Do Labor Market Networks Have An Important Spatial Dimension?

Judith K. Hellerstein*
University of Maryland and NBER

Mark J. Kutzbach*
Center for Economic Studies, U.S. Bureau of the Census

David Neumark*
UCI, NBER, and IZA

CES 12-30 September 2012

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 captures a non-wage amenity.

* We are grateful to Gilles Duranton, Kristin McCue, Erika McEntarfer, and Henry Overman for helpful comments. Any opinions and conclusions expressed herein are those of the authors 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 also stem from labor market discrimination; for example, employers with a large share of an ethnic group may 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

[2] An example is the Jobs-Plus program, discussed in more detail in Hellerstein and Neumark (2011).

networks.

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).[3] 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, 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.

More 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

[3] Pellizzari (2010) provides recent, similar evidence for many European countries.

the same Census block than are two individuals living in the same block group but not on the same block. 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.[4] This figure doubles for blacks in comparable 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, for Hispanics residence-based networks are considerably more important; for Hispanics overall, the grouping of 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

[4] This measure is explained in more detail in Section IV of the paper.

labor market, and for whom employers may have less reliable information.[5]

A potential alternative explanation of this evidence pointing to a role for labor market networks is 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 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 based on neighborhood of residence implies some racial stratification of networks. After all, given pervasive racial residential segregation in the United States, networks among neighbors have to be partially race-based. 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

[5] Hellerstein et al. (2010) present a different type of evidence pointing to the importance of labor market networks among Hispanic men.

importance of networks disregarding the race of neighbors and co-workers falls by more than 40 percent.[6]

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 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 United States 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).[7] 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 in the literature, 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

[6] 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).

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

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 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. The find that, 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.[8] Interestingly, the results are the opposite *without* the fixed firm effects, perhaps because there is more hiring of minorities in 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.[9]

---

[8] The model in Brown et al. (2012) is similar, and is tested using data from a single company which 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.

[9] 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, there is a concern that the sample 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, it is not clear that one wants to condition on worker fixed effects, since these presumably 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. Their motivation for including worker fixed effects is that there may be "selection of workers of different abilities into firms with a low or high share of workers from the same minority group" (Dustmann et al., 2011, p. 19). But that is presumably the function of networks. We can think of two reasons one might want individual fixed effects. First, there might be match-specific productivity associated with stronger network connections, even conditional on worker-specific unobservables. In other words, the same worker observed two or more times may be in a better match in workplaces in which he has more network connections. And second, 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; the individual fixed effects can control for this.

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 in that which 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 complementarities.[10]

Schmutte (2010) takes up a similar question, although he poses it somewhat differently, as 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

[10] 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, note that these results come from a cross-section, and 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. Thus, it is possible that longitudinal variation over time in the share Hispanic that speaks English poorly would give a different result, especially if poor English-speaking Hispanics are crowded into low-wage establishments, and this crowding drives the cross-sectional results.

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 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.[11]

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

[11] Studying data from a single company, Kmec (2007) studies the relationship between different kinds of informal contacts and turnover, but does not distinguish between referral and non-referral hires.

question of racial or ethnic stratification of networks.[12]

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. 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 may have important 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.[13] The LEHD program has assembled a national database with over 125 million jobs and more than 7 million employers in a typical quarter. Data series for some states begin before 1990, but the national partnership was only recently completed. This project makes use of worker and employer data from 38 states, in the years

[12] Kmec and Trimble (2009) studies 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.

[13] See Abowd et al. (2004) for a description of the LEHD program and data at the Census Bureau.

2004 through 2007.[14] 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.

Previous research on residential employment networks used the 2000 Decennial Employer-Employee Database (DEED), which was constructed based on detailed, labor-intensive matching of 2000 Long-Form Census respondents from the "Sample Edited Detail File" (SEDF) to their establishment of employment (Hellerstein and Neumark, 2003). 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," which 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. In addition, federal government workers and the self-employed are currently not covered, although there is ongoing work to add them. Second, it collects wage record data at the level of

---

[14] All 50 states and the District of Columbia have now agreed to become part of the program, but not all states' data are as yet fully incorporated into the infrastructure in terms of data availability and production of LEHD core products (Quarterly Workforce Indicators and OnTheMap commuting patterns).

an employer, identified in each state by a State Employer Identification Number (SEIN).[15] 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 some worker information available in the DEED.

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). The analysis sample is for jobs with an employer located in one of 38 states whose time series spans the entire year range. These state time-series also extend back at least to 2001, with the earlier years used to construct time-of-hire network measures, discussed below.

For each year, the sample only includes those jobs from the Employment History File (EHF) 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. Those workers with earnings in both the first and second quarter 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.[16] We make use of wage records in all available states – even those outside of the analysis sample of 38 states – to determine whether jobs in the sample are a primary job. To provide descriptive information on the sample construction, Table 1 presents the restrictions used to arrive at the analysis samples and, for a single year

[15] An SEIN defines an employer account within a state, and is not necessarily the highest level of control or ownership. To characterize firms, the LEHD program has also matched employers to firm identifiers from the Business Register, the frame for the Economic Census.

[16] Approximately 85 percent of beginning-of-quarter workers in the LEHD data only have one job at the beginning of a quarter. The DEED is similar in that respondents only provide information on the where they worked the most hours.

(2006) provides the count of workers remaining after each restriction. The initial sample consists of about 100 million workers in each year.[17]

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 establishment information in the Employer Characteristics File (ECF) for each state.[18] 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.[19] 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. Because the LEHD defines an employer as a unique SEIN, establishments may still belong to multi-unit firms broken out into different lines of business or by state.

On the worker side, we link jobs to the Individual Characteristics File (ICF), composed of demographic information from survey and administrative data. For each job record, the ICF provides age, sex, race, and ethnicity, either matched directly from federal administrative records within the Person Characteristics File, matched by personally identifying information to a response to the 2000 Decennial Census, or imputed based on available demographic information and employment history.[20] 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

[17] Because these results may be revised, the job counts presented here are rounded to the nearest 1,000. These measures will simplify disclosure avoidance analysis. The list of states included will be provided in the final version.
[18] 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.
[19] Public sector employers, such as school districts, often report as single units rather than breaking out employment by establishment.
[20] See (Abowd et al., 2011) for a discussion of the imputation model. We only use the first of ten independently drawn imputes for each worker. For this sample, 80% of jobs had no demographic information imputed, with most of the imputes being for race/ethnicity, and with age and sex only imputed in about 1% of jobs.

Decennial Census.

We restrict the set of jobs to those with a complete set of worker and employer location information, both of which are necessary for constructing the network measures and defining sub-samples. The network isolation measure defines associations by identifying sets of workers that work in the same Census tract and live in the same Census tract. We require that all jobs have at least this level of precision, dropping 3% of jobs due to employer geography and 12% due to residential geography.[21] 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, this two-worker requirement and two-employer requirement is 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 Employer Characteristics File). In contrast, the DEED sample contains, on average, only 16% of the jobs reported in matched units of the Business Register.

The network analysis sample 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 not-Hispanic, Hispanic, and Asian not-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

[21] Precision of worker and employer addresses usually exceeds this requirement, with most addresses having enough precision to be coded to a Census block.

workplace Census tract, which results in a pooled sample that is 4.1% smaller.

Although the sample derivation so far describes the sample used to construct our network measures, our model estimates are restricted to shorter-tenure workers. We impose this restriction for two reasons. First, the effects of network 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. Specifically, we retain workers whose current job has tenure of three or fewer full years prior to the April 1st employment date, which drops 37.6% of jobs. After retaining only workers in the listed race/ethnicity groups, the estimation sample consists of about 23.6 million jobs for 2006.

Finally, recall that the figures given in Table 1 are for a single year. The pooled sample for all four years is over 90 million worker/years, with almost 50 million unique workers at about 3.7 million unique employers.

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 still retains a job one year after the reference date. Job retention is measured by continued earnings from the same employer in the year following the observation.[22,23] 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

[22] In cases where earnings from an employer end before one year, it is possible that the position has been retained but that the business has restructured and uses a new SEIN for reporting. For these cases, we use the Successor Predecessor File, which lists cases where an employer most likely has workers transitioning to a new SEIN. For the 10% of workers in the sample whose job transitions to a new SEIN, we extend the job histories to include employment at the successor. Including these extra quarters increases the share of jobs lasting at least one year to 65% from 63%. Because retention is defined based on 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.

[23] In cases where an employer declines significantly or dies from one year to the next, workers whose job ends would still be considered to have turned over.

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 LEHD data. We construct indicator variables for the year of employment, as well as the year in which the worker began that job. From employer reporting in the QCEW, we construct indicator variables for industry sectors and employer size, as reported by each SEIN. From the ICF demographic variables, we include indicator variables for race/ethnicity group and female, and we construct age and age squared.

IV. Methods

Measuring the Importance of Labor Market Networks with the LEHD Data

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 recomputing the measures of the importance of labor market networks used in Hellerstein et al. (2011), but using the LEHD data instead.

Consider a single group of workers, say whites. For the sample of whites, we first compute for each worker the percentage of (white) co-workers (in the same establishment) who live in the same residential neighborhood as that worker.[24] This requires a sample restriction to establishments with at least two white workers observed. We average this percentage across workers in the sample to create the "network isolation index," denoted NI^O – the fraction of co-workers who are *observed* in our estimation sample to be residential neighbors:

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

where there are N workers, 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. Their ratio is the share of co-workers with whom each worker is co-resident. This ratio is then averaged over all workers, and to put into percentage terms can be multiplied

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

by 100.

We define residential neighborhoods as Census tracts, most importantly because Census tracts are relatively small. 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. The Census Bureau defines tracts for tabulation purposes, and designs them to be contiguous geographic units with an optimal size of about 4,000 residents.

We do not want to attribute all clustering of neighbors in the same establishment to networks (or 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 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, industry, 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 one (or 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, which is described in detail in Hellerstein et al. (2011). This gives us our maximum network isolation index, denoted 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

(2) $[(NI^O - NI^R)/(NI^M - NI^R)]$,

which we call the "effective network isolation index." It measures what share of the maximum possible network isolation that could occur in the data actually does occur in the data. As such, it provides a 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.[25]

Finally, paralleling the analysis in Hellerstein et al. (2011), we do these calculations using all

[25] 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. Moreover, the network isolation index is not our main focus in this paper. For all these reasons, we simply report the estimates below.

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.

The Effects of Networks on Turnover and Earnings

Our main analysis concerns the effects of residence-based networks on labor market outcomes – specifically turnover and earnings. The analysis of turnover of course requires longitudinal data, which is one reason why the LEHD data are so valuable.

As explained in the previous subsection, our labor market network measure is the "effective network isolation index" defined in equation (2). In that equation, we adjusted NI^O by NI^R to account for differences across Census tracts of employment that could generate variation in the extent to which neighbors work together that 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 the same tract. The NI^O, NI^R, and NI^M measures (in Hellerstein et al., 2011, and in the brief overview above) were originally constructed to measure network isolation in the aggregate. Because the random and maximum network isolation measures are defined based on shuffling workers among jobs within a Census tract, the measures 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.

One other difference between our measure of the extent of labor market networks and our analysis of the impact of networks on turnover and earnings is that for the latter we use NI^O measured 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 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. Table 2 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.[26]

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 that makes flows of workers among particular residential and work tracts more common. To control for observed network isolation that is the result of commuting tendencies rather than interpersonal

[26] 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. 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). Note that 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 (1)), based on contemporary co-workers. The percentages cited here are low because we have already restricted the subsamples by race or ethnicity to only include workers with at least one co-worker in the same group contemporaneously.

connections, some of the empirical specifications include an origin/destination network isolation measure. For each worker, TI^O gives 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 (so denoted because it captures isolation by tract) is constructed in an identical manner as NI^O,[27] except that we use the workplace Census tract rather than the establishment. Our hypothesis is that workers in jobs with greater network 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.[28] Both of these predictions are generally borne out in the data.

In a sense, TI^O substitutes for NI^R, 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 conditional on TI^O indicates the presence of a network that links neighbors to specific establishments within Census tracts.[29]

For our analysis of the importance of networks for turnover and earnings we also create a control variable for the average level of network isolation at an employer. This variable is constructed as

$$(3) \qquad NI_E{}^O = \frac{1}{N_E}\sum_{i=1}^{N_E}\frac{\sum_{j\neq i} I^R(i,j)\cdot I^E(i,j)}{\sum_{j\neq i} I^E(i,j)},$$

where the subscript of *E* on *N* indicates that the observed network isolation is averaged over workers at

[27] In particular, to be symmetric with NI^O, we define this based on time of hire.
[28] We couch this discussion in terms of commuting. But there could be other influences that lead people who live in particular neighborhoods to work in specific areas. As long as this yields utility to workers, the same argument applies.
[29] The adjustment by NI^M is less relevant to our analysis in this paper, as this adjustment was meant to scale the importance of labor market networks by asking how much of the clustering of workers in firms that could occur actually did occur. For the present analysis, we are asking 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.

the establishment.[30] 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 NI_E^O, 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, and rather than adjusting for NI^R and NI^M, which are not relevant constructs here, we include NI_E^O and TI^O 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:

$$(4) \qquad R_{ijt} = \alpha + \beta_G S^G{}_{jt} + \gamma_B B_{ijt} + \gamma_H H_{ijt} + \gamma_A A_{ijt} + X_{ijt}\theta + \varepsilon_{ijt}.$$

In this equation, S^G denotes the share of the employer's workforce that is in the individual's race or ethnic group – defined as white, black, Hispanic, or Asian.[31] Thus, β_G measures the effect of a higher share of a worker's race or ethnicity in the workforce on retention of workers. B, H, and A 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

[30] For example, the value of NI_E^O in 2006, for those hired in 2003, is computed as the average across all co-workers of NI^O in 2003, where NI^O is based on the observed clustering of neighbors in the establishment in 2003.
[31] Specifically, the groups are non-Hispanic whites, non-Hispanic blacks, Hispanics, and non-Hispanic Asians. Again, to be symmetric with the network isolation measure that gets added to this specification, and to be more consistent with the potential effect of networks on job matches, we define this at the time of hire.

one's race or ethnic group varies across groups. Of course in these analyses the race/ethnicity dummy variables drop out. The vector X contains other individual-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 (4), but using the residence-based network measure instead, as in:

$$(5) \qquad R_{ijt} = \alpha + \beta_O NI^O{}_{ijt} + \beta_E NI_E{}^O{}_{jt} + \beta_T TI^O{}_{jt} + \gamma_B B_{ijt} + \gamma_H H_{ijt} + \gamma_A A_{ijt} + X_{ijt}\theta + \varepsilon_{ijt}.$$

This specification simply asks whether worker turnover is lower when the worker has stronger residential links to co-workers, as measured by β_O.[32] 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, $NI_E{}^O$. In addition, this specification accounts for variation driven by stronger commute flows between tracts of work and tracts of residence, by also including TI^O – which captures the clustering of workers in the same employment tract as the worker in the same tracts of residence – as an additional control.

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, $NI_E{}^O$, and TI^O defined for the worker's own race or ethnic group – denoted $NI_G{}^O$, $NI_{GE}{}^O$, and $TI_G{}^O$ where the G subscript, as before indicates that this measure is specific to the group.[33] That is, we estimate:

$$(6) \qquad R_{ijt} = \alpha + \beta_{GO} NI_G{}^O{}_{ijt} + \beta_{GE} NI_{GE}{}^O{}_{ijt} + \beta_{GT} TI_T{}^O{}_{ijt} + \gamma_B B_{ijt} + \gamma_H H_{ijt} + \gamma_A A_{ijt} + X_{ijt}\theta + \varepsilon_{ijt}.$$

Thus far we have discussed specifications using 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 (4),

[32] This is defined over all workers, but because of residential segregation will be in large part driven by clustering of workers with co-residents of the same race/ethnicity.

[33] 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.

(5), and (6), so that the equation becomes:

$$(7) \qquad R_{ijt} = \alpha + \beta_G S^G{}_{jt} + \beta_O NI^O{}_{ijt} + \beta_E NI_E{}^O{}_{jt} + \beta_T TI^O{}_{jt} + \beta_{GO} NI_G{}^O{}_{ijt} + \beta_{GE} NI_{GE}{}^O{}_{ijt} + \beta_{GT} TI_T{}^O{}_{ijt} + \gamma_B B_{ijt} + \gamma_H H_{ijt} + \gamma_A A_{ijt} + X_{ijt}\theta + \varepsilon_{ijt}.$$

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 establishment'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.[34] 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.

As discussed with reference to regression equation (5), it is also of interest to estimate these models for separate race or ethnic groups, to see whether the conclusions vary across groups –

[34] Among other things, this would control for the influence of large layoffs (or a closing) affecting many workers in an establishment in the same year.

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).

Productive Networks or Amenities?

To this point, we have always 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.[35] 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. 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), and one that measures this clustering only for those of the same race or ethnicity (NI_G^O). 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

[35] Brown et al. (2012) offer this same interpretation of their finding of slower wage growth for referral hires. "Another possibility is that the eventual salary disadvantage is compensated by non-pecuniary aspects of the job match, such as a more enjoyable work environment because the referred worker has social contacts within the corporation" (p. 31).

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 more apparent 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 therefore are best 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.[36]

V. Results

We first report descriptive statistics, including new estimates of the importance of residence-based labor market networks using the LEHD. We then describe the results from the regression analysis of network measures, turnover, and earnings.

Descriptive Statistics

We present the network measures using the LEHD samples in Table 3. The estimates in Panel A of Table 3 are computed group by group. Pooling all the data, the observed network isolation index (NI^O)

---

[36] There is an empirical implication of this interpretation with regard to the estimates that should result for the two network measures whether or not we condition on the other, given that the measures are positively correlated. In the models for retention, if we just include one measure or the other, we should find larger positive effects on retention than when we include both, because in the former case the included network measure will pick up both positive job-match and positive amenity effects. In the earnings equation, the estimated effect of the overall network measure should increase when we condition on both, because we remove the negative amenity effect. And conversely the estimated effect of the same-group network measure should decrease because we remove the positive job-match effect of the overall measure (and it should become negative if this variable primarily picks up an amenity effect).

is on average 5.3, implying that workers have, on average, 5.3 percent of their co-workers living in their home Census tract.[37] 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 more than 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 (based on the definitions in Table 1). The effective network isolation index is a bit higher for whites, at 8.9 percent, but much higher 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.[38]

Note that these differences are not apparent in the first row of Table 3 – the values of the simple observed network isolation index. Because whites are such as 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 A1 reports similar calculations when we include workers in multi-establishment firms, using the imputation procedure to assign workers. In every case the observed network isolation

[37] In Table 3 (and Appendix Table A1 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.
[38] Those not tabulated in the sub-samples of Table 3 are also highly networked, with an effective network isolation index of 39.7.

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 the firm. The results in Appendix Table A1 are 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.[39] 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 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 3 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 from someone of the same sex is more valuable.

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

Paralleling this last point, Panel B of Table 3 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, blacks, and especially Hispanics, the effective network isolation index is 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.

For the full estimation sample and each race or ethnic group, Table 4 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 ($NI_{GT}{}^{O}$) are about the same as the overall measures (NI^{O}); 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 (2)), 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} – is typically about half the observed network isolation, roughly corresponding to the random network isolation measure in Table 3; 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 5 provides some information on the distributions of the various network measures and related control variables that we use. Comparing Tables 4 and 5, 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). For example, for the pooled sample the mean of NI^O 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 6 for the full sample, and in Tables 7-9 for the race or ethnic group subsamples. 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 over the 99 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 6, column (1) indicates that the share of a worker's co-workers in the same group (S^G) 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, introduce our residence-based network measure NI^O (along with the employer average, $NI_E{}^O$, and same commute average, TI^O, 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 5, 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) 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 give 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), as in the specification in column (1) of Table 6, may reflect these residence-based networks. Alternatively, it may capture networks along other lines that connect worker in the same race or ethnic group, or something different from networks altogether. The impact of the observed network isolation index (NI^O) 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 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 ($NI_G{}^O$) has an effect above and beyond the measure without regard to group (NI^O). Moreover, as we noted earlier, the results for these two network measures can also differ because they have different influences, with NI^O capturing productive network effects and $NI_G{}^O$ 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, because its estimated coefficient does not change as the latter are added to the specification.

The coefficients on the overall and group-specific observed network isolation indexes in column (3) 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 6 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 from what they are without any fixed effects, 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.[40]

[40] Also, one potential concern with the worker fixed effects specification is that the remaining variation across employers could be due to inconsistencies in worker/employer matches. For example, if the wrong employer were matched to a worker in one year, the variation in network measures due to the mismatch could overwhelm normal variation due to actual job changes, potentially causing spurious changes in the coefficients. Such occurrences are unlikely because wage records are submitted jointly with employer identifiers and because the LEHD processing

Tables 7-9 report results for the minority race/ethnicity subsamples of blacks, Hispanics, and Asians.[41] The structure of the tables and the specifications shown are the same as for Table 6, 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 and $NI_G{}^O$) 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. One difference between the overall results and the group-specific results is that the estimated impact of the overall network isolation index is larger than the own-group network isolation index in reducing turnover (except for the models for blacks and Hispanics with worker fixed effects). This general finding suggests either that for non-whites, being networked to neighbors from one's own race or ethnic group does not convey larger match-specific gains as reflected in turnover, or alternatively that being networked in this way may be capturing a non-wage amenity.

The effect of the same-group share, S^G, differs quite a bit in Tables 7-9 from its estimate in Table 6. First, for blacks, Hispanics, and Asians this estimate it is consistently negative. Its size also varies considerably across groups and across specifications within groups. The estimated coefficients of S^G are in contrast to 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 groups, from a different country, so we should not necessarily expect the same results.[42]

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

---

system specifically checks for cases of a continuing employer not reporting any wage records in a quarter. Furthermore, because we measure network associations as of the time-of-hire, the measures for a worker at an employer would not change from year-to-year.

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

[42] Also, Dustmann et al. (2011) focus on analyses of data for a single metropolitan area. But this seems unlikely to explain the difference in results. Network connections among ethnic group members in a single metropolitan area would influence our results as well, since nearly all workers are employed within the metropolitan area in which they live.

positive and robust.   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 four tables of estimates, for all workers and 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.[43]

Across all four tables (Tables 10-13), 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 10, for all groups (including whites) combined, the estimated effects of the overall (NI^O) and same-group ($NI_G{}^O$) 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 Tables 11-13, for the earnings of blacks, Hispanics and Asians. With all groups combined (in Table 10), 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.

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

However, in Tables 11-13, where we look at the minority groups separately, we always find a positive effect of the overall network measure (NI^O), across all specifications. This evidence is consistent with 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.[44,45]

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 3 (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, however, is not consistent with the same evidence of positive effects of the overall retention measure for blacks or other minorities considered separately.

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

[44] 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 crowding stemming from discrimination in hiring. This kind of crowding could explain the negative effect of the same-group share for minorities evident in Tables 11-13. However, there is no reason to expect 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, so the effect of the same-group network measure is unlikely to reflect crowding.

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

as of the time of hire. The results for retention were qualitatively similar, with the overall residence-based network measure (NI^O) and the group-specific measure ($NI_G{}^O$) having a positive effect on retention. For earnings, the results for NI^O were qualitatively the same as using the time-of-hire network measure. For $NI_G{}^O$, 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 of $NI_G{}^O$ was near zero and statistically insignificant, perhaps reflecting the ambiguous predictions for earnings because network amenity effects are in opposite directions.[46]

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. Thus, our view is that the results using the time-of-hire network measures, which we report in the tables in this paper, provide the cleanest evidence.

VI. Conclusions and Discussion

If labor market networks lead to better matches in the job market, they should reduce turnover and increase wages. These hypotheses are not new to our paper, nor are we the first to test them. Our innovation is to use matched employer-employee data with detailed information on where people 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 important, but it has assessed only the "presence" of connections between workers who are neighbors; it has not assessed any evidence asking whether these networks are productive in terms of improving labor market matches, which might

[46] We also found that the retention effects were robust to other variants of the specification, like dropping the employer average controls ($NI_E{}^O$ and $NI_{GE}{}^O$), and defining NI^O and the associated controls as of the time of hire but $NI_G{}^O$ 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.

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. In our view, it is much less clear on a priori grounds that workers have network connections simply because they are of the same race or ethnic group. Perhaps reflecting this, 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, even when employer fixed effects are included to account for the possibility that employers with a large share of minority workers are inferior employers in terms of pay or conditions of work that affect turnover. Note that 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. This raises what we view as important questions at the intersection of labor economics and urban economics. For example, we can ask what makes a neighborhood more "networked," in ways that increase the flow of labor market information and good matches among its residents. Are there policies, institutions, or other interventions that can increase this networking? Moreover, residence-based labor networks can help explain how ethnic and racial residential segregation reinforces poorer labor market outcomes for minorities – a long-standing question in urban economics – to the extent that minorities have weaker network connections to jobs held by whites. But residence-based networks may also present a two-edged sword, if they sometimes 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 whether it could reflect 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.

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.

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.

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.

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., Melissa McInerney, and David Neumark. 2010. "Spatial Mismatch, Immigrant Networks, and Hispanic Employment in the United States." *Annales d'Economie et de Statistique*, Vols. 99/100, July/December, pp. 141-67.

Hellerstein, Judith K., and David Neumark. 2011. "Employment in Black Urban Labor Markets: Problems and Solutions." NBER Working Paper No. 16986.

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.

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

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	102,830,000	NA
Job matches to establishment in Employer Characteristics File	100,469,000	2.3%
Employer is in private sector	85,620,000	14.8%
Employer is a single unit within state	52,921,000	38.2%
Employer location is precise to Census tract	51,394,000	2.9%
Worker linked to Individual Characteristics File	51,394,000	0.0%
Worker age is 18 to 64	47,666,000	7.3%
Worker linked to residence file	46,726,000	2.0%
Worker residence is precise to Census tract	40,999,000	12.3%
Worker has ≤ 25 jobs in last year	40,997,000	0.0%
At least two workers at employer	39,731,000	3.1%
Network analysis sample: At least two employers in workplace Census tract	39,722,000	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	38,109,000	4.1%
White not-Hispanic	27,745,000	NA
Black not-Hispanic	3,639,000	NA
Hispanic	4,684,000	NA
Asian not-Hispanic	1,800,000	NA
Not tabulated (not in subsamples or estimation)	242,000	NA
Tenure at job of three or fewer full years (< 4)	23,764,000	37.6%
Estimation sample for job retention model, without the "not tabulated" group	23,595,000	0.7%
Estimation sample for earnings model, with jobs lasting a full quarter	20,149,000	14.6%

Notes: The initial sample is the set of all jobs in the LEHD Employment 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 38 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 3. The estimation sample used in Table 4 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: 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.

Table 3: 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 4: 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)	0.694	0.809	0.381	0.430	0.410
Observed network isolation index (NI^O)	0.046	0.051	0.024	0.034	0.057
Employer average ($NI_E{}^O$)	0.048	0.054	0.024	0.034	0.056
Same group observed network isolation index ($NI_G{}^O$)	0.051	0.053	0.034	0.043	0.072
Same group, employer average ($NI_{GE}{}^O$)	0.053	0.056	0.034	0.045	0.074
Same commute (TI^O)	0.025	0.029	0.013	0.016	0.010
Same commute, same group ($TI_G{}^O$)	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 25 to 50	0.141	0.147	0.120	0.141	0.112
Employer size 50 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)	99.6	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 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 5: Percentiles of segregation measures

Sample	Variable (at time-of-hire)	Percentiles				
		10th	25th	50th	75th	90th
Pooled	Share of co-workers in same group (S^G)	0.194	0.491	0.790	0.949	1.000
	Observed network isolation index (NI^O)	0.000	0.000	0.000	0.023	0.121
	Employer average ($NI_E{}^O$)	0.000	0.003	0.012	0.039	0.110
	Same group observed network isolation index ($NI_G{}^O$)	0.000	0.000	0.000	0.023	0.135
	Same group, employer average ($NI_{GE}{}^O$)	0.000	0.001	0.012	0.042	0.123
	Same commute (TI^O)	0.000	0.000	0.004	0.019	0.065
	Same commute, same group ($TI_G{}^O$)	0.000	0.000	0.004	0.022	0.074
White	Share of co-workers in same group (S^G)	0.512	0.705	0.870	0.989	1.000
	Observed network isolation index (NI^O)	0.000	0.000	0.000	0.028	0.143
	Employer average ($NI_E{}^O$)	0.000	0.003	0.014	0.046	0.127
	Same group observed network isolation index ($NI_G{}^O$)	0.000	0.000	0.000	0.028	0.146
	Same group, employer average ($NI_{GE}{}^O$)	0.000	0.003	0.014	0.047	0.133
	Same commute (TI^O)	0.000	0.001	0.005	0.024	0.078
	Same commute, same group ($TI_G{}^O$)	0.000	0.001	0.005	0.025	0.081
Black	Share of co-workers in same group (S^G)	0.071	0.150	0.324	0.574	0.801
	Observed network isolation index (NI^O)	0.000	0.000	0.000	0.015	0.058
	Employer average ($NI_E{}^O$)	0.001	0.004	0.009	0.022	0.052
	Same group observed network isolation index ($NI_G{}^O$)	0.000	0.000	0.000	0.015	0.083
	Same group, employer average ($NI_{GE}{}^O$)	0.000	0.000	0.009	0.027	0.075
	Same commute (TI^O)	0.000	0.000	0.002	0.009	0.032
	Same commute, same group ($TI_G{}^O$)	0.000	0.000	0.003	0.018	0.055
Hispanic	Share of co-workers in same group (S^G)	0.073	0.174	0.379	0.658	0.886
	Observed network isolation index (NI^O)	0.000	0.000	0.000	0.016	0.081
	Employer average ($NI_E{}^O$)	0.000	0.003	0.010	0.027	0.072
	Same group observed network isolation index ($NI_G{}^O$)	0.000	0.000	0.000	0.010	0.108
	Same group, employer average ($NI_{GE}{}^O$)	0.000	0.000	0.008	0.031	0.094
	Same commute (TI^O)	0.000	0.000	0.002	0.010	0.037
	Same commute, same group ($TI_G{}^O$)	0.000	0.000	0.002	0.014	0.052
Asian	Share of co-workers in same group (S^G)	0.036	0.100	0.271	0.748	1.000
	Observed network isolation index (NI^O)	0.000	0.000	0.000	0.016	0.121
	Employer average ($NI_E{}^O$)	0.000	0.002	0.008	0.024	0.099
	Same group observed network isolation index ($NI_G{}^O$)	0.000	0.000	0.000	0.012	0.208
	Same group, employer average ($NI_{GE}{}^O$)	0.000	0.000	0.008	0.039	0.166
	Same commute (TI^O)	0.000	0.000	0.002	0.007	0.025
	Same commute, same group ($TI_G{}^O$)	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 6: 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)	(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)		(0.001)		(0.002)	(0.002)	(0.002)	(0.006)
Employer average ($NI_E{}^O$)		-0.076		-0.047	-0.022	-0.028	0.002
		(0.001)		(0.002)	(0.002)	(0.003)	(0.006)
Same commute (TI^O)		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 ($NI_G{}^O$)			(0.001)	(0.002)	(0.002)	(0.002)	(0.005)
Same-group, employer			-0.064	-0.041	-0.038	-0.043	-0.025
average ($NI_{GE}{}^O$)			(0.001)	(0.002)	(0.002)	(0.002)	(0.005)
Same commute, same			0.059	0.100	0.083	0.082	-0.006
group ($TI_G{}^O$)			(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 are 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, 25-49, 50 to 99, 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 7: Effect of network measures on job retention, blacks

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.051			-0.056	-0.343	-0.314	-0.071
in same group (S^G)	(0.004)			(0.004)	(0.004)	(0.005)	(0.002)
Observed network isolation		0.122		0.080	0.112	0.119	0.019
index (NI^O)		(0.006)		(0.008)	(0.007)	(0.007)	(0.018)
Employer average ($NI_E{}^O$)		-0.014		0.049	0.037	0.031	0.059
		(0.007)		(0.009)	(0.010)	(0.012)	(0.018)
Same commute (TI^O)		0.202		0.137	0.122	0.102	0.303
		(0.011)		(0.012)	(0.010)	(0.010)	(0.024)
Same-group observed network			0.099	0.059	0.057	0.055	0.059
isolation index ($NI_G{}^O$)			(0.004)	(0.005)	(0.005)	(0.005)	(0.012)
Same-group, employer			-0.036	-0.057	-0.041	-0.051	-0.082
average ($NI_{GE}{}^O$)			(0.005)	(0.006)	(0.006)	(0.007)	(0.012)
Same commute, same			0.125	0.056	0.056	0.062	-0.049
group ($TI_G{}^O$)			(0.005)	(0.005)	(0.004)	(0.004)	(0.012)
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)	9.7	9.7	9.7	9.7	9.7	9.7	9.7
R-squared (within for FE)	0.045	0.045	0.045	0.046	0.023	0.019	0.023
Fixed effects (millions)					0.5	1.2	5.7

Notes: Notes from Table 6 apply. There are no race and ethnicity dummy variables included. The sample includes those listed in the black subsample, pooled across the years 2004 to 2007.

Table 8: Effect of network measures on job retention, Hispanics

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.013			-0.019	-0.267	-0.267	-0.100
in same group (S^G)	(0.003)			(0.003)	(0.003)	(0.003)	(0.002)
Observed network isolation		0.250		0.137	0.110	0.110	0.046
index (NI^O)		(0.004)		(0.005)	(0.005)	(0.005)	(0.013)
Employer average ($NI_E{}^O$)		-0.103		-0.023	0.032	0.032	0.082
		(0.004)		(0.005)	(0.007)	(0.007)	(0.013)
Same commute (TI^O)		-0.021		-0.097	0.029	0.029	0.246
		(0.008)		(0.009)	(0.011)	(0.011)	(0.019)
Same-group observed network			0.191	0.118	0.123	0.123	0.047
isolation index ($NI_G{}^O$)			(0.003)	(0.004)	(0.004)	(0.004)	(0.011)
Same-group, employer			-0.090	-0.079	-0.102	-0.102	-0.049
average ($NI_{GE}{}^O$)			(0.003)	(0.004)	(0.005)	(0.005)	(0.010)
Same commute, same			0.051	0.072	0.063	0.063	-0.066
group ($TI_G{}^O$)			(0.004)	(0.004)	(0.004)	(0.004)	(0.012)
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)	11.9	11.9	11.9	11.9	11.9	11.9	11.9
R-squared (within for FE)	0.025	0.025	0.027	0.027	0.019	0.015	0.016
Fixed effects (millions)					0.9	2.0	6.5

Notes: Notes from Table 6 apply. There are no race and ethnicity dummy variables included. The sample includes those listed in the Hispanic subsample, pooled across the years 2004 to 2007.

Table 9: Effect of network measures on job retention, Asians

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.005			-0.017	-0.174	-0.134	-0.072
in same group (S^G)	(0.003)			(0.003)	(0.004)	(0.004)	(0.004)
Observed network isolation		0.299		0.200	0.162	0.161	0.139
index (NI^O)		(0.006)		(0.008)	(0.009)	(0.009)	(0.024)
Employer average ($NI_E{}^O$)		-0.166		-0.134	-0.068	-0.065	-0.014
		(0.005)		(0.008)	(0.011)	(0.011)	(0.022)
Same commute (TI^O)		0.168		0.105	0.169	0.165	0.488
		(0.024)		(0.025)	(0.023)	(0.020)	(0.048)
Same-group observed network			0.203	0.070	0.115	0.108	0.025
isolation index ($NI_G{}^O$)			(0.004)	(0.006)	(0.006)	(0.007)	(0.020)
Same-group, employer			-0.076	0.004	-0.077	-0.100	0.010
average ($NI_{GE}{}^O$)			(0.004)	(0.006)	(0.007)	(0.008)	(0.018)
Same commute, same			0.051	0.024	0.004	0.013	-0.114
group ($TI_G{}^O$)			(0.007)	(0.007)	(0.006)	(0.006)	(0.014)
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)	4.5	4.5	4.5	4.5	4.5	4.5	4.5
R-squared (within for FE)	0.027	0.031	0.031	0.032	0.019	0.015	0.016
Fixed effects (millions)					0.4	0.9	2.4

Notes: Notes from Table 6 apply. There are no race and ethnicity dummy variables included. The sample includes those listed in the Asian subsample, pooled across the years 2004 to 2007.

Table 10: 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)	(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)		(0.004)		(0.006)	(0.004)	(0.004)	(0.009)
Employer average ($NI_E{}^O$)		-0.743		-0.606	0.051	0.101	-0.263
		(0.004)		(0.006)	(0.005)	(0.005)	(0.009)
Same commute (TI^O)		-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 ($NI_G{}^O$)			(0.003)	(0.005)	(0.004)	(0.004)	(0.008)
Same-group, employer			-0.658	-0.093	0.058	0.045	0.004
average ($NI_{GE}{}^O$)			(0.004)	(0.005)	(0.004)	(0.004)	(0.008)
Same commute, same			-0.924	-0.283	0.028	0.077	0.012
group ($TI_G{}^O$)			(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 listed in the race/ethnicity subsamples, pooled across the years 2004 to 2007. Standard errors are 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, 25-49, 50 to 99, 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 11: Effect of network measures on log earnings, blacks

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.532			-0.520	-0.207	-0.419	-0.211
in same group (S^G)	(0.010)			(0.010)	(0.007)	(0.009)	(0.004)
Observed network isolation		-0.013		0.562	0.094	0.107	0.044
index (NI^O)		(0.014)		(0.018)	(0.014)	(0.016)	(0.029)
Employer average ($NI_E{}^O$)		-0.878		-0.760	0.049	0.120	-0.252
		(0.018)		(0.021)	(0.021)	(0.025)	(0.030)
Same commute (TI^O)		-0.881		-0.333	0.031	0.036	-0.1448
		(0.030)		(0.033)	(0.023)	(0.024)	(0.037)
Same-group observed network			0.055	-0.210	-0.196	-0.203	-0.032
isolation index ($NI_G{}^O$)			(0.009)	(0.010)	(0.009)	(0.010)	(0.019)
Same-group, employer			-0.563	-0.146	0.102	0.116	0.002
average ($NI_{GE}{}^O$)			(0.011)	(0.012)	(0.010)	(0.012)	(0.019)
Same commute, same			-0.612	-0.538	-0.119	-0.085	0.021
group ($TI_G{}^O$)			(0.016)	(0.016)	(0.015)	(0.017)	(0.016)
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)	8.0	8.0	8.0	8.0	8.0	8.0	8.0
R-squared (within for FE)	0.160	0.142	0.142	0.167	0.062	0.052	0.057
Fixed effects (millions)					0.5	1.2	4.7

Notes: Notes from Table 10 apply. There are no race and ethnicity dummy variables included. The sample includes those listed in the black subsample, pooled across the years 2004 to 2007.

Table 12: Effect of network measures on log earnings, Hispanics

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.443			-0.431	-0.160	-0.376	-0.130
in same group (S^G)	(0.006)			(0.006)	(0.005)	(0.005)	(0.003)
Observed network isolation		0.040		0.433	0.103	0.123	0.035
index (NI^O)		(0.010)		(0.013)	(0.009)	(0.009)	(0.019)
Employer average ($NI_E{}^O$)		-0.548		-0.474	0.013	0.080	-0.234
		(0.013)		(0.015)	(0.011)	(0.013)	(0.018)
Same commute (TI^O)		-0.783		-0.485	-0.008	-0.006	-0.295
		(0.032)		(0.035)	(0.014)	(0.015)	(0.028)
Same-group observed network			0.032	-0.206	-0.181	-0.174	-0.015
isolation index ($NI_G{}^O$)			(0.006)	(0.007)	(0.006)	(0.007)	(0.015)
Same-group, employer			-0.389	-0.049	0.077	0.068	-0.008
average ($NI_{GE}{}^O$)			(0.008)	(0.008)	(0.007)	(0.008)	(0.014)
Same commute, same			-0.525	-0.368	-0.093	-0.067	0.015
group ($TI_G{}^O$)			(0.015)	(0.010)	(0.007)	(0.008)	(0.016)
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)	10.1	10.1	10.1	10.1	10.1	10.1	10.1
R-squared (within for FE)	0.160	0.142	0.141	0.166	0.076	0.064	0.072
Fixed effects (millions)					0.8	1.9	5.7

Notes: Notes from Table 10 apply. There are no race and ethnicity dummy variables included. The sample includes those listed in the Hispanic subsample, pooled across the years 2004 to 2007.

Table 13: Effect of network measures on log earnings, Asians

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of co-workers	-0.424			-0.406	-0.129	-0.273	-0.150
in same group (S^G)	(0.014)			(0.015)	(0.007)	(0.008)	(0.005)
Observed network isolation		0.431		0.577	0.206	0.197	0.072
index (NI^O)		(0.013)		(0.022)	(0.016)	(0.017)	(0.030)
Employer average ($NI_E{}^O$)		-0.830		-0.589	-0.086	0.040	-0.211
		(0.017)		(0.025)	(0.023)	(0.020)	(0.029)
Same commute (TI^O)		-1.612		-1.364	0.008	0.017	-0.544
		(0.077)		(0.085)	(0.045)	(0.041)	(0.058)
Same-group observed network			0.273	-0.039	0.022	0.034	-0.027
isolation index ($NI_G{}^O$)			(0.010)	(0.016)	(0.013)	(0.014)	(0.025)
Same-group, employer			-0.565	-0.161	-0.144	-0.178	-0.052
average ($NI_{GE}{}^O$)			(0.012)	(0.017)	(0.015)	(0.016)	(0.024)
Same commute, same			-0.358	-0.278	-0.026	0.008	-0.034
group ($TI_G{}^O$)			(0.018)	(0.021)	(0.015)	(0.016)	(0.016)
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)	4.0	4.0	4.0	4.0	4.0	4.0	4.0
R-squared (within for FE)	0.228	0.220	0.218	0.234	0.100	0.092	0.089
Fixed effects (millions)					0.4	0.9	2.2

Notes: Notes from Table 10 apply. There are no race and ethnicity dummy variables included. The sample includes those listed in the Asian subsample, pooled across the years 2004 to 2007.

Appendix Table A1: 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.