementarities for Hispanics in the United States. Specifically, wages are not higher for poor English-speaking Hispanics when they work with a higher share of poor English-speaking Hispanics. However, these results come from a cross-section in which it is not possible to include establishment fixed effects because they would be perfectly collinear with variables describing the share of the workforce with any given set of characteristics. It is possible that crowding of poor English-speaking Hispanics into low-wage establishments biases the estimates against finding language complementarities.

to be in good matches – an implication that naturally follows from any model whereby networks lead to better job matches. And in a much earlier study, Datcher (1983) found that – for black workers and college-educated workers – quit probabilities were lower for those who knew someone at their current place of work before taking a job there, even conditioning on the wage.

Our Approach

Our analysis differs from the existing literature in four ways. First, while Dustmann et al. (2011) infer the existence of a network connection between new workers and existing workers based on common ethnicity, we instead incorporate the measure of network strength based on Hellerstein et al. (2011) – specifically, characterizing a worker's network connections with his co-workers as stronger when more of them are from the same neighborhood. This is closer to what Schmutte (2010) does, although – just like the contrast between Bayer et al. (2008) and Hellerstein et al. (2011) – the nature of the network connection is different, as we focus explicitly on neighbors who work together, for whom we believe there are more likely to be labor market network connections.

Second, we explicitly focus on the question of the racial or ethnic stratification of networks, asking whether the effects of network connections between workers appear stronger for networks defined only for those of the same race or ethnicity. Most of the existing papers do not explicitly study the question of racial or ethnic stratification of networks.[9]

Third, because we can define these residence-based network connections among workers, we can, in a sense, test the validity of these measures of network connections by seeing if theoretical predictions for how network connections affect wages and turnover hold for these networks. Moreover, we can examine whether, once we account for these types of network connections, the more general measure of the share of an employer's workforce in a particular race or ethnic group still satisfies these theoretical predictions. It may, because people can form network connections along lines other than co-residence.

[9] Kmec and Trimble (2009) study the differential effects of white, black, or Latino contacts on pay, and how these differ depending on the racial or ethnic makeup of the firm. However, their analysis does not ask, for example, how the race or ethnicity of the contact varies with the race or ethnicity of the worker, or – more importantly with respect to the productivity of networks – whether the effect of the contact on the wage varies with whether or not the contact is the same race or ethnicity as the worker.

Aside from developing a better understanding of underlying behavior, it would be useful to know the answer to this question because in many data sets we might have information on the share of an employer's workforce that is white, black, or Hispanic, but not have information on residence-based networks – the measurement of which imposes much more severe demands on the data. On the other hand, as discussed earlier, evidence that residence-based network connections are particularly important has implications for both policy and for our understanding of the spatial connections between labor and housing markets.

III. Data

This project uses employer-employee matched data to construct measures of network ties and estimates of the impact of employment networks. The core data is a set of infrastructure files produced by the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program (see Abowd et al., 2009). Previous research on residential employment networks in Hellerstein and Neumark (2003) used the 2000 Decennial Employer-Employee Database (DEED), based on matching 2000 Long-Form Census respondents from the "Sample Edited Detail File" (SEDF) to their establishment of employment. The DEED is complementary to the LEHD dataset. A strength of the DEED is that it matches workers at the establishment level, and that it includes all of the demographic information on the Census Long Form. But it has two key limitations. First, although the data set includes millions of workers matched to over one million establishments, it does not capture the universe of workers (or establishments), both because the matching process begins with the one-sixth of Census respondents who complete the Long Form, and because matching on transcribed business names and addresses entailed a process of "fuzzy matching" that leads to matching only around 30 percent of Long-Form respondents. And second, it does not provide longitudinal information on workers or employers.

The LEHD data cover essentially the universe of workers within firms, and does so longitudinally. Given the highly-detailed analyses we pursue in this paper, the coverage rates of the LEHD and the fact that it extends to recent years make it more useful than the DEED. In addition, some of our analyses – including all of our analyses of turnover – require longitudinal information. The LEHD data have two drawbacks. First, the LEHD program has not covered all workers in all states over time. Second, the LEHD program collects wage record data at the level of an employer, identified in each state by a State Employer Identification Number (SEIN).[10] For employers with more than one establishment within a state (multi-unit employers), LEHD uses an imputation model to match establishments to workers (approximately 40 percent of jobs are at multi-unit employers). And third, it does not include observed demographic information for some workers. We discuss key issues regarding our LEHD data in this section, and provide additional details in Appendix A.

As with research on employment networks using the DEED, we begin by defining a frame of jobs suitable for calculating the network measures. The analysis in this paper covers the four-year span from 2004 through 2007, a period of stable employment between the early 2000s recession and the Great Recession (which began in the December of 2007, according to NBER dates). We also use data extending back at least to 2001, with the earlier years used to construct time-of-hire network measures, discussed below. The analysis sample is for primary jobs held at the beginning of the second quarter of each year at a private-sector employer located in one of 39 states whose time series in the LEHD infrastructure files spans the entire year range. There are numerous restrictions we impose to arrive at our analysis sample. The restriction of largest consequence is to limit the set of jobs to single establishment employers, for which we know the place of work to which a worker reports. These restrictions, and their effects on the sample size for one illustrative year (2006) are listed in Table 1; additional details on the sample restrictions are described in Appendix A.

For the computation of network measures, we do not restrict the sample based on tenure. Our model estimates, however, are restricted to shorter-tenure workers, for two reasons. First, the effects of networks on the quality of job matches should be most apparent in the early years of a worker's tenure with an employer (consistent with the argument in Dustmann et al., 2011). Second, we utilize information from the characteristics of co-workers at the time a worker is hired, and the LEHD sample of establishments has much better coverage if we restrict the analysis to workers hired in or after 2000.

[10] An SEIN defines an employer within a state, and is not necessarily the highest level of control or ownership.

Specifically, we retain workers whose current job has tenure of less than four full years, which drops 37.6% of jobs. As an example, for 2006 alone, after retaining workers in the racial/ethnic groups we study – white (non-Hispanic), black (non-Hispanic), Hispanic, or Asian (non-Hispanic) – the estimation sample consists of about 23.6 million jobs.

As is explained in the methods section that follows, we estimate the effect of employment networks on two outcomes. One is an indicator for whether a worker retains a job one year after the beginning of the second-quarter reference date. Job retention is measured by continued earnings from the same employer in the year following the observation.[11,12] The second is the log of quarterly earnings, which is the highest frequency available in the LEHD. To avoid measurement error from incomplete quarters of work, we restrict the earnings analysis to quarters in which the worker is employed at the same establishment in the previous and the following quarter. Retaining only these "full quarter" jobs for the earnings models drops a further 14.6% of jobs, with a pooled sample of almost 78 million worker-years.

Beyond the network isolation measures, described below, we construct control variables from the LEHD data. We construct indicator variables for the year of employment, as well as the year in which the worker began that job. We are also able to construct indicator variables for industry sectors and employer size, as reported by each SEIN, indicator variables for race/ethnicity group and female, and age and age squared.

IV. Measuring the Importance of Labor Market Networks with the LEHD Data

Methods

To provide a point of reference between the LEHD data used in this paper, and the DEED data used in earlier research, we begin by re-computing the measures of the importance of labor market

[11] Because our job measure is a person-employer pair, one concern is that an employer identifier (SEIN) change could be perceived as a new job. We minimize this possibility by using the Successor Predecessor File, which tracks businesses that have changed administrative record number or restructured, but retain the same set of workers. By extending job histories to include employment at the successor, we increase the share of jobs lasting at least one year to 65% (from 63%). In cases where an employer declines significantly or dies from one year to the next and is not succeeded by another employer, workers whose job ends would still be considered to have turned over.

[12] Because retention is defined based on staying with an employer from a given year to the following year, retention for our 2007 observations is defined using data from 2008, when the Great Recession had begun in earnest. Nonetheless, the inclusion of year effects (and employer-year effects in some specifications) should mitigate concerns about the influence of this period on the results.

networks used in Hellerstein et al. (2011), but using the LEHD data instead.

For a population of workers, we first compute for each worker the share of co-workers (in the same establishment) who live in the same residential neighborhood as that worker.[13] This measure, known as "observed network isolation," requires a sample restriction to establishments with at least two workers observed, and is calculated as:

$$(1) \qquad NI_i^O = \frac{\sum_{j \neq i} I^R(i,j) I^E(i,j)}{\sum_{j \neq i} I^E(i,j)},$$

where there are N+1 workers in the population, indexed by i and j for all possible pairs of workers. $I^R(i, j)$ is an indicator for whether workers i and j live in the same residential neighborhood, and $I^E(i, j)$ is an indicator for whether i and j work in the same establishment. The sums in the numerator and denominator are taken over all N workers other than the worker i. We summarize the degree of network isolation across an entire population as the average of each individual's observed network isolation, defined as:

$$(2) \qquad NI^O = \frac{1}{N} \sum_i^N NI_i^O.$$

We multiply this population measure by 100 to define a network isolation index. The same measure may be calculated for a sub-population, such as those of the same race or ethnicity.

We define residential neighborhoods as Census tracts, most importantly because Census tracts are relatively small. (The Census Bureau defines tracts as contiguous geographic units with an optimal size of about 4,000 residents.) This is important for two reasons. First, on the residential side, Census tract residents have contact with each other, if not "over the back fence," then at parks, schools, churches, stores, business, and other institutions. Second, on the workplace side, because tracts are small, factors such as access to mass transportation may lead residential neighbors to work in the same tract, but should not influence at which establishment in the tract they work; within urban areas, especially, individuals can walk from any establishment to any other establishment in the tract.

We do not want to attribute all clustering of neighbors in the same establishment to networks (or

[13] We exclude the individual worker from this calculation, since it is meaningless to say that a person is his or her own neighbor.

other influences), because some co-residents would work together even if workers are assigned randomly to establishments, given that people tend to work relatively close to where they live. As a result, the observed network isolation (NI^O) can be positive simply by chance. We therefore compute the extent of network isolation that occurs due to randomness, denoted NI^R. To calculate NI^R, within a workplace Census tract we randomly assign workers to establishments, ensuring that we generate the same size distribution of establishments (in terms of matched workers) within a Census tract as we have in the sample. We generally do this using data only on individuals in the same racial, ethnic, or skill group for which we are trying to characterize the importance of networks. We do the simulation 10 times, and compute network isolation each time across all workplace Census tracts; NI^R is the mean over these simulations. The idea behind the random allocation is that workers, through their behavior, reveal the geographic areas in which they choose to work (aside from the clustering at specific establishments). These decisions may be based on proximity, public transportation, highway exits, etc. Having made these choices, some workers from the same neighborhood will end up at the same establishment even if there are no network connections between them. Our randomization is meant to capture this component of the clustering and to subtract it out from observed clustering, with the remainder arising from the systematic processes that determine the establishments at which people work.

We refer to $NI^O - NI^R$ as the "network isolation difference," which captures the excess presence of co-residents in a worker's own establishment relative to the presence of co-residents in the same work location. This computation requires a second sample restriction – that Census tracts of employment include at least two establishments with two workers. If there is only one establishment, we cannot distinguish the effect of residence-based labor market networks from random clustering.

While NI^R provides a reasonable lower bound of the extent to which workers work with neighbors, it is also important to know what the upper bound of network isolation could be in our data. Because establishments within a particular Census tract of employment often contain more workers than the number of workers from any particular Census tract of residence represented among those establishments, the upper bound of network isolation is unlikely to ever reach 100 percent; in this case workers would work *only* with their neighbors. There is no known general method for solving for the maximum index in all cases in our data. We instead approximate the maximum network isolation through a "greedy" algorithm, originally used in Hellerstein et al. (2011).

For a given Census tract of employment, we order the neighborhoods in which workers live by the number of workers from each neighborhood. Beginning with the neighborhood with the greatest number of workers, we assign as many workers as possible to one establishment, and any workers who are not assigned to that establishment are grouped together and treated as a "new" neighborhood. We then move to the second largest neighborhood from which workers originate (which could be the "new" neighborhood left from the previous pass), and assign workers from that neighborhood to the establishment that holds the maximum number of these, again keeping neighbors working together in establishments as much as possible.[14] We continue moving down the list of neighborhoods, assigning workers to establishments until all workers are assigned.[15] We do this for every Census tract in our sample where workers work, and then compute the weighted average of the maximum network isolation in each Census tract of employment, weighting by the number of workers in that Census tract, denoting this maximum network isolation index as NI^M.

The difference ($NI^M - NI^R$) then measures the maximum extent to which networks could lead to workplace sorting beyond sorting that would occur randomly, so we scale ($NI^O - NI^R$), the difference between our observed network isolation index and the random isolation index, by ($NI^M - NI^R$), yielding

(3) $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$,

which we call the "effective network isolation index." It measures the percent of the maximum possible network isolation that could occur in the data actually does occur in the data. As such, it provides a

[14] If we instead started with smaller neighborhoods, we would be more likely to end up having to distribute workers from a large neighborhood across many establishments, not achieving as high an isolation index.

[15] Computing the maximum network isolation that could occur for an arbitrary group of workers, residential Census tracts, and workplaces falls into a well-known class of problems in computer science called "n-p complete" problems. For more on n-p completeness and greedy algorithms, see, for example, Cormen et al. (2001). The greedy algorithm we use is bounded from above by the true maximum. However, there is an argument, in this context, for not being overly concerned about this problem, as the greedy algorithm we employ is one that would bound from above what might happen in the real world if employers each individually tried to hire to the maximum extent possible from only a small set of neighborhoods – and where large employers might have more resources and therefore more ability to do so. In that sense our algorithm may be a reasonable practical measure to use.

natural scaling for the importance of networks formed among co-residents in determining the establishments in which people are employed. Because the measure is normalized, it can be compared across different samples with different possible lower and upper bounds of network isolation.[16]

Finally, paralleling the analysis in Hellerstein et al. (2011), we do these calculations using all workers, but computing the network measures by race or ethnicity (and sex, for reasons described below). We also do them simply focusing on those within a race or ethnic group, to see whether the evidence points to residence-based labor market networks are racially or ethnically stratified.

Results

Table 2 reports the network measures using the LEHD samples. The estimates in Panel A of Table 2 are computed group by group. Pooling all the data, the observed network isolation index (NI^O) is on average 5.3, implying that workers have, on average, 5.3 percent of their co-workers living in their home Census tract.[17] When we randomly allocate workers to establishments within the same tract in which they work, this clustering – the random network isolation index (NI^R) – is only 2.8 percent. The difference between them, which we term the network isolation difference, is 2.5 percent, implying that the actual clustering of co-workers in the same residential tract is about twice as large as the clustering that would occur randomly based on who works somewhere in the Census tract.

We then use the greedy algorithm to compute the maximum possible network isolation index (NI^M) within a workplace Census tract, which is 32.2 percent on average. Last, we combine these measures into a composite effective network isolation index, which is 8.6 percent. To reiterate, this measures the share of the maximum possible network isolation that could occur in the data that actually does occur.

The remaining columns of the table show similar calculations for whites, blacks, Hispanics, and Asians. The effective network isolation index is a bit higher for whites, at 8.9 percent, but much higher

[16] Hellerstein et al. (2011) also use bootstrap methods to assess the statistical significance of differences in estimates of the effective network isolation index for different subsamples. The differences were always strongly significant, and the LEHD sample is even larger than the DEED sample they used. Moreover, the network isolation index is not our main focus in this paper. For all these reasons, we simply report the estimates below.

[17] In Table 2 (and Appendix Table B1 discussed below), we multiply the indexes by 100 so they can be described in percentage terms. In the other tables they are defined on a scale of zero to one.

for the other race or ethnic groups; the corresponding measures are 23.3 for blacks, 20.9 for Hispanics, and 44.4 for Asians, indicating that Asians appear to be the group most networked to residential neighbors.

Note that these differences are not always apparent in the first row of Table 2 – the values of the simple observed network isolation index. Because whites are such a large share of the population, the degree to which they live and work together is higher than for blacks and Hispanics (although not Asians). But whites' large share of the population also implies much larger values for the maximum network isolation, as reflected in the fourth row of the table, which explains why the effective network isolation index is lower for whites.

Appendix Table B1 reports similar calculations when we include workers in multi-establishment firms, using the LEHD program's imputation procedure to assign workers. In every case the observed network isolation index is smaller. Correspondingly, the effective network isolation index is also lower for each group, although the magnitudes are still sizable. These smaller estimates could reflect incorrect assignment of workers to establishments. Or they could reflect less reliance on networks in multi-unit firms or mobility of workers between establishments within firms. The results in Appendix Table B1 may be more comparable to the DEED findings in Hellerstein et al. (2011), because the latter did not exclude establishments from multi-establishment firms. The LEHD sample yields lower observed network isolation for the white, black, and Hispanic subsamples.[18] However, because the maximum isolation measure is also much smaller for each group, the effective network isolation index turns out to be similar for whites, a bit smaller for Hispanics, and larger for blacks. Overall, the magnitudes are fairly similar; in the DEED analysis, the effective network isolation index was 10.0 for whites, 9.6 for blacks, and 22.4 for Hispanics. Note, though, that these results are not completely comparable because the LEHD provides a much more representative sample of establishments.

One potential source of spurious evidence of the importance of labor market networks is that spouses may work together at a non-negligible rate (Hyatt, 2012); certainly if they met at work, we would

[18] The earlier study did not look at Asians.

not want to attribute their employment at the same workplace coupled with their co-residence to network effects. To examine whether spouses working together are responsible for much of the clustering of "neighbors" at the same workplace, the last two columns of Table 2 provide estimates that pool across race or ethnic groups, but use separate samples by sex. The effective network isolation index is actually higher for each sex considered separately – 10.9 and 9.4 for men and women, respectively, versus 8.6 in the first column. This implies that the measures are not driven by spouses working at the same location. Moreover, the fact that the index is larger when computed separately by sex suggests that residence-based labor market networks are stronger within than across sexes. This could be explained by more sharing of labor market information with neighbors of the same sex, either because of who is friends with whom, or because sex segregation across workplaces implies that information about jobs or job referrals from someone of the same sex are more valuable.

Paralleling this last point, Panel B of Table 2 reports estimates where we use the whole sample, but simply compute the averages of the indexes by group (race, ethnicity, or sex). If residence-based labor market networks are stratified by race, ethnicity, or sex, then the network isolation measures computed this way should be lower, because in these computations we capture the extent to which neighbors are clustered in the same workplace irrespective of whether those neighbors are the same race or ethnicity (or sex). For whites and more so blacks and Hispanics, the effective network isolation indexes are lower when computed across the combined sample. For Asians, and for men and women, however, the estimates are quite similar in Panels A and B.

V. The Effects of Networks on Turnover and Earnings

Methods

Our main analysis concerns the effects of residence-based networks on labor market outcomes – specifically turnover and earnings. In the regression analysis of the effects of networks, we use measures related to those that go into our "effective network isolation index" defined in equation (3), but there are some differences. For our analysis of the impact of networks on turnover and earnings we measure the extent of labor market networks for worker *i* at employer *e* in year *t* as NI^{O}_{iet} , with network isolation calculated at the time of hiring, rather than contemporaneously. The main hypothesis being explored is that stronger network connections imply better labor market matches via referrals from network connections. These matches should be reflected in who the worker was connected to when they were hired, rather than who the worker worked with at some later observation. In contrast, the contemporaneous measure of network strength could reflect different influences. For example, when we are studying retention, the effect of who one works with in the prior year could reflect connections that help one get a future job, and hence *reduce* retention – in contrast to the job-match story. Network connections among workers also could enhance productivity in ways other than through the original referral, and if they do, contemporaneous network strength could increase earnings (and perhaps reduce retention through this channel). Thus, for each worker-employer pair (job) in our estimation sample, we use the earliest available measure of that worker's network associations at that employer.

In our index, we adjusted NI^O by NI^R to account for differences in employment patterns across Census tracts that could generate variation in the extent to which neighbors work together, which might be unrelated to actual network connections. For example, transportation infrastructure in an area (like a highway or subway line) might lead to many people from one tract of residence working in a common tract. The random (NI^R) and maximum (NI^M) network isolation measures, which were used to measure network isolation in the aggregate, are defined based on shuffling workers among jobs within a Census tract, and hence are only meaningful for tabulations at the Census tract level or higher. As a result, they are not meaningful for our regression-based analysis of individual-level turnover or earnings, and the establishment-level determinants of these individual outcomes. Thus, we do not use NI^R and NI^M as explanatory variables in the measurement of the importance of residential-based labor market networks.

Although we do not use the random and maximum measures of network isolation in our main analysis, we nonetheless want to parse out the effects of potential interpersonal connections between workers and the apparent connections that can arise because of other factors that might lead employers to have concentrations of workers from specific residential areas – such as transportation infrastructure. To control for observed network isolation that is the result of commuting tendencies rather than interpersonal connections, most of the empirical specifications include an origin-destination network isolation measure. For each worker, TI^{O}_{iet} measures the share of total workers in an employment tract who reside in the same tract as that worker – i.e., having the same origin and destination tracts in their commute. TI^{O}_{iet} (so denoted because it captures isolation by tract) is constructed in an identical manner as NI^{O}_{iet} (and also is based on time of hire), except that we use the workplace Census tract rather than the establishment. Our hypothesis is that workers in jobs with greater isolation in their commute flow (that is, a greater concentration of commutes between a worker's home and work tracts) may also have higher job retention because the large commute flow may be indicative of a low commute cost, which is a non-wage amenity. By the same argument, this variable should be associated with lower wages.[19] Both of these predictions are generally borne out in the data.

In a sense, TI^{O}_{iet} substitutes for the role of NI^{R} in our network measure, by capturing the network isolation that occurs because of clustering of workers from particular residential tracts in particular work tracts. Note that the presence of other workers in a tract sharing a commute with worker *i* does not, in and of itself, negate the importance of having a residence-based co-worker network. Rather, the importance of NI^{O}_{iet} conditional on TI^{O}_{iet} indicates the presence of a network that links neighbors to specific establishments within Census tracts.[20]

We also create a control variable for the average level of network isolation at an employer. The average observed network isolation for all N_e workers at employer *e* is constructed as

(4) $$EI^{O}_{e} = \frac{1}{N_e}\sum_{j=1}^{N_e} NI^{O}_{j}.$$

Because average network isolation may differ at an employer across years, workers *i* and *j* at employer *e* may have different values of EI^{O} if they were hired in different years, we denote this variable

[19] We couch this discussion in terms of commuting. The same argument applies to other utility-enhancing factors that lead people who live in particular neighborhoods to work in specific areas.

[20] The adjustment by NI^{M} is less relevant to our analysis in this paper, which asks whether stronger network connections among an employer's workers reduce turnover. There is no reason to rescale by the maximum network isolation that could occur – which would act to reduce the measured clustering of neighbors at the same employer in cases where that clustering could be higher than at other employers. For example, if co-workers A and B each work with 20 neighbors, from the perspective of the network effects we study in this paper it should not matter that A could have worked with 50 neighbors, but B could only have worked with 30.

EI^O_{iet}. It is not evident a priori whether this measure will be positively or negatively related to job retention or earnings, once we also control for the individual worker's network connections. Higher overall networking within an employer may lead to other advantages at work that lead to lower turnover or higher earnings even of individuals not networked as strongly, or could have the opposite effect if a higher value EI^O_{iet}, conditional on the worker's own network connections, implies weaker connections to other workers. But we want to include the control to be sure we are detecting effects of the individual worker's strength of connections to co-residents.

Thus, for our baseline analysis, we focus only on NI^O_{iet}, and rather than adjusting for NI^R and NI^M, which are not relevant constructs here, we include EI^O_{iet} and TI^O_{iet} as control variables. In many of our analyses, we also construct these measures to only include workers of a given race or ethnicity.

We capture effects on turnover by estimating linear probability models for being retained by (not separating from) the employer after some interval (R). We estimate corresponding linear regressions where the dependent variable is the log of quarterly earnings. For purposes of exposition, we discuss our methods in the context of the retention regressions. The discussion carries over to the earnings analysis.

We begin by estimating linear probability models similar to Dustmann et al. (2011), with retention depending not on any network measure per se, but on the share of the employer's workforce that is of the same race or ethnic group, as in:

$$(5) \qquad R_{iet} = \alpha + \beta_S S^G_{iet} + \gamma_B B_i + \gamma_H H_i + \gamma_A A_i + X_{iet}\theta + \varepsilon_{iet}.$$

where in this equation, worker i at employer e has outcome R_{iet} in year t. S^G_{iet} denotes the share of the employer's workforce (also at the time of hire) that is in the individual's race or ethnic group – defined as (non-Hispanic) whites, (non-Hispanic) blacks, Hispanics, or (non-Hispanic) Asians. Thus, β_S measures the effect of a higher share of a worker's race or ethnicity in the employer's workforce on retention of workers. Because same group share is measured at the time a worker is hired, S^G_{iet} can vary across workers of the same group at the same employer. B_i, H_i, and A_i are dummy variables for three of the four race/ethnic groups (with whites excluded). We also estimate this model for each group separately, which allows us to see whether the effect of a larger share of the workforce in one's race or ethnic group varies across groups. Of course in these analyses the race/ethnicity dummy variables drop out. The vector X_{iet} contains other individual and employer-level control variables that may affect turnover, like age and the other controls discussed earlier.

The next step is to introduce our measure of residence-based networks. We start by estimating a similar model for turnover as equation (5), but using the residence-based network measure instead, as in:

(6) $$R_{iet} = \alpha + \beta_N NI^O{}_{iet} + \beta_E EI^O{}_{iet} + \beta_T TI^O{}_{iet} + \gamma_B B_i + \gamma_H H_i + \gamma_A A_i + X_{iet}\theta + \varepsilon_{iet}.$$

This specification asks whether worker turnover is lower when the worker has stronger residential links to co-workers, as measured by β_N. To ensure that we are identifying the effect of an individual worker's clustering with co-residents, this model also controls for the average observed network isolation of workers at the establishment, $EI^O{}_{iet}$. In addition, this specification accounts for variation driven by stronger commute flows between tracts of work and tracts of residence by including $TI^O{}_{iet}$ – which captures the clustering of workers in the same employment tract as the worker in the same tracts of residence.

This specification uses the observed network isolation measure over all workers in the establishment. To see whether networks are stronger within race or ethnic groups, consistent with racially- or ethnically-stratified networks, we next estimate this same specification substituting versions of $NI^O{}_{iet}$, $EI^O{}_{iet}$, and $TI^O{}_{iet}$ defined for the worker's own race or ethnic group – denoted $NI^{GO}{}_{iet}$, $EI^{GO}{}_{iet}$, and $TI^{GO}{}_{iet}$ where the G superscript, as before, indicates that this measure is specific to the group.[21] That is, we estimate:

(7) $$R_{iet} = \alpha + \beta_{GN} NI^{GO}{}_{iet} + \beta_{GE} EI^{GO}{}_{iet} + \beta_{GT} TI^{GO}{}_{iet} + \gamma_B B_i + \gamma_H H_i + \gamma_A A_i + X_{iet}\theta + \varepsilon_{iet}.$$

The specifications thus far use either the group share in the establishment's workforce, network isolation measures for all workers, or network isolation measures that are race- or ethnic-specific. We next estimate specifications where we include all of the covariates in equations (5), (6), and (7), so that the equation becomes:

[21] The group-specific measure capturing commuting flows may be more accurate if, for example, there are race or ethnic difference in the role of transportation infrastructure because minorities are more reliant on public transportation.

$$(8) \qquad R_{iet} = \alpha + \beta_S S^G{}_{iet} + \beta_N NI^O{}_{iet} + \beta_E EI^O{}_{iet} + \beta_T TI^O{}_{iet} + \beta_{GN} NI^{GO}{}_{iet} + \beta_{GE} EI^{GO}{}_{iet} + \beta_{GT} TI^{GO}{}_{iet}$$

$$+ \gamma_B B_i + \gamma_H H_i + \gamma_A A_i + X_{ift}\theta + \varepsilon_{iet}.$$

One can interpret this as a "horse race" between the importance of the overall network isolation measures, the group-specific network isolation measures, and the simple share of a worker's group in the employer's workforce. Including the network measures is important in understanding the effect of S^G because the latter can reflect the variation otherwise captured by the residence-based measures. For example, a black worker who works with more co-residents is also likely to have more black co-workers because of residential segregation. And comparing the relative importance of the network isolation measures that are not group-specific with those that are group-specific tests whether residence-based networks have a racial or ethnic component.

We explore the robustness of the results to the inclusion of fixed effects for employers or for employer-year interactions, as well as individual worker fixed effects. Fixed employer effects control for variation across employers in policies or other factors that can affect turnover, and the employer-year interactions allow for employer-specific shocks that affect all workers in a year. When the employer fixed effects are included, we identify the effect of network isolation on retention from variation within an employer both at a point in time and over time in the network isolation of its employees at their times of hire. And when the employer-year interactions are included we only identify differences across workers working for an employer in a given year in those workers' time-of-hire network isolation.[22] Worker fixed effects control for quite a different source of variation by accounting for individual characteristics that are fixed over time. Given that the network measures based on time of hire do not vary over observations on an individual at the same establishment, when worker fixed effects are included the effects of the networks measures are identified from observations on the same individual at multiple establishments.

We also estimate these models for separate race or ethnic groups, to see whether the conclusions vary across groups – corresponding, perhaps, to some of the apparent differences in the importance of labor market networks across groups based on the evidence in Hellerstein et al. (2011).

[22] This would control, for example, for the influence of large layoffs (or a closing) affecting many workers in an establishment in the same year.

Productive Networks or Amenities?

To this point, we have discussed the potential effects of working with one's neighbors (or simply one's co-ethnics) in terms of positive network effects via better job matches. An alternative perspective, however, is that working with neighbors is a job amenity. Moreover, given the racially- and ethnically-segregated nature of social relationships in the United States (see, e.g., Estlund, 2003), working with one's neighbors of the same race or ethnic group may be particularly likely to represent a job amenity. We might expect, then, that the estimated effects of our network measures could reflect these amenities. Whereas positive job amenities should reduce wages or earnings, they should presumably increase retention.

This has potentially important implications for the interpretation of our results. With respect to the models for retention, the predicted sign of our network-related measures is positive whether that measure reflects productivity or amenities. On the other hand, the two hypotheses have opposite implications for earnings – with network influences on job matches increasing earnings, while amenity effects lower them. Thus, the evidence for earnings may not be as clear cut, but the sign of the estimates may provide insight into which effect dominates. For example, with a positive effect of our network measures for both earnings and retention we can most confidently interpret the evidence as pointing to network effects.

In our empirical work the issue is more nuanced because we use two different residence-based network measures – one that measures the clustering of one's neighbors at one's workplace irrespective of race or ethnicity (NI^{O}_{iet}), and one that measures this clustering only for those of the same race or ethnicity (NI^{GO}_{iet}). What differences might we expect in the effects of these two network measures? We have two conjectures. First, when both measures are included in the models, we might expect the overall network measure to reflect the job-matching effect, whereas conditional on working with many of one's neighbors, working *also* with many neighbors of the same race or ethnicity may have little added effect via the job-matching channel, but have a stronger amenity effect. That is, the quality of match resulting from a referral from one's neighbor may be no different depending on the race or ethnicity of that neighbor, while working with one's neighbor of the same race or ethnicity may be what yields utility to the worker as an amenity. Second, building on Granovetter's (1974) distinction between strong and weak network ties, it may be the network connections to people of different races or ethnicities – who are more likely to be outside of an individual's circle of friends – which represent the weak ties that are more productive in the labor market. These productive weak ties may be better captured by our overall network measure, particularly when we condition on the same-group measure.

This line of reasoning suggests that we might find stronger evidence of a job-matching effect of the overall network measure, and of an amenity effect from the same-group measure. This has an observable implication for the earnings regressions, where the first effect should be positive, and the second effect negative. In that sense, then, the earnings regressions provide evidence on whether the estimated effect of the overall network measure captures a job-matching effect of residence-based networks, while the same-group network measure captures an amenity effect.

Descriptive Statistics

For the full estimation sample and each race or ethnic group, Table 3 presents the means of most of the variables used in the regression analysis. Looking first at the network measures, the large share of whites in the population results in that group having a high average share of co-workers in the same group (81 percent) and high observed network measures. The group specific measures of network isolation (NI^{GO}_{iet}) are about the same as the overall measures (NI^{O}_{iet}); note that this is not inconsistent with greater clustering of same-race or same-ethnicity neighbors in the same tract of employment based on the effective network isolation index (equation (3)), because the calculation of that index also takes account of how much clustering could occur randomly – which will tend to be lower for the group-specific measures. The measure of the share having the same origin and destination tracts in their commute – TI^{O}_{iet} – is typically about half the observed network isolation, roughly corresponding to the random network isolation measure in Table 2; the only exception is for Asians for whom it is much lower.

Turning to the dependent variables and the other controls, about 65 percent of all workers retain their primary job for one year, with the Asian subsample having the highest retention rate, at 69 percent, and the black subsample having the lowest, at 58 percent. The average number of jobs held in a year is about 1.9, although higher for blacks, consistent with their lower retention rate. As a result, primary job earnings should account for a large share of total earnings. The average censored tenure is 6.1 quarters, or just over 18 months. Compared with the DEED sample (for 2000), there are fewer workers in manufacturing employment and more in services, reflecting better representativeness of the LEHD data across industries, presumably because of better representativeness by establishment size. The industry difference may also reflect the LEHD sample we use excluding workers at multi-unit employers.

Finally, Table 4 provides some information on the distributions of the various network measures and related control variables that we use. Comparing Tables 3 and 4, the most notable feature is that the distributions of the network and related measures are quite highly skewed (for all except the simple same-group share, S^G_{iet}). For example, for the pooled sample the mean of NI^O_{iet} is 0.046, whereas the median is zero and the 75th percentile is 0.023, half of the mean. The implication is that much of the variation in these measures is in the upper tail because, perhaps not surprisingly, a large share of workers do not work with any of their neighbors. For example, the 90th percentile rises sharply to 0.121.

Regression Results

Linear probability model estimates for job retention are presented in Table 5 for the full sample, and in Table 6 for the race or ethnic group subsamples (for the main specifications). Each table of estimation results includes coefficients for the network variables along with standard errors. Because of the large sample sizes, many variables included are significant at less than the one percent level, and we therefore do not focus on the statistical significance of the coefficient estimates in the ensuing discussion, except to note where the main coefficients of interest are not significantly different from zero.

For the sample including all groups, presented in Table 5, column (1) indicates that the share of a worker's co-workers in the same group (S^G_{iet}) increases retention. The estimate implies that a 10 percentage-point increase in this share boosts the probability of retention by 0.0077, or three-quarters of a percentage point.

We next, instead of S^G_{iet}, introduce our residence-based network measure NI^O_{iet} (along with the employer average, EI^O_{iet}, and same commute average, TI^O_{iet}, as controls). The estimated effect of the observed network isolation index, as reported in column (2), is also positive and statistically significant, indicating that a 10 percentage point increase in this measure – which, referring to Table 4, is about the average across race/ethnic groups of the difference between the 50th and 90th percentiles – boosts the probability of retention by 0.0194, or just below two percentage points. Interestingly, the coefficient on the employer average of the network isolation index is negative (and much smaller), so that the mechanism by which the employer average affects retention is not one that ties workers to employers. The estimated coefficient of the control for the fraction of workers in the Census tract who come from the same neighborhood (TI^O_{iet}) is positive for retention, consistent with it capturing a non-wage commute-related amenity. In column (3) we instead use the observed network isolation measure for the workers' same race or ethnic group. The estimated coefficients are similar to those for overall observed network isolation.

Each of these three "baseline" estimates gives an indication that labor market networks may be important. However, a central goal of this paper is to assess the importance of residence-based labor market networks – a particular form of spatial labor market networks – and to understand their nature. The positive effect of the share of workers in the same group (S^G_{iet}), as in the specification in column (1) of Table 5, may reflect these residence-based networks. Alternatively, it may capture networks along other lines that connect workers in the same race or ethnic group, or something different from networks altogether. The impact of the observed network isolation index (NI^O_{iet}) for all co-workers or for one's own group is more likely to reflect network connections between neighbors. By seeing what happens to the estimated coefficients of S^G_{iet} when we add the residence-based network measures, we can better determine what kinds of connections between workers affect turnover/retention.

In addition, other evidence in Hellerstein et al. (2011) suggests that residence-based networks are racially stratified (and therefore likely also ethnically stratified as well). We can gauge this in the present context by seeing whether the observed network isolation index for the same group (NI^{GO}_{iet}) has an effect above and beyond the measure without regard to group (NI^O_{iet}). Moreover, as we noted earlier, the results for these two network measures can also differ because they have different influences, with NI^{O}_{iet} capturing productive network effects and NI^{GO}_{iet} reflecting amenity effects.

As shown in column (4), when the same-group share, overall observed network isolation measure, and the same-group observed network isolation index are all included in the retention regression simultaneously (along with the other controls), the coefficient on the same-group share is virtually unchanged from its estimate in column (1) when the network isolation controls are excluded. This does not rule out the effect of the same-group share also reflecting labor market networks, but it does suggest that the mechanism by which the same-group share affects turnover is largely orthogonal to the mechanism by which observed network isolation measures (conditional on the other controls) affect turnover.

The coefficients on the overall and group-specific observed network isolation indexes in column (4) are both smaller than their counterparts in columns (2) and (3), but both are still positive, with the estimated coefficient on the same-group measure larger than that of the overall measure (0.116 vs. 0.073). One interpretation consistent with this evidence is that network connections among all neighbors improve labor market matches, but that network connections among same-group neighbors are even more important. However, the latter network connections (in particular) could instead reflect amenity effects, because as noted above, both productivity-enhancing network effects and amenity effects have positive implications for retention. Which of these two is at work can be sorted out in the earnings estimates we present below.

Columns (5) through (7) of Table 5 add employer (establishment) fixed effects, then employer/year fixed effects, and finally worker fixed effects, respectively. As shown in columns (5) and (6), including the employer fixed effects or employer-year fixed effects reduces the estimated effect of the same-group share measure, but each has little impact on the coefficients on the network isolation measures. Including worker fixed effects in the regression (rather than establishment fixed effects) substantially reduces the impact of the overall and same-group share measures, but does not come close to eliminating the positive impact of each of the network isolation measures. One must be especially cautious in comparing columns (4) and (7), however, because sweeping out worker fixed effects removes a lot of potentially useful variation.[23]

Table 6 reports results for the minority race/ethnicity subsamples of blacks, Hispanics, and Asians.[24] The structure of the tables and the specifications shown are the same as for Table 5, so the results can be described more succinctly. Qualitatively, most of the results for the network isolation measures are similar for each group. Most importantly, perhaps, the overall and same-group network isolation indexes (NI^{O}_{iet} and NI^{GO}_{iet}) always enter into the equations with positive coefficients, although the estimated coefficient of the overall measure for blacks is not statistically significant when worker fixed effects are included. The fact that both are positive suggests that for non-whites, being networked to neighbors from one's own race or ethnic group conveys a larger reduction in turnover. However, the supplemental effect could also reflect a non-wage amenity.

The effect of the same-group share, S^{G}_{iet}, differs quite a bit in Table 6 from its estimate in Table 5. First, for blacks, Hispanics, and Asians this estimate is consistently negative. Its size also varies considerably across groups and across specifications within groups. The estimated coefficients of S^{G}_{iet} contrast with the results in Dustmann et al. (2011), who find that the effect of a higher "own share" of minority groups on turnover (the opposite of retention, our dependent variable) is positive without firm fixed effects, but negative with them. Of course they study data on different and more narrowly defined groups, from a different country, so we should not necessarily expect the same results.

The clear message from these results is that that the empirical relationship between simply working with a large share of workers from one's minority race or ethnic group and retention is not positive and robust, and is in fact negative for minority groups. In contrast, the empirical relationships with our residence-based network measures are robust across groups and across specifications.

We next turn to estimates for (the log of) earnings. We again provide two tables of estimates, one

[23] Note that even though there are four years of data, there are almost half as many worker fixed effects as worker-year observations. There is a shortfall in observations because workers progressing beyond the four-year tenure threshold depart from the estimation sample.

[24] The results for whites only are very similar to the full sample results, and hence are not reported or discussed separately.

for all workers and one reporting the main specifications for each of the separate minority subgroups. The specifications mimic those of the retention equations, but also include as additional controls dummy variables for the year-of-hire. Because there are also dummy variables for the current year in the regressions, these additional controls are essentially controls for year of tenure, as is appropriate in an earnings regression.[25]

Across Tables 7 and 8, the estimated effect of the share of a worker's co-workers in the same group always *reduces* wages. These results, again, contrast with those of Dustmann et al. (2011). More substantively, in the context of labor market models of networks, this evidence suggests that a greater representation of workers from one's own group does not reflect productive labor market network effects, such as better matches or other productivity-enhancing influences.

When we turn to the evidence on the residence-based network measures, however, there is more evidence of productive positive network effects as reflected in earnings, and in particular for all three minority groups. In Table 7, for all groups (including whites) combined, the estimated effects of the overall ($NI^{O}{}_{iet}$) and same-group ($NI^{GO}{}_{iet}$) network measures are positive in columns (2) and (3), consistent with positive job-match effects of networks. However, when we add both of these (and their associated controls) simultaneously, the estimated effect of the overall network measure becomes more positive, while the estimate effect of the same-group measure becomes negative. Note that this evidence is exactly what is predicted by the hypothesis that the overall measure captures positive network effects, and the same-group measure instead captures a non-wage amenity. Moreover, this same pattern of results occurs in Table 8, for the earnings of blacks and Hispanics, although for Asians we do not find the negative effect of the same-group measure as consistently. With all groups combined (in Table 7), the evidence of a positive impact of the overall network measure reverses (becoming weakly negative) when the employer fixed effects are included, but not when the worker fixed effects are included.

However, in Table 8, where we look at the minority groups separately, we always find a positive effect of the overall network measure ($NI^{O}{}_{iet}$), across all specifications. This evidence is consistent with

[25] Nonetheless, we verified that the estimates were insensitive to excluding the year of hire (tenure) controls.

the overall residence-based network measure reflecting positive job-match effects. In addition, we always find a negative effect of the same-group network measure in the earnings regressions for the minority groups, once we include both residence-based network measures, although these estimates are not significant in the specifications with worker fixed effects. This general pattern is consistent with the same-group measure reflecting a non-wage amenity but not a positive productivity effect.[26,27]

Note, by the way, that this non-wage amenity effect can explain greater clustering of same-race or same-ethnicity neighbors in the same workplace, which was reported in Table 2 (and in Hellerstein et al., 2011). The implication is that the refined tests of networks we implement in this paper – which look not just at who works with whom but at the effects of this clustering – potentially provide a better understanding of what it means for workers to have connections to each other that might be viewed on a priori grounds as capturing labor market networks.

One could potentially argue that the positive effect on earnings of the overall network measure conditional on the same-group measure is consistent with a *negative* non-wage amenity. That is, for example, blacks have to be paid more to work with whites. This interpretation, however, is not sufficient to explain the positive retention effects of the overall network measure for blacks or other minorities considered separately, suggesting instead that the evidence reflects a positive productivity effect. Indeed, one interpretation of the positive wage and retention effects of the overall network measure conditional on the same-group measure is that referrals of those in *other* race and ethnic groups are particularly informative to employers.

Finally, we carried out some additional sensitivity analyses beyond the many specifications we

[26] A negative relationship between a high minority share at an establishment and low wages of minorities has sometimes been interpreted as reflecting the effects of a labor market with frictions that differ between minority and majority groups. For example, models such as Lang (1986) and Black (1995) predict that there will be segregation of minorities into specific employers, and both frameworks can lead to lower turnover and lower wages for minority groups working for more segregated employers. This kind of crowding therefore could explain the negative effect of the same-group share for minorities evident in Table 8. However, it is not clear that there should be an additional negative crowding effect attributable to being employed with a large share of minorities from one's neighborhood, conditional on the overall group-share measure. As a result, we view it as unlikely that crowding explains the effect of the same-group network measure on wages.

[27] We also verified that the results were similar if we estimated the models and network measures for men only, to avoid any confounding influences of spouses working together.

have already reported in the tables. First, we re-estimated all of the models excluding the final year of data, which could have been affected by the Great Recession. The results did not appreciably change. Second, we estimated the models using the network measures computed contemporaneously, rather than as of the time of hire. The results for retention were qualitatively similar, with the overall residence-based network measure (NI^{O}_{iet}) and the group-specific measure (NI^{GO}_{iet}) having a positive effect on retention. For earnings, the results for NI^{O}_{iet} were qualitatively the same as using the time-of-hire network measure. For NI^{GO}_{iet}, however, the results were a little less clear cut. For Hispanics and Asians the same negative effect was found – which we have interpreted as reflecting amenities. But for blacks the earnings effect was near zero and statistically insignificant, perhaps reflecting the ambiguous predictions for earnings because network and amenity effects are in opposite directions.[28]

Related to the issue of using contemporaneous versus time-of-hire network measures is the question of whether the predictions for the magnitude of the productivity or amenity effects are different based on the time-of-hire and contemporaneous network measures. In terms of a pure referral effect, we might expect the time-of-hire measures to be most important, whereas the amenity effect seems most likely to be reflected in the contemporaneous measures. However, productivity effects of networks need not arise only from referrals, and hence contemporaneous network measures could have positive productivity effects. Similarly, the persistence of wages from the time of hire could still allow amenity effects to be reflected in the time-of-hire measures. Our view, though, is that the results using the time-of-hire network measures, which we report in the tables in this paper, provide the cleanest evidence on network effects.

VI. Conclusions and Discussion

If labor market networks lead to better matches in the job market, they should reduce turnover and increase wages. We use matched employer-employee data with detailed information on where people

[28] We also found that the retention effects were robust to other variants of the specification, like dropping the employer average controls (EI^{O}_{iet} and EI^{GO}_{iet}), and defining NI^{O}_{iet} and the associated controls as of the time of hire but NI^{GO}_{iet} and the associated controls contemporaneously. Again, the earnings results were more sensitive to these alternative specifications. Between the ambiguous predictions for earnings, and the fact that we are varying which of a large number of highly-correlated variables to include, this sensitivity is not surprising.

live and where they work to test this hypothesis in the context of residence-based labor market networks. Recent research has suggested that these types of labor market networks may be present, but it has not assessed any evidence asking whether these networks are productive in terms of improving labor market matches, which might well be regarded as a necessary condition to interpret an empirical connection between workers as reflecting labor market networks. Thus, more broadly, this paper emphasizes and tests for a spatial dimension of labor market networks – a dimension likely to be particularly important in urban labor markets that are dense with employers and jobs, and in which workers often live in residentially-segregated neighborhoods.

The empirical analysis provides robust evidence that workers who are more connected to their neighbors have lower turnover. Moreover, we find this effect for measures of network connectedness to one's neighbors generally, and to neighbors of the same race or ethnic group. We argue that the first effect likely reflects productive network effects, whereas the latter effect could also reflect productive network effects of networks that are racially-or ethnically-segregated. Alternatively, the effect of working with one's neighbors of the same race or ethnicity could reflect a non-wage amenity owing to preferences of workers to work with neighbors of the same race or ethnicity, with whom they are particularly likely to have friendships and other social relationships.

When we turn to earnings, we find a similar positive effect of the overall residence-based network measure, but a negative effect of the same-group measure, which is generally but not always statistically significant. Negative effects of the same-group network measure are consistent with this measure capturing a non-wage amenity, while the positive effects of the overall network measure are consistent with residence-based networks leading to better, more-productive matches. However, the evidence does not suggest that the residence-based networks that produce these better matches are racially or ethnically stratified.

Finally, the evidence we find of spatial labor market networks contrasts with past work that has used the simple share of one's co-workers of the same race or ethnic group as a measure of labor market networks, as we fail to find a robust positive relationship between either retention or earnings and simply working with a large share of workers from one's race or ethnic group. Indeed for blacks, Hispanics, and Asians the estimated relationships are consistently negative. The negative earnings effects are consistent with working with one's own race or ethnic group being a non-wage amenity; but the retention results are not.

The evidence that the residence-based network measures are important and affect labor market outcomes identifies an important spatial dimension of labor market networks, raising potentially important questions at the intersection of labor economics and urban economics. Perhaps most important, are there policies, institutions, or other interventions that can increase network connections between neighbors (or others), increasing the flow of labor market information and good matches among its residents? Moreover, residence-based labor networks can help explain how ethnic and racial residential segregation reinforces poorer labor market outcomes for minorities, to the extent that minorities have weaker network connections to jobs held by whites, although in some cases residence-based networks might help minorities to strengthen their labor market connections, such as in the example of ethnic enclaves (Edin et al., 2003).

Our results also to some extent call into question the conclusion that productive residence-based networks are racially or ethnically stratified. Whereas in this paper and Hellerstein et al. (2011) there *is* evidence of greater clustering of same-race or same-ethnicity co-residents in the workplace, the evidence on outcomes raises questions as to whether this clustering is productive, or instead reflects other influences such as preferences of people with these connections to work together. Given that there is other evidence consistent with positive effects of same-ethnicity residence-based networks (Damm, 2012), this clearly remains an open – and in our view quite interesting – question.

Appendix A: Additional Details on Data Construction

The LEHD program has assembled a national database with over 130 million jobs and more than 7 million employers in a typical quarter. Data series for some states begin as early as 1985, but the national partnership was only recently completed. States provide two files on a quarterly basis: (1) wage records for all Unemployment Insurance covered jobs including worker and employer identifiers; and (2) establishment and employer account records, known as the Quarterly Census of Employment and Wages (QCEW), and also referred to as the 'ES-202' program. The wage records cover roughly 96 percent of private non-farm wage and salary employment.

For each year, our sample only includes those jobs from the Person History File (which is derived from the Employment History File) held at the beginning of the second quarter. This sample definition is similar to that of the DEED, which includes jobs reported in the 2000 Census that were held as of April 1st of that year (the reference date for the Long Form). In contrast, because LEHD wage records only provide quarterly earnings, with no job start or end dates, it would not be possible to identify a set of jobs held on a specified date within a quarter. Rather, the LEHD program reports beginning-of-quarter job counts in the Quarterly Workforce Indicators. Those workers with earnings in both the first and second quarters of a year are assumed to be employed on April 1st, the first day of the second quarter. For workers with multiple jobs satisfying that definition, the analysis sample only includes the highest earning job in the second quarter, or the primary job.[29] We make use of wage records in all available states – even those outside of the analysis sample of 39 states – to determine whether jobs in the sample are a primary job.[30]

We augment the wage records with a set of worker and employer characteristics derived from other administrative and survey data files. On the employer side, we link jobs by SEIN to employer and

---

[29] Approximately 85 percent of beginning-of-quarter workers in the LEHD data only have one job at the beginning of a quarter.
[30] The 39 states in the analysis are (by postal code): AL, CA, CO, FL, GA, IA, ID, IL, IN, KS, KY, LA, MD, ME, MI, MN, MT, NC, ND, NE, NJ, NV, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, WV, WY.

establishment information in the Employer Characteristics File (ECF) for each state.[31] These files provide establishment ownership type, industry, reported employment size, and geocoded location. We limit the analysis to private-sector employers because hiring practices may be highly regulated in the public sector and because of some data quality issues in public-sector data.[32] We only retain single-unit employers – over 60% of the sample – in order to focus on workers whose establishment is not imputed in the LEHD. Using these imputations would introduce noise for the analysis of workplace connections. Note that an SEIN is not necessarily the highest level of control or ownership, so there may still be multiple SEIN "employers" within a larger state-wide or national firm.

On the worker side, we link jobs to the Individual Characteristics File (ICF), which provides age, sex, race, ethnicity, and other demographic information. The LEHD program constructs an ICF record for each worker with 2000 Decennial Census responses and federal administrative data, linked to the job records with personal identifying information. For the set of workers used in this analysis, over 80 percent had observed values for each of these characteristics. The LEHD program has developed an imputation model that uses available employment and demographic information to impute ICF characteristics for those workers with missing values. We use the first draw from this imputation model for any missing characteristic, so that all workers have complete demographic information.[33] We restrict the frame to workers aged 18 to 64 on April 1st of each year.

Also on the worker side, we link jobs to federal administrative records on residential location. The Composite Person Record, produced at the Census Bureau, provides a single housing unit associated with a person in a year. We link the housing unit, by way of geocoded address data files, to Census tracts, defined for tabulations of the 2000 Decennial Census. We restrict the set of jobs to those with place of work and place of residence information down to the Census-tract level, which is necessary to construct the network measures; this drops 3% of jobs due to employer geography and 12% due to residential

[31] This employer match fails for a small fraction of jobs when an establishment identifier changes, often when an employer transitions from single to multi-unit or the reverse.
[32] Public-sector employers, such as school districts, often report as single units rather than breaking out employment by establishment.
[33] See (Abowd et al., 2011) for a discussion of the imputation model.

geography. As a precaution, we also drop the small number of workers who have held more than 25 jobs in the previous year, which may be indicative of problematic administrative data records.

The network isolation measures we use are only defined for a job if the employer has at least two workers in the sample (or in the sub-sample if a restriction on race or ethnicity is imposed). Because the network measures cannot be calculated for those jobs that do not meet this criterion, they are excluded (about 3.1% of remaining jobs). In order to compute the network measures, which shuffle workers between nearby establishments, we also require that there are at least two employers in the same Census tract (dropping fewer than 0.1% of jobs). Note that because the LEHD frame is much larger than the DEED frame, the two-worker requirement and two-employer requirements are not as restrictive and allows for more small establishments, whereas in the DEED it has more serious effects on the representativeness of the sample (Hellerstein et al., 2011). In the full DEED sample, 65% of establishments have fewer than 25 employees, compared to the LEHD sample, where almost 87% are in that size class. On average, employers in the LEHD sample include over 78% of the jobs that they report having in the QCEW (as listed in the ECF). In contrast, the DEED sample contains, on average, only 16% of the jobs reported in matched units of the Business Register.

As noted in the main text, Table 1 presents the restrictions used to arrive at the analysis samples and, for a single year (2006) provides the count of workers remaining after each restriction. The initial sample consists of about 100 million workers in each year. After our sample restrictions are imposed, the sample for computing our overall network measures consists of about 40 million jobs in each year, which is 39% of the initial set of primary jobs. We define race/ethnicity sub-samples based on the ICF demographics as white not-Hispanic, black non-Hispanic, Hispanic, and Asian non-Hispanic. For these sub-samples, we again impose the requirements of having at least two workers at an employer and two such employers in a workplace Census tract, which results in a pooled sample that is 4.1% smaller. The sample is reduced further by the tenure restriction we impose. Finally, recall that the figures given in Table 1 are for a single year. The pooled sample for our regression analysis for all four years is over 90 million worker-years, with almost 50 million unique workers at about 3.7 million unique employers.

Our analysis of the effects of networks uses network measures defined at the time a worker is hired, for reasons explained in Section V. In some cases, we are not able to match a job to network associations in exactly the first year of employment. For example, if the worker's place of residence is not known in the first year of his or her job or if there were not at least two workers in the sample at the employer in that year, we cannot compute the network isolation measure for that year. Appendix Table A1 presents the availability of network association information for the year in which a worker was hired. Overall, for almost 90% of observations we are able to use time-of-hire network association measures for the year in which the worker was hired. In the worst case, for workers with three years of tenure, we are still able to match 77% to network association measures in the year they were hired, with the remainder matched from later years.[34]

Appendix Table A1: Availability of network association information in year of hire, by job tenure

		Years after time-of-hire that network information is available, by tenure				
Full years of tenure	Sample share	0	1	2	3	Row total
0	0.471	1	0	0	0	1
1	0.237	0.841	0.159	0	0	1
2	0.159	0.794	0.110	0.095	0	1
3	0.133	0.773	0.105	0.057	0.065	1
		Years after time-of-hire that network information is available, overall				
		0	1	2	3	Row total
		0.899	0.069	0.023	0.009	1

Notes: For the analysis of effects of networks, we use network association measures calculated for the year in which a worker began his or her job, or for as close to that year as we could obtain a measure. For workers with less than one year of tenure, time-of-hire network associations are the same as current network associations. For workers with one or more years of tenure, we match the worker/employer, or job, to a network association measure for the same worker/employer in the first year it is available. For about ten percent of worker/employers (see the lower panel), we could not match them to a network association in the first year of that job, and had to use a later year. The match could fail if the job did not satisfy the sample requirements for the network analysis sample in Table 1. For example, if the worker or employer did not have a known location in the first year, we could not calculate some of the network measures for that year.

[34] For a small share of cases, we are able to match a worker to a network isolation measure for his or her time of hire, but there is no group-specific information available. This occurs if, when hired, the worker was the only member of a race/ethnicity group at an employer. In our sample this occurred in less than 1% of cases overall and in less than 2.2% of cases for Asians, where it is most likely to occur. For these cases, we substitute the value zero for same group observed network isolation, even though such a case is not formally defined in equation (1). This applies to the time-of-hire network measures used in our regression analyses, but not to our measurement of the importance of networks (equation (2)), based on contemporary co-workers.

References

Abowd, John, Henry Hyatt, Mark Kutzbach, Erika McEntarfer, Kevin McKinney, Michael Strain, Lars Vilhuber, and Chen Zhao. 2011. "The New National Individual Characteristics File (ICFv4) and Quarterly Workforce Indicator (QWI) Tabulations by Worker Race, Ethnicity, and Education." Technical Memo, Center for Economic Studies, U.S. Census Bureau.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon D. Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In Dunne, T., Jensen, J. Bradford, and Roberts, Mark R., Producer Dynamics: New Evidence from Micro Data, Vol. 68, Studies in Income and Wealth. Chicago: University of Chicago Press, pp. 149-230.

Andersson, Fredrik, Monica Garcia-Perez, John Haltiwanger, Kristin McCue, Seth Sanders. 2010. "Workplace Concentration of Immigrants." Working Paper 10-39r, Center for Economic Studies, U.S. Census Bureau, revised Nov 2011.

Åslund, Olof, Lena Hensvik, and Oskar Nordström Skans. 2009. "Seeking Similarity: How Immigrants and Natives Manage at the Labor Market." IFAU Working Paper No. 24.

Bayer, Patrick, Stephen Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy*, Vol. 116, No. 6, December, pp. 1150-96.

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model", *Journal of Labor Economics*, Vol. 13, No. 2, April, pp. 309-334.

Brown, Meta, Elizabeth Setren, and Giorgio Topa. 2012. "Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System." Unpublished paper, Federal Reserve Bank of New York.

Cormen, Thomas H., Charles E. Leiserson, Robert L. Rivest, and Clifford Stein. 2001. Introduction to algorithms. 2nd ed. Cambridge, MA: MIT Press and McGraw-Hill.

Damm, Anna Piil. 2012. "Neighborhood Quality and Labor Market Outcomes: Evidence from Quasi-Random Neighborhood Assignment of Immigrants." Aarhus University Economics Working Paper 2012-18.

Datcher, Linda. 1983. "The Impact of Informal Networks on Quit Behavior." *Review of Economics and Statistics*, Vol. 65, No. 3, August, pp. 491-5.

Dustmann, Christian, Albrecht Glitz, and Uta Schönberg. 2011. "Referral-Based Job Search Networks." IZA Discussion Paper No. 5777.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic Enclaves and the Economic Success of Immigrants – Evidence from a Natural Experiment." *Quarterly Journal of Economics*, Vol. 118, No. 1, February, pp. 329-57.

Estlund, Cynthia. 2003. Working Together: How Workplace Bonds Strengthen a Diverse Democracy. New York, NY: Oxford University Press.

Giuliano, Laura, David I. Levine, and Jonathan Leonard. 2009. "Manager Race and the Race of New Hires." *Journal of Labor Economics*, Vol. 27, No. 4, October, pp. 589-632.

Granovetter, Mark S. 1974. Getting a Job: A Study of Contacts and Careers. Cambridge, MA: Harvard University Press.

Hellerstein, Judith, Melissa McInerney, and David Neumark. 2011. "Neighbors and Co-Workers: The Importance of Residential Labor Market Networks." *Journal of Labor Economics*, Vol. 29, No. 4, October, pp. 659-95.

Hellerstein, Judith K., and David Neumark. 2013. "Employment in Black Urban Labor Markets: Problems and Solutions." In Jefferson, Philip N., Oxford Handbook of the Economics of Poverty. Oxford, UK: Oxford University Press, pp. 164-202.

Hellerstein, Judith K., and David Neumark. 2008. "Workplace Segregation in the United States: Race, Ethnicity, and Skill." *Review of Economics and Statistics*, Vol. 90, No. 3, August, pp. 459-77.

Hellerstein, Judith K., and David Neumark. 2003. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set." *Annales d'Economie et de Statistique*, Vol. 71-72, July-December, pp. 19-78.

Hyatt, Henry. 2012. "A Study of the Frequency with which Couples Maintain Similar Employment." Unpublished manuscript.

Ioannides, Yannis M., and Linda Datcher Loury. 2004. "Job Information, Networks, Neighborhood Effects, and Inequality." *Journal of Economic Literature*, Vol. 42, No. 4, December, pp. 1056-93.

Kasinitz, Philip, and Jan Rosenberg. 1996. "Missing the Connection: Social Isolation and Employment on the Brooklyn Waterfront." *Social Forces*, Vol. 43, No. 2, May, pp. 180-96.

Kmec, Julie A. 2007. "Ties that Bind? Race and Networks in Job Turnover." *Social Problems*, Vol. 54, No. 4, November, pp. 483-503.

Kmec, Julie A., and Lindsey B. Trimble. 2009. "Does it Pay to Have a Network Contact? Social Network Ties, Workplace Racial Context, and Pay Outcomes." *Social Science Research*, Vol. 38, No. 2, June, pp. 266-78.

Kramarz, Francis, and Oskar Nordström Skans. 2007. "With a Little Help from My … Parents? Family Networks and Youth Labor Market Entry." Unpublished paper, CREST.

Lang, Kevin. 1986. "A Language Theory of Discrimination". *The Quarterly Journal of Economics* Vol. 101, No. 2, May, pp. 363-382

Laschever, Ron. 2009. "The Doughboys Network: Social Interactions and the Employment of World War I Veterans." Unpublished manuscript, University of Illinois at Urbana-Champaign.

Loury, Linda Datcher. 2006. "Some Contacts Are More Equal than Others: Informal Networks, Job Tenure, and Wages." *Journal of Labor Economics*, Vol. 24, No. 2, April, pp. 288-318.

Montgomery, James D. 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis." *American Economic Review*, Vol. 81, No. 5, December, pp. 1408-18.

Oyer, Paul, and Scott Schaefer. 2009. "The Personnel-Economic Geography of U.S. Law Firms and Law Schools." Unpublished manuscript, Stanford University.

Pellizzari, Michele. 2010. "Do Friends and Relatives Really Help in Getting a Good Job?" *Industrial and Labor Relations Review*, Vol. 63, No. 3, April, pp. 494-510.

Schmutte, Ian M. 2010. "Job Referral Networks and the Determination of Earnings in Local Labor Markets." Unpublished manuscript, University of Georgia.

Semyonov, Moshe, and Anya Glikman. 2009. "Ethnic Residential Segregation, Social Contacts, and Anti-Minority Attitudes in European Societies." *European Sociological Review*, Vol. 25, No. 6, December, pp. 693-708.

Simon, Curtis J., and John T. Warner. 1992. "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure." *Journal of Labor Economics*, Vol. 10, No. 3, July, pp. 306-30.

Table 1: Sample construction for one year (2006)

Condition to retain	Initial size	Percent reduction
Workers with a primary job in one of 38 states, held in 2006 at the beginning of the second quarter	102,830,093	NA
Job matches to establishment in Employer Characteristics File	100,469,098	2.3%
Employer is in private sector	85,620,476	14.8%
Employer is a single unit within state	52,920,722	38.2%
Employer location is precise to Census tract	51,394,045	2.9%
Worker linked to Individual Characteristics File	51,394,045	0.0%
Worker age is 18 to 64	47,666,356	7.3%
Worker linked to residence file	46,725,669	2.0%
Worker residence is precise to Census tract	40,999,240	12.3%
Worker has ≤ 25 jobs in last year	40,996,558	0.0%
At least two workers at employer	39,731,456	3.1%
Network analysis sample: At least two employers in workplace Census tract	39,721,695	0.0%
Network analysis subsamples: At least two workers in race/ethnicity group at employer and at least two such employers in workplace Census tract	38,109,311	4.1%
White not-Hispanic	27,744,956	NA
Black not-Hispanic	3,638,842	NA
Hispanic	4,684,102	NA
Asian not-Hispanic	1,799,887	NA
Not tabulated (not in subsamples or estimation)	241,524	NA
Tenure at job of three or fewer full years (< 4)	23,763,922	37.6%
Estimation sample for job retention model, without the “not tabulated” group	23,594,578	0.7%
Estimation sample for earnings model, with jobs lasting a full quarter	20,149,025	14.6%

Notes: The initial sample is the set of all jobs in the LEHD Person History File held at the beginning of the second quarter of each year, restricted to the highest earning job for each worker. Jobs are located in 39 states in the years 2004 to 2007. Only the jobs held at the beginning of the second quarter in 2006 are presented here. Job counts are rounded to the nearest 1,000. Jobs are linked to employer information in the Employer Characteristics File using a State Employer Identification Number. Workers are linked to demographic information in the Individual Characteristics File, which is constructed from federal administrative data and the year 2000 Decennial Census. The network analysis sample restrictions are necessary for producing the network isolation measures in Table 2. The estimation sample used in Table 3 and beyond pools observations across all four years. In subsequent tables, the race/ethnicity groups are listed as White, Black, Hispanic, and Asian.

Table 2: Network measures for one year (2006)

Variable	All	White	Black	Hispanic	Asian	Male	Female
A. Computed by sample							
Observed network isolation (NI^O) $NI^O \times 100$	5.3	6.1	3.6	4.7	8.0	4.4	4.6
Simulated random network isolation (NI^R) $NI^R \times 100$	2.8	3.0	0.9	1.0	1.0	1.8	2.0
Network isolation difference $[NI^O - NI^R] \times 100$	2.5	3.1	2.7	3.7	7.1	2.6	2.6
Maximum possible network isolation (NI^M) $NI^M \times 100$	32.2	38.1	12.5	18.6	16.9	25.8	29.5
Maximum isolation difference $[NI^M - NI^R] \times 100$	29.5	35.1	11.6	17.6	15.9	24.1	27.5
Effective network isolation index $[(NI^O - NI^R) / (NI^M - NI^R)] \times 100$	8.6	8.9	23.3	20.9	44.4	10.9	9.4
Observations (millions)	39.7	27.7	3.6	3.7	1.8	21.1	17.3
B. Combined sample							
Observed network isolation (NI^O) $NI^O \times 100$		5.9	2.6	3.9	6.1	5.1	5.5
Simulated random network isolation (NI^R) $NI^R \times 100$		3.2	1.5	1.8	1.2	2.6	2.9
Network isolation difference $[NI^O - NI^R] \times 100$		2.6	1.1	2.1	4.9	2.5	2.6
Maximum possible network isolation (NI^M) $NI^M \times 100$		34.9	16.4	18.4	12.3	28.5	29.4
Maximum isolation difference $[NI^M - NI^R] \times 100$		31.6	14.9	16.6	11.1	25.9	26.5
Effective network isolation index $[(NI^O - NI^R) / (NI^M - NI^R)] \times 100$		8.3	7.3	12.7	44.3	9.6	9.7
Observations (millions)		30.0	4.0	5.2	2.1	21.7	18.0

Notes: NI^O is the fraction of a worker's co-workers (excluding the worker) who reside in the same Census tract as the worker, averaged across all workers in the national sample. NI^R is the same fraction after workers are shuffled randomly among establishments in a workplace tract. NI^M is the same fraction after applying an allocation algorithm to simulate the maximum possible network isolation. Effective network isolation gives an index of the percentage of possible isolation that is explained by residential employment networks. Panel A does the computation using only workers of the same race, ethnicity, or sex (or all workers combined in the "all" column). Panel B does the computation using all workers, but the averages are computed by race or ethnicity. The sample here is for the year 2006 only. The "All" group has more observations than the sum of the tabulated groups because it includes those not tabulated in the listed race/ethnicity categories, and because it does not impose the requirement that a worker's workplace block include at least two establishments with two or more employees of the worker's race/ethnicity group, including the worker's own establishment (see Table 1 for sample restrictions).

Table 3: Descriptive statistics

	Pooled	White	Black	Hispanic	Asian
Variable	Mean	Mean	Mean	Mean	Mean
Retained job for one year	0.650	0.661	0.579	0.633	0.694
Quarterly earnings	8,790	9,408	6,198	6,784	10,880
Quarterly earnings, full quarter jobs	9,627	10,261	7,004	7,449	11,568
Share of co-workers in same group (S^G_{iet})	0.694	0.809	0.381	0.430	0.410
Observed network isolation index (NI^O_{iet})	0.046	0.051	0.024	0.034	0.057
Employer average (EI^O_{iet})	0.048	0.054	0.024	0.034	0.056
Same group observed network isolation index (NI^{GO}_{iet})	0.051	0.053	0.034	0.043	0.072
Same group, employer average (EI^{GO}_{iet})	0.053	0.056	0.034	0.045	0.074
Same commute (TI^O_{iet})	0.025	0.029	0.013	0.016	0.010
Same commute, same group (TI^{GO}_{iet})	0.028	0.030	0.023	0.021	0.025
Age	36.4	36.7	36.0	34.7	36.9
Female	0.466	0.463	0.526	0.433	0.465
Construction	0.099	0.105	0.048	0.138	0.022
Manufacturing	0.131	0.128	0.111	0.154	0.155
Wholesale	0.058	0.059	0.039	0.062	0.072
Retail	0.088	0.094	0.060	0.078	0.082
Transportation, utilities	0.036	0.035	0.053	0.038	0.021
FIRE	0.063	0.066	0.055	0.056	0.058
Other services	0.524	0.512	0.634	0.475	0.591
Employers of size ≤ 25	0.335	0.376	0.156	0.261	0.329
Employer size 26 to 50	0.141	0.147	0.120	0.141	0.112
Employer size 51 to 100	0.139	0.136	0.152	0.153	0.119
Employer size > 100	0.385	0.341	0.572	0.446	0.440
Quarters of tenure	6.1	6.2	5.5	5.9	6.2
Quarters retention	9.3	9.6	7.8	8.9	10.0
Quarters retention with successors	9.6	9.8	8.1	9.2	10.3
Jobs held in last year	1.9	1.8	2.2	2.0	1.7
Total earnings	35,753	38,397	25,053	26,780	43,131
Primary job earnings	27,931	30,192	18,673	20,542	35,266
Observations (millions)	90.6	64.5	9.7	11.9	4.5
Observations, full quarter jobs(millions)	77.7	55.7	8.0	10.1	4.0

Notes: Descriptive statistics (means) are for the estimation sample pooled across the years 2004 to 2007, where the “All” group is the set of those in the listed race/ethnicity subsamples. All of the network measures are calculated for the year a worker is hired and exclude the reference worker (i.e., the average share of workers in the same race/ethnicity group excludes the reference worker from the numerator and denominator) and are not multiplied by 100, as was done in Table 2. Same commute gives the fraction of employees in a worker’s Census tract of employment that also reside in that worker’s home Census tract. Earnings, job tenure, and job count variables are derived from LEHD data. Demographic variables are from the Individual Characteristics File. Employer characteristics of industry and size are reported in the Quarterly Census of Employment and Wages, and input to the LEHD infrastructure file known as the Employer Characteristics File. The “Retain job for one year” and “Quarterly earnings, full quarter jobs” variables are the outcome measures in the regression analysis. The full-quarter jobs sample only retains workers with employment in three adjacent quarters, with earnings measured in the middle quarter.

Table 4: Percentiles of network measures

Sample	Variable (at time-of-hire)	Percentiles				
		10th	25th	50th	75th	90th
Pooled	Share of co-workers in same group ($S^G{}_{iet}$)	0.194	0.491	0.790	0.949	1.000
	Observed network isolation index ($NI^O{}_{iet}$)	0.000	0.000	0.000	0.023	0.121
	Employer average ($EI^O{}_{iet}$)	0.000	0.003	0.012	0.039	0.110
	Same group observed network isolation index ($NI^{GO}{}_{iet}$)	0.000	0.000	0.000	0.023	0.135
	Same group, employer average ($EI^{GO}{}_{iet}$)	0.000	0.001	0.012	0.042	0.123
	Same commute ($TI^O{}_{iet}$)	0.000	0.000	0.004	0.019	0.065
	Same commute, same group ($TI^{GO}{}_{iet}$)	0.000	0.000	0.004	0.022	0.074
White	Share of co-workers in same group ($S^G{}_{iet}$)	0.512	0.705	0.870	0.989	1.000
	Observed network isolation index ($NI^O{}_{iet}$)	0.000	0.000	0.000	0.028	0.143
	Employer average ($EI^O{}_{iet}$)	0.000	0.003	0.014	0.046	0.127
	Same group observed network isolation index ($NI^{GO}{}_{iet}$)	0.000	0.000	0.000	0.028	0.146
	Same group, employer average ($EI^{GO}{}_{iet}$)	0.000	0.003	0.014	0.047	0.133
	Same commute ($TI^O{}_{iet}$)	0.000	0.001	0.005	0.024	0.078
	Same commute, same group ($TI^{GO}{}_{iet}$)	0.000	0.001	0.005	0.025	0.081
Black	Share of co-workers in same group ($S^G{}_{iet}$)	0.071	0.150	0.324	0.574	0.801
	Observed network isolation index ($NI^O{}_{iet}$)	0.000	0.000	0.000	0.015	0.058
	Employer average ($EI^O{}_{iet}$)	0.001	0.004	0.009	0.022	0.052
	Same group observed network isolation index ($NI^{GO}{}_{iet}$)	0.000	0.000	0.000	0.015	0.083
	Same group, employer average ($EI^{GO}{}_{iet}$)	0.000	0.000	0.009	0.027	0.075
	Same commute ($TI^O{}_{iet}$)	0.000	0.000	0.002	0.009	0.032
	Same commute, same group ($TI^{GO}{}_{iet}$)	0.000	0.000	0.003	0.018	0.055
Hispanic	Share of co-workers in same group ($S^G{}_{iet}$)	0.073	0.174	0.379	0.658	0.886
	Observed network isolation index ($NI^O{}_{iet}$)	0.000	0.000	0.000	0.016	0.081
	Employer average ($EI^O{}_{iet}$)	0.000	0.003	0.010	0.027	0.072
	Same group observed network isolation index ($NI^{GO}{}_{iet}$)	0.000	0.000	0.000	0.010	0.108
	Same group, employer average ($EI^{GO}{}_{iet}$)	0.000	0.000	0.008	0.031	0.094
	Same commute ($TI^O{}_{iet}$)	0.000	0.000	0.002	0.010	0.037
	Same commute, same group ($TI^{GO}{}_{iet}$)	0.000	0.000	0.002	0.014	0.052
Asian	Share of co-workers in same group ($S^G{}_{iet}$)	0.036	0.100	0.271	0.748	1.000
	Observed network isolation index ($NI^O{}_{iet}$)	0.000	0.000	0.000	0.016	0.121
	Employer average ($EI^O{}_{iet}$)	0.000	0.002	0.008	0.024	0.099
	Same group observed network isolation index ($NI^{GO}{}_{iet}$)	0.000	0.000	0.000	0.012	0.208
	Same group, employer average ($EI^{GO}{}_{iet}$)	0.000	0.000	0.008	0.039	0.166
	Same commute ($TI^O{}_{iet}$)	0.000	0.000	0.002	0.007	0.025
	Same commute, same group ($TI^{GO}{}_{iet}$)	0.000	0.000	0.001	0.014	0.059

Notes: For the pooled sample and for each race/ethnicity sub-sample, this table provides the values of network association measures at listed percentiles. Network association measures presented here are for the variables used in the estimation analysis, and are not multiplied by 100.

Table 5: Effect of network measures on job retention, all groups combined

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	0.077			0.071	0.038	0.041	0.030
in same group (S^{G}_{iet})	(0.001)			(0.001)	(0.000)	(0.000)	(0.001)
Observed network isolation		0.194		0.073	0.074	0.074	0.038
index (NI^{O}_{iet})		(0.001)		(0.002)	(0.002)	(0.002)	(0.006)
Employer average (EI^{O}_{iet})		-0.076		-0.047	-0.022	-0.028	0.002
		(0.001)		(0.002)	(0.002)	(0.003)	(0.006)
Same commute (TI^{O}_{iet})		0.039		-0.072	0.049	0.046	0.116
		(0.002)		(0.004)	(0.003)	(0.003)	(0.008)
Same-group observed network			0.170	0.116	0.097	0.097	0.050
isolation index (NI^{GO}_{iet})			(0.001)	(0.002)	(0.002)	(0.002)	(0.005)
Same-group, employer			-0.064	-0.041	-0.038	-0.043	-0.025
average (EI^{GO}_{iet})			(0.001)	(0.002)	(0.002)	(0.002)	(0.005)
Same commute, same			0.059	0.100	0.083	0.082	-0.006
group (TI^{GO}_{iet})			(0.002)	(0.004)	(0.003)	(0.002)	(0.007)
Worker controls	yes	yes	yes	yes	yes	yes	
Year controls	yes	yes	yes	yes	yes		yes
Employer controls	yes	yes	yes	yes			yes
Employer fixed effects					yes		
Employer/year fixed effects						yes	
Worker fixed effects							yes
Observations (millions)	90.6	90.6	90.6	90.6	90.6	90.6	90.6
R-squared (within for FE)	0.039	0.040	0.040	0.041	0.021	0.018	0.013
Fixed effects (millions)					3.8	9.7	48.7

Notes: Estimation results are for a linear probability model where the dependent variable is an indicator for whether a worker retains a job one year after the reference date (beginning of the second quarter) in each year. The sample includes those listed in the race/ethnicity subsamples, pooled across the years 2004 to 2007. Standard errors in parentheses are clustered at the employer level, except in column (7) where they are clustered at the individual level. Positive coefficients reflect an increased probability of job retention. The network and share measures range from zero to one. In columns (5)-(7) the within R^2 is reported. The worker controls are dummy variables for race, ethnicity, and sex, dummy variables for the year of the observations, and age and its square. The employer controls include broad industry dummy variables (for construction, manufacturing, wholesale, retail, transport/utilities, and FIRE, relative to the omitted group) and for employer size (≤ 25, 26-50, 51 to 100, relative to the omitted group). Employer fixed effects are at the establishment level, defined by a unique State Employer Identification Number. Employer/year fixed effects are for each establishment in each year.

Table 6: Effect of network measures on job retention – blacks, Hispanics, and Asians

	Blacks	Blacks	Blacks	Blacks	Hispanics	Hispanics	Hispanics	Hispanics	Asians	Asians	Asians	Asians
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Share of co-workers	-0.056	-0.343	-0.314	-0.071	-0.019	-0.267	-0.267	-0.100	-0.017	-0.174	-0.134	-0.072
in same group (S^{G}_{iet})	(0.004)	(0.004)	(0.005)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)
Observed network isolation	0.080	0.112	0.119	0.019	0.137	0.110	0.110	0.046	0.200	0.162	0.161	0.139
index (NI^{O}_{iet})	(0.008)	(0.007)	(0.007)	(0.018)	(0.005)	(0.005)	(0.005)	(0.013)	(0.008)	(0.009)	(0.009)	(0.024)
Employer average (EI^{O}_{iet})	0.049	0.037	0.031	0.059	-0.023	0.032	0.032	0.082	-0.134	-0.068	-0.065	-0.014
	(0.009)	(0.010)	(0.012)	(0.018)	(0.005)	(0.007)	(0.007)	(0.013)	(0.008)	(0.011)	(0.011)	(0.022)
Same commute (TI^{O}_{iet})	0.137	0.122	0.102	0.303	-0.097	0.029	0.029	0.246	0.105	0.169	0.165	0.488
	(0.012)	(0.010)	(0.010)	(0.024)	(0.009)	(0.011)	(0.011)	(0.019)	(0.025)	(0.023)	(0.020)	(0.048)
Same-group observed network	0.059	0.057	0.055	0.059	0.118	0.123	0.123	0.047	0.070	0.115	0.108	0.025
isolation index (NI^{GO}_{iet})	(0.005)	(0.005)	(0.005)	(0.012)	(0.004)	(0.004)	(0.004)	(0.011)	(0.006)	(0.006)	(0.007)	(0.020)
Same-group, employer	-0.057	-0.041	-0.051	-0.082	-0.079	-0.102	-0.102	-0.049	0.004	-0.077	-0.100	0.010
average (EI^{GO}_{iet})	(0.006)	(0.006)	(0.007)	(0.012)	(0.004)	(0.005)	(0.005)	(0.010)	(0.006)	(0.007)	(0.008)	(0.018)
Same commute, same	0.056	0.056	0.062	-0.049	0.072	0.063	0.063	-0.066	0.024	0.004	0.013	-0.114
group (TI^{GO}_{iet})	(0.005)	(0.004)	(0.004)	(0.012)	(0.004)	(0.004)	(0.004)	(0.012)	(0.007)	(0.006)	(0.006)	(0.014)
Worker controls	yes	yes	yes		yes	yes	yes		yes	yes	yes	
Year controls	yes	yes		yes	yes	yes		yes	yes	yes		yes
Employer controls	yes			yes	yes			yes	yes			yes
Employer fixed effects		yes				yes				yes		
Employer/year fixed effects			yes				yes				yes	
Worker fixed effects				yes				yes				yes
Observations (millions)	9.7	9.7	9.7	9.7	11.9	11.9	11.9	11.9	4.5	4.5	4.5	4.5
R-squared (within for FE)	0.046	0.023	0.019	0.023	0.027	0.019	0.015	0.016	0.032	0.019	0.015	0.016
Fixed effects (millions)		0.5	1.2	5.7		0.9	2.0	6.5		0.4	0.9	2.4

Notes: Notes from Table 5 apply. There are no race and ethnicity dummy variables included. The samples are the black, Hispanic, or Asian subsamples, pooled across the years 2004 to 2007.

Table 7: Effect of network measures on log earnings, all groups combined

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.370			-0.311	-0.100	-0.096	-0.093
in same group (S^{G}_{iet})	(0.004)			(0.004)	(0.001)	(0.001)	(0.001)
Observed network isolation		0.180		0.354	-0.056	-0.042	0.043
index (NI^{O}_{iet})		(0.004)		(0.006)	(0.004)	(0.004)	(0.009)
Employer average (EI^{O}_{iet})		-0.743		-0.606	0.051	0.101	-0.263
		(0.004)		(0.006)	(0.005)	(0.005)	(0.009)
Same commute (TI^{O}_{iet})		-1.086		-0.755	-0.201	-0.225	-0.329
		(0.008)		(0.015)	(0.007)	(0.008)	(0.011)
Same-group observed network			0.160	-0.146	-0.092	-0.083	-0.051
isolation index (NI^{GO}_{iet})			(0.003)	(0.005)	(0.004)	(0.004)	(0.008)
Same-group, employer			-0.658	-0.093	0.058	0.045	0.004
average (EI^{GO}_{iet})			(0.004)	(0.005)	(0.004)	(0.004)	(0.008)
Same commute, same			-0.924	-0.283	0.028	0.077	0.012
group (TI^{GO}_{iet})			(0.008)	(0.012)	(0.006)	(0.007)	(0.009)
Worker controls	yes	yes	yes	yes	yes	yes	
Year controls	yes	yes	yes	yes	yes		yes
Employer controls	yes	yes	yes	yes			yes
Employer fixed effects					yes		
Employer/year fixed effects						yes	
Worker fixed effects							yes
Observations (millions)	77.7	77.7	77.7	77.7	77.7	77.7	77.7
R-squared (within for FE)	0.217	0.224	0.223	0.230	0.121	0.111	0.072
Fixed effects (millions)					3.6	9.3	42.7

Notes: Estimation results are for a linear regression where the dependent variable is the log of quarterly earnings in the second quarter of the year. The sample includes those with full quarter employment listed in the race/ethnicity subsamples, pooled across the years 2004 to 2007. Standard errors in parentheses are clustered at the employer level, except in column (7) where they are clustered at the individual level. The network and share measures range from zero to one. In columns (5)-(7) the within R^2 is reported. The worker controls are dummy variables for race, ethnicity, and sex, dummy variables for the year of the observations, age and its square, and dummy variables for the year of hire (which implies controlling for tenure since dummy variables for the year of observation are also included). The employer controls include broad industry dummy variables (for construction, manufacturing, wholesale, retail, transport/utilities, and FIRE, relative to the omitted group) and for employer size (≤ 25, 26-50, 51 to 100, relative to the omitted group). Employer fixed effects are at the establishment level, defined by a unique State Employer Identification Number. Employer/year fixed effects are for each establishment in each year.

Table 8: Effect of network measures on log earnings – blacks, Hispanics, and Asians

	Blacks	Blacks	Blacks	Blacks	Hispanics	Hispanics	Hispanics	Hispanics	Asians	Asians	Asians	Asians
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Share of co-workers	-0.520	-0.207	-0.419	-0.211	-0.431	-0.160	-0.376	-0.130	-0.406	-0.129	-0.273	-0.150
in same group (S^{G}_{iet})	(0.010)	(0.007)	(0.009)	(0.004)	(0.006)	(0.005)	(0.005)	(0.003)	(0.015)	(0.007)	(0.008)	(0.005)
Observed network isolation	0.562	0.094	0.107	0.044	0.433	0.103	0.123	0.035	0.577	0.206	0.197	0.072
index (NI^{O}_{iet})	(0.018)	(0.014)	(0.016)	(0.029)	(0.013)	(0.009)	(0.009)	(0.019)	(0.022)	(0.016)	(0.017)	(0.030)
Employer average (EI^{O}_{iet})	-0.760	0.049	0.120	-0.252	-0.474	0.013	0.080	-0.234	-0.589	-0.086	0.040	-0.211
	(0.021)	(0.021)	(0.025)	(0.030)	(0.015)	(0.011)	(0.013)	(0.018)	(0.025)	(0.023)	(0.020)	(0.029)
Same commute (TI^{O}_{iet})	-0.333	0.031	0.036	-0.144	-0.485	-0.008	-0.006	-0.295	-1.364	0.008	0.017	-0.544
	(0.033)	(0.023)	(0.024)	(0.037)	(0.035)	(0.014)	(0.015)	(0.028)	(0.085)	(0.045)	(0.041)	(0.058)
Same-group observed network	-0.210	-0.196	-0.203	-0.032	-0.206	-0.181	-0.174	-0.015	-0.039	0.022	0.034	-0.027
isolation index (NI^{GO}_{iet})	(0.010)	(0.009)	(0.010)	(0.019)	(0.007)	(0.006)	(0.007)	(0.015)	(0.016)	(0.013)	(0.014)	(0.025)
Same-group, employer	-0.146	0.102	0.116	0.002	-0.049	0.077	0.068	-0.008	-0.161	-0.144	-0.178	-0.052
average (EI^{GO}_{iet})	(0.012)	(0.010)	(0.012)	(0.019)	(0.008)	(0.007)	(0.008)	(0.014)	(0.017)	(0.015)	(0.016)	(0.024)
Same commute, same	-0.538	-0.119	-0.085	0.021	-0.368	-0.093	-0.067	0.015	-0.278	-0.026	0.008	-0.034
group (TI^{GO}_{iet})	(0.016)	(0.015)	(0.017)	(0.016)	(0.010)	(0.007)	(0.008)	(0.016)	(0.021)	(0.015)	(0.016)	(0.016)
Worker controls	yes	yes	yes		yes	yes	yes		yes	yes	yes	
Year controls	yes	yes		yes	yes	yes		yes	yes	yes		yes
Employer controls	yes			yes	yes			yes	yes			yes
Employer fixed effects		yes				yes				yes		
Employer/year fixed effects			yes				yes				yes	
Worker fixed effects				yes				yes				yes
Observations (millions)	8.0	8.0	8.0	8.0	10.1	10.1	10.1	10.1	4.0	4.0	4.0	4.0
R-squared (within for FE)	0.167	0.062	0.052	0.057	0.166	0.076	0.064	0.072	0.234	0.100	0.092	0.089
Fixed effects (millions)		0.5	1.2	4.7		0.8	1.9	5.7		0.4	0.9	2.2

Notes: Notes from Table 7 apply. There are no race and ethnicity dummy variables included. The samples are the black, Hispanic, or Asian subsamples, pooled across the years 2004 to 2007.

Appendix Table B1: Network measures for one year (2006), sample including multi-establishment employers

Variable	All	White	Black	Hispanic	Asian	Male	Female
Computed by sample							
Observed network isolation (NI^O) $NI^O \times 100$	3.7	4.4	2.5	3.4	5.7	3.2	3.2
Simulated random network isolation (NI^R) $NI^R \times 100$	1.9	2.0	0.6	0.7	0.5	1.2	1.3
Network isolation difference $[NI^O - NI^R] \times 100$	1.8	2.3	2.0	2.7	5.1	2.0	1.9
Maximum possible network isolation (NI^M) $NI^M \times 100$	25.6	30.8	12.4	16.4	14.5	21.2	23.1
Maximum isolation difference $[NI^M - NI^R] \times 100$	23.7	28.8	11.8	15.5	14.0	20.0	21.8
Effective network isolation index $[(NI^O - NI^R) / (NI^M - NI^R)] \times 100$	7.7	8.1	16.7	17.3	36.8	10.0	8.9
Observations (millions)	66.3	45.5	7.3	8.0	2.9	34.3	30.6

Notes: Notes to Table 2, referring to Panel A of that table, apply. The only difference is that employers with multiple establishments in the same state are included.

2011

2011/1, Oppedisano, V; Turati, G.: "What are the causes of educational inequalities and of their evolution over time in Europe? Evidence from PISA"
2011/2, Dahlberg, M; Edmark, K; Lundqvist, H.: "Ethnic diversity and preferences for redistribution "
2011/3, Canova, L.; Vaglio, A.: "Why do educated mothers matter? A model of parental help"
2011/4, Delgado, F.J.; Lago-Peñas, S.; Mayor, M.: "On the determinants of local tax rates: new evidence from Spain"
2011/5, Piolatto, A.; Schuett, F.: "A model of music piracy with popularity-dependent copying costs"
2011/6, Duch, N.; García-Estévez, J.; Parellada, M.: "Universities and regional economic growth in Spanish regions"
2011/7, Duch, N.; García-Estévez, J.: "Do universities affect firms' location decisions? Evidence from Spain"
2011/8, Dahlberg, M.; Mörk, E.: "Is there an election cycle in public employment? Separating time effects from election year effects"
2011/9, Costas-Pérez, E.; Solé-Ollé, A.; Sorribas-Navarro, P.: "Corruption scandals, press reporting, and accountability. Evidence from Spanish mayors"
2011/10, Choi, A.; Calero, J.; Escardíbul, J.O.: "Hell to touch the sky? private tutoring and academic achievement in Korea"
2011/11, Mira Godinho, M.; Cartaxo, R.: "University patenting, licensing and technology transfer: how organizational context and available resources determine performance"
2011/12, Duch-Brown, N.; García-Quevedo, J.; Montolio, D.: "The link between public support and private R&D effort: What is the optimal subsidy?"
2011/13, Breuillé, M.L.; Duran-Vigneron, P.; Samson, A.L.: "To assemble to resemble? A study of tax disparities among French municipalities"
2011/14, McCann, P.; Ortega-Argilés, R.: "Smart specialisation, regional growth and applications to EU cohesion policy"
2011/15, Montolio, D.; Trillas, F.: "Regulatory federalism and industrial policy in broadband telecommunications"
2011/16, Pelegrín, A.; Bolancé, C.: "Offshoring and company characteristics: some evidence from the analysis of Spanish firm data"
2011/17, Lin, C.: "Give me your wired and your highly skilled: measuring the impact of immigration policy on employers and shareholders"
2011/18, Bianchini, L.; Revelli, F.: "Green polities: urban environmental performance and government popularity"
2011/19, López Real, J.: "Family reunification or point-based immigration system? The case of the U.S. and Mexico"
2011/20, Bogliacino, F.; Piva, M.; Vivarelli, M.: "The impact of R&D on employment in Europe: a firm-level analysis"
2011/21, Tonello, M.: "Mechanisms of peer interactions between native and non-native students: rejection or integration?"
2011/22, García-Quevedo, J.; Mas-Verdú, F.; Montolio, D.: "What type of innovative firms acquire knowledge intensive services and from which suppliers?"
2011/23, Banal-Estañol, A.; Macho-Stadler, I.; Pérez-Castrillo, D.: "Research output from university-industry collaborative projects"
2011/24, Ligthart, J.E.; Van Oudheusden, P.: "In government we trust: the role of fiscal decentralization"
2011/25, Mongrain, S.; Wilson, J.D.: "Tax competition with heterogeneous capital mobility"
2011/26, Caruso, R.; Costa, J.; Ricciuti, R.: "The probability of military rule in Africa, 1970-2007"
2011/27, Solé-Ollé, A.; Viladecans-Marsal, E.: "Local spending and the housing boom"
2011/28, Simón, H.; Ramos, R.; Sanromá, E.: "Occupational mobility of immigrants in a low skilled economy. The Spanish case"
2011/29, Piolatto, A.; Trotin, G.: "Optimal tax enforcement under prospect theory"
2011/30, Montolio, D; Piolatto, A.: "Financing public education when altruistic agents have retirement concerns"
2011/31, García-Quevedo, J.; Pellegrino, G.; Vivarelli, M.: "The determinants of YICs' R&D activity"
2011/32, Goodspeed, T.J.: "Corruption, accountability, and decentralization: theory and evidence from Mexico"
2011/33, Pedraja, F.; Cordero, J.M.: "Analysis of alternative proposals to reform the Spanish intergovernmental transfer system for municipalities"
2011/34, Jofre-Monseny, J.; Sorribas-Navarro, P.; Vázquez-Grenno, J.: "Welfare spending and ethnic heterogeneity: evidence from a massive immigration wave"
2011/35, Lyytikäinen, T.: "Tax competition among local governments: evidence from a property tax reform in Finland"
2011/36, Brülhart, M.; Schmidheiny, K.: "Estimating the Rivalness of State-Level Inward FDI"
2011/37, García-Pérez, J.I.; Hidalgo-Hidalgo, M.; Robles-Zurita, J.A.: "Does grade retention affect achievement? Some evidence from Pisa"
2011/38, Boffa, f.; Panzar. J.: "Bottleneck co-ownership as a regulatory alternative"

2011/39, González-Val, R.; Olmo, J.: "Growth in a cross-section of cities: location, increasing returns or random growth?"
2011/40, Anesi, V.; De Donder, P.: "Voting under the threat of secession: accommodation vs. repression"
2011/41, Di Pietro, G.; Mora, T.: "The effect of the l'Aquila earthquake on labour market outcomes"
2011/42, Brueckner, J.K.; Neumark, D.: "Beaches, sunshine, and public-sector pay: theory and evidence on amenities and rent extraction by government workers"
2011/43, Cortés, D.: "Decentralization of government and contracting with the private sector"
2011/44, Turati, G.; Montolio, D.; Piacenza, M.: "Fiscal decentralisation, private school funding, and students' achievements. A tale from two Roman catholic countries"

2012

2012/1, Montolio, D.; Trujillo, E.: "What drives investment in telecommunications? The role of regulation, firms' internationalization and market knowledge"
2012/2, Giesen, K.; Suedekum, J.: "The size distribution across all "cities": a unifying approach"
2012/3, Foremny, D.; Riedel, N.: "Business taxes and the electoral cycle"
2012/4, García-Estévez, J.; Duch-Brown, N.: "Student graduation: to what extent does university expenditure matter?"
2012/5, Durán-Cabré, J.M.; Esteller-Moré, A.; Salvadori, L.: "Empirical evidence on horizontal competition in tax enforcement"
2012/6, Pickering, A.C.; Rockey, J.: "Ideology and the growth of US state government"
2012/7, Vergolini, L.; Zanini, N.: "How does aid matter? The effect of financial aid on university enrolment decisions"
2012/8, Backus, P.: "Gibrat's law and legacy for non-profit organisations: a non-parametric analysis"
2012/9, Jofre-Monseny, J.; Marín-López, R.; Viladecans-Marsal, E.: "What underlies localization and urbanization economies? Evidence from the location of new firms"
2012/10, Mantovani, A.; Vandekerckhove, J.: "The strategic interplay between bundling and merging in complementary markets"
2012/11, Garcia-López, M.A.: "Urban spatial structure, suburbanization and transportation in Barcelona"
2012/12, Revelli, F.: "Business taxation and economic performance in hierarchical government structures"
2012/13, Arqué-Castells, P.; Mohnen, P.: "Sunk costs, extensive R&D subsidies and permanent inducement effects"
2012/14, Boffa, F.; Piolatto, A.; Ponzetto, G.: "Centralization and accountability: theory and evidence from the Clean Air Act"
2012/15, Cheshire, P.C.; Hilber, C.A.L.; Kaplanis, I.: "Land use regulation and productivity – land matters: evidence from a UK supermarket chain"
2012/16, Choi, A.; Calero, J.: "The contribution of the disabled to the attainment of the Europe 2020 strategy headline targets"
2012/17, Silva, J.I.; Vázquez-Grenno, J.: "The ins and outs of unemployment in a two-tier labor market"
2012/18, González-Val, R.; Lanaspa, L.; Sanz, F.: "New evidence on Gibrat's law for cities"
2012/19, Vázquez-Grenno, J.: "Job search methods in times of crisis: native and immigrant strategies in Spain"
2012/20, Lessmann, C.: "Regional inequality and decentralization – an empirical analysis"
2012/21, Nuevo-Chiquero, A.: "Trends in shotgun marriages: the pill, the will or the cost?"
2012/22, Piil Damm, A.: "Neighborhood quality and labor market outcomes: evidence from quasi-random neighborhood assignment of immigrants"
2012/23, Ploeckl, F.: "Space, settlements, towns: the influence of geography and market access on settlement distribution and urbanization"
2012/24, Algan, Y.; Hémet, C.; Laitin, D.: "Diversity and local public goods: a natural experiment with exogenous residential allocation"
2012/25, Martinez, D.; Sjögren, T.: "Vertical externalities with lump-sum taxes: how much difference does unemployment make?"
2012/26, Cubel, M.; Sanchez-Pages, S.: "The effect of within-group inequality in a conflict against a unitary threat"
2012/27, Andini, M.; De Blasio, G.; Duranton, G.; Strange, W.C.: "Marshallian labor market pooling: evidence from Italy"
2012/28, Solé-Ollé, A.; Viladecans-Marsal, E.: "Do political parties matter for local land use policies?"
2012/29, Buonanno, P.; Durante, R.; Prarolo, G.; Vanin, P.: "Poor institutions, rich mines: resource curse and the origins of the Sicilian mafia"
2012/30, Anghel, B.; Cabrales, A.; Carro, J.M.: "Evaluating a bilingual education program in Spain: the impact beyond foreign language learning"
2012/31, Curto-Grau, M.; Solé-Ollé, A.; Sorribas-Navarro, P.: "Partisan targeting of inter-governmental transfers & state interference in local elections: evidence from Spain"

2012/32, Kappeler, A.; Solé-Ollé, A.; Stephan, A.; Välilä, T.: "Does fiscal decentralization foster regional investment in productive infrastructure?"
2012/33, Rizzo, L.; Zanardi, A.: "Single vs double ballot and party coalitions: the impact on fiscal policy. Evidence from Italy"
2012/34, Ramachandran, R.: "Language use in education and primary schooling attainment: evidence from a natural experiment in Ethiopia"
2012/35, Rothstein, J.: "Teacher quality policy when supply matters"
2012/36, Ahlfeldt, G.M.: "The hidden dimensions of urbanity"
2012/37, Mora, T.; Gil, J.; Sicras-Mainar, A.: "The influence of BMI, obesity and overweight on medical costs: a panel data approach"
2012/38, Pelegrín, A.; García-Quevedo, J.: "Which firms are involved in foreign vertical integration?"
2012/39, Agasisti, T.; Longobardi, S.: "Inequality in education: can Italian disadvantaged students close the gap? A focus on resilience in the Italian school system"

2013

2013/1, Sánchez-Vidal, M.; González-Val, R.; Viladecans-Marsal, E.: "Sequential city growth in the US: does age matter?"
2013/2, Hortas Rico, M.: "Sprawl, blight and the role of urban containment policies. Evidence from US cities"
2013/3, Lampón, J.F.; Cabanelas-Lorenzo, P-; Lago-Peñas, S.: "Why firms relocate their production overseas? The answer lies inside: corporate, logistic and technological determinants"
2013/4, Montolio, D.; Planells, S.: "Does tourism boost criminal activity? Evidence from a top touristic country"
2013/5, Garcia-López, M.A.; Holl, A.; Viladecans-Marsal, E.: "Suburbanization and highways: when the Romans, the Bourbons and the first cars still shape Spanish cities"
2013/6, Bosch, N.; Espasa, M.; Montolio, D.: "Should large Spanish municipalities be financially compensated? Costs and benefits of being a capital/central municipality"
2013/7, Escardíbul, J.O.; Mora, T.: "Teacher gender and student performance in mathematics. Evidence from Catalonia"
2013/8, Arqué-Castells, P.; Viladecans-Marsal, E.: "Banking towards development: evidence from the Spanish banking expansion plan"
2013/9, Asensio, J.; Gómez-Lobo, A.; Matas, A.: "How effective are policies to reduce gasoline consumption? Evaluating a quasi-natural experiment in Spain"
2013/10, Jofre-Monseny, J.: "The effects of unemployment benefits on migration in lagging regions"
2013/11, Segarra, A.; García-Quevedo, J.; Teruel, M.: "Financial constraints and the failure of innovation projects"
2013/12, Jerrim, J.; Choi, A.: "The mathematics skills of school children: How does England compare to the high performing East Asian jurisdictions?"
2013/13, González-Val, R.; Tirado-Fabregat, D.A.; Viladecans-Marsal, E.: "Market potential and city growth: Spain 1860-1960"
2013/14, Lundqvist, H.: "Is it worth it? On the returns to holding political office"
2013/15, Ahlfeldt, G.M.; Maennig, W.: "Homevoters vs. leasevoters: a spatial analysis of airport effects"
2013/16, Lampón, J.F.; Lago-Peñas, S.: "Factors behind international relocation and changes in production geography in the European automobile components industry"
2013/17, Guío, J.M.; Choi, A.: "Evolution of the school failure risk during the 2000 decade in Spain: analysis of Pisa results with a two-level logistic mode"
2013/18, Dahlby, B.; Rodden, J.: "A political economy model of the vertical fiscal gap and vertical fiscal imbalances in a federation"
2013/19, Acacia, F.; Cubel, M.: "Strategic voting and happiness"

Universitat de Barcelona

Dos campus d'excel·lència internacional

ieb@ub.edu
www.ieb.ub.edu

Cities and Innovation

Document de treball de l'IEB