NBER WORKING PAPER SERIES

MEASURING THE IMPORTANCE OF LABOR MARKET NETWORKS

Judith K. Hellerstein
Melissa McInerney
David Neumark

Measuring the Importance of Labor Market Networks
Judith K. Hellerstein, Melissa McInerney, and David Neumark
NBER Working Paper No. 14201
July 2008
JEL No. J15,J61

ABSTRACT

We specify and implement a test for the importance of network effects in determining the establishments
at which people work, using recently-constructed matched employer-employee data at the establishment
level. We explicitly measure the importance of network effects for groups broken out by race, ethnicity,
and various measures of skill, for networks generated by residential proximity. The evidence indicates
that labor market networks play an important role in hiring, more so for minorities and the less-skilled,
especially among Hispanics, and that labor market networks appear to be race-based.

Judith K. Hellerstein
Department of Economics
Tydings Hall
University of Maryland
College Park, MD 20742
and NBER
hellerst@econ.umd.edu

Melissa McInerney
College of William and Mary
and U.S. Census Bureau
Economics Department
P.O. Box 8795
Williamsburg, VA 23187
mcinerney@econ.umd.edu

David Neumark
Department of Economics
3151 Social Science Plaza
University of California, Irvine
Irvine, CA 92697-5100
and NBER
dneumark@uci.edu

I. Introduction

Racial and ethnic disparities in labor market outcomes are well documented. There is also strong evidence of residential segregation in the United States by race and ethnicity (e.g., Iceland and Weinberg, 2002), and this segregation is correlated with poor labor market outcomes for minority groups, especially the low-skilled among them (e.g., Cutler and Glaeser, 1997). The existence of residential segregation is not itself an underlying economic mechanism for poor labor market outcomes for minorities. But there are at least two theories that can explain how residential segregation by race and ethnicity can contribute to these poorer outcomes.

Perhaps the most famous of these theories is the spatial mismatch hypothesis (e.g., Kain, 1968). The essence of the spatial mismatch hypothesis is that blacks disproportionately live in areas with poor access to jobs, and that this spatial isolation contributes to poor labor market outcomes for blacks.

The other theory that can be used to explain the link between residential segregation and labor market outcomes is network effects. In one well-known network effects model, Montgomery (1991) specifies a labor market where firms cannot observe the underlying ability of a potential worker but where 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. Hence, networks act at the establishment level to reduce search frictions. In equilibrium, individuals are more likely to receive and accept wage offers from the firms that employ others in their social network, creating stratification across firms on the basis of social networks. In Montgomery's framework, if social networks are at least partially race- or ethnic-based due to residential segregation, and white workers are initially employed at higher rates than blacks and Hispanics, 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. Alternatively, Calvó-Armengol and Jackson (2007) analyze how labor market networks can translate initial differences in employment rates

1

and wages between two groups into persistent differences; a natural application is to blacks and whites, with networks in which blacks are more connected to other blacks and whites to other whites.

In previous research, we report evidence suggesting that spatial mismatch is not the mechanism by which residential segregation leads to poor economic outcomes for blacks (Hellerstein et al., forthcoming). In particular, we find that poor employment outcomes for low-skilled blacks are not a function of a lack of jobs per se where blacks live, but rather that local blacks get these jobs only when employers are hiring other black workers. We suggested that this evidence could be driven by networks or by neighborhood effects, and that discrimination against blacks could also potentially play a role. We did not attempt to distinguish sharply between these alternatives to spatial mismatch. However, we concluded that network effects were potentially a better explanation because we found a similar relationship for whites to that which we found for blacks—with whites' employment prospects boosted by higher employment of whites where they live, but not by higher employment of blacks. In short, the combined evidence for blacks and whites is consistent with race-based labor market networks. Moreover, in other work we find evidence of substantial workplace segregation by race and ethnicity, evidence that is also consistent with race- and ethnic-based labor market networks (e.g., Hellerstein and Neumark, forthcoming).

The potential importance of stratified labor market networks has been noted previously. Granovetter (1974) is one of the early sources of evidence on the importance of informal contacts in finding employment, and subsequent work has noted the potential for labor market networks to 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.[1] More generally, Ioannides and Datcher Loury (2004) review evidence indicating widespread reliance on friends, relatives, and acquaintances to search for and find jobs. They conclude that there is very little racial (black-white) difference in the use of informal contacts in job search, but that rates of use of informal contacts are higher for low-educated workers compared to high-educated workers, and that there are substantially higher rates of use of

---

[1] See, e.g., the discussion and references in Kmec (2007).

informal contacts for Hispanics than for other groups. However, none of these studies assesses the

importance of networks in determining in which establishments people work, and as such none of them

provides direct evidence consistent with employers relying on informal networks to reduce search

frictions. Rather, the reliance on informal referrals documented in this past research simply implies that

friends, relatives, etc., are a source of labor market information.

Bayer et al. (2005a) present an analysis testing more directly for network effects. They use

confidential data from the 2000 Long Form of the Decennial Census and focus exclusively on workers in

the Boston area. In a clever method of inferring whether networks are important in explaining

employment patterns, they find that two individuals who live on the same census block are about one-

third more likely to work on the same census block than are two individuals who live in the same block

group but not on the same block.[2] To the extent that informal networks are stronger within the block than

within the block group, this evidence is consistent with labor market networks affecting hiring.[3]

We regard the Bayer et al. paper as providing the most definitive evidence to date suggesting that

residential geographic proximity affects labor market outcomes, presumably through networks that

connect people living in close proximity (see also Ioannides and Datcher Loury, 2004). Nonetheless,

there are a couple of limitations of the paper upon which we try to improve in the present study. First, the

data contain no information on the exact establishment in which the workers work, so that two individuals

who work in the same census block may work for different employers, particularly for blocks in the

central city that contain multiple employers.[4] This is significant limitation with respect to testing the

hypothesis that labor market networks reduce search frictions on the part of employers, as in

---

[2] For the Boston metropolitan area data they study, a census block corresponds roughly to a city block, and there are on average 10 blocks per block group. (They get similar results using the 10 closest blocks to each block based on physical distance.)
[3] Although the baseline rate at which these workers work together is very small to begin with—0.36 percent—this effect is estimated for any pair of workers; the authors suggest that the estimate implies a considerably higher probability that a worker works on the same block as at least one person who resides on his block.
[4] For example, focusing on central city areas in our sample of urban establishments and their employees, which is described below, we find that there are on average 2.58 establishments per block, which is an undercount by what we estimate to be about a factor of four, given that we only observe in our data a subsample of establishments. This casts doubt on Bayer et al.'s assumption that workers employed on the same block "work with" one another or "work together" (e.g., 2005a, p. 26).

Montgomery's (1991) model. Second, Bayer et al. (2005a) focus on the simple link between place of residence and place of work, and due to sample size issues (which is partly related to the demographics of Boston), cannot fully explore the differential importance of networks by race, and cannot isolate the importance of networks for Hispanics or subgroups of Hispanics. These are groups for whom it is particularly important to study networks, not only because of the evidence discussed in Ioannides and Datcher Loury (2004) that suggests differential use of networks, but also because of theoretical reasons to expect the importance of networks to vary across groups. For example, if labor market discrimination (either statistical or taste-based) by some employers raises search costs for certain groups, such as blacks or Hispanics (Black, 1995), residentially-based networks may partially mitigate these search costs by helping to identify possible employers on whom to focus a job search. And to the extent that employers stigmatize some groups of workers based on other characteristics associated with race (such as welfare use or a criminal record) or ethnicity (such as language ability), informal networks may help to overcome statistical discrimination.[5] Note that in these latter cases some degree of residential segregation may help disadvantaged groups.

In this paper we specify and implement a test for the importance of network effects in determining the establishments at which people work, using recently-constructed matched employer-employee data at the establishment level. We explicitly measure the importance of network effects for groups broken out by race (black vs. white), by ethnicity (Hispanic vs. white), and by various measures of skill (education and English language proficiency).[6] Because of our data and the resulting methods we employ, it is important to note that we are testing for a particular form of labor market networks—namely, networks that are generated by residential proximity. In that sense, our analysis parallels Bayer et al. (2005a).[7]

---

[5] The latter explanation follows more directly from Montgomery's (1991) model.

[6] As Montgomery (1991, p. 1409) notes, since referrals may serve as a screening device, workers with lower unobserved ability may be less likely to get referrals. However, this does not necessarily carry over to these observable measures of skill.

[7] We readily note that, in finding that networks are important, we cannot definitively establish that the direction of causality runs from place of residence to place of work. Bayer et al. (2005a) examine this issue in detail and conclude that it is likely that networks operate to allow people who live together to find work together.

4

The measure of labor market networks that we use 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. This type of segregation of employees across establishments would arise from residential proximity capturing the "network connectedness" that is important to the flow of labor market information between specific employers, their employees, and potential hires.[8]

In particular, we first identify all establishments within each census tract. Since 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 then compare how likely it is, on average, that an individual works in the same establishment as his/her residential neighbors relative to the likelihood that this would result if the establishment hired workers randomly from the geographic areas where *all* individuals who work in the census tract reside. Residential-based networks would predict that the probability that a worker works with neighbors would be higher—and possibly much higher—than would result from the random hiring process. Thus, the difference between these two probabilities provides a measure of the importance of residential-based labor market networks, which we rescale into an "effective" measure of the importance of networks by comparing this difference to the maximum possible extent of networks that could arise in the data (again, relative to randomness), given the distributions of workers across establishments and across residential neighborhoods.

We examine the importance of residential-based networks separately by race, ethnicity, and skill in order to establish whether the importance of networks in grouping residential neighbors into employment in specific establishments seems to be greater for Hispanics and blacks than for whites, and for less-skilled than for more-skilled workers. Differences by skill might arise because less-skilled workers have worse labor market information and may benefit most from networks through a reduction in search costs and increased information access. Race and ethnic differences in the importance of networks may exist if networks reduce employer uncertainty about productivity when it is difficult for workers to

---

[8] See Calvó-Armengol and Jackson (2007) for a model of these connections.

provide signals of productivity because of language, cultural, and educational differences (e.g., Lang, 1986), or because of statistical discrimination. In the case of Hispanic workers, in particular, networks may compensate for less-developed formal hiring networks in predominantly immigrant communities, attributable in part, perhaps, to hiring of undocumented workers.

We compute our effective measure of the importance of networks unconditionally as well as conditional on the industry of employment. By measuring networks conditional on industry we establish that our results really are the result of a process whereby networks help workers gain access to specific employers, rather than work opportunities more generally in a given type of industry; the latter would be more likely to occur if informal contacts simply tend to provide information about job opportunities in a particular sector of the economy.

The data we use for this study come from the 2000 Decennial Employer-Employee Database, which we have constructed at the U.S. Census Bureau. The DEED is a large dataset consisting of workers matched to their establishment of employment. The employer-employee matches enable us to study directly whether workers employed in the same establishment are likely to live in the same neighborhoods. This is one important dimension in which our work improves upon that of Bayer et al. (2005a) who—because they have information only on the Census block in which an employer is located—can only ask whether workers both live and work in proximity to one another. In addition, we have national data and focus on race and ethnic differences, whereas they study data for Boston and focus mostly on whites. One drawback of the DEED, however, is that the matching process used in constructing the dataset, as well as sample restrictions required to implement our empirical tests, result in the disproportionate exclusion of smaller establishments.

Overall, we find evidence that labor market networks play an important role in hiring. For blacks, we find that the grouping of workers from the same neighborhoods in the same business establishments is about 9 percent of the maximum grouping that could occur, in a sense we define later. Networks are also important for whites, although less so than for blacks. For blacks and whites, labor market networks affect employment patterns more for less-skilled than for more-skilled workers, and

6

because networks are also important within skill groups, the findings do not simply reflect residential segregation by skill. Moreover, the evidence indicates that networks are partly race-based, operating more strongly within than across races. We also find that networks are even more significant for Hispanics, for whom the grouping of workers from the same neighborhoods in the same business establishments is about 21 percent of the theoretical maximum. And among Hispanics, networks play a larger role for immigrants and those with poor English skills. Finally, the results conditioning on industry indicate that the network effects we find are largely due to the assignment of workers to specific establishments rather than simply to industries; this evidence provides support for theoretical models in which networks serve to match workers to specific employers.

II. The 2000 DEED

The analysis in this paper is based on the 2000 DEED, a data set that matches workers to their establishments, which we have created at the Center for Economic Studies at the U.S. Bureau of the Census. We have constructed a similar data set for 1990 and described in detail the process of its construction elsewhere (in particular, Hellerstein and Neumark, 2003). The construction of the 2000 DEED follows the same procedures. Thus, in this section we simply provide a brief overview.

The 2000 DEED is formed by matching workers to establishments. The workers 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 list (BR) for 2000; the BR is a database containing information for all business establishments operating in the United States in each year, which is continuously updated (see Jarmin and Miranda, 2002). The BR is a list of all business establishments with one or more employees operating in the United States. The Census Bureau uses the BR as a sampling frame for its Economic Censuses and Surveys, and continuously updates the information it contains. The BR contains the name and address of each establishment, geographic codes based on its location, its four-digit SIC code, and an identifier that allows the establishment to be linked to other establishments that are part of the same enterprise, and to other Census Bureau establishment- or firm-level data sets that contain more detailed employer

7

characteristics.

Households receiving the 2000 Decennial Census Long Form were asked to report the name and address of the employer in the previous week for each employed member of the household. The file containing this employer name and address information is referred to as the "Write-In" file, which contains the information written on the questionnaires by Long-Form respondents, but not actually captured in the SEDF. We can use employer names and addresses for each worker in the Write-In file to match the Write-In file to the BR. Because the name and address information on the Write-In file is also available for virtually all employers in the BR, nearly all of the establishments in the BR that are classified as "active" by the Census Bureau are available for matching. Finally, because both the Write-In file and the SEDF contain identical sets of unique individual identifiers, we can use these identifiers to link the Write-In file to the SEDF. Thus, this procedure yields a very large data set with workers matched to their establishments, along with all of the information on workers from the SEDF.

Matching workers and establishments is a difficult task, because we would not expect employers' names and addresses to be recorded identically on the two files. To match workers and establishments based on the Write-In file, we use MatchWare—a specialized record-linkage program. MatchWare is comprised of two parts: a name and address standardization mechanism (AutoStan); and a matching system (AutoMatch). Our method to link records using MatchWare involves two basic steps. The first step is to use AutoStan to standardize employer names and addresses across the Write-In file and the BR. Standardization of addresses in the establishment and worker files helps to eliminate differences in how data are reported. The standardization software considers a wide variety of different ways that common address and business terms can be written, and converts each to a single standard form. Once the software standardizes the business names and addresses, each item is parsed into components. The value of parsing the addresses into multiple pieces is that we can match on various combinations of these components.

The second step of the matching process is to select and implement the matching specifications. The AutoMatch software uses a probabilistic matching algorithm that accounts for missing information,

8

misspellings, and even inaccurate information. This software also permits users to control 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. Different match specifications may produce different sets of matches. Matching criteria should be broad enough to cover as many potential matches as possible, but narrow enough to ensure that only matches that are correct with a high probability are linked. Because the AutoMatch algorithm is not exact there is always a range of quality of matches, and we therefore are cautious in accepting linked record pairs. We chose matching algorithms based on substantial experimentation and visual inspection of many thousands of records.

The final 2000 DEED is an extremely large data set containing information on 4.09 million workers matched to around 1.28 million establishments, accounting for 29.1 percent of workers in the SEDF and 22.6 percent of establishments in the BR.[9] We impose additional sample restrictions for our analysis, which we discuss following the explanation of our empirical methods in the next section.

III. Measuring the Importance of Networks

We measure whether and to what extent residential networks play a role in the assignment of workers to establishments via an analysis that is based on the percentage of workers in an individual's establishment (i.e., workplace) that comes from the individual's residential neighborhood. We compute the importance of networks across a variety of subsamples, e.g., all blacks, low-educated blacks, all whites, and Hispanics who speak English poorly. For explication, here we describe in detail how we measure the role of race-based residential networks for black workers. Construction of the network measure for other subsamples follows in a parallel fashion.

We first compute for each black worker in our sample the percentage of black workers with

---

[9] For both the DEED and SEDF we have excluded individuals as follows: with missing wages; who did not work in the year prior to 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 fifty 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 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.)

which that worker works who come from the same residential neighborhood as that worker, excluding the worker him or herself. Because we exclude the individual from this calculation (since it is meaningless to say that a person is his or her own neighbor), this requires a sample restriction to establishments where we observe at least two black workers.

We then average these percentages across all the black workers in our sample to create a "network isolation index," denoted NI^O, which measures the average fraction of a worker's co-workers who are also residential neighbors of that worker. In using the phrase "network isolation index," we are borrowing from the sociology literature measuring residential segregation (often by race) by defining the "isolation index" to be the fraction of a black person's residential neighbors who are themselves black. The superscript "O" on the network isolation index emphasizes that this is the fraction of a black worker's co-workers who are *observed* in our estimation sample to be residential neighbors.

To operationalize our network isolation index, we need to define what it means for workers to be residential neighbors, which in turn requires us to define neighborhood boundaries. We define residential neighborhoods to be census tracts. There are a few reasons why this definition seems sensible to us. First, census tracts define the boundaries that are traditionally used to measure residential segregation (see, e.g., Iceland and Weinberg, 2002). Second (and related to the first), census tracts are defined by the U.S. Census Bureau in order to ensure that the tracts are "as homogeneous as possible with respect to population characteristics, economic status, and living conditions,"[10] which in itself is a reasonable definition of a neighborhood. Third, most census tracts are relatively small, so it is reasonable to think that it is quite possible that census tract residents have contact with each other, if not "over the back fence," then at parks, schools, churches, stores, etc.

To provide some idea of the size of census tracts, detailed maps of these tracts for Chicago are reproduced in Figures 1 and 2. For the PMSA (Figure 1), the median census tract was 0.57 square miles; the mean was 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 itself (Figure 2), the median was

---

[10] See http://www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed April 21, 2008).

0.17 square miles, the mean was 0.26 square miles, the minimum was again 0.02 square miles, and the

maximum was 8 square miles. The larger tracts are not problematic. For the city, the two largest tracts

are O'Hare Field (the airport) and another tract at the southern edge of the city that is mainly industrial.[11]

After these two, the next largest tract is 3.5 square miles. The much larger census tracts in outlying areas,

which are more rural, do not necessarily imply fewer social contacts, because the density of schools,

churches, etc., is of course much lower. And these tracts have few residents or establishments and hence

contribute little to the overall findings.

There are two reasons why we might observe residential neighbors working in the same

establishment that have nothing to do with residentially-based labor market networks. First, access to

mass transportation alone may lead residential neighbors to work in the same geographic place, although

not necessarily to work in the same establishments within that place. Therefore, in establishing

boundaries for the local area in which workers might be employed, we again use census tracts. Census

tracts are generally small enough that it is possible, particularly within urban areas, for individuals to be

able to walk from any establishment to any other establishment, so in that sense transportation differences

should not materially affect the distribution of workers across establishments in the tract.

Second, some clustering of residential neighbors into establishments can occur even if workers

are assigned randomly to establishments, and we are of course interested in workplace clustering of

residential neighbors that occurs systematically—i.e., that which is greater than would be expected to

result from randomness.[12] We therefore consider deviations of NI^O from what would be expected if

workers from the same residential neighborhood were randomly assigned to any establishment within the

census tract in which they are employed.

In particular, we compute the extent of network isolation that would occur simply due to

---

[11] The latter large tract represents the industrial South Deering community. The census tract is nearly 8 square miles, but Lake Calumet, wetlands areas, and the Calumet River comprise over 2 square miles. The area is largely industrial; the region's first steel mill was located here (http://www.fieldmuseum.org/calumet/SouthDeering.html, viewed May 8, 2008).

[12] This is a point that we discuss in some detail in the context of measuring workplace segregation (Hellerstein and Neumark, forthcoming), and was noted previously by others (see, e.g., Carrington and Troske, 1997).

randomness, denoted NI^R, by simulating random allocation through Monte Carlo methods.[13] Within a

census tract, we randomly assign workers to establishments, ensuring that we generate the same size

distribution of establishments within a census tract as we have in the sample. We do this simulation 100

times, and then compute the random network isolation index, NI^R, as the mean over these 100

simulations. Not surprisingly, the random network isolation measures are very precise; in all cases the

standard deviations were trivially small. We then focus on the difference $NI^O - NI^R$, measuring network

isolation above and beyond that which occurs randomly. Note that this adjustment implies that we under-

estimate the importance of residential-based labor market networks more generally in explaining where

people work if labor market networks serve to help members of the network find jobs close to those of

others in the network, in addition to helping find jobs through referrals to members' current employers.[14]

This computation requires a second sample restriction—that census tracts of employment include at least

two establishments with two black workers. Otherwise, if the only black workers in a census tract work

together in one establishment, then we cannot distinguish the effect of residential-based labor market

networks from random clustering.

The difference $NI^O - NI^R$ does not, in and of itself, give us a sense of the importance of networks

in generating the observed spatial pattern of neighbors working together in establishments. In particular,

while NI^R gives us a measure of the minimum amount of clustering of neighborhoods that would happen

with random assignment of workers to establishments within a census tract, it is also important to know

what the maximum possible network isolation could be in our data, given the distribution of individuals

across residential census tracts and given the size distribution (based on matched workers) of

establishments in each census tract. To compute this, we use an algorithm that calculates the approximate

maximum network isolation that could occur if workers were systematically assigned along with their

---

[13] For example, consider a census tract that has two establishments in it, each of which is observed to employ two black workers. Assume that two of the black workers come from neighborhood A and two from neighborhood B. Even if the four black workers are randomly allocated across the two establishments without regard to the neighborhoods in which they live, 1/3 of the time they will be working with a neighbor, so that NI^R will be 1/3.
[14] And of course because we only focus on a residential-based labor market networks, we have no measure of labor market networks that derive from non-residential sources.

neighbors to the maximum extent possible to establishments within the census tract in which they are actually observed to work. For a given census tract in which we have establishments represented in our sample, we order the neighborhoods in which workers live by the number of workers from each neighborhood. Beginning with the neighborhood with the greatest number of workers, we assign as many workers as possible to one establishment, and any workers who are not assigned to that establishment are grouped together and treated as a "new" neighborhood. We then move to the second largest neighborhood from which workers originate (which could be the "new" neighborhood left from the previous pass), and assign workers from that neighborhood to the establishment that holds the maximum number of these, again therefore keeping neighbors working together in establishments as much as possible. We continue moving down the list of neighborhoods, from those with larger to smaller numbers of workers, assigning workers to establishments until all workers are assigned. We would expect this maximum network isolation index to be less than one.[15]

This algorithm assigns workers to establishments in a way that simulates the maximum possible neighborhood isolation, by mechanically ensuring that it is more likely that workers from large neighborhoods will work together. After assigning workers from the large neighborhoods, it often is still likely that we can assign workers from small neighborhoods to work together in establishments, filling in the remaining slots in establishments that are not filled by the workers from large neighborhoods. In contrast, 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.[16] We do this for every census

---

[15] For example, consider a census tract that employs nine workers across three establishments, A, B, and C. Establishment A employs five workers, while establishments B and C each employ two workers. Six of the workers live in one census tract 1, two live in tract 2, and one lives in tract 3. Thus, our algorithm proceeds as follows. We first take the workers from the largest neighborhood, tract 1, and put five of them in establishment A, forming a new "neighborhood" (call it tract 1A) consisting of the one leftover worker from tract 1. We then take the two workers from tract 2, the next largest neighborhood, and put them in establishment B (the same result occurs if we put them in establishment C). At this point we remain with the single worker from tract 3 and the single leftover worker from tract 1A, who have to be assigned in this example to establishment C so as to preserve the size distribution of establishments. In this case, the maximum network isolation index is $\{(5/9) \cdot 1 + (2/9) \cdot 1 + (2/9) \cdot 0\} = 7/9$. To see this, the five workers in establishment A are all from the same tract, with the share of co-workers from their census tract equal to one. The same is true for the two workers in establishment B. And this share is zero for the two workers assigned to establishment C.

[16] We have not determined analytically that this always yields the maximum segregation, but experimenting at the

13

tract in our sample where workers work, and then compute the weighted average of the maximum

network isolation in each census tract of employment, weighting by the number of workers we observe to

be working in that census tract. We label this maximum network isolation number NI^M.

Finally, we turn back to $(NI^O - NI^R)$, the difference between our observed network isolation index

and the random isolation index, and we scale it by the maximum network isolation that can occur beyond

randomness, or $(NI^M - NI^R)$, yielding

$$[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100,$$

which we call the "effective network isolation index." It measures the share of the maximum possible

network isolation that is actually observed, and provides a natural scaling for the importance of networks

formed by residential networks in determining the establishments in which people are employed.

IV. Sample Characteristics and Restrictions

In Table 1 we provide descriptive statistics for the matched workers from the DEED as compared

to the SEDF. Column (1) reports summary statistics for the SEDF for the sample of white, black, or

Hispanic workers who were eligible to be matched to their establishments. Column (2) reports summary

statistics for all white, black or Hispanic workers in the full DEED sample—that is, those who we

successfully matched to their establishments. 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. These

differences result in part from the matching process, because there are many individuals who meet our

sample inclusion criteria but for whom the quality of the business address information in the Write-In file

is poor. We suspect that the differences in business address information partially reflect weaker labor

market attachment among less-skilled workers, suggesting that estimates of the importance of networks

we obtain might best be interpreted as measuring the extent of network isolation among workers who

have relatively high attachment to the labor force and to their employers. The last eight rows of the table

report on the industry distribution of workers. There is some over-representation of workers in

manufacturing in the full DEED, because larger establishments are more likely to be matched.

tract level, we have yet to uncover an algorithm that yields a higher measure.

We must make several restrictions to arrive at the samples used for our analysis of network effects. First, we only consider workers who live and work in the same Metropolitan Statistical Area/Primary Metropolitan Statistical Area (MSA/PMSA). Second, we exclude from the sample workers that have no co-workers in the same establishment who are of the same race or ethnicity as the worker; because we are interested in the impact of race- (and ethnic-) based networks on the composition of a workplace, we need to observe at least one co-worker of the same race or ethnicity to be able to identify a race- or ethnic-based workplace network. Finally, we retain only workers who work in establishments that are located in census tracts with at least one other establishment having two matched workers of the same race or ethnic group. The means for the samples of white, black, and Hispanic workers that we analyze are shown in columns (3)-(5).

In addition to comparing worker-based means, it is useful to examine the characteristics of establishments in the DEED once we make our sample restrictions. Table 2 shows descriptive statistics for establishments in the full DEED, and then for each of our final analysis samples. Because only one in six workers is sent the Decennial Census Long Form, it is more likely that large establishments will have two matched workers, especially for smaller racial or ethnic groups. One can see evidence of the bias toward larger employers by comparing the medians across the columns for total employment in the establishment as recorded in the BR. (This bias presumably also influences the distribution of workers and establishments across industries, where, for example, the DEED itself, and the final analysis samples for blacks and Hispanics even more so, over-represent workers in manufacturing establishments.) The establishments corresponding to the full 2000 DEED sample of workers have median employment of 15 workers. Once we restrict attention to establishments in metropolitan areas with at least two matched white workers, and with at least two such establishments in the census tract, as in column (2), the median employment level rises to 35. Restricting our attention to those establishments in census tracts with another establishment having at least two *black* workers matched, in column (3), median employment is even higher, at 154; and in column (4), when we consider the sample of establishments employing two matched Hispanics in census tracts with at least two such establishments, median employment is 84.

15

Later in the empirical analysis we consider the possible ramifications of these consequences of our sample selection rules.

V. Network Isolation Results

V.1. Results for Whites and Blacks

Table 3 presents network isolation index results for white and black workers with various levels of education. We begin with white workers, who make up the largest subsample of workers that we use, and serve as a good baseline for comparison. Column (1) reports results for all approximately 1.7 million white workers in our sample. These workers work in 26,470 unique census tracts and live in 46,764 census tracts.[17] The mean number of establishments in each census tract for which we observe at least two white workers in two establishments is 130; the mean number of residents from the same neighborhood working in the same census tract is 9.4. The observed network isolation for the full sample of whites is 7.87, indicating that, on average, 7.87 percent of a white worker's white co-workers live in the same census tract as that worker. When we randomize workers in this sample across establishments within census tracts, we recover a random network index of 2.97, less than 40 percent of the observed index, and the difference between the two is 4.90. The maximum possible network isolation that could be observed in the data for all white workers is 54.84. This is well below 100 (perfect sorting of workers by residential neighborhoods into establishments), because in many of our census tracts establishments have more workers observed to work in them than are drawn from any particular census tract residential neighborhood. It is also considerably above the random network isolation measure.[18]

As a result, when we scale up the difference between the observed and random network indexes by the maximum network isolation that occurs beyond randomness, we recover an effective network isolation index of 9.45. That is, approximately 9.5 percent of the maximum amount to which residential networks (at the census tract level) could contribute to the sorting of workers into establishments is

---

[17] This is out of a total of about 65,000 census tracts in 2000.
[18] Bootstrap methods show that all of the effective network isolation indices we report are statistically significantly different from zero. Indeed, the estimates of the effective network indices are quite precise, so that, in general, substantive numerical differences across columns and tables in reported network isolation indexes are also statistically significant.

actually observed in the data. Whether this is a large number or a small one is a subjective matter, of course, 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 labor market networks are quite important indeed.

An alternative explanation for the effective network isolation index in column (1) is that there is sorting of workers by both neighborhoods and establishments according to skill. For example, in a census tract with two establishments employing workers from two neighborhoods, sorting of this type could occur if one establishment hires only less-skilled workers (for example, grocery store cashiers), who tend to live together in a neighborhood where housing is cheap, and the second establishment hires only more-skilled workers (for example, lawyers), who tend to live together in a different neighborhood with more expensive housing.[19] Such sorting on skill potentially invalidates a network-based interpretation of evidence that people who live near each other tend to work together.[20]

In order to evaluate whether the results are driven by skill differentials that lead to both residential and workplace skill-based segregation across establishments, in columns (2) and (3) we report results separating whites by education level. Column (2) reports the results for only those whites who have at most a high school education. That is, we take the sample of white workers from column (1) who have no more than a high school education and compute the network isolation indexes for those workers. Specifically, we first calculate the observed network isolation index (NI^O) by averaging across the sample of low-educated white workers the fraction of each individual's white co-workers who live in that individual's residential census tract, regardless of the co-workers' education levels.[21] This number is

---

[19] In Hellerstein and Neumark (forthcoming) we report evidence of some segregation of workers across establishments based on skill levels. Bayer et al. (2005b) provide evidence of residential segregation by education for blacks.

[20] The Bayer et al. (2005a) study addresses this issue directly, arguing that the sorting of individuals on residential location is at the block group level (roughly speaking, 10-block areas), so that a finding that those living on the same individual block are more likely to work on the same block (albeit a different one) than are those living on the same block group reflects geographic proximity but not sorting. Evidence they present based on observables suggests that, in their analysis, this assumption is likely to be valid.

[21] We could do a more extreme version of this calculation where we also restrict the sample of co-workers to be those with low education. This approach yields qualitatively similar conclusions for the various groups we study, but in some cases leads to rather small samples as it causes us to discard some establishments and some census tracts

10.56, somewhat higher than the 7.87 number in column (1). We then calculate the network isolation index for low-educated white workers using the simulated sample from column (1). The random network isolation index is 4.06, also higher than its counterpart in column (1), so that the difference between the observed and random index is 6.50. The fact that both of these numbers are somewhat higher than they are when considering all white workers suggests that education is playing some role in driving the probability of working with a neighbor. Finally, to calculate the maximum possible network isolation index, we calculate what the index would be if low-educated white workers were able to work to the maximum extent possible in establishments with their neighbors of any education level, given the size distribution of establishments in our sample, and the size and residential distribution of workers in them. This number is 61.62, which is again somewhat higher in column (1). Taking all of these together, the effective network isolation index is 11.29, which is 19 percent higher than in column (1).

We perform the same exercise for whites with more than a high school degree and report the results in column (3). Across all rows, the resulting measures of network isolation are smaller than in the full sample, and the effective network isolation index for this group is 8.21. Looking across columns (1) through (3), the results suggest that networks are somewhat more important for low-educated whites, as has been suggested in previous surveys of workers' use of informal contacts by education. Returning to the original point of this inquiry, however, an equally important conclusion from these results is that effective network isolation is about as high or higher once we disaggregate by education, implying that the network results we report for the full sample in column (1) are not being spuriously driven by the joint sorting of workers by education level into neighborhoods and establishments.

In columns (4)-(6) of Table 3 we provide parallel results for black workers. As noted earlier, theoretically the importance of network effects may differ for blacks and whites. On the one hand, like Montgomery's (1991) model suggests, since 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

---

in order to meet the data requirements necessary to compute the indexes. Since we do not expect residential sorting by skill to be perfect, there is no reason to preclude labor market information flowing across workers of different skill levels.

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.[22] 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 non-discriminatory employers, one might expect network isolation to be larger for blacks than for whites. Ultimately, this is an empirical question.

Column (4) reports results for the sample of black workers. We estimate that the observed network isolation index (NI^O) for all blacks is 5.29, somewhat smaller than for whites, and the random network isolation index (NI^R) is 2.58, just slightly below that for whites. For blacks, the maximum possible network isolation index (NI^M) is 31.60, quite a bit below that for whites. All together, these numbers yield an effective network isolation index of 9.35, very close to what we find for whites. (In fact, the effective network isolation index for all blacks (9.35) is not statistically significantly different than the estimate for all whites (9.45).) These full sample results therefore imply that there is very little racial difference in the importance of residential-based networks in explaining the assignment of workers to establishments.

In column (5) of Table 3 we report results for blacks who have at most a high school education, and in column (6) we report results for blacks who have more than a high school education. In both cases, each of the observed indexes is lower than for whites, but as with the full sample results the overall effective network isolation indexes for blacks by education level are substantively similar to those for whites, and suggest that residentially-based networks are more important for less-educated blacks than for more-educated blacks. For less-educated blacks, networks generate 11.52 percent of the maximum amount of sorting by neighborhood that could occur (and this is not statistically significantly different from the estimate of 11.29 percent for less-educated whites), and for more-educated blacks, the effective index is 7.01 (compared to 8.21 for more-educated whites).

---

[22] In our full samples of whites and blacks in columns (1) and (4) of Table 3, blacks have an average of 10 black working neighbors, whereas whites have an average of 64 white working neighbors.

Overall, the results in Table 3 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 (Ioannides and Datcher Loury, 2004; Topa, 2001); evidence based on place of work and place of residence indicating stronger network effects among those with less education (Bayer et al., 2005a); and an absence of consistent evidence of race differences in the reported use of informal contacts (Ioannides and Datcher Loury, 2004).[23] As noted earlier, though, our evidence speaks specifically (and exclusively) to network effects that operate to assign individuals in the network to establishments in which other network members work.

Table 4 presents some 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, our full samples of black and white workers may well work in very different local labor markets where one might expect the importance of networks to differ as a result of labor market institutions, constraints, or other factors. Therefore, in columns (3) and (4) of Table 4 we repeat the analyses for whites and blacks, but restricting the samples to workers working in census tracts that are represented by workers in both the white and black samples. That is, we restrict the samples of whites and blacks to those 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). For comparison, columns (1) and (2) repeat the full sample results for whites and blacks from Table 3.

The results for the restricted sample of whites are reported in column (3). The sample restriction reduces the sample by about one-half, to 845,290 workers, and the numbers of workplace census tracts and residential census tracts are also reduced considerably. So for whites, we are effectively restricting the sample to those who work in places where black workers are also working. The results are different

_____

[23] Weinberg et al. (2004) study neighborhood effects on hours worked, as a manifestation of network effects, and also find stronger effects for the less-educated.

than for the full sample of whites. First, the observed network isolation index is lower at 5.68, and interestingly is closer to that for blacks. And second, the effective network isolation measure is lower (7.00 vs. 9.45 in column (1)).[24] Column (4) reports results for blacks. The effective network isolation index for the restricted sample of blacks is 9.08, quite similar to the figure of 9.35 in column (2), which is not surprising since between columns (2) and (4) the sample is reduced by fewer than 4,000 workers.[25] More interestingly, however, the effective network isolation measure for the consistent samples is 30 percent higher for blacks than for white—9.08 vs. 7.00—suggesting that networks are somewhat more important for black than for white workers, at least in places where both are employed.

Finally, while we have suggested that the network effects we find are race-based, the results we have presented thus far are indirect, in that blacks (whites) who live together are more likely than otherwise to work together with other blacks (whites) in the same establishment, and to the extent 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. Understanding the role of race in driving network effects is extremely important. Race-based networks are central to the work of Calvó-Armengol and Jackson (2007) in deriving the result that networks can perpetuate and exacerbate initial differences in employment between blacks and whites. Networks that operate along racial lines can also potentially help explain the results in Hellerstein et al. (forthcoming) that higher local job density for one's own race affects employment probabilities, but higher job density for the other race does not.[26] This, in turn, suggests that policies that solely address spatial mismatch, by

---

[24] The difference between the effective network isolation measure in columns (1) and (3) is driven in part by the higher maximum isolation measure in the latter case. This occurs because the observations in column (3) come from a much smaller number of tracts with many more establishments and more matched workers per establishment, making it possible to achieve a higher maximum amount of network isolation.

[25] We also computed results (but do not report them in a table) for the sample of black workers who work in urban establishments (that is, in census tracts that are within the borders of the main city included in the MSA or PMSA). The effective network isolation index is 8.18, a bit smaller than for the full sample of blacks. Of course, one reason that the apparent network effects appear a bit weaker in urban areas may be that residential neighborhoods in which people interact and exchange job market information are more likely to extend beyond the census tract, given how small census tracts are in urban areas (Figure 2).

[26] In that paper, however, the network effects do not necessarily operate among those living in the same neighborhood, as we estimated the effect of the density of jobs in a residential neighborhood—whether or not held by neighbors—on residents' employment.

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 finally, race-based hiring networks can help explain the establishment-level racial segregation that we have documented in Hellerstein and Neumark (forthcoming).

Therefore, in column (5) of Table 4 we provide more direct evidence on the extent to which networks are race-based. To do this, 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 co-workers to consist of blacks *or* whites. We begin by constructing a sample of black workers who we observe to work in establishments where at least one other white or black worker is matched. We then further restrict the sample of black workers to those who work in a census tract with at least two establishments meeting the selection criterion, that is, which have at least two white or black workers matched to them (i.e., at least two blacks, one white and one black, or two whites). With this sample of black workers, we again construct an effective network isolation index, but what is different now is that we construct this measure by asking whether our sample of black workers are more likely than would be predicted by randomness to work in the same establishment with a neighbor, *regardless of the race of that neighbor*. To the extent that race plays at least a partial role in determining residentially-based networks, the effective network isolation index 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.

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 co-workers who live in that individual's residential census tract, regardless of race. This number is 3.35, substantially lower than the 5.29 number in column (2). We then calculate the random network isolation index by taking all matched white or black workers in these establishments, randomizing them 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 1.41, leading to a difference between the observed and random indices

of 1.93. The fact that both of these numbers are lower than they are when restricting the sample of co-workers to blacks, in column (2), suggests that race is indeed playing a role in driving the probability of working with a neighbor. To obtain the maximum possible network isolation index, we calculate what the index would be if the blacks 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, the size and residential distribution of workers in them, and the workers' races. In computing this maximum, we first assign to establishments workers from neighborhoods that have at least one black worker, so that we are able to assign as many neighbors of black workers as possible to the same establishment. We then proceed to assign neighborhoods with only white workers to establishments. The resulting maximum network isolation number is 42.34, which is larger than in column (2) because we now have many more possible establishments to which we can first assign neighbors of black workers in attempting to maximize network isolation for blacks.[27] Taking all of these together, the effective network isolation index is only 4.73, which is about 50 percent smaller than in column (2), providing direct evidence that residential-based labor market networks have a fairly strong race-based component.[28]

V.2. Results for Hispanics

Networks may be more important determinants of employment for blacks than for whites because search frictions are larger for black workers than for white workers, or because networks serve to mitigate these frictions more for blacks than for whites. Although the evidence on black-white differences presented in the previous subsection did not point to strong differences, there may be other subgroups in the labor market for which networks may be particularly important. The results from use of referrals suggest that Hispanics use referrals in finding employment much more than blacks and whites (Ioannides

---

[27] We now consider nearly three times as many co-workers in the computation of the network isolation index (53 co-workers, versus 19 when just considering matched black co-workers); however, the mean number of neighbors blacks work with increased by only a factor of two (5.7 versus 2.6).

[28] In future work, we plan on trying to determine the extent to which race-based networks are attributable to residential segregation by race, as opposed to weaker network connections across than within races for those residing in the same neighborhood.

and Datcher Loury, 2004). Moreover, although there is limited formal large-scale empirical evidence of

this to date, it is often thought that immigrants, particularly those with limited English skills, tend to work

with others from their same country of origin or who speak the same language.[29] Immigrants 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.

In this section we turn to an analysis of results for Hispanic workers, paying particular attention

to Hispanic workers who speak English poorly (or not at all) and Hispanic workers who are immigrants.[30]

The results are presented in Table 5. Column (1) presents results for the full sample of Hispanic workers

(again with the sample restrictions that allow us to construct the network isolation indexes). 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 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 21.37, which is more than twice as large as what we find for

blacks or whites.

In column (2) we restrict the sample for which we estimate network effects to Hispanics who self-

report speaking English either "poorly" or "not at all" on the Long Form of the Decennial Census—which

together we refer to, for simplicity, as the sample of poor English speakers. For this sample, the observed

network isolation index is 20.27. This is very large—it means that, on average, for a poor-English

speaking worker, 20.27 percent of his or her co-workers who also are poor English speakers live in the

---

[29] There is only a limited amount of rigorous evidence on this. Evidence on establishment-level segregation by language is presented in Hellerstein and Neumark (forthcoming). Patel and Vella (2007) show that that new immigrants work disproportionately in occupations held by previous immigrants from the same country, and that this varies by geographic region.
[30] For a different kind of evidence of network effects for Mexican immigrants to the United States, see Munshi (2003).

same census tract! The random network isolation index is much smaller, at 4.99, and the maximum

network index is 54.32. Taken together, these numbers yield an effective network isolation index of

30.98, meaning that over 30 percent 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 three times larger than what we find for blacks and whites, and suggests to us that networks are

extremely important for Hispanics who speak English poorly.[31]

By way of contrast, in column (3) of Table 5 we report the results for the sample of Hispanics

who report speaking English "well" or "very well." The effective network isolation index is 17.47, just

over 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 exists for workers

in the United States whose English language skills are poor, at least with respect to finding employment.

In column (4) we report the results for Hispanic immigrants. The effective network isolation

index is 27.12, which is quite a bit higher than for all Hispanics.[32] In contrast, in column (5) we report the

results for non-immigrant Hispanics, for whom the effective network isolation index is 13.62, smaller

than that of the full sample of Hispanics or any of the other Hispanic subgroups. To the extent that these

Hispanic workers are most integrated into U.S. society and all have good English language skills, this

provides further evidence that what we are capturing in our measure of network isolation index is, indeed,

the important of residence-based networks that reduce search frictions in the labor market.[33]

V. 3. Results for Small Establishments

We noted earlier that our sample selection rules lead to under-representation of small

establishments for blacks, and to a lesser extent for Hispanics. To the extent that network isolation differs

in a way that is related to establishment size, the different sample compositions of establishments could

---

[31] 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 skill implies that the results for Hispanics are not driven by residential sorting on language skills.

[32] The results are similar for Mexican immigrants, who of course represent the largest share of Hispanic immigrants.

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

bias our comparisons of the extent of network isolation across racial and ethnic groups. The results of previous research might lead one to expect that thus far we have *understated* the relative importance of networks for blacks and Hispanics. In particular, evidence such as that reported in Holzer (1998) indicates that smaller establishments rely more heavily on informal referrals. To the extent that these referrals are associated with the types of network effects we capture, under-representation of small establishments in our black and Hispanic samples results in downward bias in the estimates of the importance of networks for minorities.

As confirming evidence, Table 6 reports our baseline analyses for whites, blacks, and Hispanics, but utilizing a restricted sample of workers employed only at establishments with 50 or fewer workers. Two things are apparent. First, compared to the corresponding estimates in columns (1) and (4) of Table 3 and column (1) of Table 5, networks appear much more important when we restrict the analysis to small establishments. Second, and more to the point, the racial and ethnic comparisons in Table 6, which are based on much more homogeneous samples with respect to establishment size, indicate that networks are much more important for blacks than for whites (an effective network isolation index of 38.26 vs. 20.47), and even more so for Hispanics relative to whites. Thus, the relative importance of networks for blacks and Hispanics compared to whites is greater than is suggested by the analyses of our full samples in Tables 3-5.[34]

V.4. Network Isolation Conditional on Industry

In interpreting the results to this point we have presumed that the residential-based networks we are measuring operate to help workers find jobs in particular establishments. As we have discussed, this is in line with the theoretical literature on networks (Montgomery, 1991; Calvó-Armengol and Jackson, 2007). However, networks may instead (or as well) serve to help job searchers learn of vacancies in certain industries, rather than reducing frictions that prevent workers from matching to specific

---

[34] Although not reported in the table, the analysis of blacks and whites for these restricted samples yields evidence indicating that networks are race-based, similar to that reported in column (5) of Table 4. In this case—for the smaller establishments for which networks are more important—the effective network isolation index falls from 38.26 in Table 6 to 14.27 when we ignore race, a decline of 63 percent.

establishments within industries. For example, someone who works in a retail firm may tell a neighbor of

job vacancy postings in other nearby retail establishments. Note that, in principle, this kind of mechanism

could underlie the results in Bayer et al. (2005a), since they only establish that those who live nearby are

likely to work in the same narrow geographic area, not the same establishment.

In this section we explore whether the network effects we find reflect employment at the

establishment level, or instead only at the industry level. We do this by constructing "conditional"

network isolation indexes, simulating network isolation while holding the distribution of workers across

industries fixed within a census tract of employment. We do this by modifying the procedure used

previously to construct the random network isolation index (NI^R). Instead of randomly assigning all

workers in a census tract of employment to establishments in the tract, holding the size distribution of

establishments fixed, we instead ensure that workers are *also* assigned to their industry of employment.[35]

We then once again compute the average (across the simulations) simulated fraction of co-workers who

come from a worker's own neighborhood, denoting this NI^C, and we define the extent of "conditional

effective network isolation" to be:

$$[\{NI^O - NI^C\}/\{ NI^M - NI^R\}]\times 100 ,$$

where NI^R and NI^M are defined as before, without regard to industry. A conditional effective network

isolation index of zero (when $NI^O = NI^C$) would imply that all of the effective network isolation can be

attributed to networks that help workers find employment in specific industries, but not to establishments

within industries; that is, above and beyond the clustering of employment of neighbors in the same

industry, there is no clustering in the same establishment. Conversely, a network isolation index equal to

that of the (unconditional) effective segregation measure (when $NI^C = NI^R$) would imply that all of the

effective network isolation works to help individuals find jobs in specific establishments within

industries, and that industry, per se, plays no role in sorting.

For brevity, we report in Table 7 results on conditional network isolation only for groups that are

---

[35] The industry definitions that we use are the same eight industries reported in Tables 1 and 2: mining; construction; manufacturing; transportation; wholesale trade; retail trade; FIRE; and services.

"low-skilled" in the sense that they either have low levels of education (for whites and blacks) or poor English proficiency (for Hispanics). These are the groups for which networks appear most important. In column (1) we repeat the network isolation results for less-educated whites that we previously reported in Table 3, column (2). Then, in column (2), we report results where we condition the random network isolation index on industry. The observed network isolation index, the (unconditional) random isolation index, and the maximum network isolation index all remain the same, as the conditional random isolation index does not play a role in these calculations. The simulated conditional index is 5.46, which means that when workers are randomly assigned to establishments in the same industry in which they are observed to work (and of course the same census tract), on average 5.46 percent of their co-workers will come from the same residential neighborhood. This is higher than the unconditional random isolation index of 4.06, so that that the conditional effective network isolation index of 8.86 is somewhat smaller than the unconditional effective index of 11.29 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, even after conditioning on industry, effective network isolation is still relatively high, and dividing 8.86 by 11.29, we see that 78 percent of the effective network isolation remains even after we condition on a worker's industry. This demonstrates clearly that, at least at this level of industry detail, most of the (unconditional) effective network isolation for less-educated whites cannot be explained just by a mechanism whereby residential-based networks serve only to help workers find jobs in the same industries as their neighbors.

Columns (3) and (4) of Table 7 explore this issue for less-educated blacks. Column (3) repeats the unconditional network isolation indexes from Table 3, column (5), and column (4) reports the indexes conditional on industry. The results are very similar to those for less-educated whites. The conditional random network isolation index of 4.72 as reported in column (4) is somewhat larger than the unconditional random index of 3.51, so that the conditional effective network isolation index of 7.95 is somewhat smaller than the unconditional index of 11.52. And, as the last row of column (4) reports, assignment of workers to industry alone rather than to specific establishments only explains slightly more

28

than 30 percent of effective network isolation, leaving close to 70 percent remaining.

Finally, columns (5) and (6) turn to estimates of network isolation for low-skilled Hispanics, with low skill defined as poor English proficiency. The results parallel closely the qualitative results for less-educated whites and blacks. Column (5) repeats the unconditional results for less-skilled Hispanics previously reported in Table 5, column (2). Column (6) reports the conditional indexes, where the conditional effective network isolation of 21.86, when compared to the (unconditional) effective network isolation index of 30.98, implies that 71 percent of the assignment of workers from residential neighborhoods to establishments within a census tract cannot be explained by industry alone.

We conclude from these analyses conditioning on industry that the network isolation results we find are largely due to the assignment of workers to specific establishments, providing evidence in support of theoretical models of networks where networks serve to match workers to specific establishments.[36]

VI. Conclusions

We use matched employer-employee data for the United States to measure the importance of network effects in employment. 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 all workers employed in a particular census tract reside in a subset of 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

---

[36] We also compute isolation indexes where we condition on a worker's reported occupation, rather than industry. This is useful if one is concerned that our unconditional effective network isolation indexes are driven not by networks per se but by the sorting of workers of different skills into different neighborhoods and establishments, where occupation is a proxy for those skills. For less-educated whites, occupation (broken out into six categories) explains less than 7 percent of effective network isolation; for both less-educated blacks and low-English-proficiency Hispanics it explains somewhat more, approximately 17 percent. All in all, though, the results show that sorting by occupation does not come anywhere close to explaining our effective network isolation indexes.

networks operate along the lines of race, ethnicity, and skill, we consider separately the importance of labor market networks for whites, blacks, and Hispanics, and, within each group, the relative importance of networks across the skill distribution.

Our evidence is complementary to a fairly large 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 as referrals for 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 9.4 percent of the theoretical maximum amount of grouping that could be found in the data, and most of our analyses indicate that labor market networks are more important for blacks than for whites. For both whites and blacks, 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 some evidence that networks are more important in small establishments. Finally, our results provide evidence that networks operate to some extent along racial lines, as the link between residential location and the establishment of employment is stronger for blacks when we consider only co-workers of the same race. 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.

We also find that networks are more important for Hispanics than for blacks or whites, and among Hispanics, 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.

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 operating within establishments. Aside from further attempts to construct or obtain data to study the kind of network effects we examine in this paper, a number of other important questions remain. First, along what other dimensions of social interactions—aside from residence—do networks operate to cause individuals to work together, and what types of networks are most important? Among the possibilities are schools,[37] religious institutions, and community groups, as well as existing places of employment (from which workers may move to other jobs). Second, 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 as a result? Third, are minorities who have network relationships mainly with other minorities disadvantaged relative to those that have network relationships with whites? Fourth, do networks cross racial boundaries fluidly within residential neighborhoods—suggesting that residential segregation, per se, is the key source of race-based networks—or are there racial barriers to networks even within neighborhoods in which blacks and whites live? The answers to these questions may 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. The data demands for answering many of these questions are daunting.

---

[37] Indeed Bayer et al. (2005a) show that the type of network effects they study appear to be stronger for those with children of similar ages, which could reflect social interactions of families in schools.

References

Bayer, Patrick, Stephen Ross, and Giorgio Topa. 2005a. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." NBER Working Paper No. 11019.

Bayer, Patrick, Hanming Fang, and Robert McMillan. 2005b. "Separate When Equal? Racial Inequality and Residential Segregation." Unpublished paper, Duke University.

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." *Journal of Labor Economics*, Vol. 13, No. 2, April, pp. 309-34.

Calvó-Armengol, Antoni, and Matthew O. Jackson. 2007. "Networks in Labor Markets: Wage and Employment Dynamics and Inequality." *Journal of Economic Theory*, Vol. 132, No. 1, January, pp. 27-46.

Carrington, William J., and Kenneth R. Troske. 1997. "On Measuring Segregation in Samples with Small Units." *Journal of Business & Economic Statistics*, Vol. 15, No. 4, October, pp. 402-9.

Cutler, David M., and Edward L. Glaeser. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics*, Vol. 112, No. 3, August, pp. 827-72.

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*, Vol. 71-72, July-December, pp. 19-78.

Hellerstein, Judith K., and David Neumark. "Workplace Segregation in the United States: Race, Ethnicity, and Skill." Forthcoming in *Review of Economics and Statistics*.

Hellerstein, Judith K., David Neumark, and Melissa McInerney. "Spatial Mismatch vs. Racial Mismatch?" Forthcoming in *Journal of Urban Economics*.

Holzer, Harry J. 1998. "Why Do Small Establishments Hire Fewer Blacks Than Large Ones?" *Journal of Human Resources*, Vol. 33, No. 4, August, pp. 896-914.

Iceland, John, and Daniel H. Weinberg. 2002. "Racial and Ethnic Segregation in the United States: 1980-2000." U.S. Census Bureau, Census 2000 Special Reports. Available at http://www.census.gov/hhes/www/housing/housing_patterns/pdf/censr-3.pdf (viewed April 2008).

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

Jarmin, Ron S., and Javier Miranda. 2002. "The Longitudinal Business Database." CES Working Paper No. CES-WP-02-17.

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

Lang, Kevin. 1986. "A Language Theory of Discrimination." *Quarterly Journal of Economics*, Vol. 101, No. 2, May, pp. 363-82.

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

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics*, Vol. 118, No. 2, May, pp. 549-99.

Patel, Krishna, and Francis Vella. 2007. "Immigrant Networks and Their Implication for Occupational Choices and Wages." IZA Working Paper No. 3217.

Topa, Giorgio. 2001. "Social Interactions, Local Spillovers, and Unemployment." *Review of Economic Studies*, Vol. 68, No. 2, April, pp. 261-95.

Weinberg, Bruce A., Patricia B. Reagan, and Jeffrey J. Yankow. 2004. "Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data." *Journal of Labor Economics*, Vol. 22, No. 4, October, pp. 891-934.

Figure 1: Census Tracts, Chicago PMSA
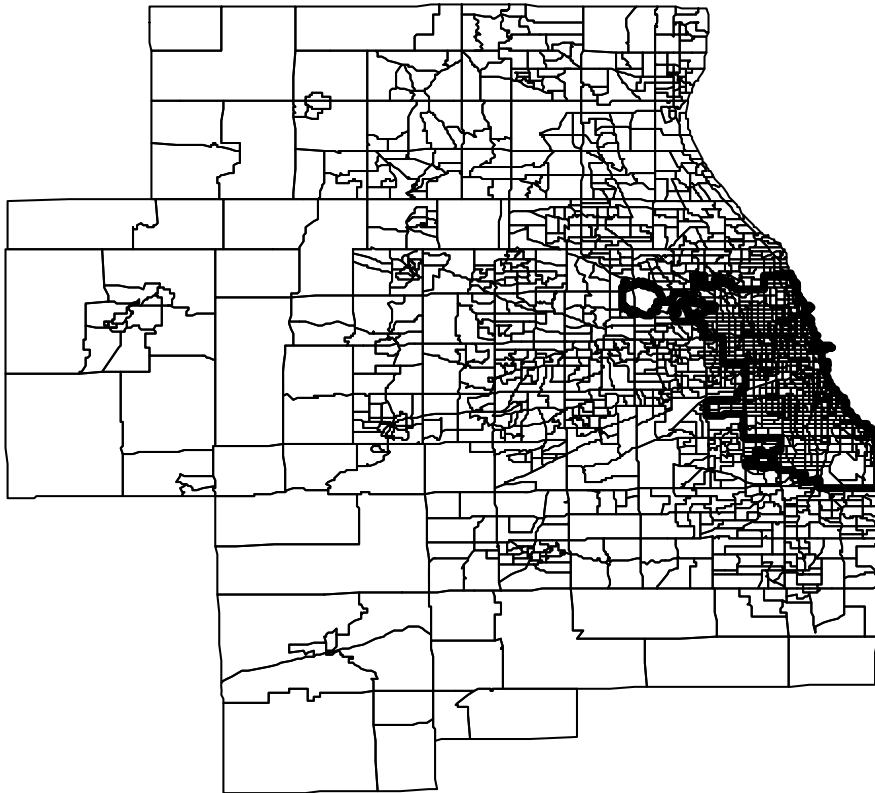
Source: U.S. Census Bureau, http://www.census.gov/geo/www.cob.tr2000.html (viewed May 5, 2008).
Note that Chicago city limits are highlighted.

Figure 2: Census Tracts, Chicago

Table 1: Individual-Level Means for Preliminary and Final Analysis Samples, White, Black, and Hispanic Workers

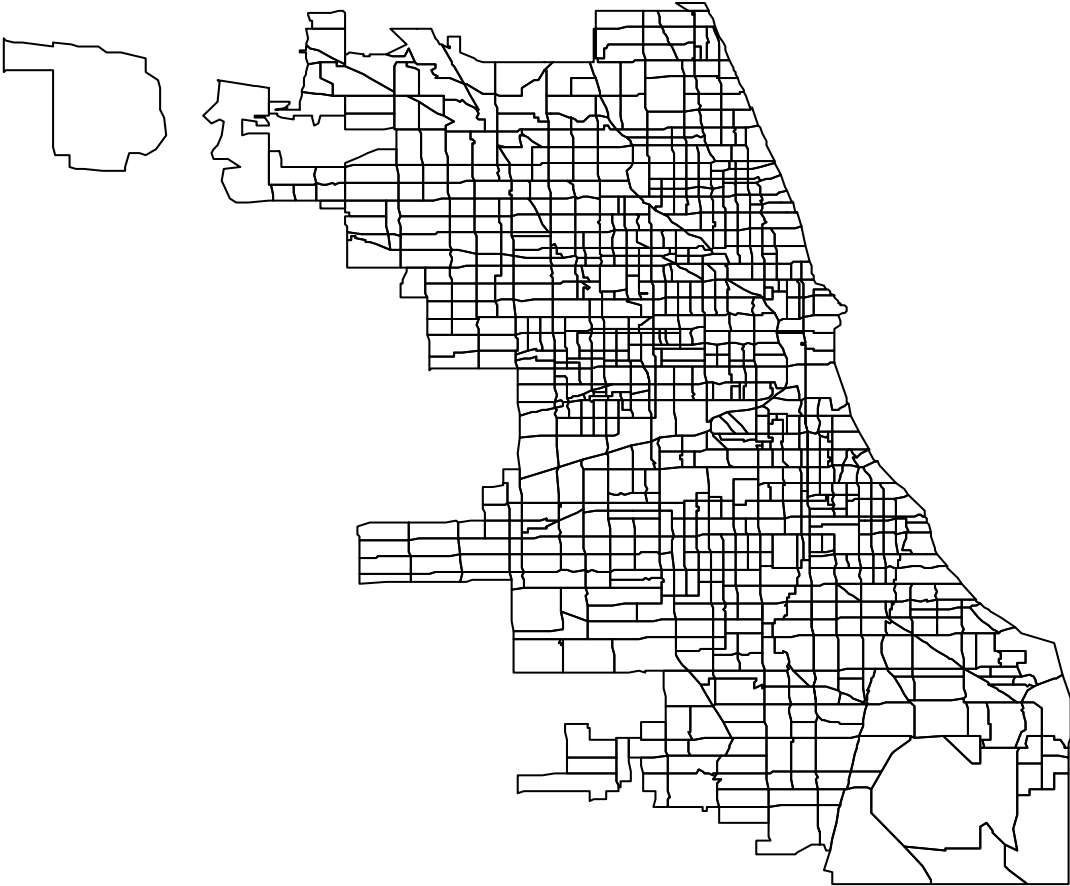|	2000 SEDF, workers eligible to be matched	2000 full DEED	2000 DEED, final analysis sample for whites	2000 DEED, final analysis sample for blacks	2000 DEED, final analysis sample for Hispanics
	(1)	(2)	(3)	(4)	(5)
Age	39.18	39.60	40.14	38.32	35.65
	(13.07)	(12.54)	(12.37)	(11.49)	(11.48)
Female	.46	.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
Number of 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

Notes: In addition to restricting by race and ethnicity, the three additional restrictions imposed in going from column (2) to columns (3) through (5) are: the individual must live and work in same MSA/PMSA; there must be at least two workers matched to establishment; and there must be at least one other establishment with two matched workers in the census tract.

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

	2000 full DEED	2000 DEED, final analysis sample for whites	2000 DEED, final analysis sample for blacks	2000 DEED, final analysis sample for Hispanics
	(1)	(2)	(3)	(4)
Total employment	49.82	102.82	412.43	258.09
	(368.46)	(344.85)	(887.34)	(670.16)
Total employment (pseudo-median)	15	35	154	84
Total matched workers (pseudo-median)	1	3	2	2
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:				
North East	.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 ($1000)	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 employees matched	.16	.14	.05	.07
Multi-unit establishment	.40	.51	.80	.61
N	1,254,718	329,943	21,872	30,343

Notes: See notes to Table 1.

Table 3: Network Isolation for Whites and Blacks, Overall and by Education

	Whites			Blacks		
	All	High school degree or less	More than high school degree	All	High school degree or less	More than high school degree
	(1)	(2)	(3)	(4)	(5)	(6)
Network isolation index, observed, NI^O	7.87	10.56	6.51	5.29	7.42	3.90
Simulated random network isolation index, NI^R	2.97	4.06	2.41	2.58	3.51	1.97
$NI^O - NI^R$	4.90	6.50	4.10	2.71	3.90	1.93
Maximum possible network isolation index, NI^M	54.84	61.62	52.32	31.60	37.40	29.55
Effective network isolation, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100$	*9.45*	*11.29*	*8.21*	*9.35*	*11.52*	*7.01*
N	1,675,412	561,370	1,114,042	97,967	38,754	59,213
# place of work tracts	26,470	25,690	26,299	4,490	4,301	4,350
# residential tracts	46,764	43,469	45,666	21,623	13,598	18,299
Mean establishments/tract	129.6	114.5	137.2	22.2	19.8	23.8
Mean matched workers/establishment	38.4	24.4	45.4	18.6	15.6	20.5
Mean number of workers in tract of employment from same tract of residence	9.4	11.2	8.5	2.6	2.9	2.4

Notes: The calculation is described in the text. NI^O is the average fraction of a worker's co-workers (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.

Table 4: Network Isolation for Blacks and Whites, for Consistent Samples of Establishments, and Ignoring Race

| | All whites | All blacks | Establishments located in tracts in both black and white samples | | Network isolation based on blacks and whites |
| | | | Whites | Blacks | All blacks |
	(1)	(2)	(3)	(4)	(5)
Network isolation index, observed, NI^O	7.87	5.29	5.68	5.25	3.35
Simulated random network isolation index, NI^R	2.97	2.58	1.48	2.55	1.41
$NI^O - NI^R$	4.90	2.71	4.20	2.71	1.93
Maximum possible network isolation index, NI^M	54.84	31.60	61.43	32.35	42.34
Effective network isolation, $[(NI^O - NI^R)/(NI^M - NI^R)] \cdot 100$	*9.45*	*9.35*	*7.00*	*9.08*	*4.73*
N	1,675,412	97,967	845,290	94,210	160,456
# place of work tracts	26,470	4,490	4,122	4,122	16,818
# residential tracts	46,764	21,623	42,533	21,459	49,586
Mean establishments/tract	129.6	22.2	229.7	23.0	125.7
Mean matched workers/establishment	38.4	18.6	46.6	17.7	52.7
Mean number of workers in tract of employment from same tract of residence	9.4	2.6	11.4	2.6	5.7

Notes: See notes to Table 3.

Table 5: Network Isolation for Hispanics, Overall and by Skill and Immigrant Status

	All	Poor English skills	Good English skills	Immigrant	Non-immigrant
	(1)	(2)	(3)	(4)	(5)
Network isolation index, observed, NI^O	11.22	20.27	8.89	15.27	7.25
Simulated random network isolation index, NI^R	3.08	4.99	2.59	3.78	2.39
$NI^O - NI^R$	8.14	15.28	6.30	11.49	4.86
Maximum possible network isolation index, NI^M	41.15	54.32	38.65	46.14	38.06
Effective network isolation, $[(NI^O - NI^R)/(NI^M - NI^R)] \cdot 100$	*21.37*	*30.98*	*17.47*	*27.12*	*13.62*
N	110,235	22,538	87,697	54,529	55,706
# place of work tracts	5,059	3,683	5,043	4,680	4,754
# residential tracts	20,716	7,703	19,920	13,802	16,548
Mean establishments/tract	59.7	67.6	57.6	66.1	53.4
Mean matched workers/establishment	9.1	6.9	9.7	7.5	10.7
Mean number of workers in tract of employment from same tract of residence	2.7	3.3	2.5	2.9	2.5

Notes: See notes to Table 3.

Table 6: Network Isolation for Whites, Blacks, and Hispanics Working in Small Establishments (50 employees or less)

	Whites	Blacks	Hispanics
	(1)	(2)	(3)
Network isolation index, observed, NI^O	15.76	12.76	22.79
Simulated random network isolation index, NI^R	4.49	4.43	5.74
$NI^O - NI^R$	11.27	8.33	17.05
Maximum possible network isolation index, NI^M	59.54	26.20	40.71
Effective network isolation, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*20.47*	*38.26*	*48.75*
N	527,430	8,706	21,952
# place of work tracts	22,162	1,097	2,093
# 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

Notes: See notes to Table 3.

Table 7: Network Isolation for Low-Skilled Workers, Conditional on Industry

	Whites, high school or less		Blacks, high school or less		Hispanics, poor English skills	
	Uncond.	Cond.	Uncond.	Cond.	Uncond.	Cond.
	(1)	(2)	(3)	(4)	(5)	(6)
Network isolation index, observed, NI^O	10.56	10.56	7.42	7.42	20.27	20.27
Simulated random network isolation index, NI^R	4.06	4.06	3.51	3.51	4.99	4.99
Simulated conditional network isolation index, NI^C		5.46		4.72		9.48
$NI^O - NI^C$		5.10		2.69		10.78
Maximum possible network isolation index, NI^M	61.62	61.62	37.40	37.40	54.32	54.32
Effective network isolation, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100$	*11.29*		*11.52*		*30.98*	
Conditional effective network isolation, $[(NI^O - NI^C)/(NI^M - NI^R)]\cdot100$		*8.86*		*7.95*		*21.86*
Share of effective isolation unexplained		*78.46*		*69.01*		*70.57*

See notes to Table 3. There are 561,370 whites with a high school degree or less; 38,754 blacks with a high school degree or less; and 22,538 Hispanics with poor English skills. The industries are mining; construction; manufacturing; transportation and warehousing; wholesale trade; retail trade; finance, insurance, and real estate; and services.