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

Judith K. Hellerstein
Department of Economics and MPRC,
University of Maryland,
and NBER

Melissa McInerney
Department of Economics,
College of William and Mary

and

David Neumark
Department of Economics,
UCI,
NBER, and IZA

July 2009

Abstract

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

*This research was supported by NICHD grant R01HD042806. We are grateful for the comments of Joel Elvery, Mark Granovetter, Lisa Hellerstein, Jed Kolko, Andrew Noymer, Michael Ransom, Cliff Stein, Etienne Wasmer, and participants in "Social Networks and Peer Effects: Theory and Applications, A Conference in Memory of Antoni Calvó-Armengol," the Western Economics Association meetings, the All-UC Labor Conference, the Society of Labor Economists Annual Meetings, and seminars at ERMES, IZA, NYU, UCI, and USC. The research in this paper was conducted while McInerney was employed by the U.S. Census Bureau. This paper has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. It has been screened to ensure that no confidential information is revealed. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau. The authors' email addresses, in order, are hellerst@econ.umd.edu, mpmcinerney@wm.edu, and dneumark@uci.edu.

I. Introduction

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

Perhaps the most famous of these theories is the spatial mismatch hypothesis (e.g., Kain, 1968). The essence of the spatial mismatch hypothesis is that blacks disproportionately live in areas with poor access to jobs, and that this spatial isolation contributes to poor labor market outcomes for blacks. In previous research, we report evidence suggesting that spatial mismatch is not the mechanism by which residential segregation leads to poor economic outcomes for blacks (Hellerstein et al., 2008a). In particular, we find that poor employment outcomes for low-skilled blacks are not a function of a lack of jobs per se where blacks live, but rather that local blacks get these jobs only when local employers are hiring other black workers. We concluded that network effects were an explanation consistent with our results because we found a similar relationship for whites to that which we found for blacks – with whites' employment prospects boosted by higher employment of whites where they live, but not by higher employment of blacks. That is, the combined evidence for blacks and whites is consistent with race-based labor market networks. Moreover, in other work we find evidence of substantial workplace segregation by race and ethnicity, evidence that is also consistent with race- and ethnic-based labor market networks (e.g., Hellerstein and Neumark, 2008).[1]

Indeed, theoretical models of labor market networks can be shown to formalize the link between

[1] Kasinitz and Rosenberg (1996) provide compelling case study evidence consistent with the results of our previous work and this interpretation of them. 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 document that many employers hire workers almost exclusively from outside of Red Hook, recruiting employees via social networks within specific (non-black) ethnic groups.

residential segregation and labor market outcomes, 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. The specific ways in which networks do this, however, vary across models. For example, Calvó-Armengol and Jackson (2007) analyze a labor market where workers face a cost of obtaining information on vacancies. Vacancy postings are known by only some individuals in the labor market, and those individuals then selectively pass the information on only to others in their own networks (if they do not want the job themselves). The model also generates predictions for how the existence of labor market networks can translate initial differences in employment rates and wages between two groups into persistent differences. It does not specify, however, how initial information about vacancies is transmitted to individuals, and in particular there are no firms per se in the model.

Montgomery (1991), on the other hand, specifies a labor market where the information imperfection is on the employer side, and where the role of the employer is explicit. 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.[2] 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, creating stratification across firms on the basis of social networks. In Montgomery's framework, if social networks are at least partially race- or ethnic-based – which may be due to residential segregation – and white workers are initially employed at higher rates than blacks and Hispanics, then the existence of a larger network of white workers will lead to more job referrals at high wages for whites searching for jobs, creating wage disparities between whites and other groups. Although Montgomery's model does not build in a reservation wage, having an option for remaining out of the labor market would, in his

[2] See also Simon and Warner (1992).

framework, lead to employment differentials across groups as well.[3]

Our goal in this paper is to provide evidence on the importance of residentially-based labor market networks in determining the assignment of workers to establishments, based on a large-scale data set covering most of the U.S. economy. In that sense, our paper is motivated directly by theoretical models of networks like Montgomery's, in which networks yield information for (or about) specific employers. We measure the extent to which the importance of these labor market networks varies across whites, blacks, and Hispanics, as well as subgroups of Hispanics. Moreover, we provide suggestive evidence that more directly addresses the question of stratification of networks, asking, in particular, whether the networks we identify are race-based, operating more strongly within than across races.

To achieve this goal, we specify and implement a test for the importance of residence-based labor market networks in determining the establishments at which people work, using recently-constructed matched employer-employee data at the establishment level. We measure the importance of network effects for groups broken out by race (black vs. white), by ethnicity (Hispanic vs. white), and by various measures of skill (education, English language proficiency, and immigrant status). Our data and methods imply that we are testing for a particular form of labor market networks – namely, networks that are generated by residential proximity. In that sense, our analysis has some parallels to Bayer et al. (2005a). We recognize that our measure of this particular type of labor market network underestimates the overall importance of networks – in that networks may be based on more than just residence, and they may serve to help workers find employment in more than one specific establishment – and that these networks may be more important for labor markets are more local.

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. This type of segregation of employees across establishments would arise from

[3] Similarly, a natural application of Calvó-Armengol's model is one where race-based networks lead to persistent differences between blacks and whites in wages and employment.

residential proximity capturing the "network connectedness" that is 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 then 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 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. Thus, the difference between these two shares provides a measure of the importance of residence-based labor market networks, which we rescale into an "effective" measure of the importance of networks by comparing this difference to the maximum possible extent of networks that we calculate could arise in the data (again, relative to randomness), given the distributions of workers across establishments and across residential neighborhoods. We of course consider influences other than networks that could give rise to this evidence.

The data we use for this study come from the 2000 Decennial Employer-Employee Database, which we have constructed at the U.S. Census Bureau. The DEED is a large dataset consisting of workers matched to their establishment of employment. The employer-employee matches enable us to study directly whether workers employed in the same establishment are likely to live in the same neighborhoods.

Overall, we find evidence that residence-based labor market networks play an important role in hiring. For blacks, we find that the grouping of workers from the same neighborhoods in the same business establishments is about 9 percent of the maximum grouping that could occur. These networks are also important for whites, although somewhat less so than for blacks. For blacks and whites, residence-based labor market networks affect employment patterns more for less-skilled than for more-skilled workers, and because these networks are also important within skill groups, the findings do not simply reflect residential segregation by skill. Moreover, the evidence indicates that the networks we study are partly race-based, operating more strongly within than across races. We also 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 21 percent of the maximum. And among Hispanics, these networks play a larger role for immigrants and those with poor English skills. Finally, results conditioning on industry indicate that the network effects we find are largely due to the assignment of workers to specific establishments rather than simply to industries. This evidence therefore provides support for theoretical models in which networks serve to match workers to specific employers.

II. Relation to Existing Literature

A large body of existing research points to the potential importance of labor market networks. Granovetter (1974) is one of the early and most-cited sources of evidence on the importance of informal contacts in finding employment. Ioannides and Datcher Loury (2004) provide a recent review of evidence that indicates widespread reliance on friends, relatives, and acquaintances to search for and find jobs. As in Granovetter's work, much of this evidence is based on surveys of workers (e.g., Corcoran et al., 1980; Blau, 1992), although evidence from a study of one firm also suggests that many hires are referred by friends (Petersen et al., 2000). Based on the survey evidence from workers, Ioannides and Datcher Loury conclude that there is very 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 compared to high-educated workers, and that there are substantially higher rates of use of informal contacts for Hispanics than for 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.[4]

Bayer et al. (2005a) present an analysis that moves beyond survey evidence, testing more directly for network effects. They use confidential data from the 2000 Long Form of the Decennial Census and focus exclusively on workers in the Boston area. In a clever method of inferring whether networks are

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

important in explaining employment patterns, they find that two individuals who live on the same census block are about one-third more likely to work on the same census block than are two individuals who live in the same block group but not on the same block.[5] To the extent that informal networks are stronger within the block than within the block group, this evidence is consistent with residence-based labor market networks affecting hiring.[6]

We regard the Bayer et al. paper as providing the most definitive evidence to date suggesting that residential geographic proximity affects labor market outcomes, presumably through networks that connect people living in close proximity (see also Ioannides and Datcher Loury, 2004). Nonetheless, there are some limitations of their study, as well as the broader literature, upon which we try to improve in the present study.[7] 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 in the same census block may work for different employers, particularly for blocks in the central city that contain multiple employers.[8] The Petersen et al. (2000) study, which analyzes data from one firm, does 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 necessarily to infer that the company used the referral method to obtain

---

[5] For the Boston metropolitan area data they study, a census block corresponds roughly to a city block, and there are on average 10 blocks per block group. (They get similar results using the 10 closest blocks to each block based on physical distance.)

[6] 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.

[7] We readily note that we cannot definitively establish that the direction of causality runs from place of residence to place of work. Bayer et al. (2005a) examine this issue in detail and conclude that it is likely that networks operate to allow people who live together to find work together. Like them, we largely treat place of residence as exogenous to the place of work decision. However, we present one analysis later on that assesses whether the potential endogeneity of residence likely drives any of our findings, and we conclude that it does not.

[8] For example, focusing on central city areas in our sample of urban establishments and their employees, which is described below, we find that there are on average 2.58 establishments per block, which is an undercount by what we estimate to be about a factor of four, given that we only observe in our data a subsample of establishments. This raises questions about Bayer et al.'s assumption that workers employed on the same block "work with" one another or "work together" (e.g., 2005a, p. 26).

information about potential hires from current employees. One exception of which we are aware, however, is the Granovetter study, which reports that informal contacts were often employed in the company where the job held by a respondent had opened up. The data and approach that we use permit us to tie network connections to employment in a particular business establishment, and thus provide more direct evidence on the hypothesis that labor market networks reduce search frictions on the part of employers, as in Montgomery's (1991) model.

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.[9] There are theoretical reasons, though, to expect the importance of networks to vary across groups. For example, if labor market discrimination (either statistical or taste-based) by some employers raises search costs for certain groups, such as blacks or Hispanics (Black, 1995), networks may partially mitigate these search costs by helping to identify possible employers on whom to focus a job search. And to the extent that employers stigmatize some groups of workers based on other characteristics associated with race (such as welfare use or a criminal record) or ethnicity (such as language ability), informal networks may help to overcome statistical discrimination.[10] Similarly, race and ethnic differences in the importance of networks may exist if networks reduce employer uncertainty about productivity when it is difficult for workers to provide signals of productivity because of language, cultural, and educational differences (e.g., Lang, 1986). 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 (e.g., Granovetter, 1974 and Bayer et al., 2005) or a single firm (e.g., Petersen

[9] Bayer et al. (2005a) study data only for Boston and focus mostly on whites.
[10] The latter explanation follows more directly from Montgomery's (1991) model.

et al., 2000). Because of the large data set we construct, and the explicit linkage of workers to establishments of employment that the data set enables, we can fully explore the importance of networks that assign workers to establishments and differentials in the importance of 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.

III. The 2000 DEED

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

The 2000 DEED is formed by matching workers to establishments. The workers are drawn from the Sample Edited Detail File (SEDF), which contains all individual responses to the 2000 Decennial Census of Population one-in-six Long Form. The establishments are drawn from the Census Bureau's Business Register list (BR) for 2000; the BR is a database containing information for all business establishments operating in the United States in each year, which is continuously updated (see Jarmin and Miranda, 2002). The BR lists all business establishments with one or more employees operating in the United States. The Census Bureau uses the BR as a sampling frame for its Economic Censuses and Surveys. 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 2000 Decennial Census Long Form were asked to report the name and address of the employer in the previous week for each employed member of the household. This employer name and address information is stored in the "Write-In" file, which contains the information written on the questionnaires by Long-Form respondents but not actually captured in the SEDF. We use employer names and addresses for each worker in the Write-In file to match the Write-In file to the BR. Because the name and address information on the Write-In file is also available for virtually all employers in the BR, nearly all of the establishments in the BR that are classified as "active" by the Census Bureau are available for matching. Finally, because both the Write-In file and the SEDF contain identical sets of unique individual identifiers, we can use these identifiers to link the Write-In file to the SEDF. Thus, this procedure yields a very large data set with workers matched to their establishments, along with all of the information on workers from the SEDF.

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

The second step of the matching process is to select and implement the matching specifications. The AutoMatch software uses a probabilistic matching algorithm that accounts for missing information, misspellings, and even inaccurate information. This 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.[11] 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 measure whether and to what extent residential networks play a role in the assignment of workers to establishments via an analysis that is based on the percentage of workers in an individual's establishment (i.e., workplace) that comes from the individual's residential neighborhood. We compute the importance of these networks across a variety of subsamples, e.g., all blacks, low-educated blacks, all whites, and Hispanics who speak English poorly. For explication, here we describe in detail how we measure the role of race-based residential networks for black workers; the construction of the network measure for other subsamples is identical.

We first compute for each black worker in our sample the percentage of black 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 requires a sample restriction to establishments where we observe at least two black workers. We then average these percentages across all the black workers in our sample to create a "network isolation index," denoted NI^O, which measures the average fraction of a worker's co-workers

[11] 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.)

who are also residential neighbors of that worker. The superscript "O" on the network isolation index emphasizes that this is the fraction of a black worker's co-workers who are *observed* in our estimation sample to be residential neighbors.[12]

Formally, the observed network isolation index is:

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

where there are N workers indexed by i (and j). $I_i^R(j)$ is an indicator for whether worker j lives in the same residential neighborhood as worker i, and $I_i^W(j)$ is an indicator for whether j works in the same establishment as worker i. The sums in the numerator and denominator are taken over all 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.

To operationalize our network isolation index, we need to define what it means for workers to be residential neighbors. We define residential neighborhoods to be census tracts. There are a few reasons why this definition seems sensible to us. First, census tracts define the boundaries that are traditionally used to measure residential segregation (see, e.g., Iceland and Weinberg, 2002). Second (and related to the first), census tracts are defined by the U.S. Census Bureau to ensure that the tracts are "as homogeneous as possible with respect to population characteristics, economic status, and living conditions,"[13] 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

[12] 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 by considering the fraction of an individual's co-workers who are of a specific race, ethnicity, or skill group. There we provide a detailed discussion of the advantages of measuring segregation this way rather than via other indexes such as the Duncan Index, but we show that results are robust to using this alternative measure.
[13] See http://www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed April 21, 2008).

fence," then at parks, schools, churches, stores, etc.

To provide some idea of the size of census tracts, detailed maps of these tracts for Chicago are reproduced in Figures 1 and 2. For the PMSA (Figure 1), the median census tract was 0.57 square miles; the mean was 2.75 square miles. The smallest census tract was 0.02 square miles, and the largest (in the most outlying areas of the PMSA) was 151 square miles. For the city itself (Figure 2), the median was 0.17 square miles, the mean was 0.26 square miles, the minimum was again 0.02 square miles, and the maximum was 8 square miles. The larger tracts are not problematic. For the city, the two largest tracts are O'Hare Field (the airport) and another tract at the southern edge of the city that is mainly industrial.[14] After these two, the next largest tract is 3.5 square miles. The much larger census tracts in outlying areas, which are more rural, do not necessarily imply fewer social contacts, because the density of schools, churches, etc., is of course much lower. In addition, these tracts have few residents or establishments and hence contribute little to the overall findings.

There are three reasons why we might observe residential neighbors working in the same establishment that have nothing to do with residentially-based labor market networks. First, access to mass transportation alone may lead residential neighbors to work in the same geographic place, although not necessarily to work in the same establishments within that place. Therefore, in establishing boundaries for the local area in which workers might be employed, we again use census tracts. Census tracts are generally small enough that it is possible, particularly within urban areas, for individuals to be able to walk from any establishment to any other establishment, so in that sense transportation differences should not materially affect the distribution of workers across establishments in the tract.

Second, people may have heterogeneous tastes such that those who like similar workplaces also like to live in similar neighborhoods. If this were the case, we would expect that individuals who have the resources to better integrate into the economy are able to exercise more choice about both where to

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

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 should be most integrated into the economy are least likely to work with their neighbors.

Third, some clustering of residential neighbors into establishments can occur even if workers are assigned randomly to establishments, and we are of course interested in workplace clustering of residential neighbors that occurs systematically – i.e., more than would be expected to result from randomness.[15]

We therefore consider deviations of NI^O from what would be expected if workers from the same residential neighborhood were randomly assigned to any establishment within the census tract in which they are employed. In particular, we compute the extent of network isolation that would occur due to randomness, denoted NI^R, by simulating random allocation through Monte Carlo methods. Within a census tract, we randomly assign workers to establishments, ensuring that we generate the same size distribution of establishments (in terms of matched workers) within a census tract as we have in the sample. We do this simulation 100 times, and then compute the random network isolation index, NI^R, as the mean over these 100 simulations. Not surprisingly, in our very large sample, the random network isolation measures are very precise; in all cases the standard deviations were trivially small. We then focus on the difference $NI^O - NI^R$, measuring network isolation above and beyond that which occurs randomly; we refer to this as the "network isolation difference." 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.

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

[15] For example, consider a census tract that has two establishments in it, each of which is observed to employ two black workers. Assume that two of the black workers come from neighborhood A and two from neighborhood B. Even if the four black workers are randomly allocated across the two establishments without regard to the neighborhoods in which they live, 1/3 of the time they will be working with a neighbor, so that NI^R will be 1/3. This is a point that we discuss in some detail in the context of measuring workplace segregation (Hellerstein and Neumark, 2008), and was noted previously by others (see, e.g., Carrington and Troske, 1997).

importance of networks in generating the observed spatial pattern of neighbors working together in establishments. In particular, while NI^R gives us a measure of the minimum amount of clustering by neighborhoods that would happen with random assignment of workers to establishments within a census tract, it is also important to know what the maximum possible network isolation could be in our data, given the distribution of individuals across residential census tracts and the size distribution of establishments in each census tract. That is, we would like to know the maximum network isolation that could occur if workers were systematically assigned along with their neighbors to the maximum extent possible to establishments within the census tract in which they are actually observed to work.

While there may be individual instances of workplace census tracts in our data where we could compute the exact maximum network isolation index, we do not know of a general method for solving for the maximum index in all cases in our data.[16] We instead approximate the maximum network isolation through a "greedy" algorithm.[17] For a given census tract in which we have establishments represented in our sample, we order the neighborhoods in which workers live by the number of workers from each neighborhood. Beginning with the neighborhood with the greatest number of workers, we assign as many workers as possible to one establishment, and any workers who are not assigned to that establishment are grouped together and treated as a "new" neighborhood. We then move to the second largest neighborhood from which workers originate (which could be the "new" neighborhood left from the previous pass), and assign workers from that neighborhood to the establishment that holds the maximum number of these, again therefore keeping neighbors working together in establishments as much as possible. We continue moving down the list of neighborhoods, from those with larger to smaller numbers of workers, assigning workers to establishments until all workers are assigned. We would expect this

[16] It turns out that 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. This means that computer scientists think that there is no algorithm that will solve all potential instances of the network isolation index in time polynomial in the size of the problem. (In other words, there is unlikely to be any algorithm that will solve all potential realizations of the network isolation index in feasible computational time, particularly given the very large number of individuals in our data.) Moreover, as far as we are aware, there is also no approximation algorithm to compute the maximum network isolation that is known to be arbitrarily close to the exact maximum (something that does exist for other n-p complete problems).

[17] For more on n-p completeness and greedy algorithms, see, for example, Cormen et al. (2001).

maximum network isolation index to be less than one.[18]

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. We do this for every census tract in our sample where workers work, and then compute the weighted average of the maximum network isolation in each census tract of employment, weighting by the number of workers we observe to be working in that census tract. We label this maximum network isolation index NI^M.

Finally, we turn back to $(NI^O - NI^R)$, the difference between our observed network isolation index and the random isolation index, and we scale it by the maximum network isolation that can occur beyond randomness, or $(NI^M - NI^R)$, yielding

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

which we call the "effective network isolation index." It measures the share of the maximum possible network isolation difference that is actually observed, and provides a natural scaling for the importance of networks formed by residential networks in determining the establishments in which people are employed.

The greedy algorithm we use is bounded from above by the true maximum, but the extent to

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

which it is an understatement is unknown. There are two reasons why this does not concern us much. First, from an empirical standpoint, the conclusions we draw by comparing effective network isolation across subsamples (e.g., whites vs. Hispanics) are almost always the same conclusions that we draw simply by comparing the network isolation difference ($NI^O - NI^R$) across subsamples – that is, by comparing the numerators of the effective isolation indexes. Second, while 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,[19] 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.

Like any measure that tries to operationalize the concept of job networks, ours is limited to networks that operate *among* particular members, affecting employment in a *particular* set of jobs. Specifically, our measure only captures the extent to which networks operate to increase the likelihood that census tract co-residents work in the same establishment. 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 (perhaps because they participate together in other institutions or organizations), 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.

V. Sample Characteristics and Restrictions

In Table 1 we provide descriptive statistics for the matched workers from the DEED as compared to the SEDF. Column (1) reports summary statistics for the SEDF for the sample of white, black, or

[19] And, in the general case, the social planner would have to be able to solve the n-p complete problem of computing the maximum isolation index.

Hispanic workers who were eligible to be matched to their establishments. Column (2) reports summary statistics for all white, black or Hispanic workers in the full DEED sample – that is, those who we successfully matched to their establishments. The individuals we successfully match in the DEED are more likely to be female, to work full time, and to have more education than those in the SEDF. These differences result in part from the matching process, because there are many individuals who meet our sample inclusion criteria but for whom the quality of the business address information in the Write-In file is poor. We suspect that the differences in business address information partially reflect weaker labor market attachment among less-skilled workers, suggesting that estimates of the importance of networks we obtain might best be interpreted as measuring the extent of network isolation among workers who have relatively high attachment to the labor force and to their employers. The last eight rows of the table report on the industry distribution of workers. There is some over-representation of workers in manufacturing in the full DEED, because larger establishments are more likely to be matched, although the over-representation is not severe.[20]

We make several restrictions to arrive at the samples used for our analysis of network effects. First, we only consider workers who live and work in the same Metropolitan Statistical Area/Primary Metropolitan Statistical Area (MSA/PMSA). Second, in order to be able to identify a race- or ethnic-based workplace network, we exclude from the sample workers that have no co-workers in the same establishment who are of the same race or ethnicity as the worker. 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 means for the samples of white, black, and Hispanic workers that we analyze are shown in columns (3)-(5).

In addition to comparing worker-based means, it is useful to examine the characteristics of establishments in the DEED once we make our sample restrictions. Table 2 shows descriptive statistics for establishments in the full DEED and for each of our final analysis samples. Because only one in six

[20] For more information on the DEED versus the SEDF, see Hellerstein et al. (2008b).

workers is sent the Decennial Census Long Form, it is more likely that large establishments will have two matched workers, especially for smaller racial or ethnic groups. One can see evidence of the bias toward larger employers by comparing the medians across the columns for total employment in the establishment as recorded in the BR. (This bias presumably also influences the distribution of workers and establishments across industries, where, for example, the DEED itself, and the final analysis samples for blacks and Hispanics even more so, over-represent workers in manufacturing establishments.) The establishments corresponding to the full 2000 DEED sample of workers have median employment of 15 workers. Once we restrict attention to establishments in metropolitan areas with at least two matched white workers, and with at least two such establishments in the census tract, as in column (2), the median employment level rises to 35. Restricting our attention to those establishments in census tracts with another establishment having at least two black workers matched, in column (3), median employment is even higher, at 154; and in column (4), when we consider the sample of establishments employing two matched Hispanics in census tracts with at least two such establishments, median employment is 84. 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 white workers, who make up by far the largest subsample of workers that we use. Column (1) reports results for all approximately 1.7 million white workers in our sample. These workers work in 26,470 unique census tracts and live in 46,764 census tracts.[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 for the full sample of whites is 7.87, indicating that, on average, 7.87 percent of a white worker's white co-workers live in the same census tract as that worker.

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

When we randomize workers in this sample across establishments within census tracts, we recover a random network index of 2.97, less than 40 percent of the observed index, and the network isolation difference (the difference between the two) is 4.90. The maximum possible network isolation that could be observed in the data for all white workers is 54.84. This is well below 100 (perfect sorting of workers by residential neighborhoods into establishments), because in many of our census tracts there are establishments with more workers observed to work in them than are drawn from any particular census tract residential neighborhood. The maximum is also considerably above the observed network isolation measure.[22] As a result, when we scale up the network isolation difference by the maximum network isolation that occurs beyond randomness, we recover an effective network isolation index of 9.45. That is, approximately 9.5 percent of the maximum amount to which residential networks (at the census tract level) could contribute to the sorting of workers into establishments is actually observed in the data. Whether this is a large number or a small one is a subjective matter, of course, and there is little with which to compare it given the sparseness of empirical evidence on the importance of labor market networks. To us, however, it seems like a large number suggesting that labor market networks are quite important.

An alternative explanation for the apparent importance of networks in column (1) is that there is sorting of workers by both neighborhoods and establishments according to skill. For example, in a census tract with two establishments employing workers from two neighborhoods, perhaps one establishment hires only less-skilled workers (for example, grocery store cashiers) who tend to live together in a neighborhood where housing is cheap, and the second establishment hires only more-skilled workers (for example, lawyers) who tend to live together in a different neighborhood with more expensive housing.[23]

---

[22] Bootstrap methods show that all of the effective network isolation measures we report are statistically significantly different from zero. Indeed, the estimates of the effective network indices are quite precise, so that, in general, substantive numerical differences across columns and tables in reported effective network isolation measures are also statistically significant. (The 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.)

[23] In Hellerstein and Neumark (2008) we report evidence of some segregation of workers across establishments

Such sorting on skill potentially invalidates a network-based interpretation of evidence that people who live near each other tend to work together.[24]

In order to evaluate whether the results are driven by skill differentials that lead to both residential and workplace skill-based segregation across establishments, in columns (2) and (3) we report results separating whites by education level. Column (2) reports the results for whites with at most a high school education. That is, we take the sample of white workers from column (1) who have no more than a high school education and compute the network isolation measures for those workers. Specifically, we first calculate the observed network isolation index (NI^O) by averaging across the sample of low-educated white workers the fraction of each individual's white co-workers who live in that individual's residential census tract, regardless of the co-workers' education levels.[25] This number is 10.56, somewhat higher than the 7.87 number in column (1). We then calculate the network isolation index for low-educated white workers using the simulated sample from column (1). The random network isolation index is 4.06, also higher than its counterpart in column (1), so that the network isolation difference is 6.50. Finally, to calculate the maximum possible network isolation index, we calculate what the index would be if low-educated white workers were able to work to the maximum extent possible in establishments with their neighbors of any education level, given the size distribution of establishments in our sample, and the size and residential distribution of workers in them. This number is 61.62, which is again somewhat higher in column (1). Taking all of these together, the effective network isolation index is 11.29, 19 percent higher than the corresponding figure in column (1).

We perform the same exercise for whites with more than a high school degree and report the

---

based on skill levels. Bayer et al. (2005b) provide evidence of residential segregation by education for blacks.

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

[25] We could do a more extreme version of this calculation where we also restrict the sample of co-workers to be those with low education. This approach yields qualitatively similar conclusions for the various groups we study, but in some cases leads to rather small samples as it causes us to discard some establishments and some census tracts in order to meet the data requirements necessary to compute the indexes. Since we do not expect residential (or establishment) sorting by skill to be perfect, there is no reason to preclude labor market information flowing across workers of different skill levels.

results in column (3). Across all rows, the resulting measures of network isolation are smaller than in the full sample; the network isolation difference for this group is 4.10, and the effective network isolation index is 8.21. Looking across columns (1) through (3), the results suggest that networks are somewhat more important for low-educated whites, as has been suggested in previous surveys of workers' use of informal contacts by education. Of course, this difference by skill may arise because residence-based networks are more important for local labor markets, which in turn may be more significant for low-skilled than for high-skilled workers. An equally important conclusion from these results, however, is that effective network isolation is about as high or higher once we disaggregate by education, implying that the network results we report for the full sample in column (1) are not being spuriously driven by the joint sorting of workers by education level into neighborhoods and establishments.

Another potential 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 co-workers in neighborhoods, but this would not be due to the operation of residence-based networks. We do not think this alternative explanation of our findings is plausible, on a priori grounds, as job mobility is much higher than residential mobility. 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 one would make to be near a co-workers – was 9 percent on an annual rate.[26] 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.

To answer this question definitively, we would need to know where workers lived when they first began working (or applied for work) at a particular employer. We of course do not have such data. However, in the Census of 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

[26] See http://www.census.gov/population/socdemo/migration/p20-538/tab01.txt (viewed July 8, 2009).

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 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 quite a bit higher – 15.61 versus 9.45 – this may be because newer establishments are smaller; as we show later (in Table 8), in general the index is higher for small establishments. Finally, in column (3) we also restrict attention to those individuals who did not change residential addresses in the past five years. When we do this, the effective index climbs substantially, to 23.53. Note that this is for a sample of larger establishments 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. 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. This evidence does not just rule out the possibility that the causality runs from workplace location to residential location rather than the other way around, but it 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.

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

As noted earlier, the importance of network effects 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.[27] On the other hand, if labor market networks serve to overcome information imperfections more for blacks than for whites, perhaps by helping to lower search costs for blacks related to finding non-discriminatory employers, one might expect network isolation to be larger for blacks than for whites. Ultimately, this is an empirical question.

In Table 5 we provide results for black workers. Column (1) reports results for the sample of black workers. We estimate that the observed network isolation index (NI^O) for all blacks is 5.29, somewhat smaller than for whites, and the random network isolation index (NI^R) is 2.58, just slightly below that for whites; the network isolation difference is therefore 2.71. However, for blacks the maximum possible network isolation index (NI^M) is quite a bit below that for whites (31.60), with the net result that the effective network isolation index of 9.35 is very close to what we find for whites. (In fact, the 9.35 estimated effective network isolation index for all blacks is not statistically significantly different from the estimate of 9.45 for all whites.) These full sample results therefore suggest little racial difference 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.[28]

In column (2) of Table 5 we report results for blacks who have at most a high school education, and in column (3) we report results for blacks who have more than a high school education. The results by education are significant both to address the issue of sorting by skill, and to 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, each of the observed indexes is lower than for whites, but as with the full sample results the overall effective network isolation indexes for blacks by education level are substantively similar to those for whites (although in each case the network isolation difference is smaller for blacks than for whites). Whether or not we scale by the

[27] In our full samples of whites and blacks in columns (1) of Tables 3 and 5, blacks have an average of 10 black working neighbors, whereas whites have an average of 64 white working neighbors.
[28] We noted earlier that we find only a few instances where comparisons of the network isolation difference, $NI^O - NI^R$, and the effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$, lead to qualitatively different conclusions. The results for blacks and whites discussed in this paragraph point 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.

maximum, though, the evidence suggests that residentially-based networks are more important for less-educated blacks than for more-educated blacks. For example, for less-educated blacks networks generate 11.52 percent of the maximum amount of sorting by neighborhood that could occur (and this is not statistically significantly different from the estimate of 11.29 percent for less-educated whites), and for more-educated blacks, the effective index is 7.01 (compared to 8.21 for more-educated whites). On the other hand, conditioning on education has no substantive effect on the black-white comparison of the importance of labor market networks.

Overall, the results in Tables 3 and 5 are in line with other results from the literature, based on quite different types of analyses, including: survey results or indirect evidence indicating greater use of informal contacts among the less-educated (Ioannides and Datcher Loury, 2004; Topa, 2001); evidence based on place of work and place of residence indicating stronger network effects among those with less education (Bayer et al., 2005a); [29] and an absence of consistent evidence of race differences in the reported use of informal contacts (Ioannides and Datcher Loury, 2004).

Table 6 presents some additional analyses for black and white workers. One issue with directly comparing the full sample results for blacks and whites is that we know these groups are not similarly distributed geographically in the United States. As a result, 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 6 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.[30] For comparison, columns (1) and (2) repeat the full sample results for whites and blacks from Table 3. For whites, the sample restriction reduces the sample by about one-half, and the numbers of workplace census tracts and residential census

[29] 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.
[30] 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).

tracts are also reduced considerably. The results are different than for the full sample of whites. First, the observed network isolation index is lower at 5.68, and interestingly is closer to that for blacks. Second, the effective network isolation measure is lower (7.00 vs. 9.45 in column (1)).[31] Column (4) reports results for blacks. The effective network isolation index for the restricted sample of blacks is 9.08, quite similar to the figure of 9.35 in column (2), which is not surprising since between columns (2) and (4) the sample is reduced by fewer than 4,000 workers.[32] The random network isolation measure is quite a bit lower for this sample of whites than for blacks (1.48 vs. 2.55), so that the network isolation difference is higher for whites than for blacks (4.20 vs. 2.71). 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. We conclude that the effective network isolation measure for the consistent samples is actually higher for blacks than for whites, although not by an amount that we deem important, especially given that the difference is driven by our calculation of the maximum.

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 job referrals is less likely to flow between black and white co-residents than between co-

[31] 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 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.

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

residents of the same race. Understanding the role of race in driving network effects is extremely important. Race-based networks are central to the work of Calvó-Armengol and Jackson (2007) in deriving the result that networks can perpetuate and exacerbate initial differences in employment between blacks and whites, although in their model racial stratification due solely to residential segregation is sufficient to generate this result. In contrast, racial stratification of networks even *within* neighborhoods is potentially more important. It can potentially explain the results in Hellerstein et al. (2008a) that higher local job density for one's own race affects employment probabilities, but higher job density for the other race does not.[33] More significantly, this type of racial stratification would imply that policies that solely address spatial mismatch, by attempting to move blacks to areas where more whites live and where more jobs (per person) are located, may fail to help blacks precisely because network connections are severed, and are less likely to be established with white neighbors.[34]

Therefore, in column (5) of Table 6 we assess more directly whether networks are race-based. To do this, we carry out the same types of sampling and computational procedures used before, except that we consider the relevant set of a black worker's neighbors and co-workers to consist of blacks *or* whites. We begin by constructing a sample of black workers 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 in the sample.. We again construct an effective network isolation index for black workers in this sample, but what is different now is that we construct this measure by asking whether our sample of black workers are more likely than would be predicted by randomness to work in the same establishment with a

[33] 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.

[34] 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 did include 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 neighbors – however defined – are not the only source of network connections.

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.

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 6, 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 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. To obtain a measure of the maximum possible network isolation index, we again use the same greedy strategy that approximates 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 30.74, which is similar to that in column (2). Taking all of these together, the effective network isolation index is only 6.94, which is about 25 percent smaller than in column (2), providing evidence that residence-based labor market networks have a fairly strong race-based component. Moreover, because in this column we only need one black worker in an establishment for the establishment to be in the sample, as opposed to column (2), which requires two black workers, column (3) 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 index in columns (2) and (3) would be even larger, bolstering the evidence that these networks are racially stratified.[35]

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 the finding that Hispanics with poor English skills pay 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.[36] Munshi (2003) presents a more-refined analysis of Mexican immigrants, tying labor market outcomes to a larger local population of immigrants from the same origin community. Patel and Vella (2007) find that new immigrants work disproportionately in occupations held by previous immigrants from the same country. And our previous work documents establishment-level segregation by English language skills, and segregation of Spanish-speaking workers from non-Spanish speaking workers among poor English speakers (Hellerstein and Neumark, 2008). Finally, perhaps the most direct evidence of these types of networks for immigrants comes from the work of Massey et al. (1987), who document the importance of networks linking recent and earlier immigrants from the same communities in Mexico.

In this section, therefore, we turn to an analysis of results for Hispanic workers, paying particular attention to Hispanic workers who speak English poorly (or not at all) and Hispanic workers who are

[35] In future work, we plan on trying to determine the extent to which race-based networks are attributable to residential segregation by race, as opposed to weaker network connections across than within races for those residing in the same neighborhood.
[36] For a similar type of evidence for Sweden, see Edin et al. (2003).

immigrants. The results are presented in Table 7. 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 21.37, which is more than twice as large as what we find for blacks or whites.[37]

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 20.27. This is very large – it means that, on average, for a poor-English speaking worker, 20.27 percent of his or her co-workers who also are poor English speakers live in the same census tract! The random network isolation index is much smaller, at 4.99, and the maximum network index is 54.32. Taken together, these numbers yield an effective network isolation index of 30.98, meaning that over 30 percent of the maximum possible establishment network isolation by census tract of residence for Hispanics who speak English poorly is actually observed in the data. This is more than three times larger than what we find for blacks and whites, and suggests to us that 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.

By way of contrast, in column (3) of Table 7 we report the results for the sample of Hispanics who report speaking English "well" or "very well." The effective network isolation index is 17.47, just over half as large as that for Hispanics who are poor English speakers. This contrast is consistent with the idea that networks are extremely important in mitigating the high search frictions that exist for workers in

[37] 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.

the United States whose English language skills are poor. In addition, 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 27.12, which is quite a bit higher than for all Hispanics.[38] In contrast, in column (5) we report the results for non-immigrant Hispanics, for whom the effective network isolation index is 13.62, smaller than for the full sample of Hispanics or any of the other Hispanic subgroups. 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 what we are capturing in our measure of network isolation is, indeed, the important of residence-based networks that reduce search frictions in the labor market.[39]

VI.5. Results for Small Establishments

We noted earlier that our sample selection rules lead to under-representation of small establishments for blacks, and to a lesser extent for Hispanics. If network isolation differs in a way that is related to establishment size, the different sample compositions of establishments could bias our comparisons of the importance of networks across racial and ethnic groups. In fact, other evidence indicates that smaller establishments rely more heavily on informal referrals (e.g., Holzer, 1998). To the extent that these referrals are associated with the types of network effects we capture, under-representation of small establishments in our black and Hispanic samples likely results in downward bias in our estimates of the importance of networks for minorities.

Table 8 reports evidence that addresses this type of bias. In particular, it reports our baseline analyses for whites, blacks, and Hispanics, but utilizing a restricted sample of workers employed only at establishments with 50 or fewer workers. Two things are apparent. First, compared to the corresponding estimates in columns (1) and (4) of Table 3 and column (1) of Table 7, networks appear much more important when we restrict the analysis to small establishments. Second, and more to the point, the racial

[38] The results are similar for only Mexican immigrants, who represent the largest share of Hispanic immigrants.
[39] 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.

and ethnic comparisons in Table 8, which are based on much more homogeneous samples with respect to establishment size, suggest that residence-based networks are much more important for blacks than for whites (an effective network isolation index of 38.26 vs. 20.47,[40] and even more so for Hispanics relative to whites. Thus, the relative importance of networks for blacks and especially Hispanics compared to whites is greater than is suggested by the analyses of our full samples in Tables 3-5.[41]

VI.6. Network Isolation Conditional on Industry

In interpreting the results to this point we have presumed that the residence-based networks we are measuring operate to help workers find jobs in particular establishments. However, networks may instead (or as well) serve to help job searchers learn of vacancies in certain industries, rather than reducing frictions that prevent workers from matching to specific establishments within industries. For example, someone who works in a retail firm may tell a neighbor of job vacancy postings in other nearby retail establishments. Note that, in principle, this kind of mechanism could underlie the results in Bayer et al. (2005a), since they only establish that those who live nearby are likely to work in the same narrow geographic area, not the same establishment.

If networks operate to increase the likelihood that census tract co-residents work in the same industry, our calculations to this point might overstate the extent to which networks determine the establishment of employment, because the clustering of workers from the same census tract of residence in the same industry within a census tract of employment will inevitably lead to some clustering in the same establishments. In this section we explore whether the network effects we find reflect employment at the establishment level, or instead only at the industry level. We do this by constructing "conditional" network isolation indexes, simulating network isolation while holding the distribution of workers across industries fixed within a census tract of employment. 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

[40] Paralleling the results in Table 6, columns (3) and (4), this conclusion for black-white differences is driven by the scaling by maximum segregation.

[41] The analysis of blacks and whites for these restricted samples also yields evidence indicating that networks are race-based, similar to that reported in column (5) of Table 6.

simulation will preserve that particular industry concentration, and by subtracting off the network isolation that occurs randomly *conditional* on industry we 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.[42]

To condition on industry, 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 establishments in the tract, holding the size distribution of establishments fixed, we instead ensure that workers are *also* assigned to their industry of employment.[43] We then once again compute the average (across the simulations) simulated fraction of co-workers who come from a worker's own neighborhood, denoting this NI^C, and we define the extent of "conditional effective network isolation" to be:

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

where NI^R and NI^M are defined as before, without regard to industry. A conditional effective network isolation index of zero (when $NI^O = NI^C$) would imply that all of the effective network isolation can be attributed to networks that help workers find employment in specific industries, but not to establishments within industries; that is, above and beyond the clustering of employment of neighbors in the same industry, there is no clustering in the same establishment. Conversely, a network isolation index equal to that of the (unconditional) effective segregation measure (when $NI^C = NI^R$) would imply that all of the effective network isolation comes from networks helping individuals find jobs in specific establishments within industries, and that industry, per se, plays no role in sorting.

For brevity, we report in Table 9 results on conditional network isolation only for groups that are

[42] Earlier, we noted that we have chosen a specific way to operationalize networks – as affecting the establishments at which people work. Here we are making a different argument. In particular, having chosen this definition of networks, we could be overstating the importance of networks if networks affect the industry of employment, because 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.

[43] The industry definitions that we use are the same eight industries reported in Tables 1 and 2: mining; construction; manufacturing; transportation; wholesale trade; retail trade; FIRE; and services. We do not want a more highly-detailed industry classification for this exercise. First, we are interested in the flow of information about nearby jobs, which need not be in the same finely-classified industry. Second, our test depends on having multiple establishments within the same industry in a census tract, which obviously would occur less frequently the more we disaggregate industries.

"low-skilled" in the sense that they either have low levels of education (for whites and blacks) or poor English proficiency (for Hispanics). These are the groups for which residence-based networks appear most important. In column (1) we repeat the network isolation results for less-educated whites that we previously reported in Table 3, column (2). Then, in column (2), we report results where we condition the random network isolation index on industry. The observed network isolation index, the (unconditional) random isolation index, and the maximum network isolation index all remain the same, as the conditional random isolation index does not play a role in these calculations. The simulated conditional index is 5.46, which means that when workers are randomly assigned to establishments in the same industry in which they are observed to work (and the same census tract), on average 5.46 percent of their co-workers will come from the same residential neighborhood. This is higher than the unconditional random isolation index of 4.06, so that that the conditional effective network isolation index of 8.86 is somewhat smaller than the unconditional effective index of 11.29 reported in column (1). The difference implies that assignment of workers from the same neighborhoods to specific industries within a census tract can explain some of the assignment of workers to specific establishments. However, even after conditioning on industry, effective network isolation is still relatively high, and dividing 8.86 by 11.29, 78 percent of the effective network isolation remains even after we condition on a worker's industry.[44] This demonstrates that, at this level of industry detail, most of the (unconditional) effective network isolation for less-educated whites cannot be explained just by a mechanism whereby residence-based networks serve only to help workers find jobs in the same industries as their neighbors.

Columns (3) and (4) of Table 9 explore this issue for less-educated blacks. Column (3) repeats the unconditional network isolation indexes from Table 3, column (5), and column (4) reports the indexes conditional on industry. The results are very similar to those for less-educated whites. The conditional

[44] In a sense, this is likely a lower bound for the percentage of the effective network isolation index that remains. If there is only a small number of establishments in an industry in a particular census tract, then what this procedure treats as sorting on industry may in fact represent sorting on establishments. In the limit, if there were 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.

random network isolation index of 4.72 reported in column (4) is somewhat larger than the unconditional random index of 3.51, so that the conditional effective network isolation index of 7.95 is somewhat smaller than the unconditional index of 11.52. And, as the last row of column (4) reports, assignment of workers to industry alone rather than to specific establishments only explains slightly more than 30 percent of effective network isolation.

Finally, columns (5) and (6) turn to estimates of network isolation for low-skilled Hispanics, with low skill defined as poor English proficiency. The results parallel closely the qualitative results for less-educated whites and blacks. Column (5) repeats the unconditional results for less-skilled Hispanics previously reported in Table 7, column (2). Column (6) reports the conditional indexes, where the conditional effective network isolation of 21.86, when compared to the (unconditional) effective network isolation index of 30.98, implies that 71 percent of the assignment of workers from residential neighborhoods to establishments within a census tract cannot be explained by industry alone.

We conclude from these analyses conditioning on industry that the network isolation results we find are largely due to the assignment of workers to specific establishments, providing evidence in support of theoretical models in which networks serve to match workers to specific establishments.[45]

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

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

establishment's workforce is consistent with labor market networks that connect individuals residing in the same neighborhood to specific business establishments. Because of recent research highlighting the potential importance of labor market networks for less-skilled workers in the labor market, and more generally positing that labor market networks operate along the lines of race, ethnicity, and skill, we consider separately the importance of these labor market networks for whites, blacks, and Hispanics, and, within each group, the relative importance of these networks for workers with different skills.

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

We interpret the evidence as indicating that labor market networks play an important role in establishment-level employment. For both whites and blacks we find that the grouping of workers from the same neighborhoods in the same business establishments exceeds by a factor of more than two what we would expect to occur randomly. For whites, we find that network isolation is about 9.4 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 some evidence that 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 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. Moreover, race-based labor market networks likely limit the ability of spatial policies – which encourage blacks to move to areas where there are more jobs but also likely more white co-residents – to improve labor market outcomes for blacks.

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.[46]

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? The DEED is likely to prove useful in trying to address these questions in future research.

[46] 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.

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 the possibilities are schools,[47] 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 actually work. 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.

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

References

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

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

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

Blau, David. 1992. "An Empirical Analysis of Employed and Unemployed Job Search Behavior." *Industrial and Labor Relations Review*, Vol. 45, No. 4, July, pp. 738-52.

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

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

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).

Corcoran, Mary, Linda Datcher, and Greg Duncan. 1980. "Information and Influence Networks in Labor Markets." In Five Thousand American Families, Vol. VIII, Greg Duncan and James Morgan, eds. University of Michigan, Institute for Social Research, pp. 1-37.

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.

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. 2008a. "Spatial Mismatch vs. Racial Mismatch?" *Journal of Urban Economics*, Vol. 64, No. 2, September, pp. 467-79.

Hellerstein, Judith, David Neumark, and Melissa McInerney. 2008b. "Changes in Workplace Segregation in the United States between 1990 and 2000: Evidence from Matched Employer-Employee Data." In The Analysis of Firms and Employees: Quantitative and Qualitative Approaches, Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter, Eds. Chicago: University of Chicago Press, pp.

163-195.

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.

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

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.

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.

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.

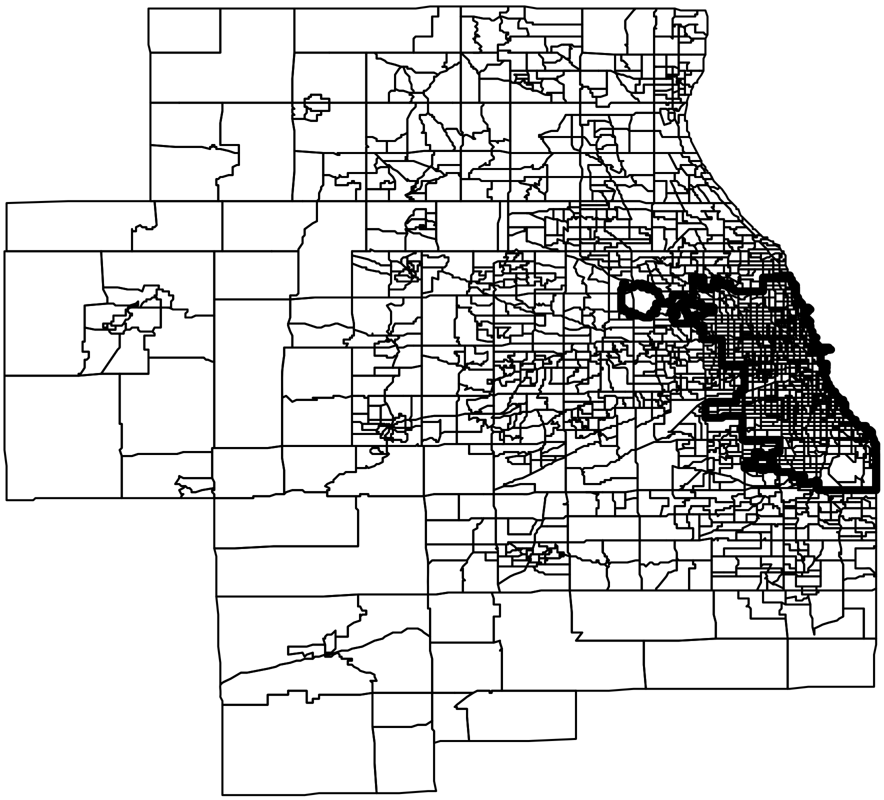
Topa, Giorgio. 2001. “Social Interactions, Local Spillovers, and Unemployment.” *Review of Economic Studies*, Vol. 68, No. 2, April, pp. 261-95.

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.

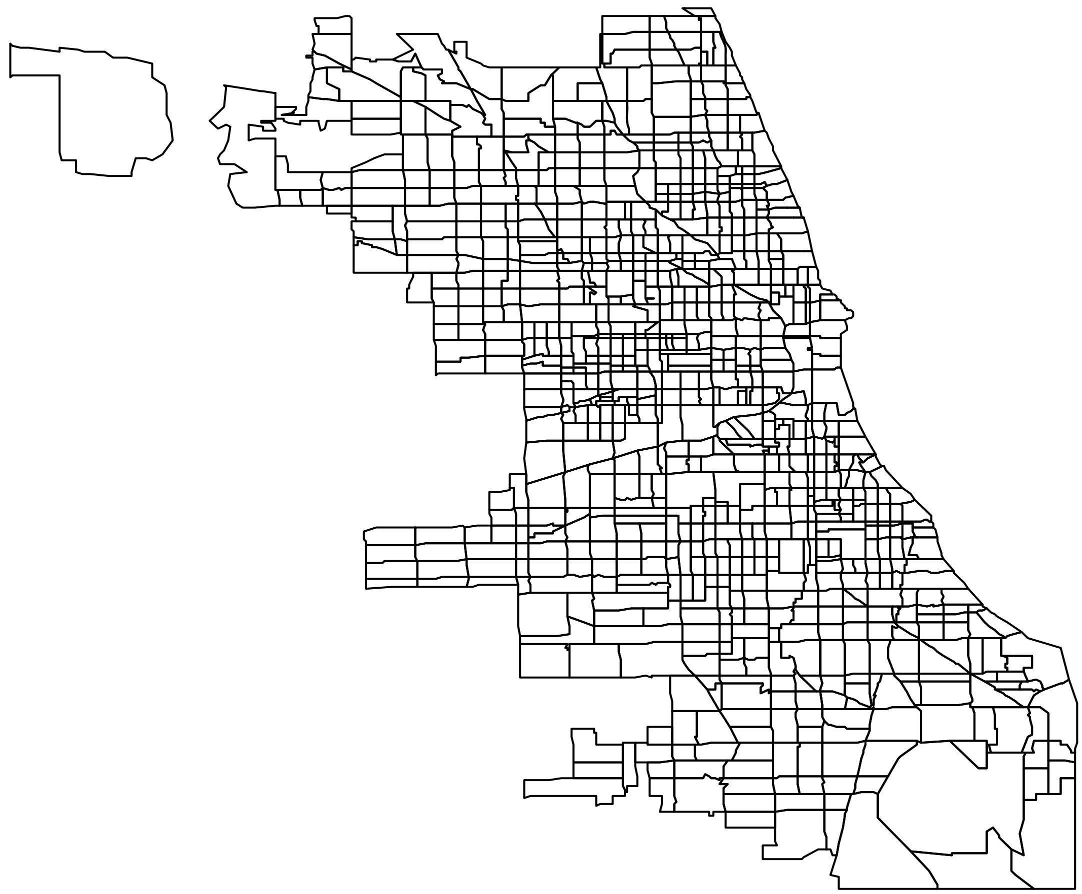
Weinberg, Bruce A., Patricia B. Reagan, and Jeffrey J. Yankow. 2004. "Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data." *Journal of Labor Economics*, Vol. 22, No. 4, October, pp. 891-934.

Figure 1: Census Tracts, Chicago PMSA



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

Figure 2: Census Tracts, Chicago



Source: U.S. Census Bureau, http://www.census.gov/geo/www.cob.tr2000.html (viewed May 5, 2008).

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

	All	High school degree or less	More than high school degree
	(1)	(2)	(3)
Network isolation index, observed, NI^O	7.87	10.56	6.51
Simulated random network isolation index, NI^R	2.97	4.06	2.41
Network isolation difference, $NI^O - NI^R$	4.90	6.50	4.10
Maximum possible network isolation index, NI^M	54.84	61.62	52.32
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)] \cdot 100$	*9.45*	*11.29*	*8.21*
N	1,675,412	561,370	1,114,042
# place of work tracts	26,470	25,690	26,299
# residential tracts	46,764	43,469	45,666
Mean establishments/tract	129.6	114.5	137.2
Mean matched workers/establishment	38.4	24.4	45.4
Mean number of workers in tract of employment from same tract of residence	9.4	11.2	8.5

Notes: The calculation is described in the text. NI^O is the average fraction of a worker's co-workers (i.e., excluding the worker) who reside in the same census tract as the worker, averaged across all workers in the sample. NI^R is the average fraction that is simulated to occur randomly. NI^M is the simulated average maximum fraction. "Effective network isolation" therefore measures the fraction of the maximum that is actually observed.

Table 4: Network Isolation for Whites, with 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	8.68	10.48
Simulated random network isolation index, NI^R	2.97	3.07	3.70
Network isolation difference, $NI^O - NI^R$	4.90	5.62	6.78
Maximum possible network isolation index, NI^M	54.84	39.03	32.51
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*9.45*	*15.61*	*23.53*
N	1,675,412	206,261	65,299
# place of work tracts	26,470	8,692	3,953
# residential tracts	46,764	38,855	26,010
Mean establishments/tract	129.6	38.0	17.2
Mean matched workers/establishment	38.4	18.4	13.6
Mean number of workers in tract of employment from same tract of residence	9.4	2.8	2.0

Notes: See notes to Table 3.

Table 5: Network Isolation for Blacks, Overall and by Education

	All	High school degree or less	More than high school degree
	(1)	(2)	(3)
Network isolation index, observed, NI^O	5.29	7.42	3.90
Simulated random network isolation index, NI^R	2.58	3.51	1.97
Network isolation difference, $NI^O - NI^R$	2.71	3.90	1.93
Maximum possible network isolation index, NI^M	31.60	37.40	29.55
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*9.35*	*11.52*	*7.01*
N	97,967	38,754	59,213
# place of work tracts	4,490	4,301	4,350
# residential tracts	21,623	13,598	18,299
Mean establishments/tract	22.2	19.8	23.8
Mean matched workers/establishment	18.6	15.6	20.5
Mean number of workers in tract of employment from same tract of residence	2.6	2.9	2.4

Notes: See notes to Table 3.

Table 6: Network Isolation for Blacks and Whites, for Consistent Samples of Establishments, and Ignoring Race

			Establishments located in tracts in both black and white samples		Network isolation based on blacks and whites
	All whites	All blacks	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	2.55	2.00
Network isolation difference, $NI^O - NI^R$	4.90	2.71	4.20	2.71	2.00
Maximum possible network isolation index, NI^M	54.84	31.60	61.43	32.35	30.74
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*9.45*	*9.35*	*7.00*	*9.08*	*6.94*
					140,083
N	1,675,412	97,967	845,290	94,210	9,094
# place of work tracts	26,470	4,490	4,122	4,122	26,768
# residential tracts	46,764	21,623	42,533	21,459	39.7
Mean establishments/tract	129.6	22.2	229.7	23.0	40.3
Mean matched workers/establishment	38.4	18.6	46.6	17.7	4.5
Mean number of workers in tract of employment from same tract of residence	9.4	2.6	11.4	2.6	2.00

Notes: See notes to Table 3.

Table 7: Network Isolation for Hispanics, Overall and by Skill and Immigrant Status

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

Notes: See notes to Table 3.

Table 8: Network Isolation for Whites, Blacks, and Hispanics Working in Small Establishments (50 employees or less)

	Whites	Blacks	Hispanics
	(1)	(2)	(3)
Network isolation index, observed, NI^O	15.76	12.76	22.79
Simulated random network isolation index, NI^R	4.49	4.43	5.74
Network isolation difference, $NI^O - NI^R$	11.27	8.33	17.05
Maximum possible network isolation index, NI^M	59.54	26.20	40.71
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*20.47*	*38.26*	*48.75*
N	527,430	8,706	21,952
# place of work tracts	22,162	1,097	2,093
# residential tracts	43,700	5,308	8,019
Mean establishments/tract	66.0	7.4	23.0
Mean matched workers/establishment	3.3	2.7	2.7
Mean number of workers in tract of employment from same tract of residence	3.9	1.4	1.8

Notes: See notes to Table 3.

Table 9: Network Isolation for Low-Skilled Workers, Conditional on Industry

	Whites, high school or less		Blacks, high school or less		Hispanics, poor English skills	
	Uncond.	Cond.	Uncond.	Cond.	Uncond.	Cond.
	(1)	(2)	(3)	(4)	(5)	(6)
Network isolation index, observed, NI^O	10.56	10.56	7.42	7.42	20.27	20.27
Simulated random network isolation index, NI^R	4.06	4.06	3.51	3.51	4.99	4.99
Simulated conditional network isolation index, NI^C		5.46		4.72		9.48
$NI^O - NI^C$		5.10		2.69		10.78
Maximum possible network isolation index, NI^M	61.62	61.62	37.40	37.40	54.32	54.32
Effective network isolation index, $[(NI^O - NI^R)/(NI^M - NI^R)]\cdot 100$	*11.29*		*11.52*		*30.98*	
Conditional effective network isolation index, $[(NI^O - NI^C)/(NI^M - NI^R)]\cdot 100$		*8.86*		*7.95*		*21.86*
Share of effective isolation index unexplained		*78.46*		*69.01*		*70.57*

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