# Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility

Ian M. Schmutte*

*Department of Economics, University of Georgia, Athens, GA, USA 30602*

**Abstract**

Most theories of the labor market imply mobility and matching should be limited: particular workers are much more likely to match with and move to particular jobs. However, the boundaries of mobility and matching have many overlapping determinants that are hard to characterize. This paper implements a novel method for analyzing complex networks to learn where the boundaries to job mobility lie. Using data from the Panel Study of Income Dynamics (PSID), I uncover densely connected matching sets in the 'realized mobility network' that connect workers to detailed industry-occupation pairs. I find first that the labor market is composed of four distinct segments between which mobility is relatively unlikely. Second, these segments are not well-described on the basis of industry, occupation, demographic characteristics, or education. Third, mobility segments are associated with earnings heterogeneity, and high-earning workers tend to be observed in high-paying jobs. Fourth, the boundaries to job mobility are counter-cyclical: workers move more freely when unemployment is low.

*Keywords:* Job Mobility, Complex Networks, Job Matching

## 1. Introduction

The flexibility with which workers move between different types of employment affects many economic outcomes, including earnings inequality (Autor and Dorn 2013), the persis-

*Tel: +1 706 542 0163

*Email address:* schmutte@uga.edu (Ian M. Schmutte)

tence of unemployment (Şahin et al. 2012), and individual earnings growth (Gathmann and Schoenberg 2010). Recognizing this, economists are devoting renewed attention to models in which workers cannot move with perfect ease between employment opportunities because of mismatch, imperfect portability of skills, imperfect information about job opportunities, or institutional and psychic barriers to mobility. Recent research has begun to focus on the complex interaction of factors that motivate and constrain mobilility (Neal 1999; Pavan 2011). However, no one has empirically characterized a common implication of such models: that there are groups of workers and groups of jobs amongst which employment mobility is relatively likely, and other groups among which mobility is relatively unlikely. In this paper, I document the network structure of job mobility, drawing on tools for complex network analysis to discover where the limits to mobility lie.

The network of connections between workers and employers evolves over time as people move from job to job. This *realized mobility network* connects workers in the economy to every employer with whom they have held a job. Models of limited mobility predict that this network will be densely connected among jobs between which it is easy to move, and sparsely connected elsewhere. In complex network analysis, the task of finding densely connected groups of nodes is known as "community structure detection". In this paper, I apply a method for detecting community structure in a general network, modularity maximization (Girvan and Newman 2002; Blondel et al. 2008), to a realized mobility network constructed from the Panel Study of Income Dynamics (PSID).

To motivate this approach, consider the basic dual labor market model (Taubman and Wachter 1986), which posits the existence of 'good' jobs which are rationed, and 'bad' jobs with low pay and job security. In that model, the vertices of the realized mobility network could be partitioned into two non-intersecting sets: the set of 'good' jobs and the set of 'bad' jobs. Some workers obtain jobs in the 'good' sector and stay there, and some are stuck moving mostly among the 'bad' jobs. In the extreme case where the probability of transition between the two segments is zero, the realized mobility network will always have two unconnected components corresponding to the two segments. This intuition is echoed in more recent models where workers and firms attempt to match on the basis of productive characteristics in the presence of search frictions (Postel-Vinay and Robin 2002; Shimer 2005; Şahin et al. 2012). Sattinger (2006) emphasizes that a common implication of such models is the presence of overlapping labor markets. Workers try to match in particular sets of jobs, but because of frictions there is overlap in the realized matching as workers occasionally move between jobs in different markets. In such a model, matching is not frictionless, and the underlying market segments can not be revealed by the component structure of the graph. However, within each market segment, the realized mobility network is more densely connected. The empirical task is to reveal labor market segments by identifying those pockets of relatively high density.

Figure 1 illustrates the empirical strategy. In the figure, nodes correspond to employers, which are connected whenever they share a worker in common. If workers are more likely to move within certain groups of jobs than others, then certain subsets of nodes are densely connected. The nodes have been colored according to the classification revealed through modularity maximization. The figure is laid out using an algorithm that heuristically minimizes the spatial distance between connected nodes. The data reveal four distinct groups of jobs and a split along two axes. I describe the method for arriving at this partition and its implications at length in Section 4.1.

While this looks interesting, labor economists already have a pretty good idea of where the boundaries to mobility lie: we expect workers to be considerably less mobile between than within industries and occupations. To a lesser extent, we also expect workers to be more likely to work with others that share their demographic characteristics. These observations suggest two questions. First, how tightly connected is the realized mobility network on the basis of known, observable characteristics? Second, and more critically, what do network analysis and modularity maximization buy us?

To demonstrate the practical value of the network approach, I adapt the concept of *homophily* (McPherson et al. 2001; Currarini et al. 2009) to directly connect the labor economics literature analyzing transition matrices to the complex network literature and community structure detection through modularity maximization. Homophily measures the strength of within-class connections (transitions) relative to their expectation under random matching. As I will show, the modularity of a network, given a particular partition, is a weighted sum of the homophily of the individual partition classes. Therefore, modularity maximization is equivalent to finding a partition of workers and jobs with the highest aggregate homophily. I show that the maximum modularity partition reveals more structure than can be revealed in an analysis of transition matrices on the basis of industry or occupation categories alone (Kambourov and Manovskii 2008; Parrado et al. 2007).

I develop three applications of network statistics to labor market and policy analysis. First, I discuss implications for the analysis of human capital specificity. Then, I discuss implications for mismatch unemployment and find that workers are less likely to cross labor market boundaries when labor market conditions are poor. Finally, I show that the groups of jobs and workers revealed on the basis of mobility patterns are distinguished by a significant amount of earnings heterogeneity. This earnings heterogeneity admits an analysis of assortative matching – who matches with whom. I find evidence that high-wage workers tend to be employed in high-paying jobs, a fact that has become rather controversial in the literature using matched employer-employee data. These applications illustrate the value of incorporating network statistics into the toolbox of empirical labor economics.

## 2. The Labor Market as a Network

Figure 2 illustrates the formation of a realized mobility network from panel data collected over three time periods in a labor market where the set of workers is $W = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and the set of employers is $J = \{A, B, C, D\}$. Workers and employers are nodes in the network, and an edge connects a worker and employer whenever there is an employment relationship between them. The example shows that while the structure of employer-employee links is sparsely connected at any point in time, the realized mobility network may, and indeed does, become densely connected very quickly. In this section, I introduce the concepts of network analysis and their application to labor market data.

### 2.1. Preliminaries

A *graph* or *network,* $G$, is defined by a set of *nodes* or *vertices*, $V(G) = \{1, ..., N\}$, and a set of *edges* that connect them, $E(G) \subset V(G) \times V(G)$. The edges are undirected $(i, j) \in E(G)$ whenever $(j, i) \in E(G)$.

This paper considers networks in which multiple edges can form between two nodes, represented by including an edge multiple times in $E(G)$. An alternative is to characterize an edge as a triple: $(i, j, \omega) \in E(G) \equiv V(G) \times V(G) \times \Omega$, where $\Omega$ is the set of whole numbers.

The set of *neighbors* of $i$ is $N_G(i) = \{j \in V(G) : ij \in E(G)\}$, the set of all nodes that are connected to $i$. The *degree of* $i$, $k_i$, is the number of $i$'s neighbors: $|N_G(i)|$.

Finally, the *adjacency matrix* representation of $G$, $A^G$, is an $N \times N$ matrix whose $ij$th entry is $\omega$ if $(i, j, \omega) \in E(G)$ and 0 otherwise. Most of the computations performed in this paper are based on manipulations of the adjacency matrix of the realized mobility network. For instance, the degree list of $G$ is simply $A^G u_N \equiv k$, where $u_N$ is the $N \times 1$ sum vector and $k = [k_1, \ldots, k_N]^T$ is the $N \times 1$ vector whose entries are the degree of each node.

### 2.2. The Realized Mobility Network

The realized mobility network is an *undirected bipartite graph* since a defining feature is that nodes representing workers only connect to nodes representing employers. To analyze matching in the realized mobility network, I work with this bipartite representation, but also with its "one-mode" projections: graphs of indirect connections between workers and employers.

#### 2.2.1. The Bipartite Representation

The realized mobility network is generated from panel data with representative entry $\{w, t, m\}$, where $w \in W$ is the index of a particular worker, $t \in T$ is the time period, and $m \in M$ identifies the employer of $w$ at time $t$. The subsample of jobs in progress at a fixed point in time, $t$, are edges in a bipartite graph, $R_t$, which is the *realized employment network* at time $t$. The vertex set is $V(R_t) = (W, M)$. An element $(w, m)$ of the edge set $E(R_t)$ indicates that worker $m$ holds a job with employer $m$. The *realized mobility network*, $R$, accumulates information about matching over the sample period, $t \in \{1, ..., T\}$, by setting $V(R) = (W, M)$ and $E(R) = \cup_{t=1}^{T} E(R_t)$. Note that the realized mobility network is, therefore, unweighted.

*2.2.2. One-Mode Projections onto Workers and Employers*

Two employers are connected in the "one-mode employer projection" of the realized mobility network whenever they share an employee in common. Many workers can hold jobs with two employers. To recover the structure of matching in the bipartite graph, we want to keep track of the number of such connections between two employers.[1] The edge connecting two employers in the one-mode projection is weighted by the number of common employees.

Formally, the one-mode employer projection graph is denoted by $R^M$. Its vertex set is $M$, the set of employers. The edge set is

$$E(R^W) = \left\{ \begin{array}{l} (m, n, \omega) : m, n \in M \wedge \\ \omega = |\{w \in W : (w, m) \wedge (w, n) \in R\}| \end{array} \right\}. \tag{1}$$

The definition of the one-mode projection onto the set of worker nodes, $R^W$, is analogous.

*2.3. Characterizing Assortativity*

This section introduces the specific measures of homophily and modularity that I use to characterize boundaries to mobility and the assortatitivity of labor market matching.

*2.3.1. Node Partitions and Partition Graphs*

Given the employer projection of the realized mobility network, $R^M$, let $\Phi^M$ be the set of all partitions of the set of employers in $M$. The representative element, $\phi^M \in \Phi^M$, is a set of collection of subsets of nodes in $M$ whose union is equal to $M$ and whose intersection is the empty set. Where the context makes it clear, I will drop the superscript denoting the underlying node set. For convenience, the elements of the partition are indexed by natural numbers. That is,

$$\phi = \{\phi_1, \ldots, \phi_L\}. \tag{2}$$

[1] Intuitively, if one worker moves between two employers, it might be a fluke. If many workers do it, it more likely indicates something common about the skill demand of the employers.

In the analysis, I represent the partition, $\phi$, of $M$ into $L$ classes by a $|M| \times L$ matrix, $\Pi^{\phi}$. The $(w, \ell)$ entry of $\Pi^{\phi}$ is 1 if $w \in \phi_{\ell}$ and zero otherwise. For example, consider the partition of the set of employer nodes into major industries, denoted $\phi^{ind}$. The associated matrix, $\Pi^{\phi^{ind}} \equiv \Pi^{ind}$ is well-known to labor economists as the design matrix of industry effects included in an analysis of industry earnings premia.

Homophily and modularity of $R^M$ are measured with respect to a given node partition, $\phi$. One way to build the intuition behind both measures is in terms of the *Partition Graph*. Given the network, $R^M$, and a partition of its nodes, $\phi$, the partition graph is a "supergraph" on "supernodes" that are the $L$ partition classes in $\phi$. There is a connection between partition classes whenever there is a connection between nodes in the underlying graph that belong to those partition classes. Under the industry partition, $\phi^{ind}$, two industries $k$ and $\ell$ are connected in the partition graph whenever a $k$ employer is connected to an $\ell$ employer in $R^M$.

Formally, the partition graph of $G$ with respect to $\phi$, denoted by $P(G, \phi)$, is defined as follows. The vertex set of $P(G, \phi)$ is $V(P(G, \phi)) = \phi$. The edge set is

$$E\left(P(G,\phi)\right) = \left\{ \begin{array}{l} (\phi_k, \phi_\ell, \omega) : \omega = \\ \left| \{(m,n) : m \in \phi_k, n \in \phi_\ell, (m,n) \in E(G)\} \right| \end{array} \right\}. \tag{3}$$

Unlike $R^M$ and $R^W$, the partition graphs may include self-edges.

The adjacency matrix of the partition graph is

$$A_{P(G,\phi)} = \left(\Pi^{\phi}\right)^T A_G \Pi^{\phi}. \tag{4}$$

In the partition graph, the degree of each supernode is

$$\begin{aligned} k_\ell = \sum_{i:\pi(i)=\ell} d_i &= e_\ell \Pi^T A_G u_I \\ &= e_\ell \Pi^T A_G \Pi u_L \\ &= e_\ell A_{P(G,\Pi)} u_L, \end{aligned} \tag{5}$$

where $e_\ell$ is the $\ell^{th}$ column of the $L \times L$ identity matrix and $u_I$ is the $I \times 1$ sum vector (a vector of ones). The total weighted degree sum of $G$ and of $P(G, \Pi)$ is $2m$, where $m$ is the sum of edge weights (in either graph).

*2.3.2. Homophily*

Intuitively, if the graph exhibits positive assortativity – homophily – it means more nodes in the same partition class are connected than would be expected under random matching. In the realized mobility network, homophily measures the extent to which a worker is more likely to be employed in jobs in the same industry than would be predicted if job assignment is random. Operationalizing this intuition requires a definition of random matching.

I adapt the approach in Currarini et al. (2009) to the labor market setting. The homophily of group $\ell$ is the share of edges that connect group members relative to the number of within-group edges that would be expected if agents connect at random. The comparison relative to random matching is important, since some social groups are larger than others, particularly with regards to majority/minority social relationships. In that case, we expect more connections within the majority group than from the majority group to the minority group simply because of their relative sizes. For the labor market application, the random matching benchmark is meaningful as some employers may be more connected simply because they constitute a larger share of total jobs. Under random matching, workers are more likely to move into a job in manufacturing than in mining, simply because of the size of the two sectors.

Let $p_{\ell,\ell'}$ be the number of edges in the partition graph between nodes in class $\ell$ and nodes in class $\ell'$. The homophily index,

$$H_\ell = \frac{p_{\ell\ell}}{k_\ell}, \tag{6}$$

is the share of the total degree of nodes in class $\ell$ accounted for by within-group edges. Next, define the *degree share of edges in class* $\ell$ as

$$w_\ell = \frac{k_\ell}{2m}. \tag{7}$$

If you pick up an edge at random, $w_\ell$ is the probability that one end is connected to a node in partition class $\phi_\ell$.

The goal is to measure, for each partition class, the deviation between the number of internal connections and the number expected under random matching. With this in mind, define the *excess homophily index* for class $\ell$ by

$$EH_\ell = H_\ell - w_\ell. \tag{8}$$

Currarini et al. (2009) define the *inbreeding homophily index* of class $\ell$ by normalizing the excess homophily:

$$IH_\ell = \frac{EH_\ell}{1 - w_\ell}. \tag{9}$$

The latter definition standardizes the excess homophily by the maximum possible level of homophily. In the empirical results, I prefer to present results for homophily using the inbreeding homophily index. The modularity function has a direct interpretation in terms of excess homophily. The results for homophily presented later are not sensitive to whether I define them in terms of $EH_\ell$ or $IH_\ell$.

*2.3.3. Modularity*

Modularity is an alternative measure of assortativity in a network with respect to a given partition of its nodes. Modularity was introduced to the literature on complex network analysis as part of a method for detecting relatively densely connected groups of vertices as a method for community structure detection. In what follows, I develop the modularity measure as an aggregation of the excess homophily index, and then show that this is equivalent to the usual definition of modularity given in the complex networks literature. The reader can therefore interpret the maximum modularity partition as the partition that maximizes aggregate homophily.

Given a graph, $G$, and a partition of its vertex set, $\phi$, the *modularity of $G$ with respect to $\phi$* is

$$\begin{aligned} Q(G,\phi) &= \frac{1}{2m}\sum_{\ell=1}^{L} k_\ell EH_\ell && (10) \\ &= \sum_{\ell=1}^{L} w_\ell EH_\ell. && (11) \end{aligned}$$

This expression is the normalized degree-weighted sum across partition classes of the inbreeding homophily index. Weighting by degree ensures that the overall influence of partitions on the graph structure are reflected in the aggregate measure. The normalization by the degree sum ensures that $Q(G, \phi) \leq 1$, with the extreme case corresponding to perfect homophily in each class.

Substituting from above, modularity can also be expressed as

$$Q(G, \phi) = \frac{1}{2m} \sum_{\ell=1}^{L} \left( p_{\ell\ell} - \frac{k_\ell^2}{2m} \right) \quad (12)$$

$$= \frac{1}{2m} \sum_{i,j} (A_{ij}^G - \frac{k_i k_j}{2m}) \delta(i,j), \quad (13)$$

where $A^G$ is the adjacency matrix of $G$, $i, j$ are nodes in $G$, $k_i$ is the degree of $i$ in $G$, and $\delta(i,j) = 1$ if $i$ and $j$ are in the same partition class and $\delta(i,j) = 0$ otherwise. The final expression is the typical characterization of modularity defined in terms of the full network, sometimes referred to as "Newman-Girvan modularity" in the complex networks literature (Good et al. 2010).

*2.3.4. Modularity Maximization*

To find the partition of employers into classes within which mobility is highly likely, I use the method of modularity maximization. Modularity maximization takes the graph as given, and finds the partition, $\phi \in \Phi$, that maximizes the modularity function. An advantage of this process is that the number of partition classes, $L$, is part of the optimization. A complete search over elements of $\phi$ is computationally intractable for all but the smallest graphs. My results are based on implementation of the 'Louvain' method for modularity maximization (Blondel et al. 2008). The Louvain algorithm uses an optimization heuristic that begins with all nodes in separate partition classes, and then finds combinations of classes that yield the best local improvement in modularity.

The Louvain heuristic is fast, has been implemented in many diverse settings (Good et al. 2010), and has excellent performance in recovering modules in networks with known structure (Blondel et al. 2008). There is a minor concern that the algorithm could be affected by the order in which classes are considered when combining them to make local increases in modularity. To address this, I computed modularity under random re-orderings with no meaningful effect on the results.

## 3. The Realized Mobility Network in the PSID

The following analysis uses a sample of heads and spouses in the Panel Study of Income Dynamics (PSID) between 1987 and 1997. The PSID records the industry and occupation in which an individual's labor is employed. I define the set of *pseudoemployers* in the PSID as the set of possible industry-occupation combinations to which a worker may be matched. The realized mobility network in the PSID is formed as workers change industry and occupation. The analysis therefore reveals the boundaries and barriers to mobility between different industry and occupation groups. As we will see, the analysis helps refine the emerging view in the literature that the nature of job mobility is related to industry and occupational history in complex ways.

### *3.1. Data*

An individual is at risk for inclusion in the analysis as long as he or she was in a family that responded to the survey in both 1987 and 1997. Individuals who never reported a primary industry-occupation pair during the sample are omitted since they do not contribute a (relevant) vertex to the network. I chose heads and spouses because the PSID consistently collects their industry and occupation data. Additional demographic variables incorporated into the analysis include race (white or not), gender, labor income, family income, state of residence, age in 1987, and education. I further restrict attention to those individuals who were over 23 years of age in 1987 and who contributed at least two years of valid industry and occupation data. The age restriction reduces the influence of mobility associated with career-shopping (Neal 1999; Pavan 2011; Yamaguchi 2010).

The number of workers in this subsample is $|W| = 7,515$ and the number of pseudoemployers (unique industry-occupation pairs) is $|M| = 6,944$. The largest connected component (CC) in the realized mobility network contains $7,432$ workers and $6,771$ pseudoemployers, highlighting that boundaries to mobility are very porous. Table 1 shows that the basic characteristics of the two samples of workers are nearly identical. Workers initiate $31,578$ unique job spells with different pseudoemployers for a total of $51,066$ job-year observations in the full panel.

Measurement error is a concern in any study using self-reported industry and occupation in the PSID. In particular, PSID coders interpret survey respondents' descriptions of industry and occupation differently, which may result in spurious transitions. Kambourov and Manovskii (2008) use PSID Retrospective Files, available before 1981, to demonstrate that this form of measurement error increases observed industry and occupational mobility. After 1981, only originally coded data are available. Like Kambourov and Manovskii (2008) (p. 72), I use the originally coded data for analyzing cross-industry and occupation mobility, with the caveat that the overall level of mobility may be too high.

*3.2. Basic Topology*

Table 2 reports the average degree, characteristic parameters of the degree distribution, and the clustering coefficient, along with basic node and edge counts. For comparison, I have also included statistics on two social networks with similar topological properties: the network of film stars and the network of co-authorship relationships in physics (Newman 2003). I have chosen these for comparison because they have similar density and clustering to the worker and pseudoemployer projections. This is not a coincidence, as the network of film stars and network of co-authors are also projection graphs: film stars are connected when they work on the same film, and authors are connected when they collaborate on a paper.

The realized mobility network contains 27,519 edges connecting 7,515 workers to 6,944 pseudoemployers. Each edge represents a unique match. The average degree on the worker nodes is 3.66 and on the pseudoemployer nodes is 3.96. The closeness of these numbers is an artifact of the fact that the sets of nodes have similar cardinality, and that the degree sum on both node sets must be equal due to its bipartite nature. The similarity in the average degree masks major differences in the degree distributions.

The worker projection induces 428,848 edges between the 7,515 workers. The pseudoemployer projection is less densely connected by an order of magnitude, with 44,840 edges representing the number of times a pseudoemployer pair shares a worker in common. The average degrees are 110.66 and 15.35 respectively, and are once again somewhat poor summaries of the actual amount of connectivity in the graph. The clustering coefficients are relatively high, at 0.65 and 0.70. The projection graphs inherit a certain amount of triadic closure, which results in high clustering whenever workers are employed in more than two jobs (for the pseudoemployer projection) and whenever pseudoemployers have more than two workers (for the worker projection). For this reason, the clustering coefficient does not give a clear measure of the extent to which workers move within tightly knit groups of jobs.[2]

### *3.3. Degree Distribution*

Figure 3 presents log-log plots of the cumulative distributions of node degree in the bipartite realized mobility network and the one-mode projections. The figure shows the empirical CDF and parametric fits to a power law, log normal, and exponential distribution.[3] The distribution of pseudoemployer degree is very fat-tailed in the realized mobility network, as well as in the pseudoemployer projection. The former is well-fit by a power law distribution, while the best fit for the latter is a power law with exponential cutoff. The worker degree is well-fit by an exponential distribution.[4]

---

[2]In the social networking context, clustering in a graph captures the idea that 'friends of my friends are also my friends'. More formally, graph clustering measures the transitivity of network relationships. The *clustering coefficient* is

$$C = \frac{1}{n} \sum_i C_i,$$

where

$$C_i = \frac{\text{\# of triangles connected to } i}{\text{\# of connected triples centered on } i}.$$

A 'connected triple' is a subgraph of three nodes, $\{j, i, k\}$, in which $j$ and $k$ are both connected to $i$, which is called the 'center' of the triple.

[3]I estimate the power law fit by maximum likelihood using the method of Clauset et al. (2009). Code and details of the estimation are available upon request.

[4]Likelihood ratio tests reject the log-normal and power law in favor of the exponential for the degree distribution of worker nodes. The evidence is less clear for the exact nature of the pseudoemployer degree distributions. For the realized mobility network, a bootstrap test fails to reject the null hypothesis that the data follow a power law, but the data do not discriminate between the power law and log-normal on the

The distribution of node degree has a strong connection to the process of network formation, particularly with respect to whether nodes connect at random or preferentially (Newman 2003). Consequently, the degree distribution in the realized mobility network should be informative about the process of labor market matching. In the context of the realized mobility network, node degree is a measure of mobility for workers and a measure of total employment for employers. The (near) power-law distribution evokes the heavy-tailed distribution of employment across firms (Axtell 2001). When workers move, they are drawn to pseudoemployers with high employment, high turnover, or both. The exponential distribution of worker degree indicates that mover-stayer heterogeneity is randomly distributed in the population of workers.

In the one-mode projections, the evidence suggests independence between worker mobility and employer size. In the worker projection, a large degree implies working in high employment sector. Fat tails would arise if some workers are always in high employment sectors and other workers are always in low employment sectors. The exponential distribution of degree in the worker projection suggests a random growth model in which the number of jobs held by a worker over his life is a Poisson random variable, but the employers with which these jobs are held are chosen by preferential attachment. This is also consistent with dynamic equilibrium matching models where workers sample employers in proportion to their employment level (Postel-Vinay and Robin 2002).

## 4. Assortativity in the Realized Mobility Network

I characterize assortativity in the realized mobility network – that is, who matches with whom – by a three step process. First, I apply modularity maximization to the weighted pseudoemployer projection of the realized mobility network. Second, I apply modularity maximization to the weighted worker projection of the realized mobility network. Third, I merge the partition classes for pseudoemployers and for workers onto the bipartite realized

basis of a likelihood ratio test. In the pseudoemployer projection, however, the bootstrap test rejects the power-law null. A better fit is obtained by a power law distribution with exponential cutoff.

mobility network. Guimerà et al. (2007) provide evidence that the three stage approach is equivalent to directly partitioning the bipartite graph.

This process results in a clear division of the labor market into separate sub-markets that attract specific groups of workers. There are three key results. The first is that the sub-markets revealed by mobility patterns are much stronger than would be predicted on the basis of grouping by occupation, by industry, or by industry-occupation pairs. Second, while there are some differences across the classes of workers that serve different market segments, there is very little homophily on the basis of race, gender, and education. Third, the revealed matching sets suggest that industry, occupation, demography and skill are relevant, but in ways that would be difficult to predict ahead of time. Taken together, these findings demonstrate the value of an inductive analysis to characterize the complex nature of labor market matching.

### *4.1. Assortativity in the Pseudoemployer Projection*

#### *4.1.1. The Maximum Modularity Partition*

Applying the Louvain algorithm of Blondel et al. (2008) to the pseudoemployer projection yields a modularity maximizing partition, $\phi^{M**}$, into 79 classes. The value of modularity at the optimizing partition is $Q(R^M, \phi^{M**}) = 0.516$[5]. The classes in $\phi_M^{**}$ are sorted by cardinality, and the first four classes account for more than 80 percent of pseudoemployer nodes and 87 percent of all matches. After the fourth partition class, the remaining classes each contain fewer than five percent of nodes each. To simplify exposition, the remaining discussion is based on a partition, $\phi^{M*}$, that collapses the small partitions of $\phi_M^{**}$ into a single class, labeled below as 'Pseudoemployer Class 5'. This simplification has no effect on the results: the modularity of the collapsed maximum modularity partition is $Q(R^M, \phi^{M*}) = 0.510$.

Figure 4 presents the sparsity pattern of the adjacency matrix of the pseudoemployer projection. The contrast between Figures 4a and 4b provides a succinct and stark exposi-

[5]Recall that at the random matching baseline, modularity is zero, and at the complete segmentation extreme, modularity is equal to 1. Values of modularity above 0.3 generally indicate substantial community structure (Blondel et al. 2008)

tion of the value of finding labor sub-markets on the basis of mobility patterns relative to looking at connectivity between industries. In Figure 4a, the rows and columns of the adjacency matrix, which each correspond to a unique pseudoemployer, are sorted by industry classification. Figure 4b shows the same matrix, but with the rows and columns sorted by modularity class. There are strong connections within industry, but the distinct blocking in the bottom panel shows that the maximum modularity classification finds much more tightly connected groups of employers.

Panel A of Table 3 presents the same information in a less dramatic, but more comprehensive form. The table reports the share of edges that end in each modularity class conditional on where they start. The entries sum to 100 down the columns. The blocking structure that appears in Figure 4b appears in the diagonal elements of the table. Between 50 and 60 percent of all edges are between nodes in the same class. Furthermore, outside of those classes, the remaining edges are distributed fairly uniformly. There is some evidence of relatively strong ties between Classes 2 and 4, a point we will return to later when discussing who matches with whom in the realized mobility network. This will be easier after I describe the industries and occupations that characterize the different classes.

#### *4.1.2. Characteristics of the Modularity Classes*

Table 4 reports the industrial and occupational composition of each class of $\phi^{M*}$. The table entries are the percent of pseudoemployers within a modularity class in a given industry (Panel A) or occupation (Panel B). The entries sum to 100 percent across the rows. For reference, the row labeled 'Pop. Share' lists the percent of pseudoemployers in each major industry or occupation in the overall population.

The most salient feature of the maximum modularity partition is the contrast between Class 1 and Class 2. Class 1 is concentrated in the Manufacturing, Construction, Transportation, and Retail industries, and in jobs as Craftsmen, Laborers, and Operatives. Class 2 is concentrated in the Professional Service, Public Administration, and FIRE industries, and overwhelmingly in Clerical and Professional Service occupations. Class 3 has an industrial composition like Class 1, but in more Professional Service, Sales, Clerical, and Managerial occupations. Class 4 is concentrated in Professional Services, Retail, and Personal Service sectors, but in Service Work occupations. Class 5, which is something of a residual class, is also concentrated in Manufacturing sector in jobs as Operatives, but also in Retail jobs and as Clerical workers.

To summarize, the classes might be labeled as:

- Class 1: "Blue Collar" jobs;
- Class 2: Clerical Service jobs;
- Class 3: "White Collar" jobs;
- Class 4: Less-skilled Service jobs;
- Class 5: Other less-skilled Service and Manufacturing jobs,

acknowledging that these are simplifications of the true industrial and occupational composition. The job segments revealed in $\phi^{M*}$ cut across both industry and occupational lines in ways that are intuitive, but not obvious *a priori*.

These descriptions make sense of the connectivity/mobility patterns in Table 3. The connectivity is weakest between Class 1 and Class 2 and between Class 3 and Class 4. Relatively few Blue Collar workers are Clerical Workers at a different point in time. Also, few 'White Collar' workers are in 'Unskilled Service' at some other point in time. These connectivity patterns are displayed visually in Figure 1, which plots Classes 1-4 of $R^M$. The nodes in each class have different colors, and the figure is laid out using an algorithm that heuristically minimizes the spatial distance between connected nodes.[6]

*4.1.3. Homophily by Industry and Occupation*

In this section, I evaluate how informative the maximum modularity approach is relative to an analysis based on industry and occupation classification. I compare the maximum

[6]Figure 1 was laid out using the implementation of the `OpenOrd` algorithm and visualization tools in the software package `Gephi`.

modularity partition ($\phi^{M*}$) with partitions based on major industry ($\phi^{ind}$), major occupation ($\phi^{occ}$), and the interaction of major industry and major occupation ($\phi^{ind\times occ}$).

Table 5 reports the modularity of the pseudoemployer projection under different partitions. Recall that $Q(R^M, \phi^{M*}) = 0.510$. By contrast, the modularity of the major industry and occupation partitions are relatively small, at $Q(R^M, \phi^{ind}) = 0.351$ and $Q(R^M, \phi^{occ}) = 0.306$ respectively. Increasing the resolution of the partition classes does not help: $Q(R^M, \phi^{ind\times occ}) = 0.223$.

From Section 2.3.3, the modularity of a partition is a weighted sum of the excess homophily index for the associated partition classes. Homophily is an arguably more direct measure of the extent to which workers who hold a job in a given industry are likely to hold another job in the same industry. It is also possible that through aggregation, the modularity function masks very large homophily in certain sectors.

Figure 5 plots the inbreeding homophily index ($IH_\ell$) against the degree share ($w_\ell$) for each class of each partition. In the figure, the triangles represent the maximum modularity partition. At every scale, the homophily in the maximum modularity partition classes is higher than the homophily associated with industry or occupation.

Recall from Section 2.3.3 that the homophily index, $H_\ell$, is the diagonal element of the partition graph, which generalizes the transition matrix. Figure 5 therefore summarizes the information that could be derived by comparing matrices that report the frequency of industry-industry or occupation-occupation transitions with the frequency of transitions across classes of the maximum modularity partition. Deriving the additional information in the maximum modularity partition would be impossible in an analysis based on industry and occupation partitions. To do so would require using more detailed industry-occupation codes, but there is not enough data to estimate the relevant probabilities.

### *4.2. Assortativity in the Worker Projection*

#### *4.2.1. The Maximum Modularity Partition*

In this section, I conduct the equivalent analysis of community structure in the worker projection of the realized mobility network. Two workers are connected whenever they have a common pseudoemployer. The modularity-maximizing partition, $\phi^{W**}$, has 86 separate classes with value $Q(R^W, \phi^{W**}) = 0.580$. The four largest partition classes in $\phi^{W*}$ account for 78 percent of workers. Again for simplicity, I work with a partition, $\phi^{W*}$, that collapses the remaining classes which each individually account for less than 5 percent of all workers, into a single class. The modularity is barely affected: $Q(R^W, \phi^{W**}) = 0.551$. I label the partition classes of the worker nodes as Class A, B, C, D, and E, with E being the residual class.

Table 6 summarizes the maximum modularity partition, after collapsing the small classes. Panel A reports the share of edges that end in each modularity class conditional on where they start. The entries sum to 100 down the columns. The community structure is even more pronounced here than in the pseudoemployer projection. Between 49 and 70 percent of all edges connect nodes in the same class.

*4.2.2. Characteristics of the Modularity Classes*

Table 7 reports the demographic and educational composition of each class in the maximum modularity partition, $\phi^{W*}$. To summarize, the modularity classes might be labeled as:

- Class A: Less-educated male workers;
- Class B: Average education female workers;
- Class C: Less-educated non-white female workers;
- Class D: More-educated white male workers;
- Class E: More-educated white female workers.

While the classes of the maximum modularity partition have relatively distinct demographic profiles, demography seems to play a very small role in labor market matching, at least relative to the amount of structure revealed in the maximum modularity partition.

### *4.2.3. Homophily by Individual Characteristics*

Table 7 reports the inbreeding homophily index for each class in $\phi^{W*}$ (last column) and for each demographic partition. Within each modularity class, the inbreeding homophily index ranges between 0.649 (for Class D) and 0.779 (for Class A). Even the residual class is very strongly connected: in Class E, the inbreeding homophily index is 0.751.

By contrast, the inbreeding homophily for white (and non-white) workers is 0.127.[7] For the education groups in $\phi^{educ}$, inbreeding homophily ranges from 0.051 and 0.228.[8] Male (and female) workers have a higher inbreeding homophily index of 0.420. These patterns are reflected in the modularity of the worker projection with respect to each demographic partition. The modularity is $Q(R^W, \phi^{white}) = 0.062$, $Q(R^W, \phi^{sex}) = 0.204$, and $Q(R^W, \phi^{educ}) = 0.098$. Therefore, while there is evidence of labor sub-markets that match workers assortatively to particular types of jobs, the attributes on which workers match skills are determined by largely unobservable factors.

### *4.3. Matching in the Realized Mobility Network*

I now characterize who matches with whom. Figure 6 illustrates labor market matching using the sparsity pattern in the adjacency matrix of the bipartite realized mobility network. Its rows are associated with workers and the columns are associated with pseudoemployers. A 1 in the $i, j$ position means worker $i$ is matched with pseudoemployer $j$. Figure 6a shows the adjacency matrix with workers and pseudoemployers sorted randomly, which has no visibly discernible structure. Figure 6b shows the same matrix with rows and columns sorted into maximum modularity partition classes. The figure shows first, that the classification of nodes obtained in the worker and pseudoemployer projections are also useful in characterizing structure in the bipartite graph from which they were derived, consistent with Guimerà et al. (2007). Second, there is a close connection between the worker classes and the pseudoemployer classes. Workers in a given modularity class tend to have jobs with

[7]For binary characteristics, the inbreeding homophily indices are equal for each type.

[8]Incidentally, across education groups, inbreeding homophily exhibits a U-shape – it is strongest for workers with the least and the most education.

pseudoemployers in a single modularity class. This allows me to give a relatively succinct summary of who matches with whom.

Table 8 provides a more complete account of the sub-markets in the realized mobility network. It reports the percent of job matches (edges) in the realized mobility network that connect workers in a given worker class to jobs in a particular pseudoemployer class. Using the characterizations in terms of industry and occupation from Section 4.1.2, and in terms of demographic characteristics from Section 4.2.2, the maximum modularity partition reveals:

- less-educated male workers are matched to "Blue Collar" jobs (16.56 percent of matches);
- female workers are evenly divided between "Clerical Service" and "Unskilled Service" jobs (17.95 percent);
- less-educated non-white female workers are matched to "Unskilled Service" jobs (9.00 percent);
- more-educated white male workers are matched to "White Collar" jobs (11.48 percent);
- more-educated white female workers are matched to "Clerical Service" jobs (10.45 percent).

These matches account for 65.44 percent of all jobs. Under random matching, the same combinations would contain 28.27 percent of all jobs. I emphasize that these characterizations are provided to facilitate intuition. My descriptive labels indicate concentrations of activity.

## 5. Implications for Labor Market and Policy Analysis

The value of using network methods to describe labor market structure has to come from enhanced analysis of labor market models and associated policy prescriptions. Here, I consider several topics that can benefit from the approach I have developed.

### *5.1. Human Capital Specificity*

The measurement of labor market mobility has traditionally focused on existing partitions of the data (by industry and occupation, primarily). My analysis confirms that there are better ways to partition workers and jobs that can reflect the underlying process of labor market mobility and matching. The comparison of homophily and modularity measures in Section 4 makes this argument precise. My results are therefore complementary to Neal (1995), Yamaguchi (2010), and Sullivan (2010), all of whom suggest that human capital specificity may depend on industry- and occupation-specific combinations that are difficult to measure using conventional methods, usually due to sample size.

Recently, the labor economics literature has moved toward cardinal measures of task-specific human capital (Lazear 2009; Gathmann and Schoenberg 2010; Yamaguchi 2012). The maximum modularity partition of pseudoemployers strongly echoes the classification of human capital across routine/non-routine and manual/analytic axes (Autor et al. 2003), but was detected on the basis of matching patterns alone, and required no assumptions about the nature or weighting of task content. The network analysis therefore confirms recent developments in the empirical analysis of human capital specificity.

*5.2. Unemployment and Mismatch*

Limitations on mobility also affect and are affected by the business cycle – a point that has been emphasized in the literature on mismatch following the recent recession (Şahin et al. 2012). Figure 7 documents how limitations on mobility, as measured by modularity, change over the period of my sample. Modularity is counter-cyclical. Figure 7 plots the four-year modularity from 1988 through 1995, along with the annual unemployment rate. Modularity increases by about 5 percent ahead of the 1990-1991 recession, then falls with the recovery. Because the modularity calculation directly controls for the amount of observed mobility, this effect is not driven by changes in the volume of mobility over the sample. It reflects changes in the extent to which workers move within and between the maximum modularity partition classes. The implication is that the boundaries to mobility are less effective during expansions than during contractions. This is consistent with cyclical upgrading of labor (McLaughlin and Bils 2001) and the counter-cyclical occupational mobility documented by Kambourov and Manovskii (2008).

The method of modularity maximization treats the realized mobility network as static, but the evolution of the labor market is, of course, a dynamic process. To produce the figure, I construct a moving four-year window on realized employment networks. Within each window, I compute the modularity of the observed pseudoemployer transition graph relative to the predicted match probabilities estimated over the entire period, using the maximum modularity partition of the pseudoemployers introduced in Section 4.1.1. The modularity estimated in each four-year window can be interpreted as a measure of the contribution of that four-year window to the overall modularity.

### *5.3. Labor Market Segments and Labor Market Earnings*

The labor market may also allocate workers to employers on the basis of general characteristics that affect productivity. The nature of that matching process has implications for individual earnings, (Gibbons et al. 2005), for aggregate inequality (Sattinger 2006; Card et al. 2013), and for labor market efficiency (Shimer 2005). In this section, I characterize mobility segments in terms of earnings heterogeneity and use this heterogeneity to describe who matches with whom. In doing so, I show that network methods can help produce information that can be used in estimation, identification, and evaluation of models of labor market matching and directed search.

I decompose labor market earnings in the PSID panel in a linear model that controls for observable individual heterogeneity, as well as for the modularity class to which a worker belongs and the modularity class in which they are employed. The analysis is structured with explicit reference to the decomposition of earnings in matched employer-employee data popularized by Abowd et al. (1999). I estimate the model

$$\ln y_{it} = \alpha + \mathbf{X_{it}}\beta + \theta_{\mathbf{i}} + \psi_{\mathbf{it}} + \varepsilon_{\mathbf{it}}, \tag{14}$$

where $\ln y_{it}$ is the log wage and $\mathbf{X}$ is a vector of observable demographic and human capital characteristics, including education, dummies for race and gender, and a gender-specific quadratic in age. The error components, $\theta$ and $\psi$, are associated with worker and firm-specific heterogeneity in the analysis of employer-employee data. Here, $\theta_i$ is the estimated coefficient on a dummy that indicates the modularity class to which a worker belongs (Worker Class

A, B, C, D, or E). $\psi_{i,t}$ is the estimated coefficient on a dummy that indicates the modularity class to which the pseudoemployer of worker $i$'s period $t$ job belongs (Pseudoemployer Class 1, 2, 3, 4, or 5).

Table 9 reports the results of estimating Equation 14 with and without controlling for modularity class heterogeneity. The key results are the coefficients on the worker and pseudoemployer modularity classes. A worker in the highest-earning modularity class earns roughly 25 percent more than a worker in the lowest modularity class, controlling for observed characteristics as well as for the type of pseudoemployer they are working for. Likewise, a worker employed in the highest paying pseudoemployer class earns 50 percent more than a worker employed in the lowest-paying class.[9] These estimates are not causal effects, as I use the earnings regression to summarize conditional expectations.

The earnings analysis shows that the maximum modularity partitions, revealed solely on the basis of mobility patterns, are strongly associated with earnings heterogeneity. This finding is consistent with these mobility patterns being driven by some latent factors that are differentially productive, and therefore differentially rewarded. One concern is that the classification is picking up heterogeneity that is usually controlled by adding industry or occupation effects. Column (3) in Table 9 shows the effect of adding controls for major industry and occupation categories to the specification in Column (2). This does attenuate the estimates of worker and pseudoemployer class heterogeneity but the overall pattern of effects remains the same, as does the estimated assortativity in matched worker and pseudoemployer types that I discuss next.

*5.3.1. Assortative Matching*

Next, to characterize who matches with whom, Table 10 reports the correlation of the components of earnings in the full sample (Panel A) and across matches (Panel B). To build the table, I assign to each observation in the PSID panel the value $X\hat{\beta}$, which is the predicted

[9] The careful reader will note that the worker and pseudoemployer modularity classes are mutually exhaustive. That is, they each produce a set of five dummy variables that will not be linearly independent of the constant term. For identification, rather than drop one class, which is the conventional approach, I restrict each effect to have mean zero across the population of matches.

contribution to earnings of observable individual characteristics. I also assign the value $\hat{\theta}$ of the modularity class to which the worker belongs, and the value $\hat{\psi}$ of the pseudoemployer to which the worker is matched.[10]

Focusing on the full sample reported in Panel A, note that the correlation between $\hat{\theta}$ and $\hat{\psi}$ is positive. There is a positive correlation between the index of observed characteristics, $X\hat{\beta}$, and both $\hat{\theta}$ and $\hat{\psi}$ as well. These three facts further characterize who matches with whom in the labor market. Workers with highly-paid observed traits are also highly-paid on the basis of unobservables that affect labor market sorting. Furthermore, higher-paying pseudoemployers tend to employ workers with highly-paid observable traits. Finally, higher-paying pseudoemployers tend to match with workers that are highly-paid on the basis of unobservables.

These observed correlations are large relative to estimates of the corresponding quantities estimated in matched employer-employee data for the U.S. (Abowd et al. 2002; 2012). A common finding in matched data is that the correlation between worker and firm heterogeneity is small and sometimes negative. However, my results are consistent with the presence of assortative matching on the basis of unobserved worker and employer-specific heterogeneity.

A common criticism of models like Equation 14 is that job mobility and assignment are endogenous (Gruetter and Lalive 2009; Gibbons et al. 2005). In a working paper, Abowd and Schmutte (2013) use a latent factor model to find groups of workers and employers that are most likely to be connected, conditional on separation. The analysis suggests an alternative approach to using non-wage information on mobility patterns to obtain additional information about latent worker and employer types that can be exploited to correct for endogenous assignment in the analysis of earnings and labor market matching.

[10] In the full sample, the mean of the variance components $\hat{\theta}$, $\hat{\psi}$ and the residual are equal to zero by assumption for identification purposed. This need not be the case in the population of matches.

## 6. Conclusion

Labor economic research runs up against the complexity inherent in individual mobility and firm employment patterns. I have shown that tools of network analysis can provide some simple descriptive statistics that characterize market complexity and that are also useful in actual applications. Using mobility patterns to identify labor market segments yields information that is not available when considering industry and occupation classifications, either together or in isolation. The data reveal a set of labor market boundaries associated with specific industry/occupation combinations that are intuitive, and bear relation to task-specific measures recently developed in the literature.

I also provide evidence of positive assortative matching on earnings heterogeneity and show that the barriers to job mobility are counter-cyclical. These findings demonstrate that network statistics are a useful source of information for cross-sectional and dynamic sorting. Combining models of network formation with dynamic equilibrium labor market models to fit these network statistics may be a productive way to integrate the rigor of formal structural models with the complex and unavoidable patterns in real-world labor market data. Research on this topic is currently underway.

### Acknowledgements

Antoine Scherrer graciously provided the MATLAB implementation of the Louvain method for community detection (http://perso.uclouvain.be/vincent.blondel/research/louvain.html). I also thank Aaron Clauset and Cosma Shalizi for making available their MATLAB implementation of the maximum-likelihood method for fitting the power-law distribution (http://tuvalu.santafe.edu/~aaronc/powerlaws/).

Abowd, J. M., Creecy, R. H., Kramarz, F., 2002. Computing person and firm effects using linked longitudinal employer-employee data. Tech. Rep. TP-2002-06, LEHD, U.S. Census Bureau.

Abowd, J. M., Kramarz, F., Margolis, D. N., Mar. 1999. High wage workers and high wage firms. Econometrica 67 (2), 251–333.

Abowd, J. M., Kramarz, F., Pérez-Duarte, S., Schmutte, I. M., March 2012. A formal test of assortative matching in the labor market, NBER Working Paper No. 15546 (revised).

Abowd, J. M., Schmutte, I. M., 2013. Endogenous mobility. Cornell University School of Industrial and Labor Relations.

Autor, D., Dorn, D., 2013. The growth of low skill service jobs and the polarization of the U.S. labor market. American Economic Review 103 (5), 1553–1597.

Autor, D., Levy, F., Murnane, R. J., 2003. The skill content of recent technological change: An empirical exploration. The Quarterly Journal of Economics 118 (4), 1279–1333.

Axtell, R. L., 2001. Zipf distribution of US firm sizes. Science 293 (5536), 1818–1820.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008.

Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of West German wage inequality. The Quarterly Journal of Economics.

Clauset, A., Shalizi, C. R., Newman, M. E., 2009. Power-law distributions in empirical data. SIAM review 51 (4), 661–703.

Currarini, S., Jackson, M. O., Pin, P., 2009. An economic model of friendship: Homophily, minorities, and segregation. Econometrica 77 (4), 1003–1045.

Gathmann, C., Schoenberg, U., 2010. How general is human capital? A task-based approach. Journal of Labor Economics 28 (1), 1–49.

Gibbons, R., Katz, L. F., Lemieux, T., Parent, D., 2005. Comparative advantage, learning and sectoral wage determination. Journal of Labor Economics 23, 681–723.

Girvan, M., Newman, M., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99, 7821–7826.

Good, B. H., de Montjoye, Y.-A., Clauset, A., 2010. Performance of modularity maximization in practical contexts. Physical Review E 81 (4), 046106.

Gruetter, M., Lalive, R., 2009. The importance of firms in wage determination. Labour Economics 16, 149–160.

Guimerà, R., Sales-Pardo, M., Amaral, L. A. N., Sep 2007. Module identification in bipartite and directed networks. Phys. Rev. E 76, 036102.

Kambourov, G., Manovskii, I., 2008. Rising occupational and industrial mobility in the United States: 1968-1997. International Economic Review 49 (1), 41–79.

Lazear, E. P., 2009. Firm-specific human capital: A skill-weights approach. Journal of Political Economy 117 (5), pp. 914–940.

McLaughlin, K. J., Bils, M., 2001. Interindustry mobility and the cyclical upgrading of labor. Journal of Labor Economics 19 (1), 94–135.

McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. Annual Review of Sociology 27 (1), 415–444.

Neal, D., 1995. Industry-specific human capital: Evidence from displaced workers. Journal of Labor Economics 13 (4), 653–77.

Neal, D., Apr. 1999. The complexity of job mobility among young men. Journal of Labor Economics 17 (2), 237–261.

Newman, M., 2003. The structure and function of complex networks. SIAM Review 45 (2), 167–256.

Parrado, E., Caner, A., Wolff, E. N., 2007. Occupational and industrial mobility in the united states. Labour Economics 14, 435–455.

Pavan, R., 2011. Career choice and wage growth. Journal of Labor Economics 29 (3), pp. 549–587.

Postel-Vinay, F., Robin, J.-M., 2002. Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica 70 (6), pp. 2295–2350.

Şahin, A., Song, J., Topa, G., Violante, G. L., Aug. 2012. Mismatch unemployment. NBER Working Papers 18265, National Bureau of Economic Research, Inc.

Sattinger, M., 2006. Overlapping labour markets. Labour Economics 13 (2), 237–257.

Shimer, R., 2005. The assignment of workers to jobs in an economy with coordination frictions. Journal of Political Economy 113, 996.

Sullivan, P., 2010. Empirical evidence on occupation and industry specific human capital. Labour Economics 17 (3), 567 – 580.

Taubman, P., Wachter, M. L., 1986. Segmented Labor Markets. Elsevier Science Publishers, Ch. 21, pp. 1183–1217.

Yamaguchi, S., 2010. The effect of match quality and specific experience on career decisions and wage growth. Labour Economics 17 (2), 407 – 423.

Yamaguchi, S., 2012. Tasks and heterogeneous human capital. Journal of Labor Economics 30 (1), 1–53.

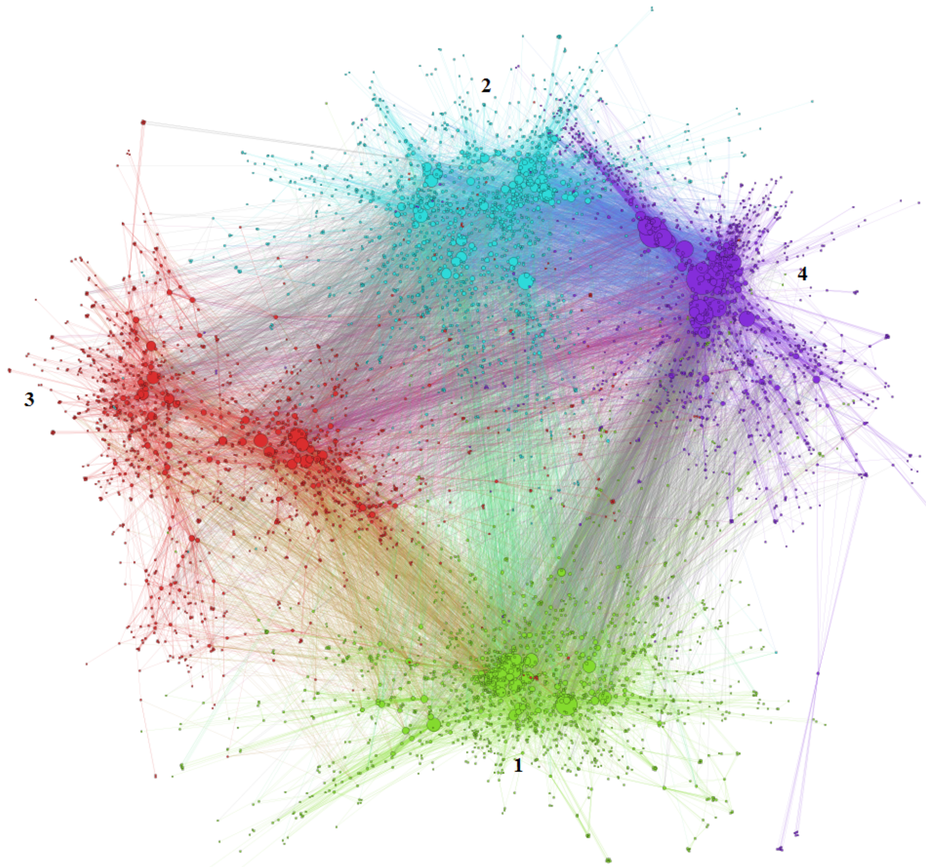



Figure 1: A detail from the projection onto pseudoemployer nodes of the realized mobility network in the PSID. The figure displays the four largest classes from the maximum modularity partition, which are labeled in bold.

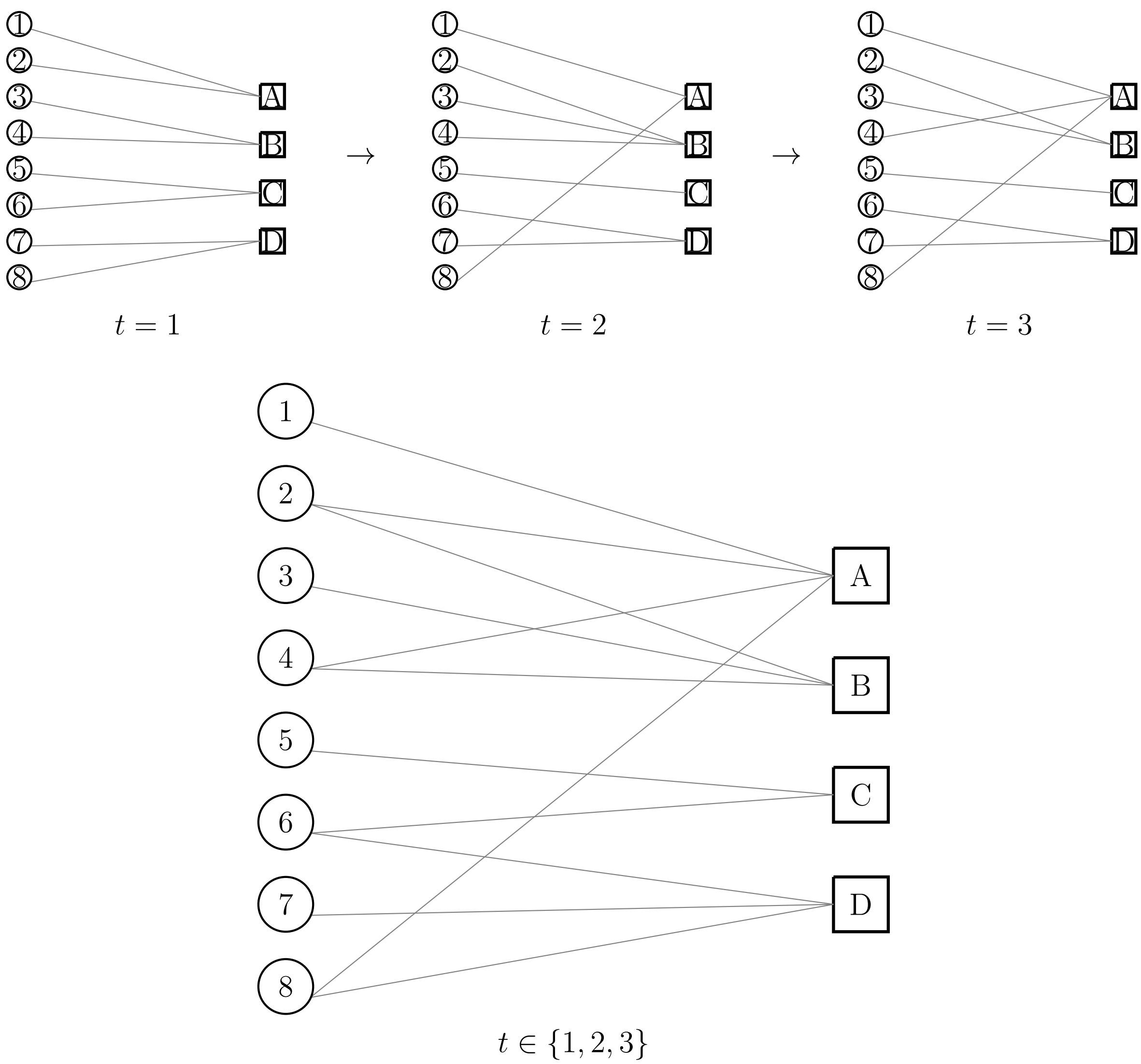
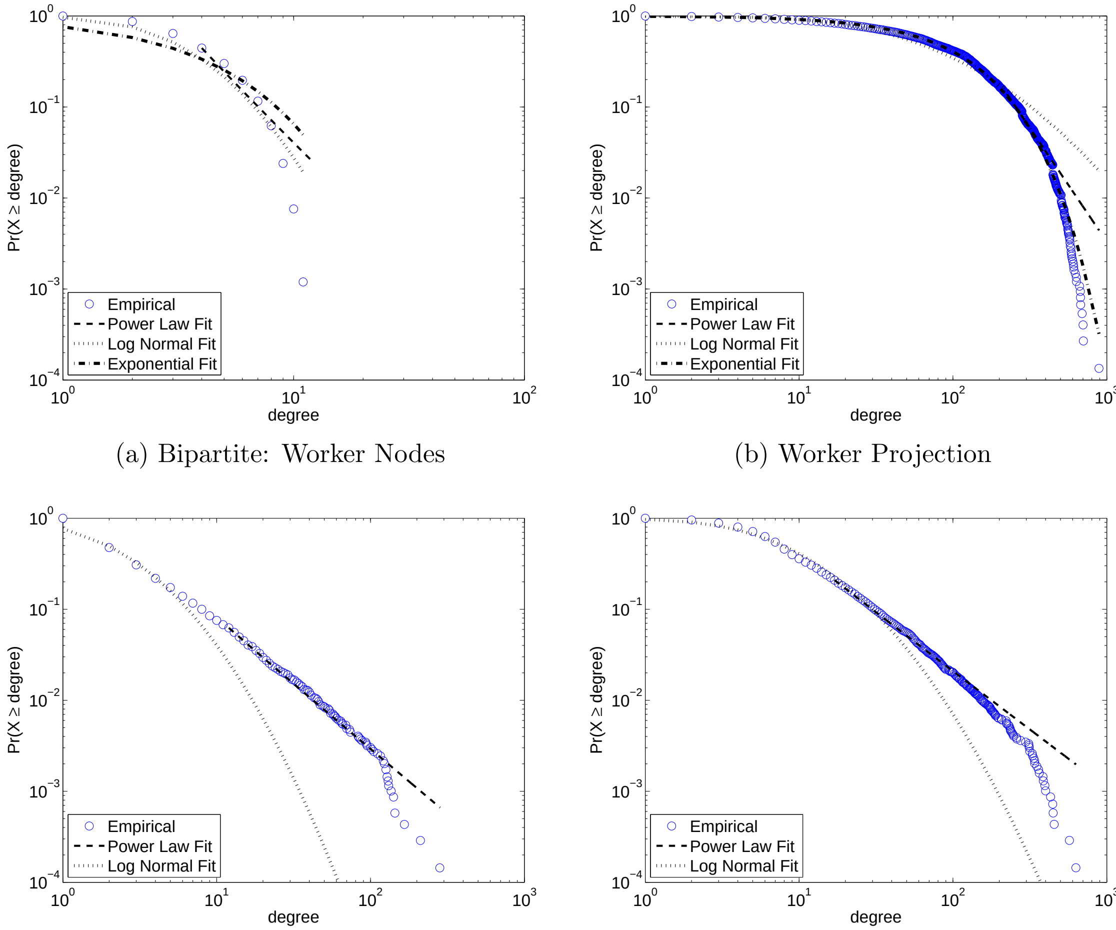



Figure 2: Evolution of the Realized Mobility Network from a Sequence of Realized Employment Networks



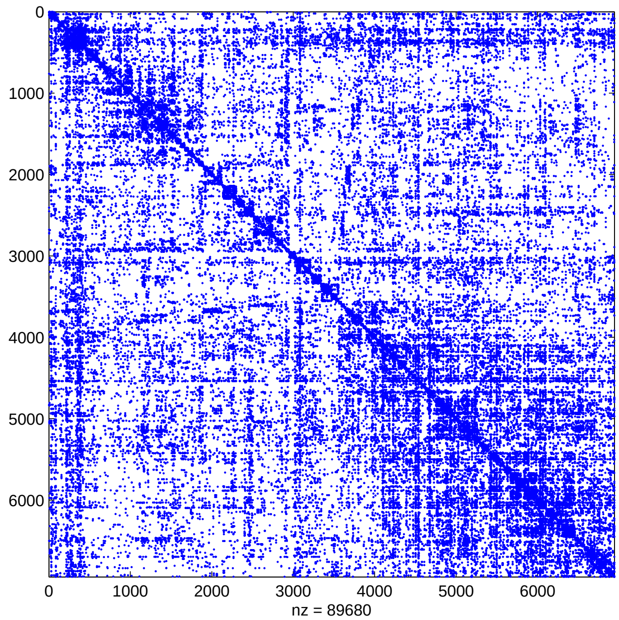

(a) Bipartite: Worker Nodes

(b) Worker Projection

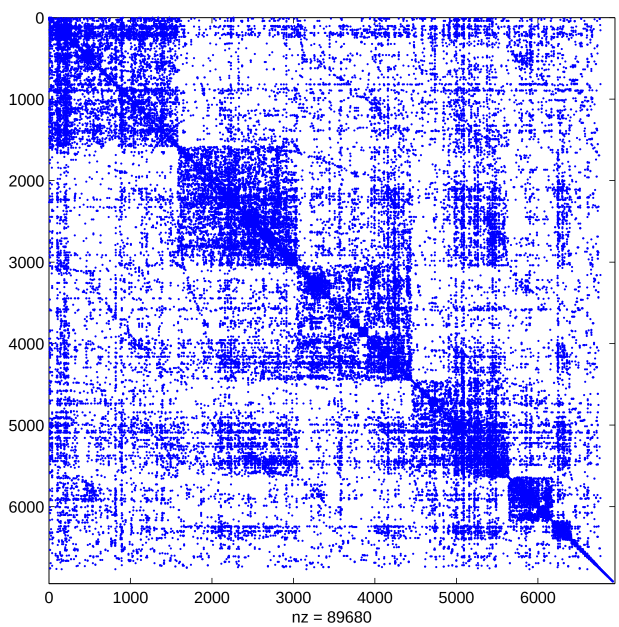
(c) Bipartite: Employer Nodes

(d) Employer Projection

Figure 3: Degree Distributions in the Realized Mobility Network




(a) Sorted by Industry Code




(b) Sorted by Modularity Class

Figure 4: Adjacency Matrix for Job Projection

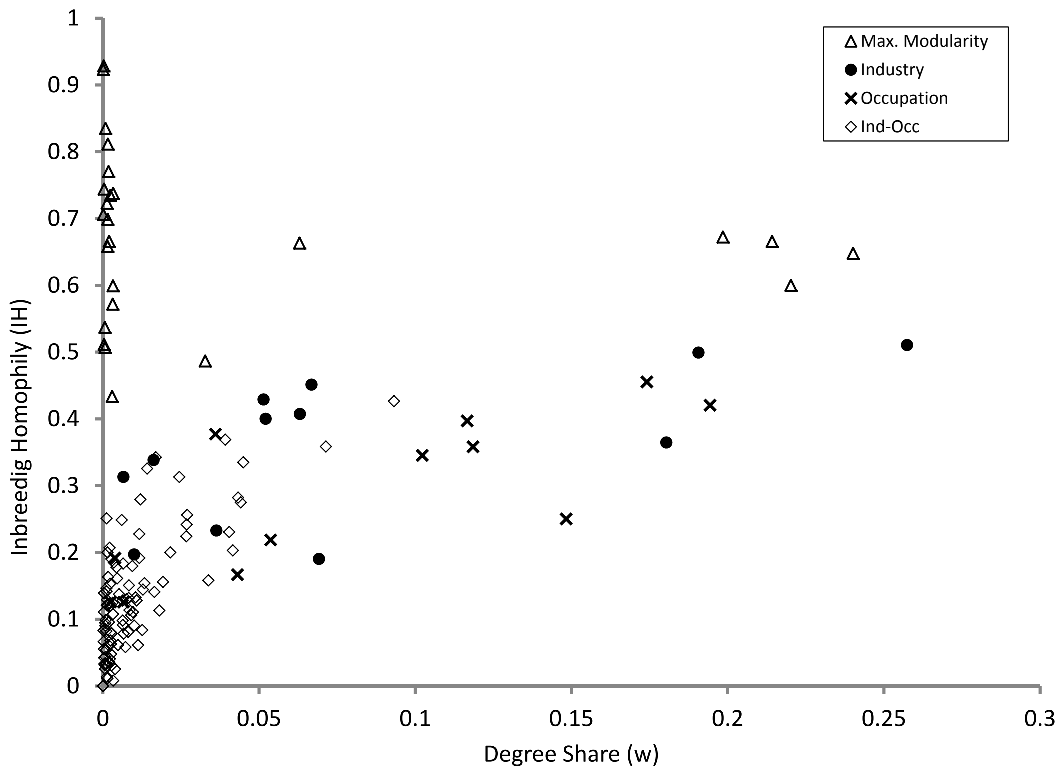
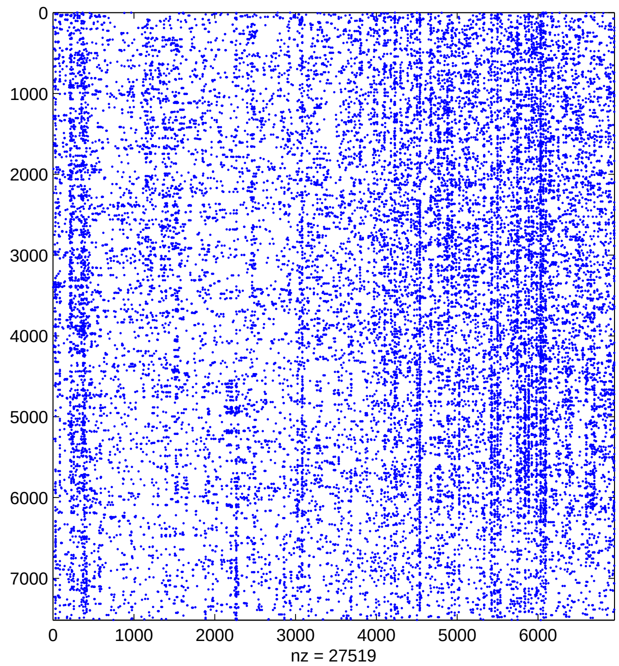
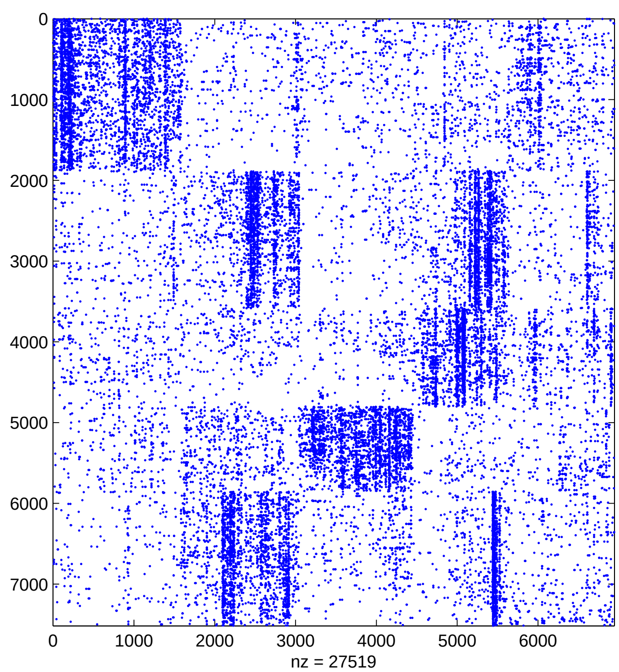



Figure 5: Homophily of Pseudoemployer Nodes By Industry, Occupation, and Maximum Modularity Classification




(a) Random Sorting




(b) Sorted by Modularity Class

Figure 6: Adjacency Matrix for the Realized Mobility Network

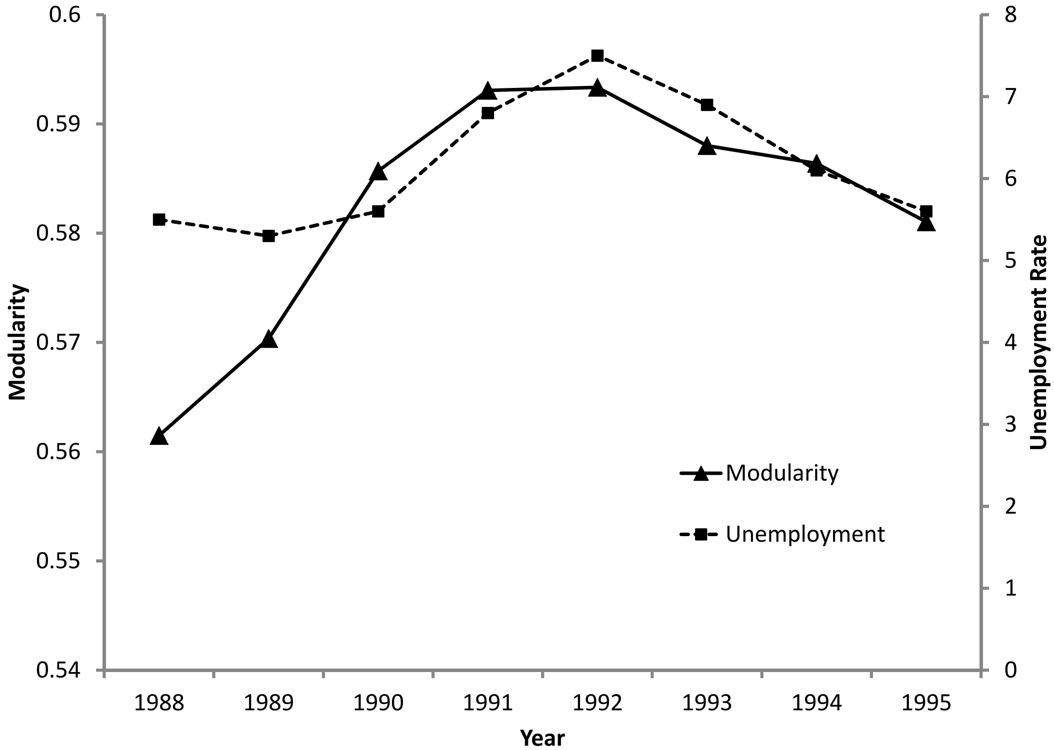



Figure 7: Time Series Variation in Modularity of the Employer Projection

Table 1: Data Summary

| | PSID Sample | CC* |
|---|---|---|
| White | 0.69 | 0.69 |
| Male | 0.51 | 0.51 |
| Number of jobs | 4.5 | 4.5 |
| Average labor earnings† | $19,900 | $19,860 |
| Age | 35.43 | 35.40 |
| Less than HS | 0.21 | 0.21 |
| High School | 0.32 | 0.32 |
| Some College | 0.23 | 0.23 |
| College | 0.14 | 0.14 |
| Postgraduate | 0.09 | 0.10 |
| Num. Workers | 7,515 | 7,432 |
| Num. Employers | 6,944 | 6,771 |

All entries are mean values in the given sample except where noted.
* Workers in the largest connected component of the realized mobility network.
† Median of average within-sample labor earnings.

Table 2: Basic Topology of the Realized Mobility Network in the PSID

| Network | | Nodes | Edges | Avg. Degree | Degree Dist. | Clustering |
|---|---|---|---|---|---|---|
| Bipartite | Worker | 7,515 | 27,519 | 3.66 | Exp. (3.66) | n/a |
| | Job | 6,944 | 27,519 | 3.96 | Power (2.42) | n/a |
| Projections | Worker | 7,515 | 428,848 | 110.66 | Exp.(110.66) | 0.65 |
| | Job | 6,944 | 44,840 | 15.35 | Power (2.28) | 0.70 |
| Comparisons | Actors† | 449,913 | 25,516,482 | 113.43 | Power(2.3) | 0.78 |
| | Physics | 52,909 | 245,300 | 9.27 | n/a | 0.56 |
| | Coauthorship† | | | | | |

NOTE – Basic topological properties of the realized mobility network in the PSID and its projections onto worker and pseudoemployer nodes. In the projection graphs, the average degree is weighted by the edge weights. For each network, the column 'Degree Dist.' reports the type of distribution with best fit along with its location parameter. The degree distribution is based on a comparison of fits between: exponential, log normal, and power law. See Figure 3 and Section 3.3 for additional details. The table also reports, for comparative purposes, the topological properties of the network of actors from the Internet Movie Database and the physics co-authorship network.
SOURCE – Author's calculations. Networks with † are reproduced from Newman (2003).

Table 3: Modularity Classes in the Pseudoemployer Projection

| | A: Edge Shares | | | | | B: Node Shares | |
|---|---|---|---|---|---|---|---|
| | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Emp. Wtd. | Not Wtd. |
| Class 1 | 58.39 | 7.21 | 11.09 | 13.10 | 14.87 | 21.36 | 22.72 |
| Class 2 | 8.11 | 57.79 | 15.41 | 17.55 | 10.24 | 24.43 | 21.08 |
| Class 3 | 10.28 | 12.69 | 58.40 | 8.09 | 8.61 | 16.58 | 20.29 |
| Class 4 | 13.99 | 16.66 | 9.32 | 52.45 | 15.16 | 25.61 | 17.12 |
| Class 5 | 9.23 | 5.65 | 5.77 | 8.82 | 51.13 | 12.02 | 18.78 |

NOTE – Table entries in Panel A are the percent of edges with one end in the modularity class indicated in the column heading and the other end in the modularity class indicated in the row heading. Table entries in Panel B are the share of nodes in each modularity class, weighted by the number of observed matches in the realized mobility network, and not weighted (raw node share). In every column of both panels, the entries sum to 100.

Table 4: Distribution of Pseudoemployer Modularity Classes by Major Industry and Occupation

| | Panel A: Major Industry | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Agr. | Mining | Constr. | Manu. | Transp. | Retail | FIRE | Bus. Svc. | Pers. Svc. | Entertain | Prof. Svc | Pub. Adm. |
| Class 1 | 7.33 | 1.33 | 24.65 | 23.12 | 12.20 | 16.09 | 1.56 | 6.09 | 1.21 | 0.82 | 2.82 | 2.79 |
| Class 2 | 1.15 | 0.40 | 2.69 | 5.28 | 2.11 | 4.30 | 18.95 | 5.09 | 0.76 | 0.39 | 41.66 | 17.22 |
| Class 3 | 0.24 | 0.44 | 1.27 | 35.49 | 14.73 | 23.63 | 0.57 | 13.31 | 0.85 | 2.72 | 5.74 | 1.01 |
| Class 4 | 0.38 | 0.09 | 0.20 | 15.07 | 2.10 | 23.71 | 0.54 | 4.51 | 14.80 | 0.95 | 36.40 | 1.25 |
| Class 5 | 0.21 | 0.88 | 0.60 | 54.81 | 5.80 | 23.85 | 0.60 | 2.42 | 0.39 | 0.33 | 3.39 | 6.71 |
| Pop. Share | 2.01 | 0.58 | 6.26 | 22.56 | 6.80 | 17.34 | 5.27 | 6.20 | 4.42 | 1.00 | 21.46 | 6.10 |
| | Panel B: Major Occupation | | | | | | | | | | | |
| | Prof.Svc. | Mgmt. | Sales | Clerical | Crafts | Oper. | Transp. | Labor. | Farm Lab. | Pvt. House | Svc. Work | Farm Mgt. |
| Class 1 | 3.61 | 13.11 | 3.18 | 4.51 | 30.40 | 12.04 | 13.27 | 11.23 | 2.86 | 0.02 | 4.12 | 1.67 |
| Class 2 | 28.19 | 17.11 | 4.06 | 40.34 | 1.80 | 2.72 | 0.40 | 0.55 | 0.01 | 0.00 | 4.73 | 0.09 |
| Class 3 | 32.49 | 25.82 | 12.36 | 13.04 | 9.73 | 4.45 | 0.28 | 0.90 | 0.02 | 0.00 | 0.90 | 0.00 |
| Class 4 | 14.32 | 8.73 | 2.75 | 12.09 | 4.24 | 12.88 | 2.14 | 2.43 | 0.01 | 3.56 | 36.84 | 0.00 |
| Class 5 | 7.01 | 6.44 | 7.10 | 16.96 | 16.72 | 33.92 | 1.96 | 6.41 | 0.06 | 0.00 | 3.42 | 0.00 |
| Pop. Share | 17.55 | 14.27 | 5.28 | 18.11 | 11.64 | 11.35 | 3.76 | 4.07 | 0.63 | 0.92 | 12.03 | 0.38 |

NOTE – In Panel A, table entries are the percent of pseudoemployers in a major industry conditional within a modularity class. In Panel B, table entries are the percent of pseudoemployers in a major occupation conditional within a modularity class. In both panels, the entries sum to 100 across the rows. The maximum modularity partition groups workers into 79 classes. ‘Class 5’ accumulates the smallest 74 classes.

Table 5: Modularity of the Pseudoemployer Projection Under Various Partitions

| | Partition | Num. Classes | Modularity |
|---|---|---|---|
| (1) | Max. Modularity | 79 | 0.516 |
| (2) | Max. Mod. Collapsed | 5 | 0.510 |
| (3) | Major Industry | 12 | 0.351 |
| (4) | Major Occupation | 12 | 0.306 |
| (5) | Maj. Ind×Occ | 144 | 0.223 |

NOTE – Table reports the modularity under each partitioning. Row (1) gives the modularity of the maximum modularity partition. Row (2) presents the modularity of the maximum modularity partition after collapsing the smallest 75 classes. Row (3) and (4) report the modularity according to the partition by major industry and major occupation. Row (5) reports modularity under a partition of jobs into a joint major industry/major occupation cell.

Table 6: Modularity Classes in the Worker Projection

| | A: Edge Shares | | | | | B: Node Shares |
|---|---|---|---|---|---|---|
| | Class A | Class B | Class C | Class D | Class E | |
| Class A | 70.12 | 7.06 | 9.78 | 12.51 | 5.05 | 25.04 |
| Class B | 10.02 | 63.94 | 19.57 | 11.42 | 14.77 | 22.65 |
| Class C | 8.30 | 11.70 | 56.36 | 10.05 | 6.12 | 16.21 |
| Class D | 4.61 | 2.97 | 4.37 | 49.72 | 4.37 | 14.00 |
| Class E | 6.95 | 14.33 | 9.93 | 16.30 | 69.70 | 22.10 |

NOTE – Table entries in Panel A are the percent of edges with one end in the modularity class indicated in the column heading and the other end in the modularity class indicated in the row heading. Table entries in Panel B are the share of workers in each modularity class. In every column of both panels, the entries sum to 100.

Table 7: Distribution of Worker Modularity Classes by Demographic Characteristics

| | White | Male | No High School | High School | Some College | College | Post Grad. | $IH$ |
|---|---|---|---|---|---|---|---|---|
| Class A | 48.94 | 86.03 | 32.78 | 37.62 | 20.19 | 6.59 | 2.55 | 0.779 |
| Class B | 45.36 | 25.73 | 21.33 | 34.31 | 24.03 | 11.69 | 8.17 | 0.694 |
| Class C | 40.23 | 39.08 | 29.89 | 39.74 | 21.43 | 5.91 | 2.38 | 0.667 |
| Class D | 66.34 | 61.98 | 9.60 | 26.71 | 27.38 | 23.86 | 12.36 | 0.640 |
| Class E | 59.90 | 39.67 | 5.96 | 22.40 | 25.17 | 23.96 | 21.91 | 0.751 |
| Inbreeding Homophily ($IH$) | 0.127 | 0.420 | 0.169 | 0.101 | 0.051 | 0.146 | 0.228 | |
| Modularity | 0.062 | 0.204 | 0.098 | | | | | |

NOTE – Table entries are the percent of workers in a modularity class/demographic characteristic cell. The table also reports the inbreeding homophily index for each partition class, as well as the modularity of the worker projection of the realized mobility network when the data are partitioned by race, sex, and education respectively.

Table 8: Cross-Tabulation of Worker and Pseudoemployer Modularity Classes Across Matches in the Realized Mobility Network

| | Pseudoemployer Modularity Class | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Worker Class A | **16.56** | 1.51 | 1.44 | 1.94 | 4.22 |
| | (5.49) | (6.27) | (4.26) | (6.58) | (3.09) |
| Worker Class B | 1.50 | 9.02 | 0.92 | 8.93 | 2.11 |
| | (4.80) | (5.49) | (3.73) | (5.76) | (2.70) |
| Worker Class C | 1.35 | 1.11 | 1.06 | **9.00** | 2.74 |
| | (3.26) | (3.73) | (2.53) | (3.91) | (1.84) |
| Worker Class D | 1.08 | 2.33 | **11.48** | 0.60 | 1.28 |
| | (3.58) | (4.10) | (2.78) | (4.30) | (2.02) |
| Worker Class E | 0.88 | **10.45** | 1.67 | 5.13 | 1.66 |
| | (4.23) | (4.84) | (3.28) | (5.07) | (2.38) |

NOTE – The main unparenthesized table entries are the percent of observed matches between worker modularity classes and pseudoemployer modularity classes. There are 27,519 unique matches, which correspond to the edges of the realized mobility network. For each match, the worker is assigned to her class in the maximum modularity partition of the weighted worker projection of the RMN. Likewise, the pseudoemployer in the match is assigned to its class in the maximum modularity partition of the weighted pseudoemployer projection of the RMN. The parenthesized values are the percent of matches in the cell under random matching, holding the number of edges in each modularity class fixed. Values in bold are those for which the observed share is at least twice as large as the predicted share.

Table 9: Linear Decomposition of Log Labor Market Earnings

| | (1) | (2) | (3) |
|---|---|---|---|
| Pseudoemployer Class Effect | | | |
| 1 | | −.057 | −.021 |
| | | (.023) | (.023) |
| 2 | | 0.186 | 0.112 |
| | | (.028) | (.030) |
| 3 | | 0.165 | 0.082 |
| | | (.032) | (.031) |
| 4 | | −.282 | −.174 |
| | | (.028) | (.028) |
| 5 | | 0.108 | 0.074 |
| | | (.026) | (.026) |
| Worker Class Effect | | | |
| A | | 0.011 | −.002 |
| | | (.023) | (.022) |
| B | | −.144 | −.060 |
| | | (.030) | (.030) |
| C | | −.007 | −.008 |
| | | (.030) | (.029) |
| D | | 0.112 | 0.090 |
| | | (.033) | (.031) |
| E | | 0.051 | 0.003 |
| | | (.029) | (.029) |
| Demographic Controls | Yes | Yes | Yes |
| Modularity Class Controls? | No | Yes | Yes |
| Ind/Occ Controls? | No | No | Yes |
| Num. Obs. | 48,049 | 48,049 | 48,049 |
| $R^2$ | 0.241 | 0.280 | 0.326 |

Standard errors are clustered within person. All models control for gender, an indicator for reporting white race, a quadratic in age interacted with gender, and educational attainment. The model in columns (2) and (3) also includes indicators for the modularity class from the worker projection and the modularity class from the pseudoemployer projection of the realized mobility network. Column (3) adds controls for major industry and occupation. Table A.1 reports estimates for the demographic controls.

Table 10: Correlation of Earnings Variance Components

| **Panel A: Full Sample (N=48,049)** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | Correlations | | | | |
| | Mean | Std. Dev. | log Earn. | $X\hat{\beta}$ | $\hat{\theta}$ | $\hat{\psi}$ | Resid. |
| log Earn. | 10.170 | 1.059 | 1 | | | | |
| $X\hat{\beta}$ | 3.307 | .458 | 0.490 | 1 | | | |
| $\hat{\theta}$ | 0 | .084 | 0.227 | 0.267 | 1 | | |
| $\hat{\psi}$ | 0 | .190 | 0.279 | 0.198 | 0.179 | 1 | |
| Resid | 0 | .899 | 0.849 | 0 | 0 | 0 | 1 |
| **Panel B: Match Sample (N=29,118)** | | | | | | | |
| | | | Correlations | | | | |
| | Mean | Std. Dev. | log Earn. | $X\hat{\beta}$ | $\hat{\theta}$ | $\hat{\psi}$ | Resid. |
| log Earn. | 10.043 | 1.109 | 1 | | | | |
| $X\hat{\beta}$ | 3.276 | 0.455 | 0.469 | 1 | | | |
| $\hat{\theta}$ | .001 | 0.086 | 0.247 | 0.268 | 1 | | |
| $\hat{\psi}$ | .006 | 0.186 | 0.294 | 0.209 | 0.224 | 1 | |
| Resid | -.101 | 0.949 | 0.864 | 0.004 | 0.026 | 0.028 | 1 |

Summary statistics and correlations of the earnings heterogeneity components estimated from Equation (14). $X\beta$ is the linear prediction of log earnings based on observable characteristics (quadratic in age interacted with sex, white or not, and highest completed education). $\theta$ is the portion of earnings variation associated with the modularity class of the worker. $\psi$ is the portion of earnings variation associated with the modularity class of the pseudoemployer. Panel A reports statistics based on the full PSID panel from 1987-1997, which implicitly weights jobs by their observed duration. Panel B reports statistics using only the first period of each new employment spell.

## Appendix

Table A.1: Linear Decomposition of Log Labor Market Earnings: Demographic Controls

| | (1) | (2) | (3) |
|---|---|---|---|
| White | .152 | .085 | .067 |
| | (0.018) | (0.017) | (0.016) |
| Male | 0.095 | 0.096 | 0.131 |
| | (0.267) | (0.256) | (0.256) |
| Age | 0.113 | 0.115 | 0.100 |
| | (0.010) | (0.009) | (0.009) |
| Age$^2$ | $-.001$ | $-.001$ | $-.001$ |
| | (0.0001) | (0.0001) | (0.0001) |
| Male×Age | 0.023 | 0.021 | 0.018 |
| | (0.013) | (0.013) | (0.012) |
| Male×Age$^2$ | $-$2E-4 | $-$2E-4 | $-$2E-4 |
| | (2E-4) | (2E-4) | (1E-4) |
| High Sch. | 0.375 | 0.315 | 0.253 |
| | (0.026) | (0.025) | (0.024) |
| Some Coll. | 0.552 | 0.417 | 0.324 |
| | (0.028) | (0.028) | (0.027) |
| College | 0.820 | 0.664 | 0.481 |
| | (0.031) | (0.033) | (0.034) |
| Postgrad. | 1.060 | 0.917 | 0.709 |
| | (0.035) | (0.037) | (0.039) |
| Constant | 6.73 | 6.862 | 7.033 |
| | (0.037) | (0.037) | (0.037) |
| Modularity Class Controls? | No | Yes | Yes |
| Ind/Occ Controls? | No | No | Yes |
| Num. Obs. | 48,049 | 48,049 | 48,049 |
| $R^2$ | 0.241 | 0.280 | 0.326 |

Estimated coefficients and standard errors of demographic controls in the linear decomposition of Table 9.