NEIGHBORS AND CO-WORKERS:
THE IMPORTANCE OF RESIDENTIAL LABOR MARKET NETWORKS[*]

Judith K. Hellerstein
University of Maryland and NBER

Melissa McInerney
The College of William and Mary

David Neumark
UCI, NBER, and IZA

College of William and Mary
Department of Economics
Working Paper Number 101

March 2010

COLLEGE OF WILLIAM AND MARY
DEPARTMENT OF ECONOMICS
WORKING PAPER 101
March 2010

NEIGHBORS AND CO-WORKERS:
THE IMPORTANCE OF RESIDENTIAL LABOR MARKET NETWORKS

Abstract

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

Judith K. Hellerstein
Department of Economics and MPRC,
University of Maryland,
and NBER
hellerst@econ.umd.edu

Melissa McInerney
Department of Economics,
College of William and Mary
Williamsburg, VA 23187-8795,
mpmcinerney@wm.edu

David Neumark
Department of Economics,
UCI,
NBER, and IZA
dneumark@uci.edu

I. Introduction

Racial and ethnic disparities in labor market outcomes in the United States are well documented. In addition to differences in wages and employment levels, there is evidence of workplace segregation by race and ethnicity (e.g., Hellerstein and Neumark, 2008). There is also strong evidence of residential segregation by race and ethnicity in the United States (e.g., Iceland and Weinberg, 2002), and this segregation is correlated with poor labor market outcomes for minority groups, especially the low-skilled among them (e.g., Cutler and Glaeser, 1997).

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

Our goal in this paper is to provide evidence on the existence and importance of labor market

---

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

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

We specify and implement a test for the importance of residence-based labor market networks in determining the establishments at which people work, using matched employer-employee data at the establishment level. Our measure of labor market networks captures the extent to which employees of a business establishment come disproportionately from the same sets of residential neighborhoods (defined as census tracts), relative to the residential locations of other employees working in the same census tract but in different establishments. We measure the importance of network effects for groups broken out by race (black and white), Hispanic ethnicity, and various measures of skill (education, English language proficiency, and immigrant status). Finally, we provide evidence on the stratification of networks, asking whether the networks we study are race-based, operating more strongly within than across races.

The workplace segregation of employees across establishments by residential neighborhood that we measure arises from residential proximity that captures the "network connectedness" important to the flow of labor market information between specific employers, their employees, and potential hires. In particular, we first identify all establishments within each census tract in our sample. Because we have matched employer-employee data, we have a sample of workers in each establishment, and we know the census tracts in which they live. We compute the share of an individual's co-workers who are his or her residential neighbors, relative to the share that would result if the establishment hired workers randomly

2

from the geographic areas where *all* individuals who work in the census tract reside. Residence-based networks would predict that the share of neighbors among a worker's co-workers would be higher – and possibly much higher – than would result from the random hiring process.

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

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

Overall, we find that residence-based labor market networks play an important role in hiring. For whites, we find that the grouping of workers from the same neighborhoods in the same business establishments is about 10 percent of the maximum grouping that could occur. These networks are largely due to the assignment of workers to specific establishments, rather than the sorting of workers in local labor markets by skill groups, occupations, or industries. It is the case, though, that networks are more important for less-skilled whites, and in small establishments. We find similar results for blacks, and some indications that residence-based labor market networks are more important for blacks than for whites. Moreover, the evidence indicates that the networks we study are partly race-based, operating

3

more strongly within than across races. We find that residence-based networks are even more significant for Hispanics, for whom the grouping of workers from the same neighborhoods in the same business establishments is about 22 percent of the maximum. And among Hispanics, these networks play a larger role for immigrants and those with poor English skills.

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

II. Relation to Existing Literature

Ioannides and Datcher Loury (2004) review a large body of evidence that is consistent with labor market networks, emphasizing survey evidence of widespread reliance on friends, relatives, and acquaintances to search for and find jobs. They conclude that there is little difference between blacks and whites in the use of informal contacts in job search, but that rates of use of informal contacts are higher for low-educated workers than for high-educated workers, and that there are substantially higher rates of use of informal contacts among Hispanics than other groups. 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 (Kmec, 2007).[2]

---

[2] Kasinitz and Rosenberg (1996) provide case study evidence consistent with this type of network. They study the Red Hook section of Brooklyn, an area of high unemployment that is populated largely by low-income blacks (and to some extent Hispanics), but with a large number of local jobs in the shipping industry. They find that many employers hire workers almost exclusively from outside of Red Hook, recruiting employees via social networks

Bayer et al. (2008) move beyond survey evidence and look for evidence of network effects among neighbors using confidential Long-Form Census data on Boston-area workers. They find that two individuals living on the same census block are one-third more likely to work on the same census block than are two individuals living in the same block group but not on the same block. As long as informal networks are stronger within blocks than within block groups, this evidence is consistent with residence-based labor market networks affecting hiring, the same type of network connections that we study.[3]

There are, however, some limitations of the evidence in Bayer et al.'s study, as well as in the broader literature, upon which we try to improve. First, most of the existing work on networks does not relate employment in the same business to network connections between employees of that business. This is certainly true of most of the evidence based on surveys of workers. Similarly, the data used by Bayer et al. contain no information on the exact establishment in which the workers work, so that two individuals who work on the same census block may work for different employers, particularly for blocks in the central city that contain multiple employers.[4] One exception is Granovetter's seminal study (1974), which finds that informal contacts were often employed in the company where the job held by a respondent had opened up. Our data and approach permit us to tie network connections to employment in a particular business establishment, and thus provide evidence more directly related to the hypothesis that labor market networks reduce search frictions on the part of employers, as in Montgomery's (1991) model. To the extent that networks reduce search frictions (as in the model), they enhance efficiency, although if networks simply act to allocate "good jobs" to some people (as in Schmutte, 2009) and not others they play only a distributional role.

within specific (non-black) ethnic groups.

[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, there are on average 2.58 establishments per block, which is likely an undercount by a factor of about four, given that we only observe in our data a subsample of establishments. This raises questions about Bayer et al.'s assumption that workers employed on the same block "work with" one another or "work together" (e.g., 2008, p. 26). Petersen et al. (2000) analyze data from one firm, and report that a large share of hires indicate that a referral from a friend was responsible for the match. However, the data do not indicate whether the friend who made the referral was a current employee of the company, so there is no reason to infer that the company used referrals from current employees to obtain information about potential hires.

Second, previous work is unable to speak convincingly about differences in the importance of networks by race and Hispanic ethnicity, and among subgroups of Hispanics.[5] In the case of Hispanic workers, in particular, networks may compensate for less-developed formal hiring channels in predominantly immigrant communities, attributable in part, perhaps, to hiring of undocumented workers. (Note that if residence-based networks are important for a group, then some degree of residential segregation may be advantageous.) Some of the survey evidence discussed in Ioannides and Datcher Loury (2004) suggests differential use of networks across racial and ethnic groups, but a good deal of the most-cited evidence cannot be used to address this question because it either covers narrow subsets of workers or a single firm. Because of the large data set we construct, and the explicit linkage of workers to establishments that the data set permits, we can more fully explore the importance of networks that assign workers to establishments and differentials in the importance of these networks by race, by Hispanic ethnicity, and across subgroups of Hispanics.

The final advantage of our approach is that it lets us assess the extent to which networks are race-based – operating more strongly within than across races. Theoretical work has suggested that racial stratification of networks may be important in perpetuating racial differences in labor market outcomes.

Our study does share two limitations with previous work, however. First, as in Bayer et al., because the data contain no direct measures of network contacts, such as information on whether an individual found work directly via a network connection in their neighborhood, our measure of networks remains an indirect one.[6] Second, as with much of the previous literature that ties networks to neighbors or to friends, the data contain no direct information on how or why the networks of neighbors that we measure are formed, or what makes some networks stronger than others. We can provide some

---

[5] Bayer et al. (2008) study data only for Boston and focus mostly on whites.
[6] In contrast, Laschever (2009) studies the employment experiences of Veterans who served together in World War I, who clearly shared war-time experience and most likely had subsequent interactions through various Veterans' groups. Laschever finds that there is a positive correlation between employment (and unemployment) of those who served together. A similar approach is taken by Cingano and Rosolia (2009), who study the re-employment experiences of workers displaced from the same firm who previously worked together; they find similar evidence of correlated experiences. Like much of the other evidence on networks, these studies do not have evidence regarding those in the network working for the same employer.

6

suggestive evidence along these lines, but our results should primarily be seen as providing descriptive evidence on the strength of networks at the level of residential neighborhoods.

III. The 2000 DEED

The analysis in this paper is based on matched employer-employee data from the 2000 DEED. We have constructed a similar data set for 1990 and described its construction in detail elsewhere (Hellerstein and Neumark, 2003). The construction of the 2000 DEED follows the same procedures, so in this section we simply provide a brief overview. Workers in the DEED are drawn from the Sample Edited Detail File (SEDF), which contains all individual responses to the 2000 Decennial Census of Population one-in-six Long Form. The establishments are drawn from the Census Bureau's Business Register list (BR) for 2000; the BR is a database containing information for all business establishments (with one or more employees) operating in the United States in each year, which is continuously updated (see Jarmin and Miranda, 2002). The Census Bureau uses the BR as a sampling frame for its Economic Censuses and Surveys. The BR contains the name and address of each establishment, geographic codes based on its location, its four-digit SIC code, and a unique establishment identifier.

Households receiving the Long Form were asked to report the name and address of the employer in the previous week for each employed household member. This name and address is stored in the "Write-In" file but not captured in the SEDF. We use employer names and addresses in the Write-In file to try to match each worker to the BR. Because the name and address information on the Write-In file is 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, these identifiers 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 difficult because employers' names and addresses can be recorded differently on the two files. We match workers and establishments based on the Write-In file using MatchWare – a specialized record-linkage program. MatchWare includes a name and address

standardization mechanism (AutoStan), and a matching system (AutoMatch). We first use AutoStan to standardize employer names and addresses across the Write-In file and the BR, to eliminate differences in how the same data items 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, so that we can match on various combinations of these components. The second step is to select and implement the matching specifications, using AutoMatch's probabilistic matching algorithm to account for missing information, misspellings, and even inaccurate information. The AutoMatch software 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 1.28 million establishments, accounting for 29.1 percent of workers in the SEDF and 22.6 percent of establishments in the BR.[7] We impose additional sample restrictions for our analysis, which we discuss following the explanation of our empirical methods in the next section.

IV. Measuring the Importance of Networks

We first study the importance of residence-based networks for whites, who make up the majority

---

[7] 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.)

of our sample. We perform a series of analyses of whites to examine alternative interpretations of the findings and to identify factors associated with variation in the importance of these networks. By focusing first on whites we avoid confounding our results with other potential influences on the sorting of workers into firms by race and ethnicity, such as labor market discrimination. However, we then examine the importance of networks for these other groups and subsamples of them.[8] In explaining our methods below we refer to whites, for simplicity; the analysis is done in the same way for the other groups.

To calculate the importance of residence-based labor market networks for the white workers in our sample, we first compute for each white worker in our sample the percentage of white workers with which that worker works who come from the same residential neighborhood as that worker. We exclude the individual worker from this calculation, since it is meaningless to say that a person is his or her own neighbor. This calculation requires a sample restriction to establishments where we observe at least two white workers. We average this percentage across all the white 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. The superscript "O" on the network isolation index emphasizes that this is the fraction of a white worker's co-workers who are *observed* in our estimation sample to be residential neighbors.[9]

---

[8] In only one analysis – discussed later – do we combine groups.

[9] The phrase "network isolation index" borrows from the sociology literature that measures 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. "Segregation" and "isolation" have the same meaning – that members of a particular group tend to interact with other members of the same group, rather than members of other groups. Of course, many other segregation measures have been used to measure the agglomeration of individuals of similar (usually two) types together in society. Measuring network segregation via the simple concept of isolation is transparent and compelling in our view.

Hellerstein and Neumark (2008) measure workplace segregation in a similar manner to this paper, considering the fraction of an individual's co-workers who are of a specific race, ethnicity, or skill group. That paper provides a detailed discussion of the advantages of measuring segregation this way rather than via other indexes such as the Duncan index. In the present study, we chose to use an isolation index rather than a Duncan Index for two reasons. First, we would have had to use a multidimensional version of the Duncan index. Such an index exists, but it is much less widely used in measuring segregation. Second, Duncan indexes are invariant to growing populations of a given group. For example, in measuring black-white residential segregation across census tracts, the Duncan index is invariant to a doubling of the black population in each census tract. It is not clear that invariance is a desirable property in the context of our study. For example, if a residential census tract becomes home to more poor English speaking Hispanics, it seems reasonable to allow the size of the resulting network in that census tract to grow and to contribute to higher network isolation.

Formally, the observed network isolation index is:

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

where there are N workers, indexed by i and j for all possible pairs of workers. $I^R(i,j)$ is an indicator for whether workers i and j live in the same residential neighborhood, and $I^W(i,j)$ is an indicator for whether i and j work in the same establishment. The sums in the numerator and denominator are taken over all N workers other than the worker i. Their ratio is the share of co-workers with whom each worker is co-resident. This ratio is then averaged over all workers and multiplied by 100.

To operationalize this index, we define residential neighborhoods as census tracts. There are a few reasons why this definition seems sensible. 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 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 and might make it more likely that co-residents interact. Third, most census tracts are relatively small, so it is reasonable to think that it is quite possible that many census tract residents have contact with each other, if not "over the back fence," then at parks, schools, churches, stores, business, and other institutions.

To provide some idea of the size of census tracts, Figure 1 displays detailed maps of these tracts for Chicago. For the PMSA, 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, the median was 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.[11]

There are a number of factors that might lead to residential neighbors working in the same

---

[10] See http://www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed April 21, 2008).
[11] The two large tracts within the city limits are O'Hare Field and another mainly industrial tract at the southern edge of the city.

establishment, which have nothing to do with residentially-based labor market networks. We want to do everything we can to preclude these factors from influencing our measurement of 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 transportation infrastructure should not materially affect the distribution of workers across establishments in the tract.

Second, people may have heterogeneous tastes such that those who choose to work in similar workplaces have similar preferences over neighborhoods. We would expect, however, that individuals who have greater economic resources are better able to exercise choice about both where to live and where to work. We therefore are careful to disaggregate our analyses across groups that a priori have differential abilities to make choices about where to live and work, and we show that those who have more resources, on average, are least likely to work with their neighbors.

Third, if there is labor market discrimination against some groups of workers, those workers generally will be sorted into jobs with non-discriminatory employers. Because of the potential for discrimination against blacks and Hispanics, in particular, we focus much of our analysis on white non-Hispanics before examining the extent of networks for these other groups.

Fourth, some clustering of residential neighbors into establishments would occur even if workers are assigned randomly to establishments, given that people tend to work relatively close to where they live. Observed network isolation (NI^O) alone therefore does not give a true sense of whether residential networks exist, because it can be greater than zero simply owing to neighbors ending up working together by chance. We therefore compute the extent of network isolation that occurs due to randomness, denoted NI^R.[12] In order to calculate NI^R, within a census tract we randomly assign workers to establishments,

---

[12] 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.

11

ensuring that we generate the same size distribution of establishments (in terms of matched workers) within a census tract as we have in the sample. Note that we generally do this using data only on individuals in the same racial, ethnic, or skill group for which we are trying to characterize the importance of networks. We conduct this simulation 100 times, and compute the random network isolation index, NI^R, as the mean over these simulations.[13]

The idea behind the random allocation is as follows. Workers, through their behavior, reveal the geographic areas in which they choose to work (aside from the clustering at specific establishments). These decisions may be based on proximity, public transportation, highway exits, industry, etc. Having made these choices, some workers from the same neighborhood will end up at the same establishment even if there are no network connections between them. Our randomization is meant to capture this component of the clustering and to subtract it out from observed clustering, yielding a remainder that arises from the systematic processes that determine the establishments at which people work.

The difference $NI^O - NI^R$ measures network isolation above and beyond that which occurs randomly, which we refer to as the "network isolation difference." This difference captures the excess presence of co-residents in a worker's own establishment relative to the presence of co-residents in the same work location. This computation requires a second sample restriction – that census tracts of employment include at least two establishments with two black workers. Otherwise, if the only black workers in a census tract work together in one establishment, we cannot distinguish the effect of residence-based labor market networks from random clustering.

While NI^R provides a measure of the lower bound of the extent to which workers work with neighbors,[14] it is also important to know what the upper bound of network isolation could be in our data given the distribution of individuals across residential census tracts and the size distribution of

---

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.

[13] Not surprisingly, in our very large sample the random network isolation measures are very precise; in all cases the standard deviations were trivially small.

[14] This is not the true lower bound, of course. A social planner could force the network isolation index to go below NI^R by deliberately disbursing neighbors across establishments to minimize neighbor-to-neighbor contact in the workplace, but this counterfactual is one that seems unlikely to be generated in practice.

establishments in each census tract. Because establishments within a particular census tract of employment often contain more workers in them than there are workers from any particular census tract of residence represented among those establishments, the upper bound of network isolation is unlikely to ever reach 100 percent (full segregation, such that workers work *only* with their neighbors).

There may be individual instances of workplace census tracts in our data where we could compute the exact maximum network isolation index, but there is no known general method for solving for the maximum index in all cases in our data. We instead approximate the maximum network isolation through a "greedy" algorithm. For a given census tract 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.[15] We would expect this maximum network isolation index to be less than 100 percent.[16,17] We do this for every census tract in

---

[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] 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. 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). This leaves a 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, yielding a maximum network isolation index of $\{(5/9) \cdot 1 + (2/9) \cdot 1 + (2/9) \cdot 0\} = 7/9$.

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 index NI^M. The difference ($NI^M - NI^R$) then provides a measure of the maximum extent to which networks could lead to workplace sorting above and beyond sorting that would occur randomly.

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

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

which we call the "effective network isolation index." It measures what share of the maximum possible network isolation that could occur in the data actually does occur in the data.[18] As such, it provides a natural scaling for the importance of networks formed by residential networks in determining the establishments in which people are employed, and because it is a scaled measure it can be compared across different samples with different possible lower and upper bounds of network isolation. For these reasons, the effective network isolation index is our preferred measure of network isolation. Nonetheless,

---

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

[18] Our analysis has some parallels to Oyer and Schaefer's (2009) study of the concentration of lawyers from particular schools in different law firms, which may be due to networks. Like us, they consider deviations of the observed concentration of students from particular law schools relative to what would occur randomly, although measured using the Ellison-Glaeser (1997) index of concentration. They do this for exactly the same reason: "under random selection of attorneys by firms, we would still observe some concentration in realized law-school shares" (p. 9). There are some differences in our analysis, however. First, we randomly assign workers to establishments within the Census tract in which they work, rather than nationally; that is we do the randomization repeatedly for a very large number of Census tracts. Second, when we do the randomization we preserve the size distribution of establishments, because otherwise changes in this size distribution could themselves affect the network isolation measures; we are not aware of how this would translate into the Ellison-Glaeser framework. And third, because of the nature of what we are trying to measure, there is a maximum amount of network isolation that could occur, of which (as just described) we take account by the rescaling in the above equation.

we also report and discuss results for NI^O, NI^R, and NI^M separately.

Bootstrap methods are used to assess the statistical significance of our estimates of the importance of residence-based networks themselves, of differences across different demographic groups, and of different analyses (with and without conditioning), etc.[19] We always find that the effective network isolation measures we report are strongly statistically significantly different from zero, and are estimated very precisely. The tables report the statistical significance of the important comparisons across estimates; because the estimates of the effective network indices are quite precise, almost all differences across columns and tables in the estimated effective network isolation measures are also statistically significant.

V. Sample Characteristics and Restrictions

Table 1 reports descriptive statistics for the matched workers from the SEDF (column (1)) and from each of our DEED analysis samples (columns (2)-(5)). We make several restrictions to arrive at the samples used for our analysis. First, we only consider workers who live and work in the same Metropolitan Statistical Area/Primary Metropolitan Statistical Area (MSA/PMSA). Second, in order to be able to identify a race- or ethnic-based workplace network, we exclude from the sample workers with no co-workers in the same establishment who are of the same race or ethnicity. Finally, we retain only workers 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, so that we can meaningfully consider the distribution of workers across establishments in a census tract.

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. The differences in business address information likely reflect weaker labor market attachment among less-skilled workers,

---

[19] Confidence intervals are constructed from bootstrap replications in which we bootstrap the entire sample, and then compute each of the measures in the tables that follow. Thus, we obtain bootstrap replications of the differences between any pair of effective isolation measures within or across tables.

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. As shown in the last eight rows of the table, there is some over-representation of

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

although the over-representation is not severe.

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 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; this is reflected in the medians for total employment in the establishment as recorded in the

BR, and how these medians increase for the different subsamples we analyze.[20] Later in the empirical

analysis we consider the possible ramifications for the estimates of these consequences of our sample

selection rules.

VI. Network Isolation Results

VI.1. Results for Whites

Table 3 presents results for the approximately 1.7 million white workers in our sample, who work

in 26,470 unique census tracts and live in 46,764 census tracts.[21] 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 index (NI^O) for the full sample of whites, reported in column (1),

is 7.87, indicating that, on average, a white worker works in an establishment where 7.87 percent of

his/her white co-workers live in the same census tract as that worker. When we randomize workers in

this sample across establishments within census tracts, the random network isolation index (NI^R) is 2.97,

---

[20] 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.

[21] This is out of a total of about 65,000 census tracts in 2000.

and the network isolation difference (the difference between the two) is 4.90. Our calculated maximum possible network isolation index (NI^M) indicates that white workers in our data could work in establishments in which at most 52.06 percent of their co-workers are neighbors. This is well below an average of 100 (perfect sorting of workers by residential neighborhoods into establishments), because in many of our census tracts there are establishments with more workers than are drawn from any particular census tract residential neighborhood. When we rescale the network isolation difference by the maximum network isolation that occurs beyond randomness, we recover an effective network isolation index of 9.99. That is, about 10 percent of the maximum amount to which residential networks (at the census tract level) *could* contribute to the sorting of workers into establishments *is* actually reflected in the sorting of workers into establishments. Whether this is a large number or a small one is a subjective matter, and there is little with which to compare it given the sparseness of empirical evidence on the importance of labor market networks. To us, however, it seems like a large number suggesting that residential labor market networks are quite important.

This evidence could reflect sorting of workers by both neighborhoods and establishments according to skill, rather than residential networks. For example, one establishment in a census tract may hire only less-skilled workers who tend to live together in a neighborhood where housing is cheap, while a second establishment hires only more-skilled workers who tend to live together in a different neighborhood with more expensive housing.[22] To evaluate this alternative, columns (2)-(4) report network isolation results conditional on various proxies for skill, based on conditional network isolation indexes that simulate network isolation while holding fixed the skill distribution of workers within establishments.

To construct these conditional indexes, we modify 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 any job in any establishment in the tract, holding the size distribution of establishments

---

[22] In Hellerstein and Neumark (2008) we report evidence of some segregation of workers across establishments based on skill levels. Bayer et al. (2005) provide evidence of residential segregation by education for blacks.

17

fixed, we instead ensure that workers are assigned (randomly) only to jobs that are observed in the real data to be held by a worker with the same level of skill. We then once again compute the average (across the simulations) simulated fraction of co-workers who come from a worker's own neighborhood, denoting this NI^C, and we define the "conditional effective network isolation" index to be:

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

where NI^R and NI^M are defined as before (i.e., without regard to skill level). When $NI^O = NI^C$, the conditional effective network isolation index is zero, implying that all of the observed network isolation is actually attributable to the sorting of workers by residential census tract into establishments on the basis of skill alone; that is, above and beyond the clustering of employment of neighbors in the same skill group, there is no clustering in the same establishment. Conversely, when $NI^C = NI^R$, the conditional and unconditional network isolation indexes are equal, implying that all of the effective network isolation comes from networks helping individuals find jobs in specific establishments, and that skill (as we measure it) plays no role in sorting.

In column (2) we construct the index conditional on two education categories: high school degree or less, and more than high school. The simulated conditional network isolation index (NI^C) that results is 3.02, only a little higher than the random (unconditional) network isolation index in column (1) of 2.97. As a result, the conditional network isolation difference ($NI^O - NI^C$) is 4.84, and the conditional effective network isolation index is 9.86, very similar to the unconditional index of 9.99. Thus, even conditional on the assignment of workers to establishments on the basis of whether or not they have attended college at all, 9.86 percent of the maximum amount of clustering of workers into establishments on the basis of census tract of residence is observed in the data. In column (3) we refine the education groups to condition on four different levels of educational attainment: less than high school, high school degree, some college, and bachelor's degree or more. Here the simulated conditional isolation index is slightly bigger, at 3.13, but the conditional effective network isolation index is still 9.65, so that overall, conditioning on education only reduces the effective isolation index by less than 4 percent.

In columns (4) and (5) we use two different proxies for skill. In column (4), we condition on six

18

occupation categories. The results again do not suggest much of a role for skill sorting in explaining effective network isolation: the simulated conditional network isolation index is 3.25, so that the network isolation difference is 4.61, and the resulting effective network isolation index 9.40. In column (5), we use a measure of house prices as a proxy for skill, given that skill can manifest in this dimension of the accumulation of wealth.[23] The results of this exercise suggest a slightly larger role for this measure of skill in explaining network isolation, but again the results are similar; the simulated conditional network isolation index is 3.49, and the resulting effective network isolation index is 8.92.

Across columns (2)-(5), then, we find that sorting of workers across census tracts by a variety of proxies for skill explains very little (at most 11 percent) of the overall (unconditional) effective network isolation.[24] Because our measures of skill are only proxies and are not all-inclusive, this evidence is by nature not definitive. Nonetheless, given that the results are robust across three different dimensions of skill (education, occupation, and housing values), we view the results as providing strong evidence that skill does not play a major role in driving our measure of network isolation.[25]

We can also examine whether network isolation varies *across* skill groups, which we do in the last two columns of Table 3. Column (6) reports results for whites with at most a high school education, computed for the sample of white workers who have no more than a high school education. The observed network isolation index, averaging across the sample of low-educated white workers the fraction of each

---

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

[24] The Bayer et al. (2008) study addresses the issue of sorting on skill in a different manner. They argue 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.

[25] It is possible that our results are driven not by networks, but by family connections, where family members, especially spouses, might work together and live together in the same household. In unreported results, we restrict the sample to white males only, so that spouses are excluded. The effective network isolation index declines by about one-third (from 9.99 to 6.44) when we do this. However, this sample selection rule also has other substantial effects on the sample (reducing mean establishments per tract and mean matched workers per establishment by about 40 percent, as well as, of course, dropping females), so the results are not strictly comparable.

19

individual's white low-educated co-workers who live in that individual's residential census tract, is 10.95, somewhat higher than the 7.87 number for the full sample of whites as reported in column (1). We calculate the simulated network isolation index by randomly assigning workers to jobs held by low-educated white workers in establishments in their census tract of work, and then, for low-educated white workers, calculating the fraction of simulated co-workers who live in the same residential census tract. The random network isolation index is 4.81, with the net effect that the network isolation difference of 6.15 is also higher than for the full sample. Finally, the calculated maximum possible network isolation index for this low-education subsample is 52.01, very similar to that of column (1). Thus, the effective network isolation index is 13.02, 30 percent higher than the corresponding figure in column (1).

Column (7) reports results from the same exercise for whites with more than a high school degree. The resulting measures of network isolation are smaller than in the full sample; the network isolation difference for this group is 3.66, and the effective network isolation index is 8.82. Comparing columns (1), (6), and (7), the results suggest that residence-based networks are more important for low-educated whites, as has been suggested in previous surveys of workers' use of informal contacts by education. This difference by skill may arise because residence-based networks are more important for local labor markets, which are more significant for low-skilled than for high-skilled workers. The evidence of residence-based networks for both schooling groups also suggests, like the evidence in columns (2) through (5), that the full-sample results in column (1) are not being spuriously driven by the joint sorting of workers by education level into neighborhoods and establishments.

Another potential alternative explanation of our results is that rather than residential neighborhood influencing where one works – via residence-based networks – place of work determines where one lives. If, for example, co-workers recommend neighborhoods or houses to which workers then move, then we would see clustering of neighbors in the same establishments, but this would not be due to the operation of residence-based networks. This alternative explanation of our findings does not seem

plausible, on a priori grounds, as job mobility is much higher than residential mobility.[26] We could

answer this question definitively if we knew where workers lived when they first began working (or

applied for work) at a particular employer, but of course we do not have such data. However, in the

Census data we know whether a person has changed addresses in the past five years, and in the Business

Register we know establishment age. Thus, if we restrict attention to residents who have *not* moved in

the past five years who work in establishments that are fewer than five years old, then we know that our

measure of the importance of networks has to be based on a predetermined choice of residential location,

since that decision necessarily preceded the decision to work at a new establishment.[27]

     This analysis is reported in Table 4. Column (1) shows the baseline estimates from Table 3. In

column (2), we make the additional restriction that all workers in the sample work in establishments that

are less than five years old (as of the 2000 Census). The effective network isolation index is quite a bit

higher – 15.21 versus 9.99. This may be because newer establishments are smaller; we show later that the

index is higher for small establishments. Finally, column (3) restricts attention to those individuals who

did not change residential addresses in the past five years, and the effective index climbs substantially, to

22.87. Thus, when we focus on those for whom the residential location decision is exogenous to the

workplace location decision, we find *stronger* evidence of residence-based networks.[28] This evidence

actually strengthens the interpretation of the results as reflecting residence-based networks, because we

would expect such network connections to be stronger among those who have lived in their

neighborhoods for longer amounts of time, and who therefore are more likely to have connections to

neighbors through any of a number of channels.

     To this point we have presumed that residence-based networks help workers find jobs in

---

[26] For example, based on March CPS data for 1999-2000, the rate of within-county residential mobility – which might roughly correspond to the type of move that would put one near a co-worker – was 9 percent on an annual rate. (See http://www.census.gov/population/socdemo/migration/p20-538/tab01.txt (viewed July 8, 2009).) But the *monthly* job-to-job mobility rate is in the range of 2.7-3.2 percent (Moscarini and Thomsson, 2008); if this is independent across months, it implies an annual job mobility rate of around 30 percent.

[27] Bayer et al. (2008) do a related analysis, for the same reason.

[28] Establishments in column (3) are larger, on average, than those in column (2), because imposing the requirement that there be two matched workers per establishment with a more stringent criterion on workers to begin with (in this case, non-movers) makes small establishments less likely to be included in the sample.

particular establishments. However, networks may instead (or as well) help job searchers find jobs in certain industries. For example, a worker in a retail firm may tell a neighbor of job vacancies in nearby retail businesses. If networks increase the likelihood that neighbors work in the same industry, our calculations might overstate the extent to which networks determine the establishment of employment, because the clustering of neighbors in the same industry within a census tract of employment will inevitably lead to some clustering in the same establishments.[29] Of course if residence-based networks largely operate at the industry rather than establishment level, they would still be important.

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

Table 5 reports results conditioning on eight industry categories, and then results conditioning on two education categories as well, to allow for the possibility that industry-specific networks operate more within than across skill groups. The simulated index conditioning on industry is 4.02 (column (2)), so that when workers are randomly assigned to establishments in the same industry (and census tract) in which they are observed to work, on average 4.02 percent of their co-workers come from the same residential neighborhood. This is higher than the unconditional random isolation index of 2.97, so that the conditional effective network isolation index of 7.83 is smaller than the unconditional effective index of 9.99 reported in column (1). The difference implies that assignment of workers from the same neighborhoods to specific industries within a census tract can explain some of the assignment of workers

---

[29] That is, two workers employed in the same census tract who work in the same industry are more likely to work in the same establishment than are two randomly chosen workers employed in the same census tract.

to specific establishments. However, 78.5 percent of the effective network isolation remains even after we condition on a worker's industry.[30] Thus, at this level of industry detail, most of the (unconditional) effective network isolation for whites cannot be explained just by a mechanism whereby residence-based networks serve only to help workers find jobs in the same industries as their neighbors. In column (3) we condition on two education categories (high school degree or less vs. some college or more) as well as industry. The simulated conditional network isolation index is again higher, at 4.12, so that the conditional effective index is 7.63. While this conditional effective network isolation index is slightly lower than the results conditional on industry only, 76.5 percent of the unconditional effective index remains after this conditioning.

We noted earlier that constructing the DEED and imposing our sample restrictions leads our analysis samples to consist disproportionately of large establishments. If networks are more important in small establishments than in large establishments, this will bias our network isolation index downward. There are two reasons to think this could be the case. First, networks might operate primarily among individuals who work in the same occupation, and if large establishments have more occupational heterogeneity, our measure of network strength may understate the importance of networks in large establishments because it treats all workers in establishments as equally eligible to be in networks together.[31] Second, other evidence (e.g., Holzer, 1998) suggests that small establishments rely more on informal referrals in hiring, and if this is associated with the kinds of networks we measure, then network isolation should vary across establishment size. Therefore, in Table 6 we disaggregate the results by splitting the sample of establishments into four size categories.

Column (2) reports the results for establishments with reported employment from the Business

---

[30] This is likely a lower bound for the percentage of the effective network isolation index that remains. If there are only a few establishments in an industry in a particular census tract, then what this procedure treats as sorting on industry may represent sorting on establishments. In the limit, with only one establishment in the industry in the census tract we could not distinguish between sorting on industry or establishment, whereas the conditioning procedure used in this subsection attributes the sorting to industry first, and only the residual to establishment.
[31] However, given that conditioning on occupation in the overall sample of white workers had little influence on the effective network isolation index (Table 3), we doubt that stratification of networks by occupation within establishments plays a major role.

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

VI.2. Results for Blacks, and Black-White Differences

The importance of networks may differ for blacks and whites. On the one hand, as Montgomery's (1991) model suggests, because whites make up a greater fraction of the working population, if networks are race-based one might expect white individuals searching for employment to be able to take advantage of a larger network of white working neighbors, making it more likely that whites will work together in the same establishment than will blacks, above and beyond what would be predicted by random allocation.[32] 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, or overcoming stigma, one might expect network isolation to be larger for blacks than for whites.

In Table 7 we report results for black workers, restricting the sample over which we compute network isolation to census tracts of employment where we observe at least two blacks working in at least two establishments. We restrict the sample only to black workers because if we include both blacks and

---

[32] In our full samples of whites and blacks in columns (1) of Tables 3 and 7 (discussed below), blacks have an average of 10 black working neighbors, whereas whites have an average of 64 white working neighbors.

whites and there is discrimination in the labor market, we would expect to observe sorting of workers by race across establishments. Then, because there is residential segregation by race, we would spuriously generate a finding of network isolation. In contrast, by considering only blacks we ask whether there is segregation of blacks by neighborhood of residence across workplaces above and beyond the segregation of workers across establishments as a function of race alone (or something correlated with race alone).

Column (1) reports results for the full sample of black workers. The observed network isolation index (NI^O) for all blacks is 5.29, meaning that on average, of a black worker's black co-workers, 5.29 percent come from that worker's neighborhood. This is somewhat smaller than for whites, and the random network isolation index (NI^R) is 2.58, just slightly below that for whites; the network isolation difference is therefore 2.71. However, for blacks the maximum possible network isolation index (NI^M) is 30.71, quite a bit below that for whites, with the net result that the effective network isolation index of 9.63 is very close to what we find for whites (although statistically different from it). These results, which underscore the importance of scaling the network isolation difference by the extent to which network isolation is possible in the data, suggest little racial difference in our full samples of blacks and whites in the importance of residence-based networks in explaining the assignment of workers to establishments, once we rescale the network isolation difference by the maximum.[33]

In column (2) we report results for blacks who have at most a high school education, and in column (3) for blacks who have more than a high school education. The results by education address the issue of sorting by skill, and examine whether black-white differences reflect race differences in education coupled with variation across skill groups in the importance of residence-based labor market networks. In both cases, the observed and simulated indexes are lower for blacks than for whites, as are the network isolation differences. When we scale the results by the maximum possible network isolation differences, the results reverse somewhat for low-educated blacks, so that the effective network isolation

---

[33] We find only a few instances where comparisons of the network isolation difference, $NI^O - NI^R$, and the effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100$, lead to qualitatively different conclusions. The results for blacks and whites discussed in this paragraph are one such instance. Indeed, it is *only* for the black-white comparisons that the conclusions are ever sensitive to scaling by the maximum possible segregation

index for low-educated blacks is somewhat larger than for whites (and statistically different as well) – 15.02 versus 13.02 – whereas for more-educated blacks, the effective index is somewhat smaller than for whites – 7.30 versus 8.82. Whether or not we scale by the maximum, though, the evidence suggests that residentially-based networks are more important for less-educated blacks than for more-educated blacks.

Overall, the results in Tables 3 and 7 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., 2008); [34] and an absence of consistent evidence of race differences in the reported use of informal contacts (Ioannides and Datcher Loury, 2004).

Table 8 presents additional analyses for black and white workers. One issue with directly comparing the full sample results for blacks and whites is that we know these groups are not similarly distributed geographically in the United States. As a result, black and white workers in our sample may work in different labor markets where the importance of networks could differ as a result of labor market institutions, constraints, or other factors. Therefore, in columns (3) and (4) of Table 8 we present estimates for whites and blacks restricting the samples to workers working in census tracts that are represented by workers in both the white and black samples. [35] For whites, the sample restriction reduces the sample by about one-half, and the numbers of workplace census tracts and residential census tracts are also reduced considerably. The results are somewhat different than for the full sample of whites. First, the observed network isolation index falls to 5.68, and is closer to that for blacks. Second, the effective network isolation measure is lower (7.38 vs. 9.99 in column (1)). [36] Column (4) reports results for blacks.

---

[34] 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.

[35] 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).

[36] The same is true of the network isolation difference. Note that 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

The effective network isolation index for the restricted sample of blacks is 9.41, quite similar to the figure of 9.63 in column (2), which is not surprising since between columns (2) and (4) the sample is reduced by fewer than 4,000 workers. The random network isolation measure is quite a bit lower for this sample of whites than for blacks (1.48 vs. 2.54), so that the network isolation difference is higher for whites than for blacks (4.20 vs. 2.72). However, the maximum network isolation index for whites is almost double that for blacks, so that when we scale by our measure of maximum possible isolation, we reverse the relative magnitudes of isolation for blacks and whites. Thus, there is suggestive evidence that networks are more important for blacks than whites in comparable areas.

VI.3. Racial Stratification of Networks

The results we have presented thus far suggest that residence-based networks are important, and in that sense point to racially-stratified networks. After all, given that there is pervasive racial residential segregation in the United States (Iceland and Weinberg, 2002), networks that are predicated on residential "connectedness" have to be partially race-based. However, it is important also to consider whether there is racial stratification of networks even *within* neighborhoods – that is, the idea that labor market information and especially job referrals are less likely to flow between black and white co-residents than between co-residents of the same race.[37]

Therefore, in column (5) of Table 8 we assess more directly whether networks are race-based. We carry out the same types of sampling and computational procedures used before, except that we consider the relevant set of a black worker's neighbors and co-workers to consist of blacks *or* whites. We begin by constructing a sample of black workers and their neighbors, regardless of race, who we observe to work in establishments where at least one other black or white worker is matched. We then further restrict the sample to those who work in a census tract with at least two establishments that have workers

case. This occurs because the observations in column (3) come from a much smaller number of tracts with many more establishments, making it possible to achieve a higher maximum amount of network isolation.

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

in the sample. We again construct an effective network isolation index for black workers in this sample, but we now construct this measure by asking whether our sample of black workers 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*. If networks among co-residents are racially stratified, then the network isolation that results when we measure how likely it is that a black works with a neighbor regardless of race should be smaller than when we measure how likely it is that a black works with a black neighbor.[38]

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. As shown in the last column of Table 8, this number is 3.99, substantially lower than the 5.29 number in column (2). We then calculate the random network isolation index by taking all workers in this sample, randomizing them (as before) across establishments in that sample, and calculating the network isolation index for the black workers in this simulated sample. The random network isolation index is 2.00, leading to a network isolation difference of 2.00. The fact that this difference is lower than when we restrict the sample of co-workers to blacks, in column (2), suggests that race is indeed playing a role in driving the probability of working with a neighbor. We obtain a measure of the maximum possible network isolation index by approximating what the index would be for blacks if the workers who make up this sample were able to work to the maximum extent possible in establishments with their neighbors of any race, given the size distribution of establishments in our sample, the residential distribution of workers in them, and the workers' races.

The resulting maximum network isolation number is 35.83, which is somewhat larger than that in column (2). Taking all of these together, the effective network isolation index as reported in column (5) is only 5.90, which is about 40 percent smaller than in column (2), providing evidence that residence-

---

[38] That said, this differential network effect does not rule out the possibility that whites and blacks within a neighborhood share information on job openings equally, but that this information is more productive for blacks in matching blacks to establishments when the information serves to link a black neighbor to a job in an establishment held by another black neighbor. In other words, we cannot distinguish whether network connections between black and white neighbors are fewer in number or less productive; all we can do is characterize the relative importance of these network connections.

based labor market networks have a fairly strong race-based component. Moreover, because in this column we only need one black worker in an establishment for the establishment to be in the sample, as opposed to column (2), which requires two black workers, column (5) includes smaller establishments for which, as already noted, residence-based networks are more important. Thus, on a comparable basis the difference between the effective network isolation indexes in columns (2) and (5) would be even larger, bolstering the evidence that these networks are racially stratified.

VI.4. Results for Hispanics

Survey evidence suggests that Hispanics use referrals in finding employment much more than do blacks or whites (Ioannides and Datcher Loury, 2004). Immigrants and poor English speakers in particular may suffer from high search costs in the labor market, both because their limited understanding of U.S. labor markets and of English may make it hard for them to search widely in the labor market, and because potential employers may have a difficult time inferring the ability of these workers. Finding employment through informal networks of other immigrants and those who speak one's native language may therefore be particularly important for these groups. There is some indirect evidence consistent with this conjecture. For example, evidence of "enclave effects," such as Hispanics with poor English skills paying less of a penalty for those poor skills when they live in a county or SMSA with a larger Hispanic population (McManus, 1990), might reflect network effects, although it could also reflect higher productivity from a greater ability to work with Spanish speakers in the enclave.[39] Munshi (2003) presents a more-refined analysis of Mexican immigrants, tying labor market outcomes to a larger local population of immigrants from the same origin community. Patel and Vella (2007) find that new immigrants work disproportionately in occupations held by previous immigrants from the same country. And our previous work documents establishment-level segregation by English language skills, and segregation of Spanish-speaking workers from non-Spanish speaking workers among poor English speakers (Hellerstein and Neumark, 2008). Finally, perhaps the most direct evidence of these types of networks for immigrants comes from the work of Massey et al. (1987), who document the importance of

---

[39] For a similar type of evidence for Sweden, see Edin et al. (2003).

networks linking recent and earlier immigrants from the same communities in Mexico.

In this section, therefore, we turn to results for Hispanic workers, paying particular attention to Hispanic workers who speak English poorly (or not at all), and immigrants. The results are presented in Table 9. Column (1) presents results for the full sample of Hispanic workers (again with the sample restrictions that allow us to construct the network isolation index). The observed network isolation index is 11.22, quite a bit larger than for blacks or whites, and the random network isolation index is 3.08. Once we scale the difference between these by the difference between the maximum possible network isolation index and the random index, we find that the effective network isolation index for Hispanic workers is 22.36 – more than twice as large as what we find for the full samples of blacks or whites.[40]

In column (2) we restrict the sample to Hispanics who self-report speaking English either "poorly" or "not at all" – which together we refer to, for simplicity, as the sample of poor English speakers. For this sample, the observed network isolation index is 22.51. This is very large – it means that, on average, for a poor-English speaking Hispanic worker, 22.51 percent of his or her Hispanic co-workers who also are poor English speakers live in the same census tract! The simulated random network isolation index is much smaller, at 6.56, and the maximum network index is 42.65. Taken together, these numbers yield an effective network isolation index of 44.20, meaning that over 40 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 four times larger than what we find for blacks and whites, and suggests that residence-based labor market networks are extremely important for Hispanics who speak English poorly. In addition, paralleling our results by education level for blacks, the fact that the importance of networks goes up when we focus on those with poor language skills implies that the overall results for Hispanics are not driven by residential sorting on language skills.[41]

---

[40] The comparisons of Hispanics to whites and blacks, and between different groups of Hispanics, are similar for the effective network isolation index and the network isolation difference (its numerator), so in the ensuing discussion we focus on the effective network isolation index.

[41] We recognize the possibility that these results may be driven partially by reverse causality, where networks that are formed on the job serve to help low-skilled Hispanics find housing. The sample sizes for Hispanics are too small for us to conduct the same kind analysis of this issue that we conducted for whites.

By way of contrast, in column (3) of Table 9 we report the results for the sample of Hispanics who report speaking English "well" or "very well." The effective network isolation index is 18.39, just 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 exist for workers in the United States whose English language skills are poor. At the same time, the finding that the effective network isolation index is much higher for Hispanics who speak good English than for whites suggests that the overall Hispanic-white differences are not driven solely by skills.

In column (4) we report the results for Hispanic immigrants. The effective network isolation index is 33.91, which is quite a bit higher than for all Hispanics. In contrast, in column (5) we report the results for non-immigrant Hispanics, for whom the effective network isolation index is 16.62, half the size of the index for immigrant Hispanics and smaller than for any other Hispanic group in the table. To the extent that the Hispanic workers in column (5) are most integrated into U.S. society and have good English language skills, this provides further evidence that our measure of network isolation captures the important of residence-based networks that reduce search frictions in the labor market.[42]

VI.5. Results by Race and Ethnicity for Small Establishments

We noted earlier that our sample selection rules lead to under-representation of small establishments, and that this is especially true for blacks and Hispanics. Moreover, Table 6 makes it clear that network isolation is more important for whites who work in small establishments than in large establishments. Therefore, the different sample compositions of establishments could bias our comparisons of the importance of networks across racial and ethnic groups in the full samples of each of these groups. In Table 10 we report our measures of (unconditional) network isolation for whites, blacks, and Hispanics, where we utilize restricted samples only of workers employed at establishments with 50 or fewer workers in total. Column (1) contains the results for whites, as previously reported in column (2) of Table 6. Column (2) contains the results for blacks, and column (3) the results for Hispanics. The results,

---

[42] 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.

which are based on much more homogeneous samples with respect to establishment size than those for the full samples, suggest that residence-based networks are much more important for blacks than for whites (an effective network isolation index of 39.27 vs. 22.87,[43] and even more so for Hispanics (52.20) relative to whites. Thus, the relative importance of networks for blacks and especially Hispanics compared to whites appears greater than is suggested by the analyses of our full samples.

VII. Conclusions

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

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

---

[43] Paralleling the results in Table 8 columns (3) and (4), this conclusion for black-white differences is driven by the scaling by maximum segregation.

actually make it more likely that workers are employed in the same business.

We interpret the evidence as indicating that labor market networks play an important role in establishment-level employment. For both whites and blacks we find that the grouping of workers from the same neighborhoods in the same business establishments exceeds by a factor of more than two what we would expect to occur randomly. For whites, we find that network isolation is about 10 percent of the theoretical maximum amount of grouping that could be found in the data, and many of our analyses indicate that residence-based labor market networks are more important for blacks than for whites. For both whites and blacks, these labor market networks appear more important for workers who have low levels of education – a high school degree or less – than for more-educated workers. There is also evidence that these networks are more important in small establishments.

Our results also provide evidence that networks operate to some extent along racial lines, above and beyond the racial stratification of networks that comes from residential segregation by race. In particular, the link between residential location and the establishment of employment is stronger for blacks when we consider only co-workers of the same race, consistent with more (or more productive) labor market information and referrals flowing across co-residents of the same race than of the other 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. More significantly, this type of racial stratification would imply that policies that solely address spatial mismatch, by attempting to move blacks to areas where more whites live and where more jobs (per person) are located, may fail to help blacks precisely because network connections are severed, and are less likely to be established with white neighbors.[44]

---

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

33

We also find that residence-based networks are more important for Hispanics than for blacks or whites, and among Hispanics, these networks are especially important for immigrants and those with poor language skills. The results for Hispanics give credence to the idea that informal labor market networks may be particularly important for those workers who are not as well-integrated into the labor market, and for whom employers may have less reliable information.[45]

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

On the other hand, matched employer-employee data sets such as the DEED have some limitations in terms of what they can teach us about networks. First, it is obviously important to consider along what other dimensions of social interactions – aside from residence – networks operate to cause individuals to work in the same establishment, and what types of networks are most important. Among

---

neighbors – however defined – are not the only source of network connections.
[45] As noted previously, the alternative interpretation of our results – that they reflect heterogeneous tastes such that people who like similar workplaces also like similar neighborhoods – is not consistent with the patterns of evidence that we find whereby residence-based networks are more important for those least integrated into the economy, such as those with less education, Hispanics, and especially Hispanic immigrants.

the possibilities are schools,[46] religious institutions, and community groups, as well as existing places of employment (from which workers may move to other jobs). Second, the DEED provides little scope for understanding the dynamics of how networks are formed and how they operate. Are all members of the network equally important? What kinds of information get shared within the network? The data demands for answering many of these questions are daunting, but the answers can provide clues regarding how important it is for individuals, communities, and other institutions to foster network relationships so as to improve economic outcomes, and what types of networks are most effective.

---

[46] Indeed Bayer et al. (2008) 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. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy*, Vol. 116, No. 6, December, pp. 1150-96.

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

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.

Cingano, Federico, and Alfonso Rosolia. 2009. "People I Know: Job Search and Social Networks." Unpublished paper, Bank of Italy.

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

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.

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

Ellison, Glenn, and Edward L. Glaeser. 1997. "Geographic Concentration in Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy*, Vol. 102, No. 5, October, pp. 889-927.

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

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

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

Hellerstein, Judith K., David Neumark, and Melissa McInerney. 2008. "Spatial Mismatch vs. Racial Mismatch?" *Journal of Urban Economics*, Vol. 64, No. 2, September, pp. 467-79.

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.

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

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

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

Massey, Douglas, Rafael Alarcón, Jorge Durand, and Humberto González. 1987. Return to Aztlan: The Social Process of International Migration from Western Mexico (Berkeley, CA: UC Berkeley Press).

McManus, Walter S. 1990. "Labor Market Effects of Language Enclaves: Hispanic Men in the United States." *Journal of Human Resources*, Vol. 25, No. 2, Spring, pp. 228-52.

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.

Moscarini, Giuseppe, and Kaj Thomsson. 2008. "Occupational and Job Mobility in the US." *Scandanavian Journal of Economics*, Vol. 109, No. 4, March, pp. 807-36.

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.

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

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

Petersen, Trond, Ishak Saporta, and Marc-David L. Seidel. 2000. "Offering a Job: Meritocracy and Social Networks." *American Journal of Sociology*, Vol. 106, No. 3, November, pp. 763-816.

Schmutte, Ian M. 2009. "Job Referral Networks and the Determination of Earnings in Local Labor Markets." Unpublished paper, Cornell University.

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

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

Turner, Kristin, Susan Clampet-Lundquist, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2006. "Neighborhood Effects on Barriers to Employment: Results from a Randomized Housing Mobility Experiment in Baltimore." *Brookings-Wharton Papers on Urban Affairs*, pp. 137-72.

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.

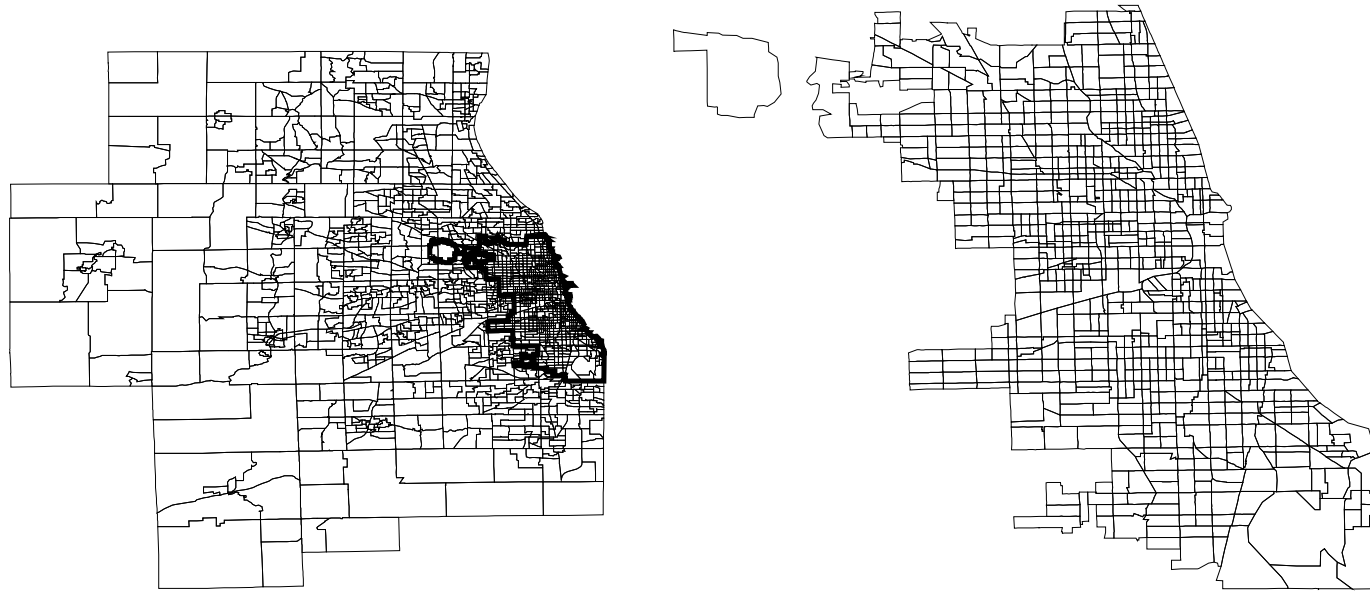Figure 1: Census Tracts, Chicago PMSA and Chicago City Limits
Source: U.S. Census Bureau, http://www.census.gov/geo/www.cob.tr2000.html (viewed May 5, 2008). Note that Chicago city limits are highlighted in the left-hand figure, and shown magnified in the right-hand figure.

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 (approximate median)	15	35	154	84
Establishment size:				
1-25	.65	.39	.11	.18
26-50	.15	.20	.11	.16
51-100	.10	.17	.14	.20
101+	.10	.22	.62	.45
Industry:				
Mining	.004	.003	.002	.003
Construction	.078	.070	.013	.053
Manufacturing	.133	.186	.226	.310
Transportation	.050	.052	.077	.051
Wholesale	.067	.074	.039	.060
Retail	.284	.265	.231	.266
FIRE	.081	.077	.082	.049
Services	.303	.272	.331	.209
In MSA/PMSA	.792	1.0	1.0	1.0
Census region:				
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. The approximate median is an average of the median and some observations to either side of the median, to preserve confidentiality.

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

	All	All, conditional on two education categories	All, conditional on four education categories	All, conditional on six occupation categories	All, conditional on housing price quartiles	High school degree or less only	More than high school degree only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Network isolation index, observed, NI^O	7.87	7.87	7.87	7.87	7.87	10.95	5.94
Simulated random network isolation index, NI^R	2.97					4.81	2.28
Network isolation difference, $NI^O - NI^R$	4.90					6.15	3.66
Maximum possible network isolation index, NI^M	52.06	52.06	52.06	52.06	52.06	52.01	43.74
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100$	*9.99*					*13.02*	*8.82****
Simulated conditional network isolation index, NI^C		3.02	3.13	3.25	3.49		
$NI^O - NI^C$		4.84	4.74	4.61	4.38		
Conditional effective network isolation index, $[(NI^O - NI^C)/(NI^M - NI^R)]\cdot100$		*9.86****	*9.65****	*9.40****	*8.92****		
N	1,675,412	1,675,412	1,675,412	1,675,412	1,675,412	441,133	1,014,931
# place of work tracts	26,470	26,470	26,470	26,470	26,470	16,073	20,378
# residential tracts	46,764	46,764	46,764	46,764	46,764	40,754	45,311
Mean establishments/tract	129.6	129.6	129.6	129.6	129.6	47.1	98.0
Mean matched workers/establishment	38.4	38.4	38.4	38.4	38.4	10.8	41.0
Mean number of workers in tract of employment from same tract of residence	9.4	9.4	9.4	9.4	9.4	6.2	5.7

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. In column (2) the categories are high school degree or less and more than high school. In column (3) they are less than high school degree, high school degree, some college, and bachelor's degree or higher. In column (4) the occupation categories are managerial and professional specialty; technical, sales, and administrative support; service; farming, forestry, and fishery; precision production, craft, and repair; and operators, fabricators, and laborers. In column (5), the housing price quartiles are constructed for the MSA. The sample sizes in columns (6)-(7) add up to less than the sample size in column (1) because the sample exclusion restrictions are imposed for each education size category. All effective measures of network isolation are statistically significantly different from zero at the 1-percent level. The table also reports results for statistical significance of the difference between the following pairs of estimates: (a) conditional on two education categories (col. 2) vs. all whites (col. 1); (b) conditional on four education categories (col. 3) vs. all whites (col. 1); (c) conditional on 1-digit occupation (col. 4) vs. all whites (col. 1); (d) conditional on housing value (col. 5) vs. all whites (col. 1); and (e) more than a high school degree (col. 7) vs. less than a high school degree (col. 6). In each case, the differences are statistically significant at the 1-percent level (denoted by ***).

Table 4: Network Isolation for Whites, Residential Location Exogenous to Job Location

	All	Working in establishments born 1996 or after	Working in newer establishments *and* did not move 1995-2000
	(1)	(2)	(3)
Network isolation index, observed, NI^O	7.87	8.47	10.25
Simulated random network isolation index, NI^R	2.97	3.01	3.64
Network isolation difference, $NI^O - NI^R$	4.90	5.46	6.61
Maximum possible network isolation index, NI^M	52.06	38.88	32.53
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \cdot 100$	*9.99*	*15.21*	*22.87*
N	1,675,412	216,920	68,821
# place of work tracts	26,470	8,886	4,074
# residential tracts	46,764	39,209	26,622
Mean establishments/tract	129.6	38.8	17.7
Mean matched workers/establishment	38.4	20.2	14.3
Mean number of workers in tract of employment from same tract of residence	9.4	2.9	2.0

Notes: See notes to Table 3. All effective network isolation measures are statistically significantly different from zero at the 1-percent level.

Table 5: Network Isolation for Whites, Conditional on Industry and on Industry and Education Simultaneously

	Unconditional	Conditional on eight industry categories	Conditional on two education categories and eight industry categories
	(1)	(2)	(3)
Network isolation index, observed, NI^O	7.87	7.87	7.87
Simulated random network isolation index, NI^R	2.97		
Network isolation difference, $NI^O - NI^R$	4.90		
Simulated conditional network isolation index, NI^C		4.02	4.12
$NI^O - NI^C$		3.85	3.75
Maximum possible network isolation index, NI^M	52.06	52.06	52.06
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100$	*9.99*		
Conditional effective network isolation index, $[(NI^O - NI^C)/(NI^M - NI^R)]\cdot100$		*7.83****	*7.63****
Share of effective isolation index unexplained after conditioning or additional conditioning		*78.5*	*76.5*

See notes to Table 3. The industries are mining; construction; manufacturing; transportation and warehousing; wholesale trade; retail trade; finance, insurance, and real estate; and services. The education categories are high school degree or less and more than high school. All effective network isolation measures are statistically significantly different from zero at the 1-percent level. The table also reports results for statistical significance of the difference between the following pairs of estimates: (a) conditional on eight industry categories (col. 2) vs. unconditional results for all whites (col. 1); and (b) conditional on two education categories and eight industry categories (col. 3) vs. unconditional results for all whites (col. 1). In each case, the differences are statistically significant at the 1-percent level (denoted by ***).

Table 6: Network Isolation for Whites, by Establishment Size

		Establishment size			
	All	≤ 50	51-100	101-250	> 250
	(1)	(2)	(3)	(4)	(5)
Network isolation index, observed, NI^O	7.87	15.76	6.08	4.59	2.82
Simulated random network isolation index, NI^R	2.97	4.55	2.86	2.58	2.03
Network isolation difference, $NI^O - NI^R$	4.90	11.21	3.22	2.01	0.79
Maximum possible network isolation index, NI^M	52.06	53.56	36.93	32.50	19.95
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot100$	*9.99*	*22.87*	*9.45*	*6.70*	*4.41*
N	1,675,412	527,430	208,507	262,437	448,155
# place of work tracts	26,470	22,162	9,055	4,448	3,790
# residential tracts	46,764	43,700	39,425	40,663	42,603
Mean establishments/tract	129.6	66.0	29.2	26.8	19.2
Mean matched workers/establishment	38.4	3.3	6.1	11.0	96.2
Mean number of workers in tract of employment from same tract of residence	9.4	3.9	2.5	3.2	6.6

Notes: See notes to Table 3. The sample sizes in columns (2)-(5) add up to less than the sample size in column (1) because the sample exclusion restrictions are imposed for each establishment size category. All effective network isolation measures are statistically significantly different from zero at the 1-percent level.

Table 7: Network Isolation for Blacks, Overall and by Education

	All	High school degree or less only	More than high school degree only
	(1)	(2)	(3)
Network isolation index, observed, NI^O	5.29	8.22	3.41
Simulated random network isolation index, NI^R	2.58	4.11	1.74
Network isolation difference, $NI^O - NI^R$	2.71	4.11	1.67
Maximum possible network isolation index, NI^M	30.71	31.48	24.67
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \cdot 100$	*9.63*[+++]	*15.02*[+++]	*7.30*[***,+++]
N	97,967	27,001	47,281
# place of work tracts	4,490	1,923	2,278
# residential tracts	21,623	10,685	16,107
Mean establishments/tract	22.2	10.2	17.4
Mean matched workers/establishment	18.6	8.2	15.8
Mean number of workers in tract of employment from same tract of residence	2.6	2.2	1.9

Notes: See notes to Table 3. All effective network isolation measures are statistically significantly different from zero at the 1-percent level. The table also reports results for statistical significance of the difference between pairs of estimates. First, there is test of differences across estimates within this table, in particular blacks with a high school degree or less (col. 2) vs. blacks with more than a high school degree (col. 3). The difference is statistically significant at the 1-percent level (denoted by [***]). Second, there are tests of differences across estimates in this table and other tables, in particular: (1) all blacks (col. 1) vs. all whites (Table 3, col. 1); (2) blacks with a high school degree or less (col. 2) vs. whites with a high school degree or less (Table 3, col. 6); and (3) blacks with more than a high school degree (col. 3) vs. whites with more than a high school degree (Table 3, col. 7). In each case, the differences are statistically significant at the 1-percent level (denoted by [+++]).

Table 8: 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	3.99
Simulated random network isolation index, NI^R	2.97	2.58	1.48	2.54	2.00
Network isolation difference, $NI^O - NI^R$	4.90	2.71	4.20	2.72	2.00
Maximum possible network isolation index, NI^M	52.06	30.71	58.37	31.43	35.83
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]{\cdot}100$	*9.99*	*9.63*	*7.38*	*9.41****	*5.90****
N	1,675,412	97,967	845,290	94,210	140,083
# place of work tracts	26,470	4,490	4,122	4,122	9,094
# residential tracts	46,764	21,623	42,533	21,459	26,768
Mean establishments/tract	129.6	22.2	229.7	23.0	39.7
Mean matched workers/establishment	38.4	18.6	46.6	17.7	40.3
Mean number of workers in tract of employment from same tract of residence	9.4	2.6	11.4	2.6	4.5

Notes: See notes to Table 3. All effective network isolation measures are statistically significantly different from zero at the 1-percent level. The table also reports results for statistical significance of the difference between the following pairs of estimates: (a) whites working in establishments located in tracts in both the black and white samples (col. 3) vs. blacks working in establishments located in tracts in the black and white samples (col. 4); and (b) blacks when network isolation is based on blacks and whites (col. 5) vs. all blacks (col. 1). In each case, the differences are statistically significant at the 1-percent level (denoted by ***).

Table 9: 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	22.51	8.30	16.68	7.20
Simulated random network isolation index, NI^R	3.08	6.56	2.47	4.46	2.50
Network isolation difference, $NI^O - NI^R$	8.14	15.95	5.83	12.22	4.70
Maximum possible network isolation index, NI^M	39.48	42.65	34.17	40.50	30.78
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*22.36[+++,§§§]*	*44.20*	*18.39[***]*	*33.91*	*16.62[***]*
N	110,235	12,451	80,661	42,712	43,763
# place of work tracts	5,059	900	4,090	2,508	2,563
# residential tracts	20,716	4,576	19,027	11,321	14,063
Mean establishments/tract	59.7	23.3	46.3	43.2	28.8
Mean matched workers/establishment	9.1	4.8	9.4	5.7	9.7
Mean number of workers in tract of employment from same tract of residence	2.7	2.3	2.2	2.5	2.0

Notes: See notes to Table 3. All effective network isolation measures are statistically significantly different from zero at the 1-percent level. The table also reports results for statistical significance of the difference between pairs of estimates. First, there is test of differences across estimates within this table, in particular: Hispanics with good English skills (col. 3) vs. Hispanics with poor English skills (col. 2); and (3) non-immigrant Hispanics (col. 5) vs. Immigrant Hispanics (col. 4). In each case, the difference is statistically significant at the 1-percent level (denoted by [***]). Second, there are tests of differences across estimates in this table and other tables, in particular: all Hispanics (col. 1) vs. all whites (Table 3, col. 1); and all Hispanics (col. 1) vs. all blacks (Table 7, col. 1). In each case, the differences are statistically significant at the one percent level (denoted by [+++] in the first case, and [§§§] in the second).

Table 10: 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	4.55	4.34	5.72
Network isolation difference, $NI^O - NI^R$	11.21	8.42	17.07
Maximum possible network isolation index, NI^M	53.56	25.77	38.41
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \cdot 100$	*22.87*	*39.27*** *	*52.20*** *
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. All effective network isolation measures are statistically significantly different from zero at the 1-percent level. The table also reports results for statistical significance of the difference between the following pairs of estimates: (a) blacks working in small establishments (col. 2) vs. whites working in small establishments (col. 1); and (b) Hispanics working in small establishments (col. 3) vs. whites working in small establishments (col. 1). In each case, the differences are statistically significant at the one percent level (denoted by ***).