Estimating Labor Mobility from Professional Social Network data

15.579 – Applied Network Theory and Analysis

Sebastian Steffen

# Section I - Introduction

Labor mobility is a key economic indicator of healthy labor markets. A highly mobile labor force enables a more efficient and competitive allocation of resources. And while labor mobility can be measured in several different ways, recent papers that estimate these different measures such as occupational (Xu 2017), employer-to-employer (E-E) (Mercan 2017), and geographic (Molloy, Smith, and Wozniak 2014) transition rates , unanimously report that labor mobility has been slowing down over the past three decades. This may be evidence of systemic impediments, such as unattainability or unavailability of better or entry-level jobs, Monopsony power of firms in the labor market, or inefficient legal limitations. However, these decreases can also be interpreted in a more favorable manner.

Some have argued that Information Technologies may have made the labor market matching function more efficient (Mercan 2017). A more efficient matching process allows workers to learn more about the distribution of jobs and benefits and lets them find their optimal job match more quickly, such that jobs become ‘inspection goods’ rather than ‘experience goods’ for workers. Similarly, firms are able to find

Transition rates may have stayed constant for some demographic groups and the overall decrease may conceal positive compositional trends, such as a reduction in youth workers (Bosler and Petrosky-Nadeau 2016), or labor market trends towards the disappearance of short-duration jobs (Hyatt and Spletzer 2013).

Within-firm transitions may have risen as well through increases in promotions or lateral career moves (Moscarini and Thomsson 2007). Geographic mobility rates may have been historically inflated as people moved predominantly to more productive urban areas (Chetty et al. 2014). Finally, there may have been measurement and data errors in previous papers that biased estimates (Kambourov and Manovskii 2013).

This paper will contribute to the literature in several ways. First, I will unify the different labor mobility measures by framing aggregated worker transitions as weighted, directed edges in a network of ‘entity nodes’. Second, I will use this framework to estimate transition rates as well as clustering and centrality measures from a novel, large-scale employer-employee matched dataset, and discuss its significant advantages for estimating labor mobility over previously-used datasets. Besides overall transitions rates and demographic decompositions (Moscarini and Vella 2003) this data is rich enough to study transition rates decomposed by detailed industries, schools, and firms.

The rest of the paper is organized as follows. In section II I review the literature labor mobility with a focus on its three main measures: occupational, employer-employer (E-E), and geographic mobility. Section III reviews the advantages and disadvantages of previously-used and our datasets and discusses the measurement methodology. *In section IV I present my labor mobility estimates as well as my clustering and centrality measures.* Section V concludes.

# Section II – Literature Review

## Occupational Transitions

Moscarini and Thomsson (2007) find that between 1976 and 2006 an average of about 3.5 percent of male workers employed in two consecutive months report different three-digit occupations.

Xu (2017) relies on the CPS and the Survey of Income and Program Participation (SIPP) to estimate occupational transition rates. He finds similar results as Moscarini and Thomsson of about 3.5 percent but documents a decline to 2 percent in 2010 before it slowly rises again.

Bosler and Petrosky-Nadeau (2016) use the SIPP and similarly report that job transitions have decreased substantially among young workers (16-24) since the late 1990s and that this decrease accounts for a large part of the overall decline in job transitions. In fact, for older workers labor markets are as dynamic as they were 20 years ago. They report rates of about 2.7 percent in 1997 and 2.0 percent in 2013, which is more or less in line with previous estimates.

## Employer-Employer (E-E) Transitions

The first reliable E-E transition measures come from Fallick and Fleischman (2004). They used the redesigned 1994 Current Population Survey (CPS), which asked all respondents who had reported to be employed in the previous month, whether they were still working for the same employer in the current month. If they were, the previous month’s answers to employer-related questions were carried forward. Otherwise, respondents had to answer additional questions about their industry, class, and occupation. From these responses, the authors create monthly Markov matrices with state space S = {Employed, Unemployed, Not in Labor Force} from 1994 until the end of 2003. They find that an average of 2.6 percent of employed persons change employers each month. Somewhat surprisingly, this number is similar to the number of people who move from an employer out of the workforce (2.4 percent). Moves from employment into unemployment are lower at about 1.3 percent.

Bjelland et al. (2011) use the Longitudinal Employer Household Dynamics (LEHD) data to estimate quarterly employer-employer transitions. They estimate E-E transitions of about 4 percent per quarter, which is lower than the monthly Fallick and Fleischman estimates. These transitions account for 4 percent of new employment starts and 29 percent of job separations.

## Geographic Transitions (Migration)

Significantly lower geographic mobility has been reported for older, less educated (Long 1988), and black workers. Substantial mobility responses to demand shocks have been documented for the 1980s (Bartik 1991; Blanchard et al. 1992) and beyond (Notowidigdo 2011). However, there is significant heterogeneity by education and race - less-educated as well as black workers display lower migration rates than their college-educated and white counterparts (Bound and Holzer 2000; Notowidigdo 2011).

Molloy, Smith, and Wozniak (2014) identify the decreased net benefit to changing employers as the main driver of the declines in geographic transitions. Thus outside options seem to have become relatively less valuable. Notowidigdo (2011) finds that migration responses are asymmetric – positive local labor demand shocks increase the local population more than negative shocks decrease it, which implies that populations are ‘sticky’.

# Section III - Data

The majority of the papers discussed above relies on the CPS. It is representative of the entire civilian US population and contains a very large sample – about 65,000 households are interviewed monthly. It is the source of the official unemployment and labor force participation measures published by the Bureau of Labor Statistics (BLS) which makes its raw numbers and definitions historically consistent and interpretable. Each household is interviewed once a month for four consecutive months in one year and then again for the same four months in the following year, resulting in 8 monthly surveys for that household. Each month a new rotational group is added, which replaces the group that rotates out in that month. This means that in any given month 8 different rotational groups are surveyed. During the week that contains the 19th of the month, the survey asks each person of age 15 years or older in each household about their labor force activity, employment status, occupation, and industry in the previous week. As described in Kambourov and Manovskii (2013), in months 2-4 and 6-8 respondents, who had reported to be employed in months 1 and 5, were asked whether they were still working for the same employer.

There are several disadvantages to the CPS. First, and this concern is shared among most data sets, its temporal resolution may be low and erroneously label multiple state transitions in a short time frame as a single transition. Thus, intervening unemployment periods or multiple employers may not be captured. This concern is particularly serious for data sources that only employ quarterly or yearly surveys such as the PSID.

Kambourov and Manovskii (2013) cautions against using the CPS and PSID to study worker mobility. They argue that the transition rates based on the CPS do not measure transition rates annually, but at a much shorter period instead. The CPS involves a considerable number of missing data imputations and mobility measures are very sensitive to them. They also raise significant doubts for using the Panel Study of Income Dynamics (PSID). Since most of its occupational affiliation data is only gathered at an annual rate, the inability to identify multiple transitions likely severely underestimates the true transition rates. The PSID also excludes immigrants arriving in the US after 1968 and is thus not representative of the US population. CPS and PSID based occupational transition rates differ significantly. PSID-based measures hover between 15 and 20 percent between 1980 and 2000, while CPS-based measured generally stay below 10 percent over the same time frame.[[1]](#footnote-1)

Bjelland et al. (2011) point out the high rate of attrition in the CPS as well, because it does not follow people who move and because of the lack of employer characteristics. Instead they rely on the Longitudinal Employer Household Dynamics (LEHD) data. These data are based on comprehensive administrative records of workers that are in employments covered by state unemployment insurance systems. It roughly covers 6-7 million jobs and 200,000 to 360,000 E-E flows per quarter. It contains data on a large fraction of the US workforce and US employers, includes detailed employer characteristics and can be matched to the census to get additional demographic characteristics.

Unlike the previously-discussed data sets the LEHD has detailed employee-employer matches, which allows the identification of rich workers’ employment histories as well as firm dynamics. Thus, these data can be used to study transition patterns due to firm characteristics, such as the firm level of turn-over or size.

On the other hand, LEHD has some weaknesses as well. First, employment can only be identified via workers’ quarterly earnings. Thus, transition undercounting is likely a significant concern. Second, cross-state migration as well as transitions into employment that is not covered by unemployment insurance may be missed.

Finally there is the Survey of Income and Program Participation (SIPP), which is a household survey conducted by the US Census Bureau. It gathers data through waves of panels, each with 14,000 to 37,000 participating households and with a time span of ranging between 2.5 and 4 years. Each wave covers a four-month period and provides information on a person’s employment status, as well as start and end dates for up to two employers.

I propose to use the largest professional social network, LinkedIn, to study labor mobility. It contains detailed employment and education histories of over 500 million people, out of which over 100 million live in the US. This data is uniquely suited to study labor mobility because it can be used to construct a long and detailed balanced career panel that contains employment as well as education.

Of course there are disadvantages to these data as well. First, there may be considerable reporting bias as people can choose what and how to report on their profiles. For example, they may choose not to report the start or end year or month for a particular position. People frequently lie about or at least favorably misrepresent themselves. Given the large number of members, I focus on the subset of workers who list start and end months for their positions. More data cleaning may be required. For example certain occupations such as ‘self-employed’ or freelancer are significantly more common on LinkedIn than reported in BLS occupation counts. One way to mitigate this may be to normalize by BLS occupation counts or to code certain occupations as unemployment. While the total number of transitions may only be somewhat affected, mistaking the type of transition as employment to employment instead of employment to unemployment is a more serious concern. Nevertheless, the richness of these data will help to understand the empirical patterns of recent changes in labor mobility.

# Section IV - Methodology

I have constructed a balanced, monthly panel of employment and education from 2000 until 2018. Thus, for each member I have 216 monthly observations of what jobs and educational training positions they are holding, which firms and schools they are attending, and what general geographic area they live in along with additional demographic data. For each of the three labor mobility measures I define a different state space. Notably, for occupational transitions I also include education programs and for E-E transitions I analogously include schools:

To identify transitions in each state space over time, I employ a similar methodology as Bjelland et al. (2011). Let be an indicator for individual i listing employer (occupation, region) j in time period t. Then an Employer-Employer transition happens when individual i moves from employer j to employer k during time period t, i.e.:

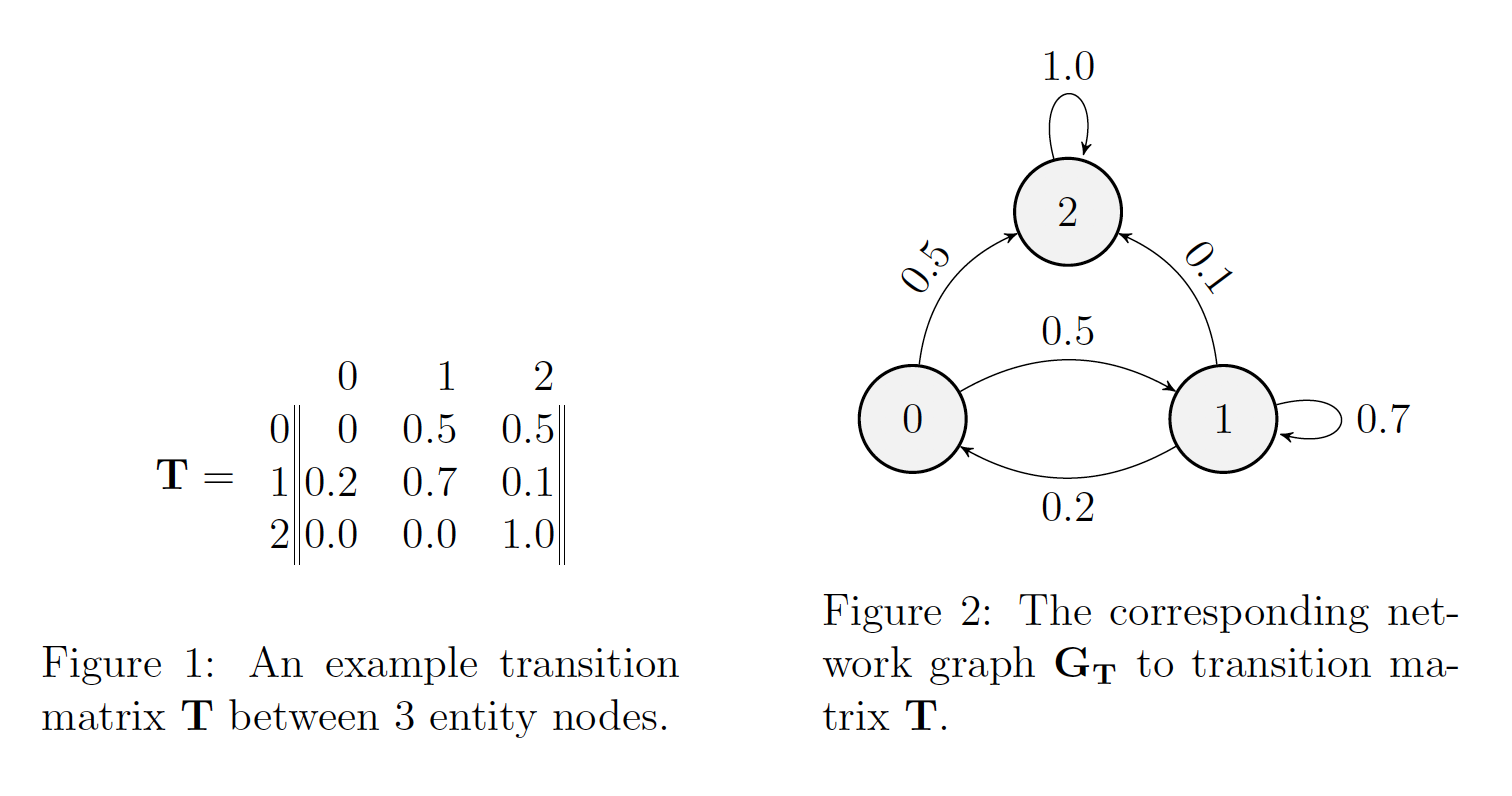
This definition of a transition is quite intuitive. However, as Bjelland et al. (2011) point, this definition cannot account for (i) multiple job holdings nor (ii) transitions that include an unemployment spell in the interim. Therefore, they instead use a more conservative definition and take not just one, but two periods before and after the transition into account, to more reliably avoid undercounting transitions:

To get to the network graph of entities, we count the number of transitions happening between each pair of nodes, i.e.:

Notably, we do not disallow the case where , i.e. an individual continuing to be in the same state. In the network, this would be represented as a self-loop on entity node j. Transitions in which are ‘true’ transitions between two different entity nodes, from j to k.

To calculate overall transition rates we sum up all transitions, i.e. non-loop edges, and divide by the number of all transitions, i.e.:

The methodology is similar for occupations and geographies. We have thus defined a probability (or Markov) matrix with probabilities of a transition happening between each state. The graphical equivalent of such a matrix is a weighted network graph.



Now that we have graphs with weighted, directed edges we can also define centrality and clustering coefficients. It may be necessary to prune the network and delete nodes or edges along which very transitions happen. Link salience and other filtering methods discussed in Fotouhi et al. (2018) may be necessary in order to calculate these coefficients, since there are several centrality and clustering measures and their correlation is not always high (Valente et al. 2008). It would also be promising to study the network over time and compare yearly labor market flow patterns. In particular, shocks such as the financial crisis as well as rising automation should result in higher transition patterns into lower paying, less automatable jobs, education programs, or unemployment.

# Section V - Conclusion

I have proposed to use LinkedIn to study labor mobility, in particular occupational, employer, and geographic transitions. By framing all three within a Markov chain setting, which is the same as a network of directed, weighted transition flows, this paper attempts to provide a clearer methodology to study labor mobility. The network setting also lends itself to defining additional measures of labor mobility and its impediments, such as network clustering and centrality. These measures can help to identify polarization of opportunities and systemic factor that exacerbate the rising inequality.

Several important research questions can be answered with the data within this framework. While the description of detailed labor mobility patterns is interesting in and of itself due to the lack of suitable data so far, other questions from the labor literature should be answerable. This includes the existence of stepping stone jobs, which are defined as low-paying jobs that may lead to high-paying ones. Similarly, the existence of sheepskin effects may be studied as well to answer whether signaling or learning are more important. I am excited to continue working with this data.

# Bibliography

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, Mich. : W.E. Upjohn Institute for Employment Research, 1991.

Bjelland, Melissa, Bruce Fallick, John Haltiwanger, and Erika McEntarfer. 2011. “Employer-to-Employer Flows in the United States: Estimates Using Linked Employer-Employee Data.” *Journal of Business & Economic Statistics* 29(4): 493–505. http://www.jstor.org.libproxy.mit.edu/stable/23243750.

Blanchard, Olivier Jean, Lawrence F Katz, Robert E Hall, and Barry Eichengreen. 1992. “Regional Evolutions.” *Brookings Papers on Economic Activity* 1992(1): 1–75. http://www.jstor.org.libproxy.mit.edu/stable/2534556.

Bosler, Canyon, and Nicolas Petrosky-Nadeau. 2016. “Job-to-Job Transitions in an Evolving Labor Market.” *FRBSF Economic Letters* November. https://www.frbsf.org/economic-research/files/el2016-34.pdf (November 16, 2018).

Bound, John, and Harry J Holzer. 2000. “Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s.” *Journal of Labor Economics* 18(1): 20–54. http://www.jstor.org/stable/10.1086/209949.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States\*.” *Quarterly Journal of Economics* 129(4): 1553–1623. http://10.0.4.69/qje/qju022.

Fallick, Bruce, and Charles A Fleischman. 2004. “Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows.” *Finance and Economics Discussion Series* 34. https://www.federalreserve.gov/pubs/feds/2004/200434/200434pap.pdf (November 4, 2018).

Fotouhi, Babak, Naghmeh Momeni, Maria A Riolo, and David L Buckeridge. 2018. “Statistical Methods for Constructing Disease Comorbidity Networks from Longitudinal Inpatient Data.” *Applied Network Science* 3. https://doi.org/10.1007/s41109-018-0101-4 (November 8, 2018).

Hyatt, Henry R, and James R Spletzer. 2013. “The Recent Decline in Employment Dynamics.” *IZA Journal of Labor Economics* 2(1): 5. https://doi.org/10.1186/2193-8997-2-5.

Kambourov, Gueorgui, and Iourii Manovskii. 2013. “A Cautionary Note on Using (March) Current Population Survey and Panel Study of Income Dynamics Data to Study Worker Mobility.” *Macroeconomic Dynamics* 17(01): 172–94. http://www.journals.cambridge.org/abstract\_S1365100510000350 (November 16, 2018).

Long, Larry E. 1988. *Migration and Residential Mobility in the United States.* New York : Russell Sage Foundation, 1988.

Mercan, A Yusuf. 2017. “Fewer but Better : The Decline in Job Mobility and the Information Channel.” *Working Paper*.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak. 2014. “Declining Migration within the U.S.: The Role of the Labor Market.” *NBER Working Paper*. https://www.nber.org/papers/w20065.pdf (December 15, 2018).

Moscarini, Giuseppe, and Kaj Thomsson. 2007. “Occupational and Job Mobility in the US.” *Scandinavian Journal of Economics* 109(4): 807–36. http://10.0.4.87/j.1467-9442.2007.00510.x.

Moscarini, Giuseppe, and Francis Vella. 2003. “Aggregate Worker Reallocation and Occupational Mobility in the United States: 1971-2000.” *Ssrn* (203): 1971–2000.

Notowidigdo, Matthew J. 2011. *The Incidence of Local Labor Demand Shocks*. https://www.nber.org/papers/w17167.pdf (December 15, 2018).

Valente, Thomas W, Kathryn Coronges, Cynthia Lakon, and Elizabeth Costenbader. 2008. “How Correlated Are Network Centrality Measures?” *Connections (Toronto, Ont.)* 28(1): 16–26. https://www.ncbi.nlm.nih.gov/pubmed/20505784.

Xu, Ming. 2017. “Understanding the Decline in Occupational Mobility.” *Working Paper*: 1–59.

1. See figure 1 of (Kambourov and Manovskii 2013) [↑](#footnote-ref-1)