# Neighbors and Coworkers: The Importance of Residential Labor Market Networks

Judith K. Hellerstein, *University of Maryland and NBER*

Melissa McInerney, *College of William and Mary*

David Neumark, *University of California, Irvine; NBER; and IZA*

We specify and implement a test for the presence and importance of labor market networks based on residential proximity, in determining the establishments at which people work. Using matched employer-employee data at the establishment level, we measure the importance of these network effects for groups broken out by race, ethnicity, and measures of skill. The evidence indicates that these types of labor market networks do exist and play an important role in determining the establishments where workers work; that they are more important for minorities and the less skilled, especially among Hispanics; and that they appear to be race based.

## I. Introduction

Racial and ethnic disparities in labor market outcomes in the United States are well documented. In addition to differences in wages and em-

This research was supported by National Institute of Child Health and Human Development grant R01HD042806. We are grateful for the comments of Joel Elvery, Mark Granovetter, Lisa Hellerstein, Jed Kolko, Andrew Noymer, Michael Ransom, Cliff Stein, Chris Taber, Etienne Wasmer, and participants in "Social Networks and Peer Effects: Theory and Applications; A Conference in Memory of Antoni Calvó-Armengol," the Western Economics Association meetings, the

[*Journal of Labor Economics,* 2011, vol. 29, no. 4]

0734-306X/2011/2904-0001$10.00

ployment levels, there is evidence of workplace segregation by race and ethnicity (e.g., Hellerstein and Neumark 2008). There is also strong evidence of residential segregation by race and ethnicity in the United States (e.g., Iceland and Weinberg 2002), and this segregation is correlated with poor labor market outcomes for minority groups, especially for the low skilled among them (e.g., Cutler and Glaeser 1997).

Theoretical models of labor market networks can formalize the link between residential segregation and labor market outcomes such as workplace segregation and disparities in wages and employment, when networks are partially or fully described by links between residential neighbors. Underlying all network models is some form of information imperfection in which networks serve at least partially to mitigate these imperfections. In the model developed by Montgomery (1991), which most directly motivates our analysis, the information imperfection is on the employer side. Firms with vacancies cannot observe the underlying ability of a potential worker, but firms can infer something about a potential worker's ability if (and only if) the firm currently employs individuals from that worker's social network, where social networks are at least partially stratified by ability.[1] Hence, networks act at the establishment level to reduce employer search costs. In equilibrium, individuals are more likely to receive and accept wage offers from firms that employ others in their social network. In this framework, if social networks are at least partially race or ethnic based—potentially as the result of residential segregation—and white workers are initially employed at higher rates than other groups, then the existence of a larger network of white workers will lead to more job referrals at high wages for whites searching for jobs, creating wage disparities between whites and other groups. Although Montgomery's model does not build in a reservation wage, having an option for remaining out of the labor market would, in his framework, lead to employment differentials across groups as well.

Our goal in this article is to provide evidence on the existence and

All-UC Labor Conference, the Society of Labor Economists annual meetings, and seminars at Équipe de Recherche sur les Marchés; Georgetown Public Policy Institute; Harvard University; the Institute for the Study of Labor (IZA); the National Bureau of Economic Research Summer Institute; New York University; University of California, Irvine; University of Maryland; University of Southern California; and Washington University. We thank Kyle Handley for outstanding research assistance. The research in this article was conducted while Melissa McInerney was a Special Sworn Status researcher of the U.S. Census Bureau and the Census Bureau Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau. This article has been screened to ensure that no confidential data are revealed. Contact the corresponding author, David Neumark, at dneumark@uci.edu.

[1] See also Simon and Warner (1992) and Calvó-Armengol and Jackson (2007).

importance of labor market networks in determining the assignment of workers to establishments, on the basis of a large-scale data set covering most of the U.S. economy. Any study of networks has to specify the channels along which network connections may flow. We study networks connecting neighbors. Labor market network connections among neighbors are plausible, as neighbors may interact in a variety of ways, and existing evidence suggests that labor market connections among neighbors may be important (Bayer, Ross, and Topa 2008). We focus in particular on the extent to which these network connections among neighbors lead to their employment at the same establishment. Our measure of this particular type of labor market network likely underestimates the overall importance of networks—in that networks may be based on more than just residence and may help workers find employment in more than one specific establishment—and these types of networks may be more important for labor markets that are more local.

We specify and implement a test 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. Our 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. We measure the importance of network effects for groups broken out by race (black and white), Hispanic ethnicity, and various measures of skill (education, English language proficiency, and immigrant status). Finally, we provide evidence on the stratification of networks, asking whether the networks we study are race based, operating more strongly within than across races.

The workplace segregation of employees across establishments by residential neighborhood that we measure arises from residential proximity that captures the "network connectedness" important to the flow of labor market information among specific employers, their employees, and potential hires. In particular, we first identify all establishments within each census tract in our sample. Because we have matched employer-employee data, we have a sample of workers in each establishment, and we know the census tracts in which they live. We compute the share of an individual's coworkers who are his or her residential neighbors, relative to the share that would result if the establishment hired workers randomly from the geographic areas where all individuals who work in the census tract reside. Residence-based networks would predict that the share of neighbors among a worker's coworkers would be higher—and possibly much higher—than would result from the random hiring process.

While random hiring represents a reasonable lower-bound baseline for the sorting of workers by neighborhoods across establishments, we also consider what the upper bound would be. In particular, if establishments are larger than networks, perfect sorting by residence-based networks across establishments cannot occur. We therefore operationalize our measure of the importance of residential labor market networks by calculating what fraction of the difference between the lower bound and the upper bound of the extent to which a worker can work with neighbors is actually observed in the data. Because we measure the fraction of the difference between the lower bound and the maximum possible sorting that could occur, the magnitude of our network measure can be compared across various subpopulations. We also consider influences other than networks that could give rise to the observed patterns of sorting across establishments, by residential location, that we observe in the data.

The data we use for this study come from the 2000 Decennial Employer-Employee Database (DEED), a large data set we have constructed consisting of workers matched to their establishment of employment. The strength of these employer-employee matches is that they enable us to study directly whether workers employed in the same establishment are likely to live in the same neighborhoods.

Overall, we find that residence-based labor market networks play an important role in hiring. For whites, we find that there is "excess" grouping of workers from the same neighborhoods in the same business establishments, with about a 4.9 percentage-point gap between the observed grouping and what a random grouping would yield, which translates into about 10% of the maximum grouping that could occur. These networks are largely due to the assignment of workers to specific establishments, rather than the sorting of workers in local labor markets by skill groups, occupations, or industries. It is the case, though, that networks are more important for less skilled whites and in small establishments. We find similar results for blacks, and some indications that residence-based labor market networks are more important for blacks than for whites. Moreover, the evidence indicates that the networks we study are partly race based, operating more strongly within than across races. We find that residence-based networks are even more significant for Hispanics, for whom the grouping of workers from the same neighborhoods in the same business establishments is about 8.1 percentage points higher than what would occur randomly, or about 22% of the maximum. And among Hispanics, these networks play a larger role for immigrants and those with poor English skills.

Like any measure that tries to capture job networks, ours is limited to measuring networks that operate among particular members, affecting employment in a specific set of jobs—in our case, the extent to which census tract coresidents working in a specific census tract of employment also work in the same establishment within that tract. To the extent that networks also increase the flow of information about jobs near the employers of network members (or jobs in other places entirely), and to the extent that networks connect people who live in different census tracts, we will understate the importance of labor market networks. At the same time, the role of residence-based labor market networks is significant in its own right, in regard to how spatially based labor market policies might either take advantage of or, instead, inadvertently weaken or sever valuable network connections between neighbors.

## II. Relation to Existing Literature

Ioannides and Datcher Loury (2004) review evidence on labor market networks, emphasizing survey findings documenting widespread reliance on friends, relatives, and acquaintances to search for and find jobs. They conclude that blacks and whites use informal contacts at similar rates but that less educated workers and especially Hispanics use them much more heavily. Subsequent work suggests 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).[2] Bayer et al. (2008) test for network effects among neighbors using confidential long-form census data on Boston-area workers. They find that two individuals living on the same census block are more likely to work on the same census block than are two individuals living in the same block group but not on the same block. If networks are stronger within blocks than within block groups, this evidence is consistent with residence-based labor market networks, the same type of network connections that we study.

There are, however, some limitations of the existing evidence. First, most existing work does not relate employment in the same business to network connections between employees of that business. This is certainly true of most evidence based on surveys of workers, and the data used by Bayer et al. (2008) contain no information on the exact establishment in which the workers work, so that two individuals who work on the same census block may work for different employers.[3] Granovetter's (1974) evidence that informal contacts were often employed in the company where the job held by a respondent had opened up is an exception. With our data, we can tie network connections to employment in a particular business establishment and, thus, provide evidence more directly related

[2] Kasinitz and Rosenberg (1996) provide case study evidence consistent with this type of network.

[3] This is particularly true for blocks in the central city. For example, focusing on central city areas in our sample, there are on average 2.58 establishments per block, which is likely an undercount by a factor of about four, given that we only observe a subsample of establishments. This raises questions about Bayer et al.'s assumption that workers employed on the same block "work with" one another or "work together" (e.g., 2008, 26).

to the hypothesis that labor market networks reduce search frictions on the part of employers, as in Montgomery's (1991) model.[4]

Second, data limitations preclude previous work from quantifying differences in the importance of networks by race and Hispanic ethnicity and among subgroups of Hispanics.[5] For Hispanics, networks may compensate for less developed formal hiring channels, especially in immigrant communities and for undocumented workers. Our large data set and its explicit linkage of workers to establishments lets us explore the importance of networks that assign workers to establishments and differentials in the importance of networks by race and Hispanic ethnicity and for subgroups of Hispanics. We can also assess the extent to which networks are race based—operating more strongly within than across races.

Our study does share two limitations with previous work, however. First, because the data contain no direct measures of network contacts, such as information on whether an individual found work directly via a network connection in his or her neighborhood, our measure of networks remains indirect.[6] Second, the data contain no direct information on how or why the networks of neighbors that we measure are formed or what makes some networks stronger than others. We can provide some suggestive evidence along these lines, but our results primarily should be seen as providing descriptive evidence on the strength of networks at the level of residential neighborhoods.

## III. The 2000 DEED

The analysis in this article is based on matched employer-employee data from the 2000 DEED. In this section, we provide only a brief overview of the DEED.[7] Workers in the DEED are drawn from the Sample Edited Detail File (SEDF), which contains all individual responses to the 2000 Decennial Census of Population one-in-six long form. The establishments are drawn from the Census Bureau's Business Register (BR) for 2000, a continuously updated database containing information for all business

[4] To the extent that networks reduce search frictions (as in the model), they enhance efficiency, although if networks simply act to allocate "good jobs" (Schmutte 2010) they play only a distributional role.

[5] For example, Bayer et al. (2008) study data only for Boston and focus mostly on whites. And although some of the survey evidence discussed in Ioannides and Datcher Loury (2004) suggests differential use of networks across racial and ethnic groups, a good deal of the most cited evidence covers narrow subsets of workers or a single firm.

[6] In contrast, Laschever (2009) studies the employment experiences of veterans who served together in World War I, and Cingano and Rosolia (2009) study the reemployment experiences of workers displaced from the same firm.

[7] We have constructed a similar data set for 1990, described in detail elsewhere (Hellerstein and Neumark 2003). The construction of the 2000 DEED follows the same procedures.

establishments (with one or more employees) operating in the United States in each year (Jarmin and Miranda 2002). The BR contains the name and address of each establishment, geographic codes based on its location, its four-digit standard industrial classification code, and a unique establishment identifier.

Long-form respondents were asked to report the name and address of the employer in the previous week for each employed household member, which is stored in the "write-in" file. We use employer names and addresses in the write-in file to try to match each worker to the BR. Finally, because both the write-in file and the SEDF contain identical individual identifiers, the two can be linked, yielding a very large data set with workers matched to their establishments, along with all of the information on workers from the SEDF.

Because employers' names and addresses can be recorded differently on the two files, we match workers and establishments on the basis of the write-in file using MatchWare—a specialized record-linkage program. MatchWare includes a name and address standardization mechanism (AutoStan) and a matching system (AutoMatch). We first use AutoStan to standardize how data items in employer names and addresses are reported in the write-in file and the BR. Once the names and addresses are standardized, each item is parsed into components, and we match on various combinations of these components, using AutoMatch's probabilistic matching algorithm to account for missing information, misspellings, and even inaccurate information. The AutoMatch software permits choices about which matching variables to use, how heavily to weight each matching variable, and how similar two addresses must be in order to constitute a match. Our strategy was to be cautious in accepting linked record pairs. We chose matching algorithms based on substantial experimentation and visual inspection of many thousands of records to avoid erroneous links.

The final 2000 DEED is an extremely large data set containing information on 4.09 million workers matched to 1.28 million establishments, accounting for 29.1% of workers in the SEDF and 22.6% of establishments in the BR. We impose additional sample restrictions for our analysis, which we discuss after the explanation of our empirical methods in the next section.

## IV. Measuring the Importance of Networks

We first study the importance of residence-based networks for whites. We perform a series of analyses of whites to examine alternative interpretations of the findings and to identify factors associated with variation in the importance of these networks. By focusing first on whites, we avoid confounding our results with other potential influences on the sorting of workers into firms by race and ethnicity, such as labor market discrimination. We then turn to evidence on networks for these other groups. In explaining our method, we refer to whites, for simplicity; the analysis is done in the same way for the other groups.

For the sample of whites, we first compute for each worker the percentage of (white) coworkers (in the same establishment) who live in the same residential neighborhood as that worker.[8] This requires a sample restricted 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 coworkers who are observed in our estimation sample to be residential neighbors:[9]

$$\mathrm{NI}^{\mathrm{O}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j\neq i}\mathrm{I}^{\mathrm{R}}(i,j)\times\mathrm{I}^{\mathrm{W}}(i,j)}{\sum_{j\neq i}\mathrm{I}^{\mathrm{W}}(i,j)}\times 100,$$

where there are $N$ workers, indexed by $i$ and $j$ for all possible pairs of workers; $\mathrm{I}^{\mathrm{R}}(i, j)$ is an indicator for whether workers $i$ and $j$ live in the same residential neighborhood; and $\mathrm{I}^{\mathrm{W}}(i, j)$ 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 coworkers with whom each worker is coresident. This ratio is then averaged over all workers and multiplied by 100.

We define residential neighborhoods as census tracts for a few reasons. First, census tracts define the boundaries that are traditionally used to measure residential segregation (e.g., Iceland and Weinberg 2002). Second (and related to the first), census tracts are defined by the U.S. Census Bureau to ensure that the tracts are "as homogeneous as possible with respect to population characteristics, economic status, and living condi-

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

[9] The phrase "network isolation index" borrows from the sociology literature, in which in the context of residential segregation the "isolation index" is the fraction of a group's residential neighbors who are themselves black. "Segregation" and "isolation" have the same meaning—that members of a particular group tend to interact with other members of the same group, rather than members of other groups. Of course, many other segregation measures have been used to measure the agglomeration of individuals of similar (usually two) types together in society. Measuring network segregation via the simple concept of isolation is transparent and compelling in our view. In particular, we think it is preferable to the commonly used alternative Duncan Index for two reasons. First, we would have had to use a multidimensional version of the Duncan index, which is much less widely used in measuring segregation. Second, Duncan indexes are invariant to growing populations of a given group. For example, in measuring black-white residential segregation across census tracts, the Duncan index is invariant to a doubling of the black population in each census tract. It is not clear that invariance is a desirable property in the context of our study. See Hellerstein and Neumark (2008) for more details.

tions" (U.S. Department of Commerce 1994, 10-1), which is a reasonable definition of a neighborhood and makes it more likely that coresidents interact. Third, most census tracts are relatively small, so many census tract residents have contact with one another, if not "over the back fence," then at parks, schools, churches, stores, businesses, and other institutions.[10]

A number of factors unrelated to residentially based labor market networks could lead to residential neighbors working in the same establishment, and we want to preclude these factors from influencing our measurement of labor market networks. First, access to mass transportation may lead residential neighbors to work in the same geographic place, although not necessarily to work in the same establishment. Because census tracts are generally small enough that, particularly within urban areas, individuals can walk from any establishment to any other establishment, transportation infrastructure has little effect on the distribution of workers across establishments in the tract. Second, heterogeneous tastes may imply that those who choose to work in similar workplaces have similar preferences over neighborhoods. We would expect individuals with greater economic resources to be better able to choose where to live and where to work. However, our disaggregated analyses across groups that a priori have varying abilities to make choices about where to live and work show that those who have more resources, on average, are least likely to work with their neighbors. Third, groups that suffer from labor market discrimination will be sorted into jobs with nondiscriminatory employers. Because of the potential for discrimination against blacks and Hispanics, in particular, we focus much of our analysis on white non-Hispanics before examining the extent of networks for these other groups.

Finally, some coresidents would work together, even if workers are assigned randomly to establishments, especially given that people tend to work relatively close to where they live. As a result, 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 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

[10] To document the size of census tracts, we did computations for Chicago, using information at http://www.census.gov/geo/www/cob/tr2000.html. For the primary metropolitan statistical area (PMSA), the median (mean) census tract was 0.57 (2.75) square miles. The smallest census tract was 0.02 square miles, and the largest (in the most outlying areas of the PMSA) was 151 square miles. For the city, the median (mean) was 0.17 (0.26) square miles. The minimum was again 0.02 square miles, and the maximum was 8 square miles. There are only two large tracts within the city limits—O'Hare Field and an industrial tract at the southern edge of the city.

in the same racial, ethnic, or skill group for which we are trying to characterize the importance of networks. We do the simulation 100 times, and $NI^R$ is the mean over these simulations.[11] 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, and so on. 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.

The difference $NI^O - NI^R$ measures network isolation above and beyond that which occurs randomly, which we refer to as the "network isolation difference." This difference captures the excess presence of coresidents in a worker's own establishment relative to the presence of coresidents in the same work location. This computation requires a second sample restriction—that census tracts of employment include at least two establishments having at least two matched 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 lower bound of the extent to which workers work with neighbors,[12] it is also important to know what the upper bound of network isolation could be in our data. Because establishments within a particular census tract of employment often contain more workers than the number of workers from any particular census tract of residence represented among those establishments, the upper bound of network isolation is unlikely to ever reach 100% (in which case workers would work with their neighbors only). 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. For a given census tract of employment, we order the neighborhoods in which workers live by the number of workers from each neighborhood. Beginning with the neighborhood with the greatest number of workers, we assign as many workers as possible to one establishment, and any workers who are not assigned to that establishment are grouped together and treated as a "new" neighborhood. We then move to the second largest

[11] In our very large sample, the random network isolation measures had trivially small standard deviations.

[12] This is not the true lower bound, of course. A social planner could force the network isolation index to go below $NI^R$ by distributing neighbors across establishments to minimize neighbor-to-neighbor contact in the workplace.

neighborhood from which workers originate (which could be the "new" neighborhood left from the previous pass) and assign workers from that neighborhood to the establishment that holds the maximum number of these, again keeping neighbors working together in establishments as much as possible.[13] We continue moving down the list of neighborhoods, assigning workers to establishments until all workers are assigned.[14] We do this for every census tract in our sample where workers work, and then we compute the weighted average of the maximum network isolation in each census tract of employment, weighting by the number of workers in that census tract and denoting this maximum network isolation index as $NI^M$.

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

$$\frac{NI^O - NI^R}{NI^M - NI^R} \times 100,$$

which we call the "effective network isolation index." It measures the share of the maximum possible network isolation that could occur in the data that actually does occur in the data.[15] As such, it provides a natural

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

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

[15] Our analysis has some parallels to Oyer and Schaefer's (2009) study of the concentration of lawyers from particular schools in different law firms, which may be due to networks. Like us, they consider deviations of the observed concentration of students from particular law schools relative to what would occur randomly (for the same reason), although measured using the Ellison and Glaeser (1997) index of concentration. There are some differences, however. First, we randomly assign workers to establishments within the census tract in which they work, rather than nationally. Second, we preserve the size distribution of establishments because otherwise changes in this size distribution could themselves affect the network isolation measures; we do not know how this would translate

scaling for the importance of networks formed by residential connections in determining the establishments in which people are employed, and because it is a scaled measure, it can be compared across different samples with different possible lower and upper bounds of network isolation. For these reasons, the effective network isolation index is our preferred measure of network isolation. Nonetheless, we also report and discuss results for $NI^O$, $NI^R$, $NI^O - NI^R$, and $NI^M$ separately.

Bootstrap methods are used to assess the statistical significance of our estimates of the importance of residence-based networks, of differences across demographic groups, and of different analyses (with and without conditioning), and so on.[16] The effective network isolation measures we report are always strongly statistically significantly different from zero and are estimated very precisely. The notes to the tables report the statistical significance of the important comparisons across estimates; because the estimates of the effective network indexes are quite precise, almost all differences across columns and tables in the estimated effective network isolation measures are also statistically significant.

## V. Sample Characteristics and Restrictions

Table 1 reports descriptive statistics for the matched workers from the SEDF and from each of our DEED analysis samples. The individuals we successfully match in the DEED are more likely to be female, to work full time, and to have more education than those in the SEDF, differences that stem in part from variation in the quality of the business address information in the write-in file. Because there appears to be lower-quality business address information for less skilled workers with weaker labor market attachment, our estimates might best be interpreted as measuring the importance of networks among workers who have relatively high attachment to the labor force and to their employers. Table 1 shows that there is some overrepresentation of workers in manufacturing in the full DEED because larger establishments are more likely to be matched, although the overrepresentation is not severe.

Table 2 shows descriptive statistics for establishments in the full DEED and for each of our final analysis samples. Because only one in six workers is sent the long form, it is more likely that large establishments will have two matched workers, especially for smaller racial or ethnic groups; this

into the Ellison-Glaeser framework. And third, we can define a maximum amount of network isolation that could occur, which (as just described) we take account of by rescaling.

[16] Confidence intervals are constructed from replications in which we bootstrap the entire sample and then compute each of the measures in the tables that follow. (For each bootstrap sample, we recomputed $NI^O$ and $NI^R$ for each analysis in which an individual appears.) Thus, we obtain bootstrap replications of the differences between any pair of effective isolation measures within or across tables.

**Table 1**
**Individual-Level Means for Preliminary and Final Analysis Samples**

| | 2000 SEDF, Workers Eligible to Be Matched | 2000 Full DEED | 2000 DEED, Final Analysis Samples | | |
|---|---|---|---|---|---|
| | | | White | Black | Hispanic |
| Age | 39.18 | 39.60 | 40.14 | 38.32 | 35.65 |
| | (13.07) | (12.54) | (12.37) | (11.49) | (11.48) |
| Female | .46 | .50 | .50 | .63 | .46 |
| Married | .58 | .61 | .62 | .42 | .57 |
| White | .81 | .86 | 1.0 | . . . | . . . |
| Black | .09 | .06 | . . . | 1.0 | . . . |
| Hispanic | .09 | .07 | . . . | . . . | 1.0 |
| Full time | .78 | .82 | .83 | .84 | .83 |
| No. kids (if female) | .77 | .76 | .69 | .87 | 1.12 |
| | (1.06) | (1.04) | (.99) | (1.12) | (1.22) |
| High school diploma | .31 | .29 | .26 | .28 | .25 |
| Some college | .33 | .36 | .36 | .41 | .28 |
| BA | .15 | .17 | .22 | .14 | .07 |
| Advanced degree | .06 | .07 | .09 | .05 | .03 |
| Speaks English well | .97 | .92 | 1.00 | .99 | .80 |
| Immigrant | .08 | .06 | .03 | .09 | .49 |
| Log(hourly wage) | 2.54 | 2.62 | 2.74 | 2.56 | 2.41 |
| | (.73) | (.69) | (.69) | (.66) | (.64) |
| Hours worked in 1999 | 40.22 | 40.71 | 41.04 | 40.39 | 40.58 |
| | (11.73) | (11.08) | (10.98) | (9.56) | (9.39) |
| Weeks worked in 1999 | 47.28 | 48.43 | 48.95 | 47.64 | 47.01 |
| | (10.53) | (9.22) | (8.55) | (10.15) | (10.54) |
| Earnings in 1999 | 33,444 | 37,091 | 42,669 | 31,090 | 26,682 |
| | (42,952) | (47,220) | (53,413) | (31,108) | (29,589) |
| Industry: | | | | | |
| Mining | .006 | .004 | .003 | .001 | .002 |
| Construction | .081 | .048 | .041 | .007 | .040 |
| Manufacturing | .207 | .257 | .266 | .242 | .353 |
| Transportation | .075 | .052 | .053 | .074 | .052 |
| Wholesale | .047 | .052 | .054 | .025 | .050 |
| Retail | .210 | .212 | .195 | .146 | .212 |
| FIRE | .070 | .068 | .072 | .079 | .043 |
| Services | .304 | .306 | .316 | .425 | .249 |
| *N* | 13,456,402 | 3,924,714 | 1,675,412 | 97,967 | 110,235 |

NOTE.—Decennial Employer-Employee Database (DEED) and Sample Edited Detail File (SEDF) samples exclude individuals who have missing wages, who did not work in the year before the survey year or in the reference week for the long form of the census, who did not report positive hourly wages, who did not work in one of the 50 states or the District of Columbia (whether or not the place of work was imputed), who were self-employed, who were not classified in a state of residence, or who were employed in an industry that was considered "out-of-scope" in the Census Bureau's Business Register (BR). (Out-of-scope industries do not fall under the purview of Census Bureau surveys. They include many agricultural industries, urban transit, the U.S. Postal Service, private households, schools and universities, labor unions, religious and membership organizations, and government/public administration. The Census Bureau does not validate the quality of BR data for businesses in out-of-scope industries.) In addition to restricting data by race and ethnicity, the three restrictions imposed in going from the full DEED to the final analysis samples are that the individual must live and work in same metropolitan statistical area/primary metropolitan statistical area, there must be at least two workers matched to the individual's establishment, and there must be at least one other establishment with two matched workers in the census tract. Standard deviations are in parentheses; FIRE = finance, insurance, and real estate.

**Table 2**
**Establishment-Level Descriptive Statistics for Preliminary and Final Analysis Samples**

| | 2000 Full DEED | 2000 DEED, Final Analysis Samples | | |
|---|---|---|---|---|
| | | White | Black | Hispanic |
| Total employment | 49.82 | 102.82 | 412.43 | 258.09 |
| | (368.46) | (344.85) | (887.34) | (670.16) |
| Total employment (approximate median) | 15 | 35 | 154 | 84 |
| Establishment size: | | | | |
| 1–25 | .65 | .39 | .11 | .18 |
| 26–50 | .15 | .20 | .11 | .16 |
| 51–100 | .10 | .17 | .14 | .20 |
| 101+ | .10 | .22 | .62 | .45 |
| Industry: | | | | |
| Mining | .004 | .003 | .002 | .003 |
| Construction | .078 | .070 | .013 | .053 |
| Manufacturing | .133 | .186 | .226 | .310 |
| Transportation | .050 | .052 | .077 | .051 |
| Wholesale | .067 | .074 | .039 | .060 |
| Retail | .284 | .265 | .231 | .266 |
| FIRE | .081 | .077 | .082 | .049 |
| Services | .303 | .272 | .331 | .209 |
| In MSA/PMSA | .792 | 1.0 | 1.0 | 1.0 |
| Census region: | | | | |
| Northeast | .053 | .048 | .013 | .015 |
| Mid-Atlantic | .135 | .148 | .122 | .084 |
| East North Central | .199 | .232 | .204 | .088 |
| West North Central | .092 | .089 | .039 | .010 |
| South Atlantic | .166 | .157 | .332 | .052 |
| East South Central | .050 | .043 | .081 | .002 |
| West South Central | .102 | .090 | .136 | .218 |
| Mountain | .061 | .057 | .012 | .084 |
| Pacific | .142 | .135 | .061 | .448 |
| Payroll ($1,000) | 2,103 | 5,303 | 19,061 | 11,905 |
| | (146,515) | (281,585) | (67,785) | (56,649) |
| Payroll / total employment | 37.14 | 47.69 | 37.63 | 35.16 |
| | (2,285) | (2,716) | (50.53) | (77.83) |
| Share of employees matched | .16 | .14 | .05 | .07 |
| Multiunit establishment | .40 | .51 | .80 | .61 |
| *N* | 1,254,718 | 329,943 | 21,872 | 30,343 |

NOTE.—See note to table 1. The approximate median is an average of the median and some observations on either side of the median, to preserve confidentiality. Standard deviations are in parentheses; DEED = Decennial Employer-Employee Database; FIRE = finance, insurance, and real estate; MSA = metropolitan statistical area; PMSA = primary metropolitan statistical area.

is reflected in the medians for total employment in the establishment as recorded in the BR and how these medians increase for the different subsamples we analyze. Later in the empirical analysis, we consider the possible ramifications for the estimates of these consequences of our sample selection rules.

## VI. Network Isolation Results

### A. Results for Whites

Table 3 presents results for whites. The observed network isolation index ($NI^O$) for the full sample of whites, reported in column 1, indicates that, on average, a white worker works in an establishment where 7.87% of his or her white coworkers live in the same census tract. The random network isolation index ($NI^R$) is 2.97, and the network isolation difference ($NI^O - NI^R$) is 4.90. Our calculated maximum possible network isolation index ($NI^M$) indicates that white workers in our data could, on average, work in establishments in which at most 52.06% of their coworkers are neighbors. The effective network isolation index, which rescales the network isolation difference by the maximum network isolation that occurs beyond randomness, is 9.99. That is, about 10% of the maximum amount to which residential networks (at the census tract level) could contribute to the sorting of workers into establishments is actually reflected in the sorting of workers into establishments. Whether this is a large number or a small one is subjective, and there is little with which to compare it, given the sparseness of empirical evidence on the importance of labor market networks. To us, however, it seems like a large number, suggesting that residential labor market networks are quite important.

This evidence could reflect sorting of workers into both neighborhoods and establishments according to skill, rather than residential networks. For example, one establishment in a census tract may hire less skilled workers who tend to live together in a neighborhood where housing is cheap, while a second hires more skilled workers who tend to live together in a different neighborhood with more expensive housing.[17] To evaluate this alternative, table 3 columns 2–4 report network isolation results conditional on various proxies for skill, based on conditional network isolation indexes that simulate network isolation while holding fixed the skill distribution of workers within establishments.

These conditional indexes are constructed by modifying the procedure used to construct the random network isolation index ($NI^R$). Instead of randomly assigning all workers in a census tract of employment to any job in any establishment in the tract, holding the size distribution of establishments fixed, we instead ensure that workers are assigned (randomly) only to jobs that are observed in the real data to be held by a worker with the same level of skill. We then once again compute the average simulated fraction of coworkers who come from a worker's own neighborhood, denoting this $NI^C$, and we define the "conditional effective network isolation" index as

$$\frac{NI^O - NI^C}{NI^M - NI^R} \times 100,$$

where $NI^R$ and $NI^M$ are defined as before (i.e., without regard to skill

[17] Hellerstein and Neumark (2008) report evidence of some segregation of workers across establishments based on skill levels. Bayer, Fang, and McMillan (2005) provide evidence of residential segregation by education for blacks.

**Table 3**
**Network Isolation for Whites Overall and by Education or Skill-Related Measures**

| | | All, Conditional on | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All (1) | Two Education Categories (2) | Four Education Categories (3) | Six Occupation Categories (4) | Housing Price Quartiles (5) | High School Degree or Less Only (6) | More than High School Degree Only (7) | All, Conditional on Eight Industry Categories (8) |
| Network isolation index observed, $NI^O$ | 7.87 | 7.87 | 7.87 | 7.87 | 7.87 | 10.95 | 5.94 | 7.87 |
| Simulated random network isolation index, $NI^R$ | 2.97 | | | | | 4.81 | 2.28 | |
| Network isolation difference, $NI^O - NI^R$ | 4.90 | | | | | 6.15 | 3.66 | |
| Maximum possible network isolation index, $NI^M$ | 52.06 | 52.06 | 52.06 | 52.06 | 52.06 | 52.01 | 43.74 | 52.06 |
| Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \times 100$ | 9.99 | | | | | 13.02 | 8.82 | |
| Simulated conditional network isolation index, $NI^C$ | | 3.02 | 3.13 | 3.25 | 3.49 | | | 4.02 |
| $NI^O - NI^C$ | | 4.84 | 4.74 | 4.61 | 4.38 | | | 3.85 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Conditional effective network isolation index, $[(NI^{O} - NI^{C})/(NI^{M} - NI^{R})] \times 100$ | | 9.86 | 9.65 | 9.40 | 8.92 | | | 7.83 |
| $N$ | 1,675,412 | 1,675,412 | 1,675,412 | 1,675,412 | 1,675,412 | 441,133 | 1,014,931 | 1,675,412 |
| No. place-of-work tracts | 26,470 | 26,470 | 26,470 | 26,470 | 26,470 | 16,073 | 20,378 | 26,470 |
| No. residential tracts | 46,764 | 46,764 | 46,764 | 46,764 | 46,764 | 40,754 | 45,311 | 46,764 |
| Mean establishments / tract | 129.6 | 129.6 | 129.6 | 129.6 | 129.6 | 47.1 | 98.0 | 129.6 |
| Mean matched workers / establishment | 38.4 | 38.4 | 38.4 | 38.4 | 38.4 | 10.8 | 41.0 | 38.4 |
| Mean number of workers in tract of employment from same tract of residence | 9.4 | 9.4 | 9.4 | 9.4 | 9.4 | 6.2 | 5.7 | 9.4 |

NOTE.—Calculation is described in the text; $NI^{O}$ is the average fraction of a worker's coworkers (i.e., excluding the worker) who reside in the same census tract as the worker, averaged across all workers in the sample; $NI^{R}$ is the average fraction that is simulated to occur randomly; $NI^{M}$ is the simulated average maximum fraction. "Effective network isolation" therefore measures the fraction of the maximum that is actually observed. In col. 2, the categories are high school degree or less and more than high school. In col. 3, they are less than high school degree, high school degree, some college, and bachelor's degree or higher. In col. 4, the occupation categories are managerial and professional specialty; technical, sales, and administrative support; service; farming, forestry, and fishery; precision production, craft, and repair; and operators, fabricators, and laborers. In col. 5, the housing price quartiles are constructed for each metropolitan statistical area (MSA). The sample sizes in cols. 6–7 sum to less than the sample size in col. 1 because the sample exclusion restrictions are imposed for each education size category. All effective measures of network isolation are statistically significantly different from zero at the 1% level. Across the columns, the following pairs of estimates of effective or conditional effective network isolation are statistically significantly different from each other at the 1% level: (*a*) conditional on two education categories (col. 2) vs. all whites (col. 1), (*b*) conditional on four education categories (col. 3) vs. all whites (col. 1), (*c*) conditional on one-digit occupation (col. 4) vs. all whites (col. 1), (*d*) conditional on housing value (col. 5) vs. all whites (col. 1), (*e*) more than a high school degree (col. 7) vs. high school degree or less (col. 6), and (*f*) conditional on eight industry categories (col. 8) vs. unconditional results for all whites (col. 1).

level). When $NI^O = NI^C$, the conditional effective network isolation index is zero, implying that the observed network isolation is fully attributable to the sorting of workers by residential census tract into establishments on the basis of skill alone. Conversely, when $NI^C = NI^R$, the conditional and unconditional network isolation indexes are equal, implying that all of the effective network isolation comes from networks helping individuals find jobs in specific establishments and that skill (as we measure it) plays no role in sorting.

In table 3, column 2, we construct the index conditional on two education categories: high school degree or less and more than high school. The simulated conditional network isolation index ($NI^C$) that results is 3.02, only a little higher than the random (unconditional) network isolation index, in column 1, of 2.97. As a result, the conditional network isolation difference ($NI^O - NI^C$) is 4.84, and the conditional effective network isolation index is 9.86, very similar to the unconditional index of 9.99. Thus, even conditional on the assignment of workers to establishments on the basis of whether they have attended college at all, 9.86% of the maximum amount of clustering of workers into establishments on the basis of their census tract of residence is observed in the data. In column 3, we refine the education groups to condition on four different levels of educational attainment: less than high school, high school degree, some college, and bachelor's degree or more. Here the simulated conditional isolation index is slightly bigger, at 3.13, but the conditional effective network isolation index is still 9.65, so that overall, conditioning on education only reduces the effective isolation index by less than 4%.

In table 3, columns 4 and 5, we use two different proxies for skill. In column 4, we condition on six occupation categories. The results again do not suggest much of a role for skill sorting in explaining effective network isolation: the simulated conditional network isolation index is 3.25, so that the network isolation difference is 4.61, and the resulting effective network isolation index is 9.40. In column 5, we use a measure of house prices as a proxy for skill, given that skill can manifest in this dimension of the accumulation of wealth.[18] The results of this exercise suggest a slightly larger role for this measure of skill in explaining network isolation, but again the results are similar; the simulated conditional net-

[18] We calculate the average house price for each census tract in our sample as reported by responding homeowners in the SEDF. Then, within each metropolitan statistical area (MSA), we calculate the quartiles of average house prices across census tracts and assign to each worker in our sample an indicator for the quartile of the within-MSA house price distribution in which that worker's census tract of residence falls. We then construct a simulated network isolation index by randomizing workers to a job within their census tract of employment held by someone whose residential census tract falls within the same quartile of the house price distribution.

work isolation index is 3.49, and the resulting effective network isolation index is 8.92.

Across table 3 columns 2–5, then, we find that sorting of workers across census tracts by a variety of proxies for skill explains very little (at most 11%) of the overall (unconditional) effective network isolation. Because our measures of skill are only proxies and are not all inclusive, this evidence is by nature not definitive. Nonetheless, given that the results are robust across three different dimensions of skill (education, occupation, and housing values), we view the results as providing strong evidence that skill does not play a major role in driving our measure of network isolation.[19]

We can also examine whether network isolation varies across skill groups, in the next two columns of table 3. Column 6 reports results for whites with at most a high school education, computed for the sample of white workers who have no more than a high school education. The observed network isolation index, averaging across the sample of low-educated white workers the fraction of each individual's white low-educated coworkers who live in that individual's residential census tract, is 10.95, somewhat higher than the 7.87 number for the full sample of whites as reported in column 1. We calculate the simulated network isolation index by randomly assigning workers to jobs held by low-educated white workers in establishments in their census tract of work and then, for low-educated white workers, calculating the fraction of simulated coworkers who live in the same residential census tract. The random network isolation index is 4.81, with the net effect that the network isolation difference of 6.15 is also higher than for the full sample. Finally, the calculated maximum possible network isolation index for this low-education subsample is 52.01, very similar to that of column 1. Thus, the effective network isolation index is 13.02, 30% higher than the corresponding figure in column 1.

Table 3, column 7, reports results from the same exercise for whites with more than a high school degree. The resulting measures of network isolation are smaller than in the full sample; the network isolation difference for this group is 3.66, and the effective network isolation index is 8.82. Comparing columns 1, 6, and 7, the results suggest that residence-based networks are more important for low-educated whites, as has been suggested in previous surveys of workers' use of informal contacts by

[19] It is possible that our results are driven by connections between family members, for example, spouses working together. We restricted the sample to white males only, which excludes spouses, and the effective network isolation index declined by about one-third (from 9.99 to 6.44). However, this sample selection rule also has other substantial effects on the sample (e.g., reducing mean matched workers per establishment by about 40%, as well as, of course, dropping females), so the results are not strictly comparable.

education. This difference by skill may arise because residence-based networks are more important for local labor markets, which are more significant for low-skilled than for high-skilled workers. The evidence of residence-based networks for both schooling groups also suggests, like the evidence in columns 2–5, that the full-sample results in column 1 are not being spuriously driven by the joint sorting of workers by education level into neighborhoods and establishments.

Networks may help job searchers find jobs in certain industries, rather than in certain establishments. For example, a worker in a retail firm may tell a neighbor of job vacancies in nearby retail businesses. If networks increase the likelihood that neighbors work in the same industry, our calculations might overstate the extent to which networks determine the establishment of employment because the clustering of neighbors in the same industry within a census tract of employment will inevitably lead to some clustering in the same establishments. Of course if residence-based networks largely operate at the industry rather than the establishment level, they would still be important.

To explore whether the network effects we find reflect employment at the establishment level, or instead at the industry level only, we construct conditional network isolation indexes in which we simulate network isolation while holding the distribution of workers across industries fixed within a census tract of employment. Intuitively, if a particular residential census tract has a lot of workers employed in a specific industry, then the random allocation of workers in the simulation will preserve that industry concentration, and by subtracting off the network isolation that occurs randomly conditional on industry, it will isolate the extent to which the clustering of census tract coresidents in the same establishments exceeds the clustering that is driven by them working in the same industry.

Column 8 of table 3 reports results conditioning on eight industry categories. The simulated index conditioning on industry is 4.02, so that when workers are randomly assigned to establishments in the same industry (and census tract) in which they are observed to work, on average 4% of their coworkers come from the same residential neighborhood. This is higher than the unconditional random isolation index of 2.97, so that the conditional effective network isolation index of 7.83 is smaller than the unconditional effective index of 9.99 reported in column 1. The difference implies that assignment of workers from the same neighborhoods to specific industries within a census tract can explain some of the assignment of workers to specific establishments. However, 78.5% of the effective network isolation remains, even after we condition on a worker's industry.[20] Thus, at this level of industry detail, most of the (unconditional)

[20] This is likely a lower bound for the percentage of the effective network isolation index that remains. If there are only a few establishments in an industry

effective network isolation for whites cannot be explained just by a mechanism whereby residence-based networks serve only to help workers find jobs in the same industries as their neighbors or by which workers are simply sorted across neighborhoods and establishments on the basis of education, but other than that there is no clustering of neighbors in the same workplace.

Another potential alternative explanation of our results is that rather than residential neighborhood influencing where one works—via residence-based networks—place of work determines where one lives. If, for example, coworkers 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 labor market networks. We do not have data on where workers lived when they first began working (or applied for work) at a particular employer. However, the census data indicate whether a person changed addresses in the past 5 years, and the BR has establishment age. Thus, we can restrict attention to residents who have not moved in the past 5 years and who work in establishments that are fewer than 5 years old, for whom the choice of residential location necessarily preceded the decision to work at a new establishment. When we did this, we found stronger evidence of residence-based networks.[21] This evidence actually strengthens the interpretation of the results as reflecting residence-based networks because we would expect such network connections to be stronger among those who have lived in their neighborhoods longer and who therefore are more likely to have connections to neighbors through any of a number of channels.

We noted earlier that constructing the DEED and imposing our sample restrictions leads our analysis samples to consist disproportionately of large establishments. If networks are more important in small establishments than in large establishments, this will bias our network isolation index downward. There are two reasons to think this could be the case. First, networks might operate primarily among individuals who work in the same occupation, and if large establishments have more occupational heterogeneity, our measure of network strength may understate the importance of networks in large establishments because it treats all workers in establishments as equally eligible to be in networks together.[22] Second, other evidence (e.g., Holzer 1998) suggests that small establishments rely

in a particular census tract, then what this procedure treats as sorting on industry may represent sorting on establishments. The results were similar when we conditioned on both industry and education.

[21] The results are available on request.

[22] However, we doubt that stratification of networks by occupation within establishments plays a major role because conditioning on occupation for white workers had little influence on the effective network isolation index (table 3).

**Table 4**
**Network Isolation for Whites, by Establishment Size**

| | | Establishment Size | | | |
|---|---|---|---|---|---|
| | All (1) | ≤50 (2) | 51–100 (3) | 101–250 (4) | >250 (5) |
| Network isolation index observed, $NI^O$ | 7.87 | 15.76 | 6.08 | 4.59 | 2.82 |
| Simulated random network isolation index, $NI^R$ | 2.97 | 4.55 | 2.86 | 2.58 | 2.03 |
| Network isolation difference, $NI^O - NI^R$ | 4.90 | 11.21 | 3.22 | 2.01 | .79 |
| Maximum possible network isolation index, $NI^M$ | 52.06 | 53.56 | 36.93 | 32.50 | 19.95 |
| Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \times 100$ | 9.99 | 22.87 | 9.45 | 6.70 | 4.41 |
| *N* | 1,675,412 | 527,430 | 208,507 | 262,437 | 448,155 |
| No. place-of-work tracts | 26,470 | 22,162 | 9,055 | 4,448 | 3,790 |
| No. residential tracts | 46,764 | 43,700 | 39,425 | 40,663 | 42,603 |
| Mean establishments / tract | 129.6 | 66.0 | 29.2 | 26.8 | 19.2 |
| Mean matched workers / establishment | 38.4 | 3.3 | 6.1 | 11.0 | 96.2 |
| Mean number of workers in tract of employment from same tract of residence | 9.4 | 3.9 | 2.5 | 3.2 | 6.6 |

Note.—See note to table 3. The sample sizes in cols. 2–5 sum to less than the sample size in col. 1 because the sample exclusion restrictions are imposed for each establishment size category. All effective network isolation measures are statistically significantly different from zero at the 1% level.

more on informal referrals in hiring, and if this is associated with the kinds of networks we measure, then network isolation should vary across establishment size. Therefore, in table 4, we disaggregate the results by splitting the sample of establishments into four size categories.

Table 4, column 2, reports the results for establishments with reported employment from the BR of 50 employees or fewer. The observed network isolation index, $NI^O$, is 15.76, much larger than 7.87 for the full sample. The simulated random network isolation difference is also somewhat bigger than for the full sample (4.55 vs. 2.97), but the implied network isolation difference is still much larger for the small establishments. Because the maximum possible indexes are similar for the full sample and for the sample of small establishments, the resulting effective network isolation index for small establishments is 22.87, more than twice as large as for the full sample (9.99). Moving across the columns, where establishment size categories rise to 51–100 employees in column 3, to 101–250 in column 4, and to greater than 250 employees in column 5, the effective network isolation index decreases monotonically, although the biggest drop comes when moving from the smallest establishments to those with at least 51 employees. In total, the results quite consistently indicate that networks are more important in small establishments.[23]

[23] If we only looked at observed isolation ($NI^O$), it is possible that the higher observed percentage of coworkers who are neighbors in small establishments masks a more equal number of coworkers who are neighbors. However, we always

## B. Results for Blacks and Black-White Differences

The importance of networks may differ for blacks and whites. On the one hand, as Montgomery's (1991) model suggests, because whites make up a greater fraction of the working population, if networks are race based one might expect white individuals searching for employment to be able to take advantage of a larger network of white working neighbors, making it more likely that whites will work together in the same establishment than will blacks, above and beyond what would be predicted by random allocation.[24] On the other hand, if labor market networks serve to overcome information imperfections more for blacks than for whites, perhaps by helping to lower search costs for blacks related to finding nondiscriminatory employers, or overcoming stigma, one might expect network isolation to be larger for blacks than for whites.

In table 5, we report results for black workers, restricting the sample over which we compute network isolation to census tracts of employment in which we observe at least two establishments, each having at least two black workers matched. We restrict the sample to black workers only because if we include both blacks and whites and there is discrimination in the labor market, we would expect to observe sorting of workers by race across establishments. Then, because there is residential segregation by race, we would spuriously generate a finding of network isolation. In contrast, by considering only blacks we ask whether there is segregation of blacks by neighborhood of residence across workplaces above and beyond the segregation of workers across establishments as a function of race alone (or something correlated with race alone).

Table 5, column 1, reports results for the full sample of black workers. The observed network isolation index ($NI^{O}$) for all blacks is 5.29, meaning that on average, of a black worker's black coworkers, 5.29% come from that worker's neighborhood. This is somewhat smaller than for whites, and the random network isolation index ($NI^{R}$) is 2.58, just slightly below that for whites; the network isolation difference is therefore 2.71. However, for blacks the maximum possible network isolation index ($NI^{M}$) is 30.71, quite a bit below that for whites, with the net result that the effective network isolation index of 9.63 is very close to what we find for whites (although statistically different from it). These results, which underscore the importance of scaling the network isolation difference by the extent to which network isolation is possible in the data, suggest little racial

look at observed isolation relative to randomness ($NI^{R}$), and the size of the establishment affects $NI^{R}$. Thus, the difference between them (whether scaled or not) always provides a meaningful comparison.

[24] In our full samples of whites and blacks in cols. 1 of tables 3 and 5 (discussed below), blacks have an average of 10 black working neighbors, whereas whites have an average of 64 white working neighbors.

**Table 5**
**Network Isolation for Blacks Overall and by Education**

| | All (1) | High School Degree or Less Only (2) | More than High School Degree Only (3) | Establishments Located in Tracts in Both Black and White Samples: White (4) | Establishments Located in Tracts in Both Black and White Samples: Black (5) | Network Isolation Based on Blacks and Whites: All Blacks (6) |
|---|---|---|---|---|---|---|
| Network isolation index observed, $NI^O$ | 5.29 | 8.22 | 3.41 | 5.68 | 5.25 | 3.99 |
| Simulated random network isolation index, $NI^R$ | 2.58 | 4.11 | 1.74 | 1.48 | 2.54 | 2.00 |
| Network isolation difference, $NI^O - NI^R$ | 2.71 | 4.11 | 1.67 | 4.20 | 2.72 | 2.00 |
| Maximum possible network isolation index, $NI^M$ | 30.71 | 31.48 | 24.67 | 58.37 | 31.43 | 35.83 |
| Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \times 100$ | 9.63 | 15.02 | 7.30 | 7.38 | 9.41 | 5.90 |
| $N$ | 97,967 | 27,001 | 47,281 | 845,290 | 94,210 | 140,083 |
| No. place-of-work tracts | 4,490 | 1,923 | 2,278 | 4,122 | 4,122 | 9,094 |
| No. residential tracts | 21,623 | 10,685 | 16,107 | 42,533 | 21,459 | 26,768 |
| Mean establishments / tract | 22.2 | 10.2 | 17.4 | 229.7 | 23.0 | 39.7 |
| Mean matched workers / establishment | 18.6 | 8.2 | 15.8 | 46.6 | 17.7 | 40.3 |
| Mean number of workers in tract of employment from same tract of residence | 2.6 | 2.2 | 1.9 | 11.4 | 2.6 | 4.5 |

Note.—See note to table 3. Columns 4 and 5 are restricted to whites and blacks who work in census tracts in which we observe at least two establishments employing at least two white workers each and two establishments employing at least two black workers each (where these latter two or more establishments could overlap with those for which we observe white employment as well). All effective network isolation measures are statistically significantly different from zero at the 1% level. Across the columns of this and other tables, the following pairs of estimates of effective network isolation are statistically significantly different from each other at the 1% level: (*a*) blacks with a high school degree or less (table 5, col. 2) vs. blacks with more than a high school degree (table 5, col. 3), (*b*) all blacks (table 5, col. 1) vs. all whites (table 3, col. 1), (*c*) blacks with a high school degree or less (table 5, col. 2) vs. whites with a high school degree or less (table 3, col. 6), (*d*) blacks with more than a high school degree (table 5, col. 3) vs. whites with more than a high school degree (table 3, col. 7), (*e*) whites working in establishments located in tracts in both the black and white samples (table 5, col. 4) vs. blacks working in establishments located in tracts in the black and white samples (table 5, col. 5), and (*f*) blacks when network isolation is based on blacks and whites (table 5, col. 6) vs. all blacks (table 5, col. 1).

difference in our full samples of blacks and whites in the importance of residence-based networks in explaining the assignment of workers to establishments, once we rescale the network isolation difference by the maximum.[25]

In table 5, column 2, we report results for blacks who have at most a high school education, and in column 3, for blacks who have more than a high school education. The results by education address the issue of sorting by skill and examine whether black-white differences reflect race differences in education coupled with variation across skill groups in the importance of residence-based labor market networks. In both cases, the observed and simulated indexes are lower for blacks than for whites, as are the network isolation differences. When we scale by the maximum possible network isolation differences, the results reverse somewhat for low-educated blacks, so that the effective network isolation index for low-educated blacks is somewhat larger than for whites (and statistically different as well)—15.02 versus 13.02—whereas for more educated blacks, the effective index is somewhat smaller than for whites (7.30 vs. 8.82). Whether or not we scale by the maximum, though, the evidence suggests that residentially based networks are more important for less educated blacks than for more educated blacks.

Overall, the results in tables 3 and 5 are in line with other results from the literature, based on quite different types of analyses, including survey results or indirect evidence indicating greater use of informal contacts among the less educated (Topa 2001; Ioannides and Datcher Loury 2004), evidence based on place of work and place of residence indicating stronger network effects among those with less education (Weinberg, Reagan, and Yankow 2004; Bayer et al. 2008), and an absence of consistent evidence of race differences in the reported use of informal contacts (Ioannides and Datcher Loury 2004).

The last three columns of table 5 present additional analyses for black and white workers. One issue with directly comparing the full sample results for blacks and whites is that we know these groups are not similarly distributed geographically in the United States. As a result, black and white workers in our sample may work in different labor markets where the importance of networks could differ as a result of labor market institutions, constraints, or other factors. Therefore, in columns 4 and 5, we present estimates for whites and blacks, restricting the samples to workers working in census tracts that are represented by workers in both

[25] We find only a few instances in which comparisons of the network isolation difference, $NI^{O} - NI^{R}$, and the effective network isolation index, $[(NI^{O}-NI^{R}) / (NI^{M}-NI^{R})] \times 100$, lead to qualitatively different conclusions. The results for blacks and whites discussed in this paragraph are one such instance. Indeed, it is only for the black-white comparisons that the conclusions are ever sensitive to scaling by the maximum possible segregation

the white and the black samples. For whites, the sample restriction reduces the sample by about one-half, and the numbers of workplace census tracts and residential census tracts are also reduced considerably. The results are somewhat different from those for the full sample of whites. First, the observed network isolation index falls to 5.68 and is closer to that for blacks. Second, the effective network isolation measure is lower (7.38 vs. 9.99 in table 3, col. 1).[26] Table 5, column 5, reports results for blacks. The effective network isolation index for the restricted sample of blacks is 9.41, quite similar to the figure of 9.63 in column 1, which is not surprising since between columns 1 and 5 the sample is reduced by fewer than 4,000 workers. For these subsamples, the random network isolation measure is quite a bit lower for this sample of whites than for blacks (1.48 vs. 2.54), so that the network isolation difference is higher for whites than for blacks (4.20 vs. 2.72). However, the maximum network isolation index for whites is almost double that for blacks, so that when we scale by our measure of maximum possible isolation, we reverse the relative magnitudes of isolation for blacks and whites. Thus, there is suggestive evidence that networks are more important for blacks than whites in comparable areas.

### C. Racial Stratification of Networks

The results we have presented thus far suggest that residence-based networks are important and, in that sense, point to racially stratified networks. After all, given that there is pervasive racial residential segregation in the United States (Iceland and Weinberg 2002), networks that are predicated on residential "connectedness" have to be partially race based. However, it is important also to consider whether there is racial stratification of networks even within neighborhoods, that is, the idea that labor market information and especially job referrals are less likely to flow between black and white coresidents than between coresidents of the same race.[27]

Therefore, in column 6 of table 5, we assess more directly whether networks are race based. We carry out the same types of sampling and computational procedures used before, except that we consider the relevant set of a black worker's neighbors and coworkers to consist of blacks

[26] The same is true of the network isolation difference. Note that the difference between the effective network isolation measure in table 5, col. 4, and table 3, col. 1, is driven in part by the higher maximum isolation measure in the latter case. This occurs because the observations in table 5, col. 4, come from a much smaller number of tracts with many more establishments, making it possible to achieve a higher maximum amount of network isolation.

[27] Racial stratification of networks within neighborhoods can potentially explain the results in Hellerstein, Neumark, and McInerney (2008) that higher local job density for one's own race affects employment probabilities, but higher job density for the other race does not.

or whites. We begin by constructing a sample of black workers and their neighbors, regardless of race, whom we observe to work in establishments where at least one other black or white worker is matched. We then further restrict the sample to those who work in a census tract with at least two establishments that have workers in the sample. We again construct an effective network isolation index for black workers in this sample, but we now construct this measure by asking whether our sample of black workers is more likely than would be predicted by randomness to work in the same establishment with a neighbor, regardless of the race of that neighbor. If networks among coresidents are racially stratified, then the network isolation that results when we measure how likely it is that a black works with a neighbor regardless of race should be smaller than when we measure how likely it is that a black works with a black neighbor.[28]

For each black worker in our sample, we first calculate an observed network isolation index by averaging across the sample (of black workers) the fraction of each individual's coworkers who live in that individual's residential census tract, regardless of race. As shown in column 6 of table 5, this number is 3.99, substantially lower than 5.29 in column 1. We then calculate the random network isolation index by taking all workers in this sample, randomizing them (as before) across establishments in that sample, and calculating the network isolation index for the black workers in this simulated sample. The random network isolation index is 2.00, leading to a network isolation difference of 2.00. The fact that this difference is lower than when we restrict the sample of coworkers to blacks, in column 1, suggests that race is indeed playing a role in driving the probability of working with a neighbor. We obtain a measure of the maximum possible network isolation index by approximating what the index would be for blacks if the workers who make up this sample were able to work to the maximum extent possible in establishments with their neighbors of any race, given the size distribution of establishments in our sample and the residential distribution of workers in them.

The resulting maximum network isolation number is 35.83, which is somewhat larger than that in table 5, column 1. Taking all of these together, the effective network isolation index as reported in column 6 is only 5.90, which is about 40% smaller than in column 1, providing evidence that residence-based labor market networks have a fairly strong race-based component. Moreover, because in this column we only need one black worker in an establishment for the establishment to be in the sample, as opposed to column 1, which requires two black workers, column 6 in-

[28] That said, this differential network effect does not distinguish whether network connections between black and white neighbors are fewer in number or less productive; we can only characterize their relative importance.

cludes smaller establishments for which, as already noted, residence-based networks are more important. Thus, on a comparable basis the difference between the effective network isolation indexes in columns 1 and 6 would be even larger, bolstering the evidence that these networks are racially stratified.

### D. Results for Hispanics

Survey evidence suggests that Hispanics use referrals in finding employment much more than do blacks or whites (Ioannides and Datcher Loury 2004). Immigrants and poor English speakers, in particular, may suffer from high search costs in the labor market, both because their limited understanding of U.S. labor markets and of English may make it hard for them to search widely in the labor market and because potential employers may have a difficult time inferring the ability of these workers. Finding employment through informal networks of other immigrants and those who speak one's native language may therefore be particularly important for these groups. There is some indirect evidence consistent with this conjecture. For example, evidence of "enclave effects," such as Hispanics with poor English skills paying less of a penalty for those poor skills when they live in a county or MSA with a larger Hispanic population (McManus 1990), might reflect network effects, although they could also reflect higher productivity from a greater ability to work with Spanish speakers in the enclave.[29] Munshi (2003) presents a more refined analysis of Mexican immigrants, tying labor market outcomes to a larger local population of immigrants from the same origin community. Patel and Vella (2007) find that new immigrants work disproportionately in occupations held by previous immigrants from the same country. And our previous work documents establishment-level segregation by English language skills and segregation of Spanish-speaking workers from non-Spanish-speaking workers among poor English speakers (Hellerstein and Neumark 2008). Finally, perhaps the most direct evidence of these types of networks for immigrants comes from the work of Massey et al. (1987), who document the importance of networks linking recent and earlier immigrants from the same communities in Mexico.

In this section, therefore, we turn to results for Hispanic workers, paying particular attention to Hispanic workers who speak English poorly (or not at all) and immigrants. The results are presented in table 6. Column 1 presents results for the full sample of Hispanic workers (again with the sample restrictions that allow us to construct the network isolation index). The observed network isolation index is 11.22, quite a bit larger than for blacks or whites, and the random network isolation index is 3.08. Once

[29] For a similar type of evidence for Sweden, see Edin, Fredriksson, and Åslund (2003).

**Table 6**
**Network Isolation for Hispanics Overall and by Skill and Immigrant Status**

| | All (1) | Poor English Skills (2) | Good English Skills (3) | Immigrant (4) | Nonimmigrant (5) |
|---|---|---|---|---|---|
| Network isolation index observed, $NI^O$ | 11.22 | 22.51 | 8.30 | 16.68 | 7.20 |
| Simulated random network isolation index, $NI^R$ | 3.08 | 6.56 | 2.47 | 4.46 | 2.50 |
| Network isolation difference, $NI^O - NI^R$ | 8.14 | 15.95 | 5.83 | 12.22 | 4.70 |
| Maximum possible network isolation index, $NI^M$ | 39.48 | 42.65 | 34.17 | 40.50 | 30.78 |
| Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \times 100$ | 22.36 | 44.20 | 18.39 | 33.91 | 16.62 |
| *N* | 110,235 | 12,451 | 80,661 | 42,712 | 43,763 |
| No. place-of-work tracts | 5,059 | 900 | 4,090 | 2,508 | 2,563 |
| No. residential tracts | 20,716 | 4,576 | 19,027 | 11,321 | 14,063 |
| Mean establishments / tract | 59.7 | 23.3 | 46.3 | 43.2 | 28.8 |
| Mean matched workers / establishment | 9.1 | 4.8 | 9.4 | 5.7 | 9.7 |
| Mean number of workers in tract of employment from same tract of residence | 2.7 | 2.3 | 2.2 | 2.5 | 2.0 |

NOTE.—See note to table 3. All effective network isolation measures are statistically significantly different from zero at the 1% level. Across the columns of this and other tables, the following pairs of estimates of effective network isolation are statistically significantly different from each other at the 1% level: (*a*) Hispanics with good English skills (table 6, col. 3) vs. Hispanics with poor English skills (table 6, col. 2), (*b*) nonimmigrant Hispanics (table 6, col. 5) vs. immigrant Hispanics (table 6, col. 4), (*c*) all Hispanics (table 6, col. 1) vs. all whites (table 3, col. 1), and (*d*) all Hispanics (table 6, col. 1) vs. all blacks (table 5, col. 1).

we scale the difference between these by the difference between the maximum possible network isolation index and the random index, we find that the effective network isolation index for Hispanic workers is 22.36—more than twice as large as what we find for the full samples of blacks or whites.[30]

In table 6, column 2, we restrict the sample to Hispanics who self-report speaking English either "poorly" or "not at all"—which together we refer to, for simplicity, as the sample of poor English speakers. For this sample, the observed network isolation index is 22.51. This is very large—it means that, on average, for a poor-English-speaking Hispanic worker, 22.51% of his or her Hispanic coworkers who also are poor English speakers live in the same census tract. The simulated random network isolation index is much smaller, at 6.56, and the maximum network index is 42.65. Taken together, these numbers yield an effective network isolation index of 44.20, meaning that over 40% of the maximum possible establishment network isolation by census tract of residence for Hispanics who speak English poorly is actually observed in the data. This is more than four times larger than what we find for blacks and whites, and it suggests that residence-based labor market networks are extremely important for Hispanics who speak English poorly. In addition, paralleling our results by education level for blacks, the fact that the importance of networks goes up when we focus on those with poor language skills implies that the overall results for Hispanics are not driven by residential sorting on language skills.[31]

By way of contrast, in column 3 of table 6 we report the results for the sample of Hispanics who report speaking English "well" or "very well." The effective network isolation index is 18.39, just under half as large as that for Hispanics who are poor English speakers. This contrast is consistent with the idea that networks are extremely important in mitigating the high search frictions that exist for workers in the United States whose English language skills are poor. At the same time, the finding that the effective network isolation index is much higher for Hispanics who speak good English than for whites suggests that the overall Hispanic-white differences are not driven solely by skills.

In table 6, column 4, we report the results for Hispanic immigrants. The effective network isolation index is 33.91, which is quite a bit higher than for all Hispanics. In contrast, in column 5 we report the results for nonimmigrant Hispanics, for whom the effective network isolation index

[30] The comparisons of Hispanics to whites and blacks, and between different groups of Hispanics, are similar for the effective network isolation index and for the network isolation difference, so we focus on the effective index.

[31] Networks formed on the job may also serve to help low-skilled Hispanics find housing. The sample sizes for Hispanics are too small for us to conduct the same analysis of this issue that we conducted for whites.

**Table 7**
**Network Isolation for Whites, Blacks, and Hispanics Working in Small Establishments (50 Employees or Less)**

| | Whites (1) | Blacks (2) | Hispanics (3) |
|---|---|---|---|
| Network isolation index observed, $NI^O$ | 15.76 | 12.76 | 22.79 |
| Simulated random network isolation index, $NI^R$ | 4.55 | 4.34 | 5.72 |
| Network isolation difference, $NI^O - NI^R$ | 11.21 | 8.42 | 17.07 |
| Maximum possible network isolation index, $NI^M$ | 53.56 | 25.77 | 38.41 |
| Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \times 100$ | 22.87 | 39.27 | 52.20 |
| *N* | 527,430 | 8,706 | 21,952 |
| No. place-of-work tracts | 22,162 | 1,097 | 2,093 |
| No. residential tracts | 43,700 | 5,308 | 8,019 |
| Mean establishments / tract | 66.0 | 7.4 | 23.0 |
| Mean matched workers / establishment | 3.3 | 2.7 | 2.7 |
| Mean number of workers in tract of employment from same tract of residence | 3.9 | 1.4 | 1.8 |

NOTE.—See note to table 3. All effective network isolation measures are statistically significantly different from zero at the 1% level. Across the columns, the following pairs of estimates of effective network isolation are statistically significantly different from each other at the 1% level: (*a*) blacks working in small establishments (col. 2) vs. whites working in small establishments (col. 1) and (*b*) Hispanics working in small establishments (col. 3) vs. whites working in small establishments (col. 1).

is 16.62, half the size of the index for immigrant Hispanics and smaller than for any other Hispanic group in the table. To the extent that the Hispanic workers in column 5 are the most integrated into U.S. society and have good English language skills, this provides further evidence that our measure of network isolation captures the important of residence-based networks that reduce search frictions in the labor market.[32]

### E. Results by Race and Ethnicity for Small Establishments

We noted earlier that our sample selection rules lead to underrepresentation of small establishments and that this is especially true for blacks and Hispanics. Moreover, table 4 makes it clear that network isolation is more important for whites who work in small establishments than in large establishments. Therefore, the different sample compositions of establishments could bias our comparisons of the importance of networks across racial and ethnic groups in the full samples of each of these groups. In table 7, we report our measures of (unconditional) network isolation for whites, blacks, and Hispanics, where we use restricted samples only of workers employed at establishments with 50 or fewer workers in total. Column 1 contains the results for whites, as previously reported in column 2 of table 4. Table 7, column 2, contains the results for blacks, and column 3 the results for Hispanics. The results, which are based on much more homogeneous samples with respect to establishment size than those for

[32] Immigrant status and language skills are strongly related. A bit under half of the immigrant sample consists of poor English speakers, while the nonimmigrant sample is nearly entirely good English speakers.

the full samples, suggest that residence-based networks are much more important for blacks than for whites (an effective network isolation index of 39.27 vs. 22.87),[33] and even more so for Hispanics (52.20) relative to whites. Thus, the relative importance of networks for blacks and especially Hispanics compared to whites appears greater than is suggested by the analyses of our full samples.

## VII. Conclusions

We use matched employer-employee data for the United States to measure the importance of residence-based labor market networks in the allocation of jobs. The core of our approach is to look at business establishments in a census tract and to ask whether the workers at each establishment are disproportionately clustered in particular residential neighborhoods, relative to what we would expect to occur randomly given that most workers employed in a particular census tract reside in a subset of nearby census tracts. Evidence of this kind of disproportionate residential concentration of a business establishment's workforce is consistent with labor market networks that connect individuals residing in the same neighborhood to specific business establishments. Because of recent research highlighting the potential importance of labor market networks for less skilled workers in the labor market and more generally positing that labor market networks operate along the lines of race, ethnicity, and skill, we consider separately the importance of these labor market networks for whites, blacks, and Hispanics and, within each group, the relative importance of these networks for workers with different skills.

Our evidence is complementary to an existing body of research on labor market networks and the use of informal labor market contacts that are thought to characterize networks. What is unique about our evidence, however, is that it looks directly at potential network effects for workers employed at the same business establishment. Given that many theories of the importance of labor market networks emphasize the gains to employers from using their current employees to refer other employees, it seems particularly useful to test whether network connections among workers—in our case based on residential location—actually make it more likely that workers are employed in the same business.

We interpret the evidence as indicating that labor market networks play an important role in establishment-level employment. For both whites and blacks, we find that the grouping of workers from the same neighborhoods in the same business establishments exceeds by a factor of more than two what we would expect to occur randomly. For whites, we find that network isolation is about 10% of the theoretical maximum amount

[33] Paralleling the earlier results, this conclusion for black-white differences is driven by the scaling.

of grouping that could be found in the data, and many of our analyses indicate that residence-based labor market networks are more important for blacks than for whites. For both whites and blacks, these labor market networks appear more important for workers who have low levels of education—a high school degree or less—than for more educated workers. There is also evidence that these networks are more important in small establishments.

Our results also provide evidence that networks operate to some extent along racial lines, above and beyond the racial stratification of networks that comes from residential segregation by race. In particular, the link between residential location and the establishment of employment is stronger for blacks when we consider only coworkers of the same race, consistent with more (or more productive) labor market information and referrals flowing across coresidents of the same race than between coresidents of different races. As emphasized in recent theoretical work by Calvó-Armengol and Jackson (2007), race-based labor market networks may prevent the convergence of black and white labor market outcomes—and can even exacerbate the differences. More significantly, this type of racial stratification would imply that policies that solely address spatial mismatch, by attempting to move blacks to areas where more whites live and where more jobs (per person) are located, may fail to help blacks precisely because network connections are severed and are less likely to be established with white neighbors.[34]

We also find that residence-based networks are more important for Hispanics than for blacks or whites, and among Hispanics, these networks are especially important for immigrants and those with poor language skills. The results for Hispanics give credence to the idea that informal labor market networks may be particularly important for those workers who are not as well integrated into the labor market and for whom employers may have less reliable information.[35]

[34] The evidence from Moving to Opportunity (MTO) is consistent with this conclusion (see, e.g., Turney et al. 2006). Qualitative interviews of experimentals and controls in MTO suggest that both groups rely heavily on network connections to find jobs and that experimentals who moved from public housing to lower-poverty neighborhoods had less access to neighbors with jobs in the sectors in which they had been previously employed (largely retail and health care). Interestingly, though, the connections through which both experimentals and controls reported finding jobs were not "immediate neighbors" but included associates from school, church, past jobs, etc. We take this to indicate that, for those represented by this sample, residential networks may be better captured by broader geographic areas like census tracts than by blocks; however, it also emphasizes that neighbors—however defined—are not the only source of network connections.

[35] As noted previously, the alternative interpretation of our results—that they reflect heterogeneous tastes such that people who like similar workplaces also like

As the discussion of the data requirements for this study indicates, it is difficult to obtain evidence on labor market networks. Although the notion of networks has been around for many decades, there are only a handful of studies providing evidence that networks affect labor market outcomes, and this study is the first to document the importance of labor market networks in determining the establishments in which workers work. Aside from further attempts to construct or obtain data to study the kind of network effects we examine in this article, a number of other important questions remain. First, what are the consequences of labor market networks that match workers in a network to specific establishments? Do those who find employment in establishments with others in their networks actually have better labor market outcomes (e.g., higher wages or more job security) as a result? Second, are minorities who have network relationships mainly with other minorities disadvantaged relative to those that have network relationships with whites? Third, how does the strength of networks vary across residential locations that vary by geography, socioeconomic status, and so on?

Data sets, like the DEED, that match workers to establishments are likely to prove useful in trying to address these questions in future research. That said, matched employer-employee data sets have limitations in terms of what they can teach us about networks. First, it is important to consider along what other dimensions of social interactions—aside from residence—networks operate to cause individuals to work in the same establishment and what types of networks are most important. Among the possibilities are schools, religious institutions, and community groups, as well as existing places of employment (from which workers may move to other jobs).[36] Second, these data sets provide little scope for understanding the dynamics of how networks are formed and how they operate. Are all members of the network equally important? What kinds of information get shared within the network? The data demands for answering many of these questions are daunting, but the answers can provide clues regarding how important it is for individuals, communities, and other institutions to foster network relationships so as to improve economic outcomes and what types of networks are most effective.

---

similar neighborhoods—is not consistent with the patterns of evidence that we find, whereby residence-based networks are more important for those least integrated into the economy, such as the less educated, Hispanics, and especially Hispanic immigrants.

[36] Indeed Bayer et al. (2008) show that the type of network effect they study appears to be stronger for those with children of similar ages, which could reflect social interactions of families in schools.

## References

Bayer, Patrick, Hanming Fang, and Robert McMillan. 2005. Separate when equal? Racial inequality and residential segregation. Unpublished manuscript, Department of Economics, Duke University.

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* 116, no. 6 (December): 1150–96.

Calvó-Armengol, Antoni, and Matthew O. Jackson. 2007. Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory* 132, no. 1 (January): 27–46.

Cingano, Federico, and Alfonso Rosolia. 2009. People I know: Job search and social networks. Unpublished manuscript, Bank of Italy, Rome.

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

Cutler, David M., and Edward L. Glaeser. 1997. Are ghettos good or bad? *Quarterly Journal of Economics* 112, no. 3 (August): 827–72.

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* 118, no. 1 (February): 329–57.

Ellison, Glenn, and Edward L. Glaeser. 1997. Geographic concentration in manufacturing industries: A dartboard approach. *Journal of Political Economy* 102, no. 5 (October): 889–927.

Granovetter, Mark S. 1974. *Getting a job: A study of contacts and careers*. Cambridge, MA: Harvard University Press.

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* 71–72 (July–December): 19–78.

———. 2008. Workplace segregation in the United States: Race, ethnicity, and skill. *Review of Economics and Statistics* 90, no. 3 (August): 459–77.

Hellerstein, Judith K., David Neumark, and Melissa McInerney. 2008. Spatial mismatch vs. racial mismatch? *Journal of Urban Economics* 64, no. 2 (September): 467–79.

Holzer, Harry J. 1998. Why do small establishments hire fewer blacks than large ones? *Journal of Human Resources* 33, no. 4 (August): 896–914.

Iceland, John, and Daniel H. Weinberg. 2002. *Racial and ethnic segregation in the United States, 1980–2000*. Washington, DC: U.S. Department of Commerce. http://www.census.gov/hhes/www/housing/housing_patterns/pdf/censr-3.pdf.

Ioannides, Yannis M., and Linda Datcher Loury. 2004. Job information, networks, neighborhood effects, and inequality. *Journal of Economic Literature* 42, no. 4 (December): 1056–93.

Jarmin, Ron S., and Javier Miranda. 2002. The longitudinal business database. Working Paper no. CES-WP-02-17, Center for Economic Studies, U.S. Census Bureau.

Kasinitz, Philip, and Jan Rosenberg. 1996. Missing the connection: Social isolation and employment on the Brooklyn waterfront. *Social Forces* 43, no. 2 (May): 180–96.

Kmec, Julie A. 2007. Ties that bind? Race and networks in job turnover. *Social Problems* 54, no. 4 (November): 483–503.

Laschever, Ron. 2009. The Doughboys network: Social interactions and the employment of World War I veterans. Unpublished manuscript, Department of Economics, University of Illinois at Urbana-Champaign.

Massey, Douglas, Rafael Alarcón, Jorge Durand, and Humberto González. 1987. *Return to Aztlan: The social process of international migration from western Mexico*. Berkeley: University of California Press.

McManus, Walter S. 1990. Labor market effects of language enclaves: Hispanic men in the United States. *Journal of Human Resources* 25, no. 2 (Spring): 228–52.

Montgomery, James D. 1991. Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review* 81, no. 5 (December): 1408–18.

Munshi, Kaivan. 2003. Networks in the modern economy: Mexican migrants in the U.S. labor market. *Quarterly Journal of Economics* 118, no. 2 (May): 549–99.

Oyer, Paul, and Scott Schaefer. 2009. The personnel-economic geography of US law firms and law schools. Unpublished manuscript, Graduate School of Business, Stanford University.

Patel, Krishna, and Francis Vella. 2007. Immigrant networks and their implications for occupational choices and wages. Working Paper no. 3217, IZA, Bonn.

Schmutte, Ian M. 2010. Job referral networks and the determination of earnings in local labor markets. Unpublished manuscript, Department of Economics, University of Georgia.

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* 10, no. 3 (July): 306–30.

Topa, Giorgio. 2001. Social interactions, local spillovers, and unemployment. *Review of Economic Studies* 68, no. 2 (April): 261–95.

Turney, Kristin, Susan Clampet-Lundquist, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2006. Neighborhood effects on barriers to employment: Results from a randomized housing mobility experiment in Baltimore. In *Brookings-Wharton Papers on Urban Affairs*, ed. Gary Burtless and Janet Rothenberg Pack, 137–72. Washington, DC: Brookings Institution.

U.S. Department of Commerce. 1994. Geographic areas reference manual. Washington, DC: U.S. Government Printing Office. http://www.census.gov/geo/www/GARM/Ch10GARM.pdf.

Weinberg, Bruce A., Patricia B. Reagan, and Jeffrey J. Yankow. 2004. Do neighborhoods affect hours worked? Evidence from longitudinal data. *Journal of Labor Economics* 22, no. 4 (October): 891–934.