Skill polarization and economic inequality

Ahmad Alabdulkareem[a,b,1], **Morgan R. Frank**[c,1], **Lijun Sun**[c], **Bedoor AlShebli**[d], **César Hidalgo**[c], **and Iyad Rahwan**[a,c,2]

[a]Institute for Data, Systems, & Society, Massachusetts Institute of Technology, Cambridge, MA, USA 02139; [b]Center for Complex Engineering Systems, King Adbulaziz City for Science and Technology, Riyadh, SAU 12371; [c]Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA 02139; [d]Electrical Engineering and Computer Science, Masdar Institute of Science and Technology, Abu Dhabi, UAE

This manuscript was compiled on April 28, 2017

Economic inequality is one of the biggest challenges facing society today. Employment in high- and low-wage occupations has grown, leading to a 'hollowing' of the middle class. However, a fundamental puzzle remains: how do skills distinguish high- and low-wage occupations and does this distinction constrain career mobility? Using tools from network science, we show that skills exhibit a striking polarization into two clusters representing socio-cognitive and sensory-physical skills. This polarization predicts wage disparity, and is strongly correlated with education. Moreover, worker mobility between occupations is constrained in a predictable way by the polarization of skills, with low-skilled workers 'stuck' in their low-wage cluster. Together, these results provide a new explanation for the persistence of economic inequality, and inform strategies to mitigate the negative effects of automation and off-shoring on employment.

labor | career mobility | complex systems | inequality

Economic inequality is on the rise, making it one of the central challenges facing policy-makers today (1–3). For example, consider that absolute income mobility–the fraction of children who earn more than their parents–has fallen dramatically in the U.S. from 90% to 50% (4). This fading of the 'American dream' (5, 6) highlights the growing need to characterize low- and high-wage occupations, and to identify the constraints on career mobility between the two.

Occupational polarization–the increasing proportion of high- and low-wage employment at the expense of middle-wage employment (7–9)–presents itself both in national statistics and in a majority of U.S. metropolitan areas (10). Mechanisms driving this trend, such as off-shoring (11) and automation (12–14), alter the demand for workers capable of specific workplace skills, tasks, knowledge, and ability (hereafter, "skills"). Since the demand for skills shapes the demand for occupations, how do skills distinguish between low- and high-wage occupations?

While education facilitates explicit career advancement (15, 16), workers also leverage complementarity between their existing skills and the skill requirements of occupations to make smaller career changes (17). Similarly, cities leverage complementarity between industries to optimize productivity and increase their competitiveness in a global economy (18, 19). Studies have identified the aggregate effects of skill complementarity on labor dynamics, such as the redefinition of skills comprising each occupation (14), but mapping the space of skills has remained elusive. In particular, does the structure of skill complementarity constrain the career mobility of individual workers and the adaptability of entire urban workforces?

In this study, we introduce a method for mapping skill complementarity as a network. We demonstrate a deep connection between occupational polarization and an underlying polarization of workplace skills. After identifying skills that are indicative of high-wage occupations, we confirm the positive role of educational attainment as a mechanism for career advancement, and we demonstrate that the skill complementarity network captures individual worker's transitions between occupations. Finally, we combine our observations to demonstrate the detrimental effects of skill polarization on career mobility.

Materials and Methods

Data Sets. The O*NET program by the U.S. Department of Labor annually produces the publicly available O*NET database detailing the importance of 161 workplace skills, knowledge, and abilities to the completion of each of the 672 occupations recognized under the Standard Occupational Classification (SOC) System. The O*NET database is updated regularly allowing for annual snapshots of the relationships between occupations and skills through continual survey of workers from each occupation. We use annual O*NET data from the years 2010 through 2015. We denote the importance of skill $s \in S$ to occupation $j \in J$ using $onet(j, s) \in [0, 1]$ where $onet(j, s) = 1$ indicates that s is essential to j, while $onet(j, s) = 0$ indicates that workers of occupation j need not possess or perform s.

The Bureau of Labor Statistics (BLS) annually produces publicly available data detailing the distribution of SOC occupations in each U.S. metropolitan statistical area (MSA). MSA's represent an entire urban system including areas with large proportions of commuters employed in the city proper. We use the terms MSA and "city" interchangeably. Along with the numbers of workers of each occupation, BLS provides additional details about the annual salary of each occupation in each city.

The U.S. Census Bureau and BLS produce a monthly Current Population Survey (CPS) through a continuous survey process that produces representative samples of the U.S. population. Providing high resolution labor statistics is one of the primary goals of CPS, and, in particular, CPS records changes in occupations of survey participants over a 1.5 year period for which that participant is an active contributor to the survey. For our purpose, we are interested only in participants who reported one occupation when they were first surveyed in 2014 and reported working a different occupation

Significance Statement

The "hollowing" of middle-class jobs is one of the most-cited causes for inequality. But we know little about how this process takes place. The traditional study of labor operates at high levels of granularity, without attention to the complex structure of skills and tasks that make up jobs. We introduce a new methodology, based on network science, to analyze skill interrelatedness in the labor market. We make three discoveries: (i) workplace skills are polarized into two distinct clusters, separating socio-cognitive skills from sensory-physical skills; (ii) this polarization has serious consequences on wages; (iii) polarization also constrains workers' ability to move up the economic ladder. These results inform strategies to mitigate the negative effects of automation and off-shoring on employment.

Author contributions: A.A., M.R.F., & L.S. performed calculations. A.A. & M.R.F. produced figures. A.A., B.A., & L.S. constructed online data visualization. A.A., M.R.F., I.R., & C.H. wrote the manuscript.

The authors declare no conflict of interest.

[1]A.A. (Ahmad Alabdulkareem) and M.R.F. (Morgan R. Frank) contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: irahwan@mit.edu

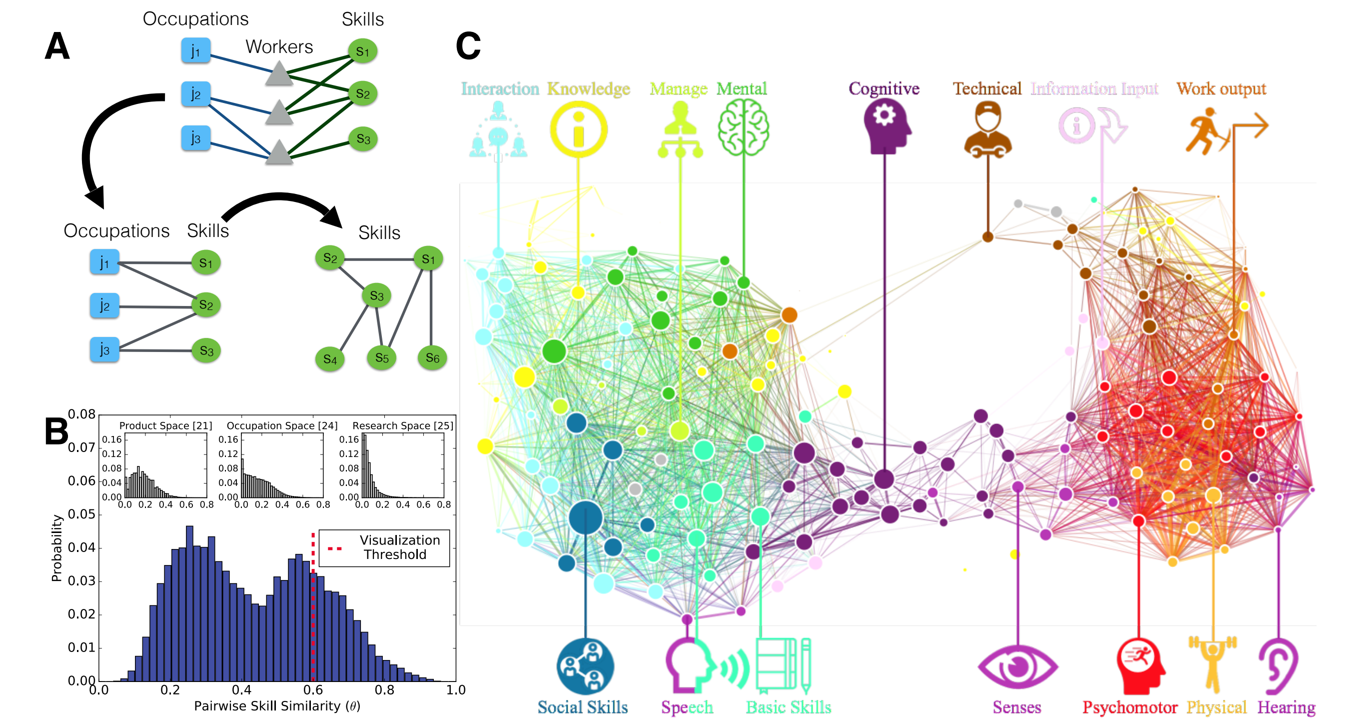

Fig. 1. Constructing the Skillscape. **(A)** An occupation is identified through the skills of workers of that occupation. The bipartite network connecting occupations to required skills is a result of an underlying tripartite network containing workers as a conduit between occupations and skills. Relationships between skills are determined from their co-occuring importance across occupations. **(B)** Unlike previous applications of RCA (insets), the Skillscape contains a bimodal distribution of pairwise skill complementarity. **(C)** The Skillscape thresholded according to a minimum skill similarity (i.e. $\theta > 0.6$) visibly reveals two communities of complementary skills and respects expertly-derived O*NET categories (colors). Node sizes reflect the total skill similarity shared between that skill and all other skills.

when they were surveyed one year later in 2015. There are several methods for joining different time periods of the CPS data (20), so we employed strict merging criteria including participant ID, gender, sex, state of residency, and age to verify the validity of our occupation transitions. The result is a dataset of 5,400 occupation transitions for individual U.S. workers from 2014 to 2015.

Mapping Skill Complementarity. Occupations are typically the unit of interest when discussing labor dynamics, but the labor requirements that define an occupation are reflected in the skills possessed by workers of that occupation (see Fig. 1A). These skill requirements represent key features that uniquely identify occupations. However, raw O*NET data do not control for ubiquitous skills, such as “Identifying Objects" and “Communicating with Supervisors and Peers" (see SI Appendix, Fig. S1). Therefore, we focus on skills that are over-expressed in an occupation by calculating the revealed comparative advantage (21–23) (RCA) of each skill in an occupation according to

$$rca(j,s) = \frac{onet(j,s)/\sum_{s' \in S} onet(j,s')}{\sum_{j' \in J} onet(j',s)/\sum_{j' \in J, s' \in S} onet(j',s')}. \qquad [1]$$

Occupations are distinguishable from each other according to their “effective use" of skills (i.e. $rca(j,s) > 1$). Skill *complementarity* (17) (denoted θ) is then the minimum of the conditional probabilities of a pair of skills being effectively used by the same occupation. The distribution of complementarity values is provided in Figure 1B. This novel methodology identifies skill pairs that co-occur across occupations and represent key occupational features. Co-occurrence captures how a pair of skills support each other, either by boosting the productivity of a worker who possesses both skills, or by the ease of acquiring both skills simultaneously; our definition of complementarity is agnostic to the exact source of the complementarity. We call the resulting network of skill complementarity the “Skillscape" (see Fig. 1C. See SI Appendix, Section 1 for visualizations of this methodology and a visualization of the Skillscape as a skill-to-skill complementarity matrix).

The Skills of Urban Workforces. Combining the relationships between occupations and skills according to the O*NET database with the distributions of occupations in cities according to BLS captures how strongly each urban workforce relies on each skill. Denoting the number of workers in city c with occupation j using $bls(c,j)$, we combine the two data sets according to

$$CS(c,s) = \sum_{j \in J} bls(c,j) \cdot onet(j,s), \qquad [2]$$

where $CS(c,s)$ denotes city c's reliance on workplace skill s.

As with the raw O*NET data, certain jobs and certain skills are ubiquitous across many cities. We again apply RCA to $CS(c,s)$ to identify which skills are effectively used in each city according to

$$rca(c,s) = \frac{CS(c,s)/\sum_{s \in S} CS(c,s)}{\sum_{c \in C} CS(c,s)/\sum_{c \in C, s \in S} CS(c,s)}. \qquad [3]$$

Similar to occupations, $rca(c,s) > 1$ indicates the effective use of s in c. We provide some additional visualizations in SI Appendix, Section 4.

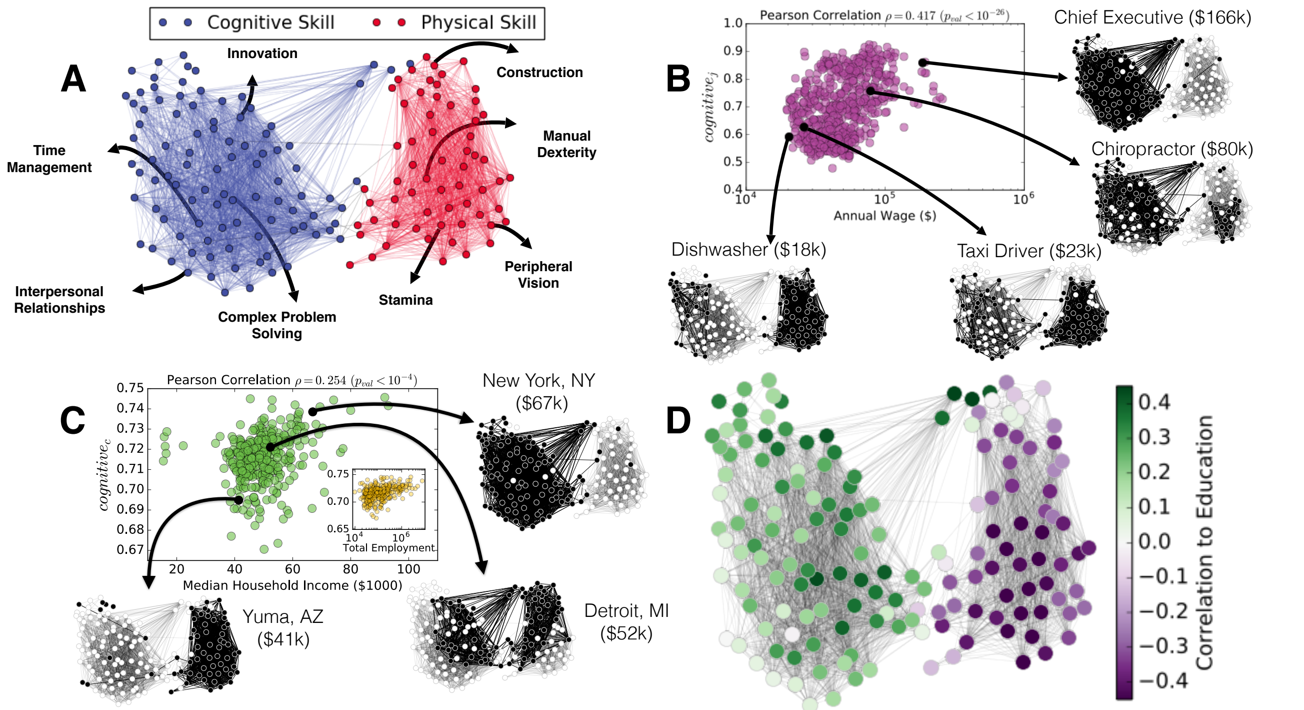

Fig. 2. The polarized Skillscape explains occupational wage polarization and economic well-being of urban workforces. **(A)** Community detection on the complete Skillscape network (i.e. no minimum θ) reveals two communities of complementary skills: socio-cognitive skills (blue) and sensory-physical skills (red). The displayed network is filtered ($\theta > 0.6$) for visualization purposes. **(B)** Occupations relying on socio-cognitive skills tend to make higher annual salaries. **(C)** Large cities increasingly rely on socio-cognitive skills (inset) yielding higher median household income. In (B) & (C), example occupations (cities), along with their annual wages (median household income), are projected onto the Skillscape using black nodes for effectively used skills. **(D)** The skill network colored by correlation between $onet(j, s)$ and the average educational degree requirement across occupations.

Skillscape Proximity. Does skill complementarity (i.e. θ) correspond to "nearby" skills in the real-world? We explore this question through occupational re-definition and the occupation transitions of individual workers. Recalling that a skill is effectively used by an occupation if $rca(j, s) > 1$, predicting which skills will be acquired by an occupation or used by a worker becomes more difficult as we restrict to skills with lower RCA values in the year of interest. Let $\lambda \in [0, 1]$ denote the RCA upper limit for skills to consider. Given the set of skills which are effectively used by occupation j in year y (denoted $E_y(j) = \{s \in S \mid rca_y(j, s) > 1\}$), we consider the unweighted proximity of each remaining skill (denoted $\bar{E}_{y,\lambda}(j) = \{s \in S \setminus E_y(j) \mid rca_y(j, s) < \lambda\}$) to the skills in $E_y(j)$. For each $s' \in \bar{E}_{y,\lambda}(j)$, we calculate

$$\overline{proximity}(s') = \sum_{s \in E_y(j)} \theta(s', s). \quad [4]$$

We control for occupations relying on ubiquitous skills by calculating weighted Skillscape proximity (hereafter "proximity") according to

$$proximity(s') = \frac{\sum_{s \in E_y(j)} \theta(s', s)}{\sum_{s \in S} \theta(s', s)}, \text{ for each } s' \in \bar{E}_{y,\lambda}(j). \quad [5]$$

Note that analogous calculations can determine Skillscape proximity from urban workforces by considering $rca(c, s)$ instead of $rca(j, s)$.

We rank each skill in $\bar{E}_{y,\lambda}(j)$ according to their proximity to skills in $E_y(j)$, and examine $\bar{E}_{y,\lambda}(j) \cap E_{2015}(j)$. This produces a receiver operating characteristic (ROC) curve from which we calculate the area under the ROC curve (commonly referred to as AUROC). AUROC allows us to assess the predictive power of Skillscape proximity without prescribing a minimum proximity threshold, and allows us to compare several different predictors with potentially different ranges of values. SI Appendix, Figure S10 provides a cartoon explanation of this calculation.

Results and Discussion

While applications of RCA to other systems reveal unimodal edge-weight distributions (21, 24, 25), the Skillscape contains a bimodal distribution of skill complementarity (see Fig. 1B). The skill network (see Fig. 1C) is polarized into two skill communities according to Louvain community detection (26) (see SI Appendix, Section 1 for list of O*NET skills comprising each cluster). One community is comprised of socio-cognitive skills and one of sensory-physical skills (see Fig. 2A). This separation between traditionally "technical" and "non-technical" skills defines a polarization of skills underlying the U.S. labor system (see SI Appendix, Section 3 for polarization validation).

To what extent does the polarization of socio-cognitive and sensory-physical skills explain occupational polarization? We quantify this according to occupation j's cognitive skill fraction, which is given by

$$cognitive_j = \frac{\sum_{s \in C} onet(j, s)}{\sum_{s \in S} onet(j, s)}, \quad [6]$$

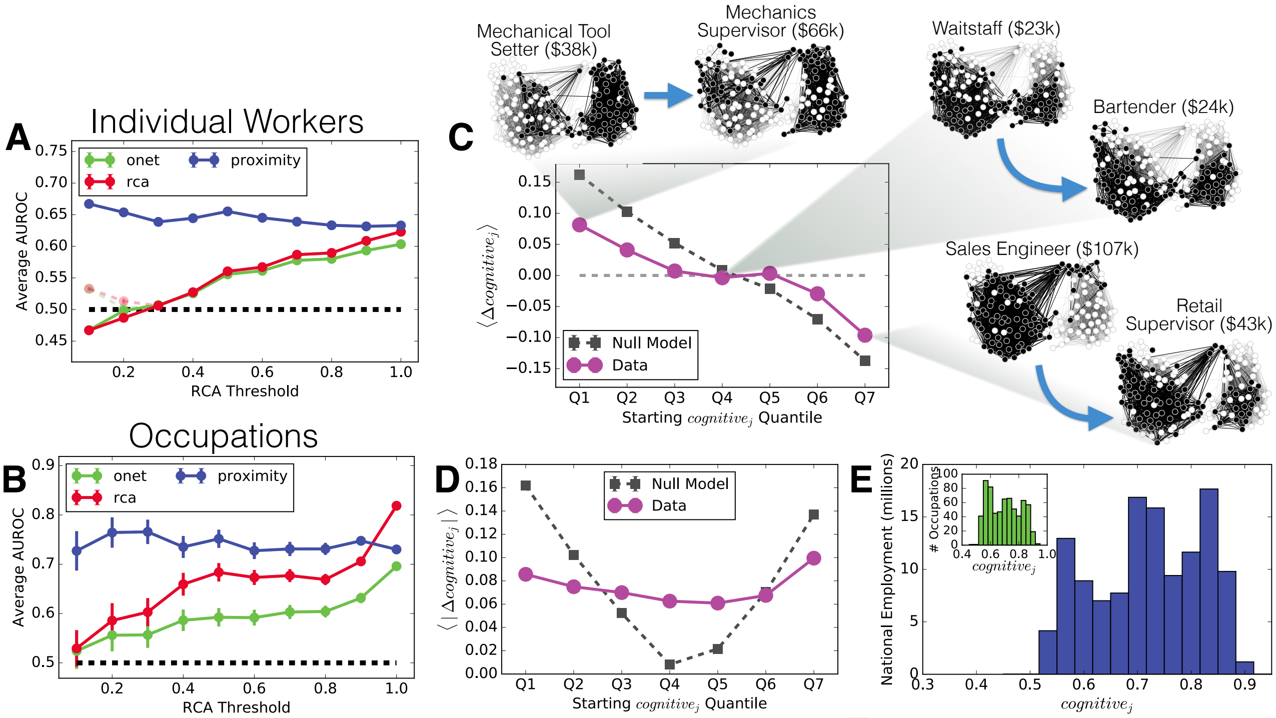

Fig. 3. Occupation redefinition and worker mobility are constrained by skill polarization. **(A)** Skillscape proximity (blue) better predicts the skill requirements of a worker's new occupation in 2015 from the skills of their old occupation in 2014 compared to O*NET data (green), RCA (red), and a null model (black). Each point is the average performance of each predictor, and standard error bars are provided, but may be negligible. **(B)** Similar to (A), the effectively used skills comprising each occupation in 2015 compared to 2010 are best predicted by Skillscape proximity. Binning by the $cognitive_j$ of the worker's occupation in 2014 reveals the **(C)** expected cognitive change and **(D)** the expected magnitude of cognitive change when workers change occupations. Random occupation selection is considered as a null model (grey). Standard error bars are provided, but are small. Actual occupation transitions are provided as examples in (C). **(E)** The national distribution of employment by $cognitive_j$ with the distribution of individual occupations as an inset.

where C denotes the set of socio-cognitive skills according to Louvain community detection. An increased $cognitive_j$ corresponds to greater annual wages for occupations (Pearson correlation $\rho = 0.42$, $p_{val} < 10^{-26}$, Fig. 2B) suggesting that skill polarization indeed relates to the division between high- and low-wage occupations.

Moreover, occupations are often considered as part of an urban workforce (27–29) because cities represent humanity's hubs for innovation (30–32) and economic growth (33, 34). Does skill polarization explain economic well-being across cities? Inspired by work detailing the occupations and labor complementarity of cities (24, 35, 36), we use employment distributions for U.S. MSAs compiled by the BLS. Similar to occupations, cities that employ larger workforces tend to require more socio-cognitive skills ($\rho = 0.40$, $p_{val} < 10^{-13}$) and hence have higher median household incomes ($\rho = 0.25$, $p_{val} < 10^{-4}$) according to the 2014 American Community Survey (see Fig. 2C).

Since socio-cognitive skills indicate economic welfare for individuals and cities, how do workers obtain socio-cognitive skills? Often, education is thought to play an essential role in preparing workers for promising careers (15, 16). For example, it has been shown that educational institutions act as a social "sorting machine" (15) when students begin their careers. However, educational attainment has not kept pace with financial returns on education in recent years (16). We explore the influence of education through the correlation between $onet(j, s)$ and the average degree requirement for each occupation. Skills in the socio-cognitive cluster indicate higher education requirements across occupations, while occupations with more lenient degree requirements tend to rely on sensory-physical skills (see Fig. 2D. Further examples in SI Appendix, Section 2). Although the Skillscape is polarized, education allows workers to obtain cognitive skills, which are essential to occupations with higher incomes.

Skill acquisition through explicit education can be costly and time consuming, so, more commonly, workers transition between occupations based on the similarity of their skill set and the skill requirements of each occupation. Given the network structure of skill complementarity, we hypothesize that skill polarization *constrains* the occupational mobility of individual workers. This hypothesis is not directly testable because we do not understand the precise mechanisms for worker adaptation, nor do we understand that mechanism's interplay with other market equilibrium dynamics (9). However, the hypothesis has three testable predictions. First, the skills of an individual worker's occupation change locally on the Skillscape with that worker's transition to new occupations. Second, workers are more likely to transition to occupations relying on skills in the same skill community. Third, skill polarization represents a bottleneck in workers' upward mobility towards high-wage occupations. This should lead to disproportionately high employment below a certain $cognitive_j$ threshold, rather than a smooth distribution of employment across the range of $cognitive_j$ values. In the remainder, we demonstrate these predictions empirically.

Does skill polarization constrain the career mobility of individual workers (12) and the redefinition of occupational skill requirements (9)? The U.S. Census Bureau's Current Population Survey (CPS) details the occupation transitions of individual workers from occupation j_A in 2014 to occupation j_B in 2015. We note the skills which are not effectively used by j_A (i.e. $rca_{2014}(j_A, s) \leq 1$) and predict the effectively used skills of j_B (i.e. $rca_{2015}(j_B, s) > 1$). Ordering these skills according to $onet_{2014}(j_A, s)$, $rca_{2014}(j_A, s)$, or Skillscape proximity, produces receiver operating characteristic curves from which the area under the curve (AUROC) represents the ability of each predictor. Recalling that a skill is not effectively used if $rca(j, s) \leq 1$, the problem of predicting which skills will be acquired becomes more difficult as we restrict ourselves to skills with lower $rca(j, s)$ in the year of interest (e.g. consider only $rca_{2014}(j, s) < .9$, see Fig. 3A). Similarly, we use the 2010 and 2015 O*NET data to test Skillscape proximity's ability to predict changes in the constituent skills of each occupation (see Fig. 3B). Skillscape proximity outperforms $rca(j, s)$, $onet(j, s)$, and a random model at predicting skill acquisition of both occupations and individual workers, which suggests that worker mobility and occupational redefinition occur locally on the Skillscape. Additional figures detailing proximity's predictive ability for several subsets of workers are provided in SI Appendix, Section 5.

Since occupational transitions represent local changes in workers' skill requirements, the polarized network of skills should constrain mobility between low-wage sensory-physical occupations and high-wage socio-cognitive occupations. We capture this explicitly by binning occupation transitions into quantiles (each representing 780 transitions) according to $cognitive_{j_A}$ and examining the average cognitive change (i.e. $\Delta cognitive = cognitive_{j_B} - cognitive_{j_A}$, see Fig. 3C) and average magnitude of cognitive change (see Fig. 3D) for each bin. We consider workers selecting there new occupations at random as a null model for comparison. See SI Appendix, Section 5.1 for a discussion of alternative null models, including randomizing the selection of "cognitive skills". Workers transitioning from low-quantile physical occupations tend towards new occupations with higher cognitive skill fraction, but the magnitude of change is less than would be expected under random occupation selection. Similarly, workers transitioning from high-quantile cognitive occupations tend towards occupations with lower $cognitive_j$, but again make smaller changes in magnitude by comparison to random occupation selection. By contrast, workers transitioning from mid-quantile occupations, which represent starting occupations which effectively use cognitive and physical skills evenly, exhibit larger magnitudes of change in $cognitive_j$ compared to the null model. In conclusion, workers of occupations relying strongly on one skill community tend towards other occupations within the same skill community, thus validating the second prediction.

For the third prediction of our hypothesis, first note that the definition of skill complementarity (17) indicates increasing returns to combining skills within each skill community. Therefore, skill communities may be explained by the easy acquisition of related skills, or by production efficiencies offered by workers who have such complementary skills. However, this also means that workers relying on sensory-physical skills will face difficulty acquiring socio-cognitive occupations because they are unprepared to exploit large proportions of the socio-cognitive skills. Until they have a sufficient proportion of socio-cognitive skills, they are bottle-necked by the polarized skill structure. If true, then we expect disproportionately high employment in occupations under some threshold of $cognitive_j$.

Indeed, binning national employment according to $cognitive_j$ yields a trimodal distribution (see Fig. 3E. Additional years and binning, and city employment distributions, are provided in SI Appendix, Section 5.2.). The upper and lower modes of the distribution correspond to workers who are effectively exploiting the skill complementarity *within* each of their respective skill communities. The presence of a third mode in the middle suggests that skill polarization constrains workers from obtaining attractive socio-cognitive skills.

The Skillscape maps the structure of workplace skill complementarity, and connects urban workforces and occupations to their constituent skills. This transparency explains the polarization of high- and low-skill occupations as a separation between occupations relying on sensory-physical skills and occupations relying on socio-cognitive skills. Furthermore, the structure of skill complementarity accurately captures path-dependent occupation transitions of individual workers and the dynamics of occupation redefinition. Therefore, the Skillscape indeed relates to the dynamics of labor demands. Combined, these results demonstrate the Skillscape's usefulness as a tool to identify obtainable workplace skills from the current abilities of an individual worker, and highlights industries and occupations which urban planners and policy makers should accommodate based on the existing labor infrastructure in their city. At scalable.media.mit.edu/skillscape (password: workforce), we provide an online interactive tool for exploring occupations and urban workforces on the Skillscape.

ACKNOWLEDGMENTS. This work was supported by the Center for Complex Engineering Systems (CCES) at King Abdulaziz City for Science and Technology (KACST) and the Massachusetts Institute of Technology (MIT).

1. Kochhar R, Fry R, Rohal M (2015) The american middle class is losing ground.
2. Alichi A, Kantenga K, Solé JA (2016) Income polarization in the united states.
3. Cingano F (2014) Trends in income inequality and its impact on economic growth.
4. Chetty R, et al. (2016) The fading american dream: Trends in absolute income mobility since 1940, (National Bureau of Economic Research), Technical report.
5. Putnam RD (2016) *Our kids: The American dream in crisis.* (Simon and Schuster).
6. Johnson HB (2014) *The American dream and the power of wealth: Choosing schools and inheriting inequality in the land of opportunity.* (Routledge).
7. Autor DH, Dorn D (2013) The growth of low-skill service jobs and the polarization of the us labor market. *The American Economic Review* 103(5):1553–1597.
8. Autor DH, Katz LF, Kearney MS (2008) Trends in us wage inequality: Revising the revisionists. *The Review of Economics and Statistics* 90(2):300–323.
9. Acemoglu D, Autor D (2011) Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics* 4:1043–1171.
10. (2016) America's shrinking middle class: A close look at changes within metropolitan areas, (Pew Research), Technical report.
11. Ebenstein A, Harrison A, McMillan M, Phillips S (2014) Estimating the impact of trade and offshoring on american workers using the current population surveys. *Review of Economics and Statistics* 96(4):581–595.
12. MacCrory F, Westerman G, Alhammadi Y, Brynjolfsson E (2014) Racing with and against the machine: Changes in occupational skill composition in an era of rapid technological advance. in *ICIS*.
13. Autor DH (2015) Why are there still so many jobs? the history and future of workplace automation. *The Journal of Economic Perspectives* 29(3):3–30.
14. Bessen JE (2015) How computer automation affects occupations: Technology, jobs, and skills. *Boston Univ. School of Law, Law and Economics Research Paper* (15-49).
15. Kerckhoff AC (2001) Education and social stratification processes in comparative perspective. *Sociology of Education* pp. 3–18.
16. Autor D (2010) The polarization of job opportunities in the us labor market: Implications for employment and earnings. *Center for American Progress and The Hamilton Project.*
17. Brynjolfsson E, Milgrom P (2013) Complementarity in organizations. *The Handbook of Organizational Economics* pp. 11–55.
18. Porter ME (1998) *Clusters and the new economics of competition.* (Harvard Business Review Boston) Vol. 76.
19. Porter ME (2000) Location, competition, and economic development: Local clusters in a global economy. *Economic development quarterly* 14(1):15–34.
20. Madrian BC, Lefgren L (1999) A note on longitudinally matching current population survey (cps) respondents.
21. Hidalgo CA, Klinger B, Barabási AL, Hausmann R (2007) The product space conditions the development of nations. *Science* 317(5837):482–487.
22. Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. *proceedings of the national academy of sciences* 106(26):10570–10575.
23. Hausmann R, Hidalgo CA (2011) The network structure of economic output. *Journal of Economic Growth* 16(4):309–342.
24. Shutters ST, Muneepeerakul R, Lobo J (2015) Constrained pathways to a creative urban economy. *Urban Studies.*
25. Guevara MR, Hartmann D, Aristarán M, Mendoza M, Hidalgo CA (2016) The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* 109(3):1695–1709.
26. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10):P10008.
27. Youn H, et al. (2016) Scaling and universality in urban economic diversification. *Journal of The Royal Society Interface* 13(114):20150937.
28. Fields GS (1975) Rural-urban migration, urban unemployment and underemployment, and job-search activity in ldcs. *Journal of Development Economics* 2(2):165–187.

29. (2015) World migration report 2015 - migrants and cities: New partnerships to manage mobility. *International Organization for Migration* 8(1561-5502).
30. Pan W, Ghoshal G, Krumme C, Cebrian M, Pentland A (2013) Urban characteristics attributable to density-driven tie formation. *Nature communications* 4.
31. Bettencourt LM, Samaniego H, Youn H (2014) Professional diversity and the productivity of cities. *Scientific reports* 4.
32. Florida R (2004) *The rise of the creative class.* (Basic books New York).
33. Glaeser E (2011) *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.* (Penguin).
34. Quigley JM (1998) Urban diversity and economic growth. *The Journal of Economic Perspectives* 12(2):127–138.
35. Kok S, Weel Bt (2014) Cities, tasks, and skills. *Journal of Regional Science* 54(5):856–892.
36. Bacolod M, Blum BS, Strange WC (2009) Skills in the city. *Journal of Urban Economics* 65(2):136–153.

Skill polarization and economic inequality

Ahmad Alabdulkareem[1,2,†], Morgan R. Frank[3,†], Lijun Sun[3],
Bedoor AlShebli[4], César Hidalgo[3] & Iyad Rahwan[1,3,*]
[1]Institute for Data, Systems, & Society, Massachusetts
Institute of Technology, USA
[2]Center for Complex Engineering Systems at KACST & MIT
[3]Media Laboratory, Massachusetts Institute of Technology, USA
[4]Electrical Engineering and Computer Science,
Masdar Institute of Science and Technology, UAE
[†]These authors contributed equally to this work.
[*]Corresponding Author: irahwan@mit.edu

April 28, 2017

Contents

1 Exploring Occupations and their Constituent Skills

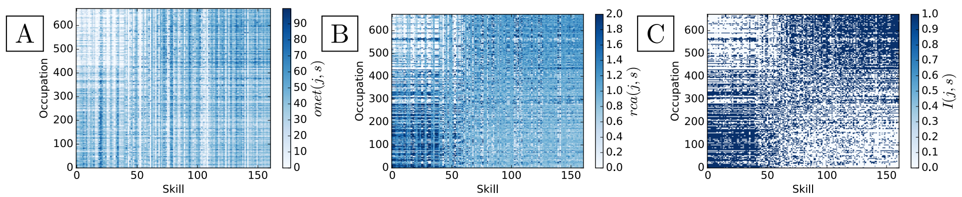

Figure 1: **(A)** The raw occupation-skill matrix, $onet(j, s)$, **(B)** the RCA occupation-skill matrix, $rca(j, s)$, and **(C)** the thresholded RCA job-skill matrix, $I(j, s)$, for 2014. Here, $I(j, s) = 1$ if and only if $rca(j, s) > 1$. Occupations (y-axis) are ordered by the sum of threshold RCA skill values, and skills (x-axis) are ordered by the correlation of their thresholded RCA values across occupations to the occupational sums.

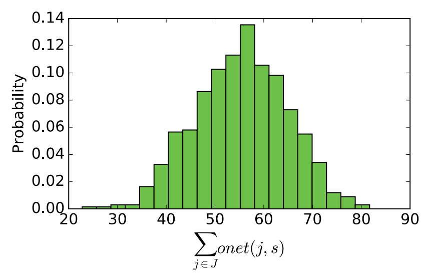

Figure 2: The distribution of aggregate skill importance by summing raw O*NET values of each occupation.

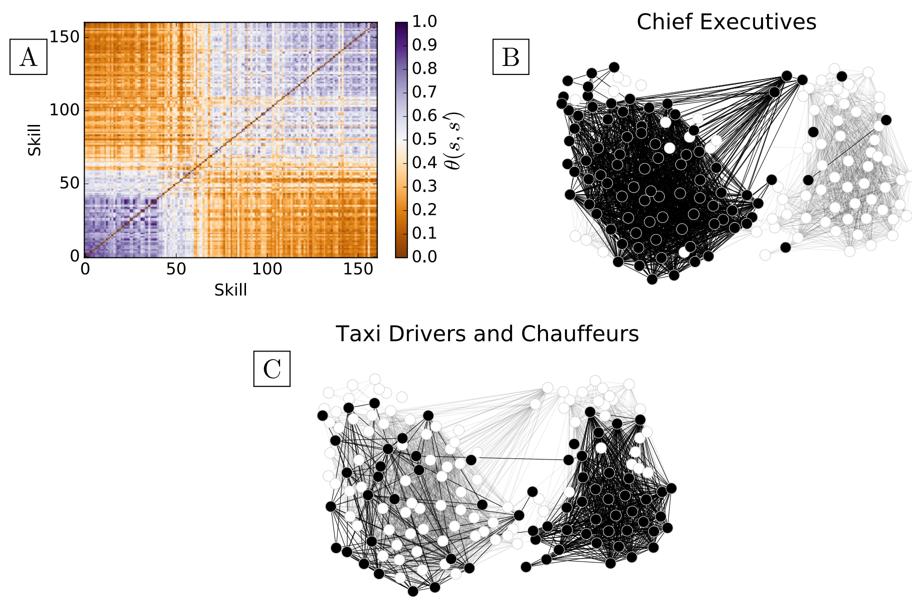

Figure 3: The skill-skill matrix **(A)** defining the Skillscape from 2014 data. The projections of the occupations of Chief Executive **(C)** and Taxi Driver **(D)** onto the Skillscape based on their effectively used skills (black).

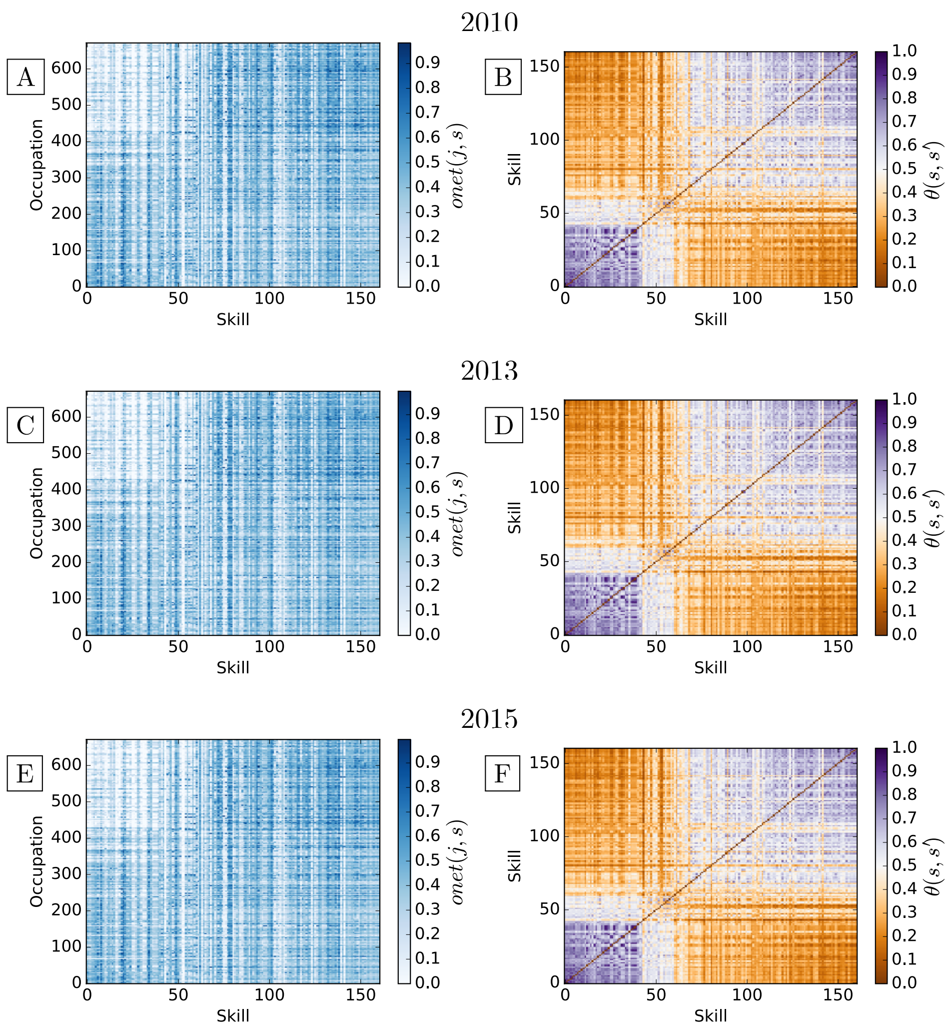

Figure 4: A comparison of raw O*NET data (left column) and the resulting Skillscape matrix (right column) for 2010, 2013, and 2015. The order of occupations and skills is preserved across plots for easy comparison. The polarized skill structure is stable across years.

Skill Type	O*NET Skill
Socio-Cognitive Skills	Active Learning, Active Listening, Complex Problem Solving, Coordination, Critical Thinking, Instructing Judgment and Decision Making, Learning Strategies, Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Mathematics, Monitoring, Negotiation, Operations Analysis, Persuasion, Programming, Reading Comprehension, Science, Service Orientation, Social Perceptiveness, Speaking, Systems Analysis, Systems Evaluation, Time Management, Writing, Category Flexibility, Deductive Reasoning, Fluency of Ideas, Inductive Reasoning, Information Ordering, Mathematical Reasoning, Memorization, Near Vision, Number Facility, Oral Comprehension, Oral Expression, Originality, Problem Sensitivity, Speech Clarity, Speech Recognition, Speed of Closure, Written Comprehension, Written Expression, Administration and Management, Biology, Clerical, Communications and Media, Computers and Electronics, Customer and Personal Service, Economics and Accounting, Education and Training, English Language, Fine Arts, Foreign Language, Geography, History and Archeology, Law and Government, Mathematics Knowledge, Medicine and Dentistry, Personnel and Human Resources, Philosophy and Theology, Psychology, Sales and Marketing, Sociology and Anthropology, Telecommunications, Therapy and Counseling, Analyzing Data or Information, Assisting and Caring for Others, Coaching and Developing Others, Communicating with Persons Outside Organization, Communicating with Supervisors, Peers, or Subordinates, Coordinating the Work and Activities of Others, Developing Objectives and Strategies, Developing and Building Teams, Documenting/Recording Information, Establishing and Maintaining Interpersonal Relationships, Evaluating Information to Determine Compliance with Standards, Getting Information, Guiding, Directing, and Motivating Subordinates, Identifying Objects, Actions, and Events, Interacting With Computers, Interpreting the Meaning of Information for Others, Judging the Qualities of Things, Services, or People, Making Decisions and Solving Problems, Monitoring and Controlling Resources, Organizing, Planning, and Prioritizing Work, Performing Administrative Activities, Performing for or Working Directly with the Public, Processing Information, Provide Consultation and Advice to Others, Resolving Conflicts and Negotiating with Others, Scheduling Work and Activities, Selling or Influencing Others, Staffing Organizational Units, Thinking Creatively, Training and Teaching Others, Updating and Using Relevant Knowledge

Sensory-Physical Skills	Equipment Maintenance, Equipment Selection, Installation, Operation Monitoring, Operation and Control, Quality Control Analysis, Repairing, Technology Design, Troubleshooting, Arm-Hand Steadiness, Auditory Attention, Control Precision, Depth Perception, Dynamic Flexibility, Dynamic Strength, Explosive Strength, Extent Flexibility, Far Vision, Finger Dexterity, Flexibility of Closure, Glare Sensitivity, Gross Body Coordination, Gross Body Equilibrium, Hearing Sensitivity, Manual Dexterity, Multilimb Coordination, Night Vision, Perceptual Speed, Peripheral Vision, Rate Control, Reaction Time, Response Orientation, Selective Attention, Sound Localization, Spatial Orientation, Speed of Limb Movement, Stamina, Static Strength, Time Sharing, Trunk Strength, Visual Color Discrimination, Visualization, Wrist-Finger Speed, Building and Construction, Chemistry, Design, Engineering and Technology, Food Production, Mechanical, Physics, Production and Processing, Public Safety and Security, Transportation, Controlling Machines and Processes, Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment, Estimating the Quantifiable Characteristics of Products, Events, or Information, Handling and Moving Objects, Inspecting Equipment, Structures, or Material, Monitor Processes, Materials, or Surroundings, Operating Vehicles, Mechanized Devices, or Equipment, Performing General Physical Activities, Repairing and Maintaining Electronic Equipment, Repairing and Maintaining Mechanical Equipment

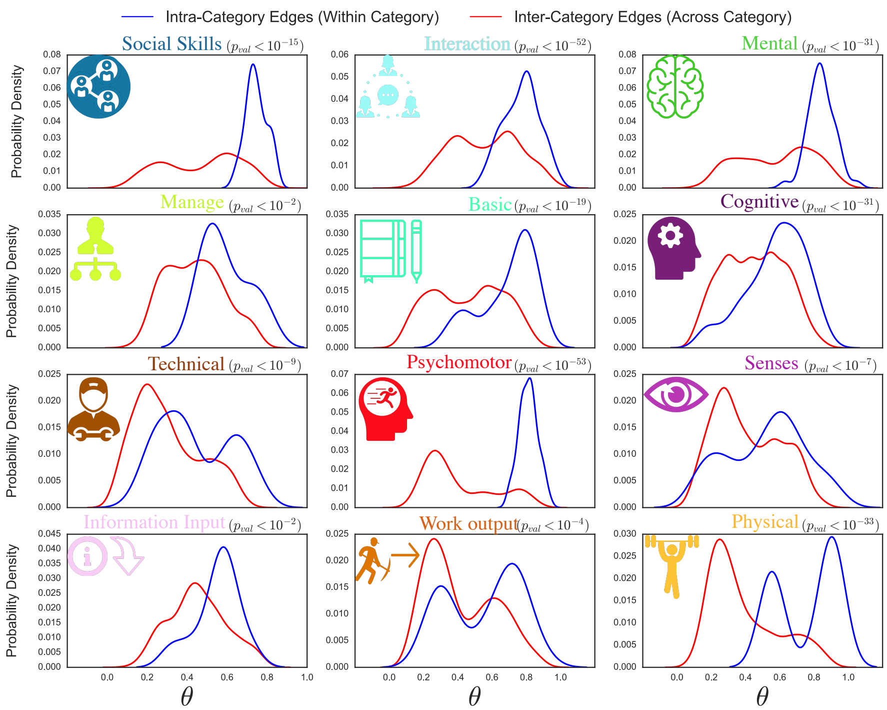

Figure 5: The Skillscape network respects skill categorization from experts. For each O*NET skill category, we measure the distribution of θ's for pairs of skills within a category (blue) and compare to the distribution of θ's for each edge connecting a skill within the category to a skill outside of the category (red). The complementarity for skills within a category is significantly stronger according the the KS statistic (title) then the complementarity for inter-category pairs of skills.

2 How Educational Requirements Relate to Skill Requirements for Occupations

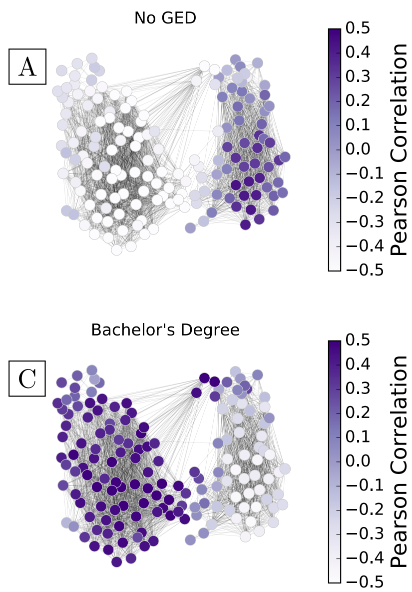

Figure 6: The skill requirements of an occupation indicate the education required. In each panel, we plot the Skillscape network thresholding edges with $\theta > 0.6$. Nodes (or skills) are colored according to the Pearson correlation between $onet(j, s)$ and the proportion of workers of each occupation with a given degree (title).

3 Validating Skill Polarization

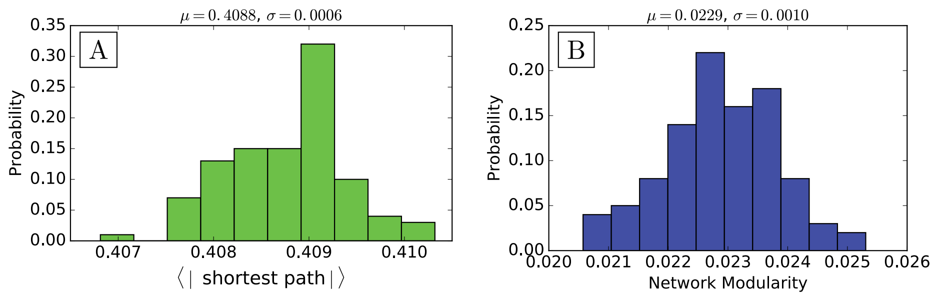

Figure 7: Testing the significance of Skillscape polarization. We perform 100 trials of randomly shuffling Skillscape edge weights (i.e. θ for modularity, and $1-\theta$ for shortest path) and measuring the resulting **(A)** average shortest path length and **(B)** network modularity. Empirically, the average shortest path is 0.567 and the network modularity is 0.159; both clear outliers given the values resulting from randomization.

4 Projecting Urban Workforces onto the Skillscape

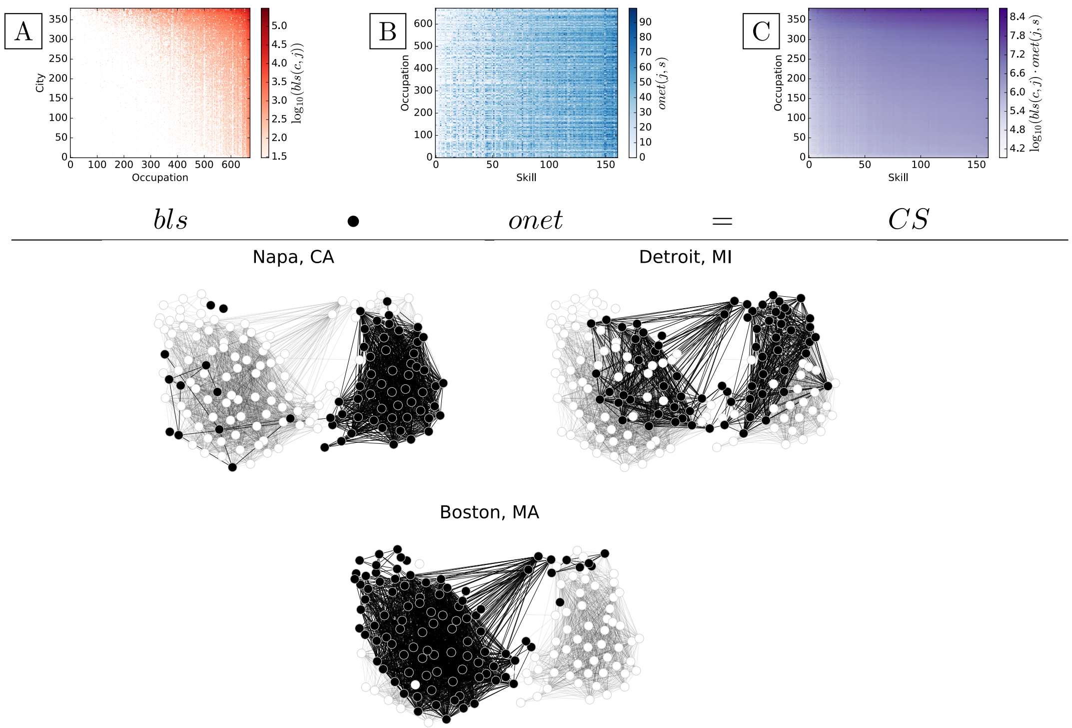

Figure 8: Using 2014 data to construct a matrix relating skills to cities from the dot product of the **(A)** city-job matrix with the **(B)** job-skill matrix to produce a **(C)** city-skill matrix. Note that (B) is the same matrix as Fig. 1A only reordered. Logarithmic values are presented in (A) and (C) only for visual appeal. Finally, we project Napa, Detroit, and Boston onto the Skillscape based on their effectively used skills.

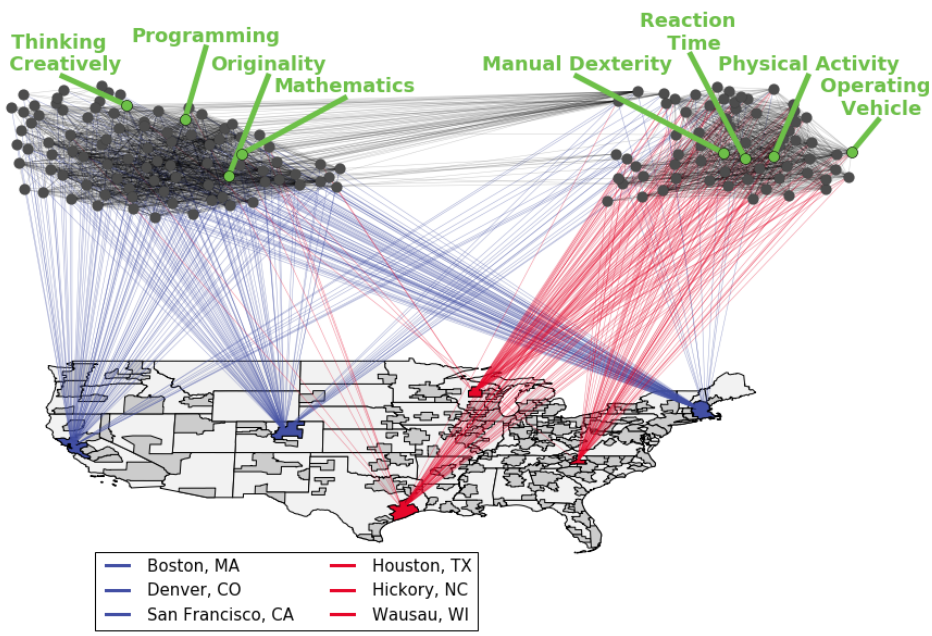

Figure 9: Example cities projected onto the Skillscape according to effective use of skills. Blue cities rely on socio-cognitive skills, while red cities rely on sensory-physical skills.

5 Using Skillscape Proximity to Predict Labor Dynamics

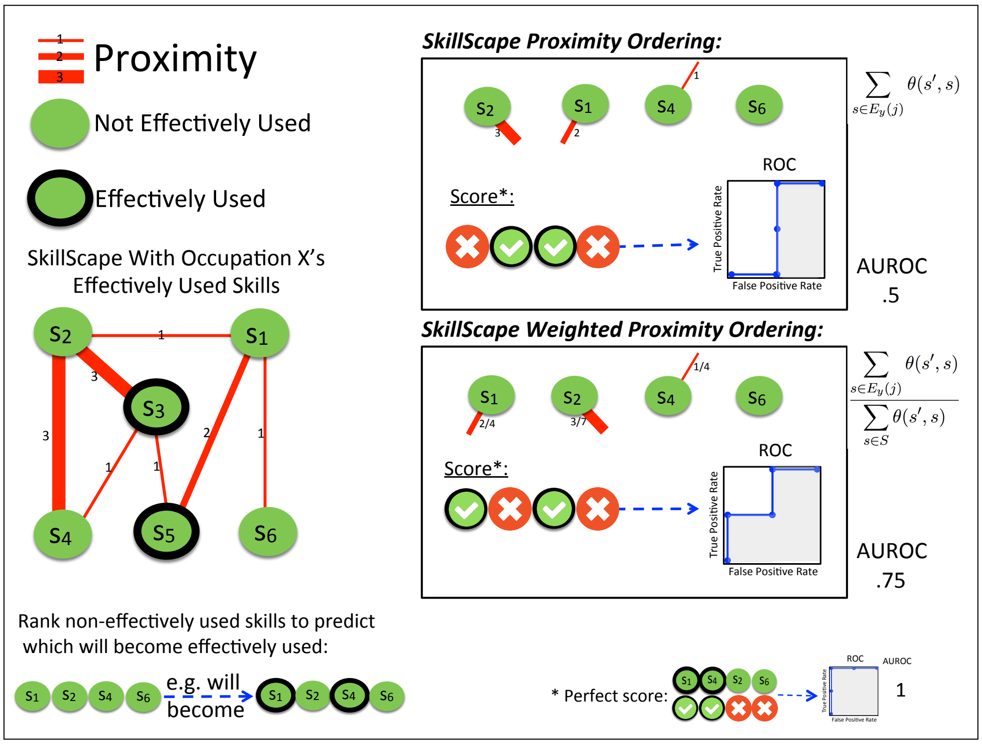

Figure 10: A cartoon example of AUROC calculation.

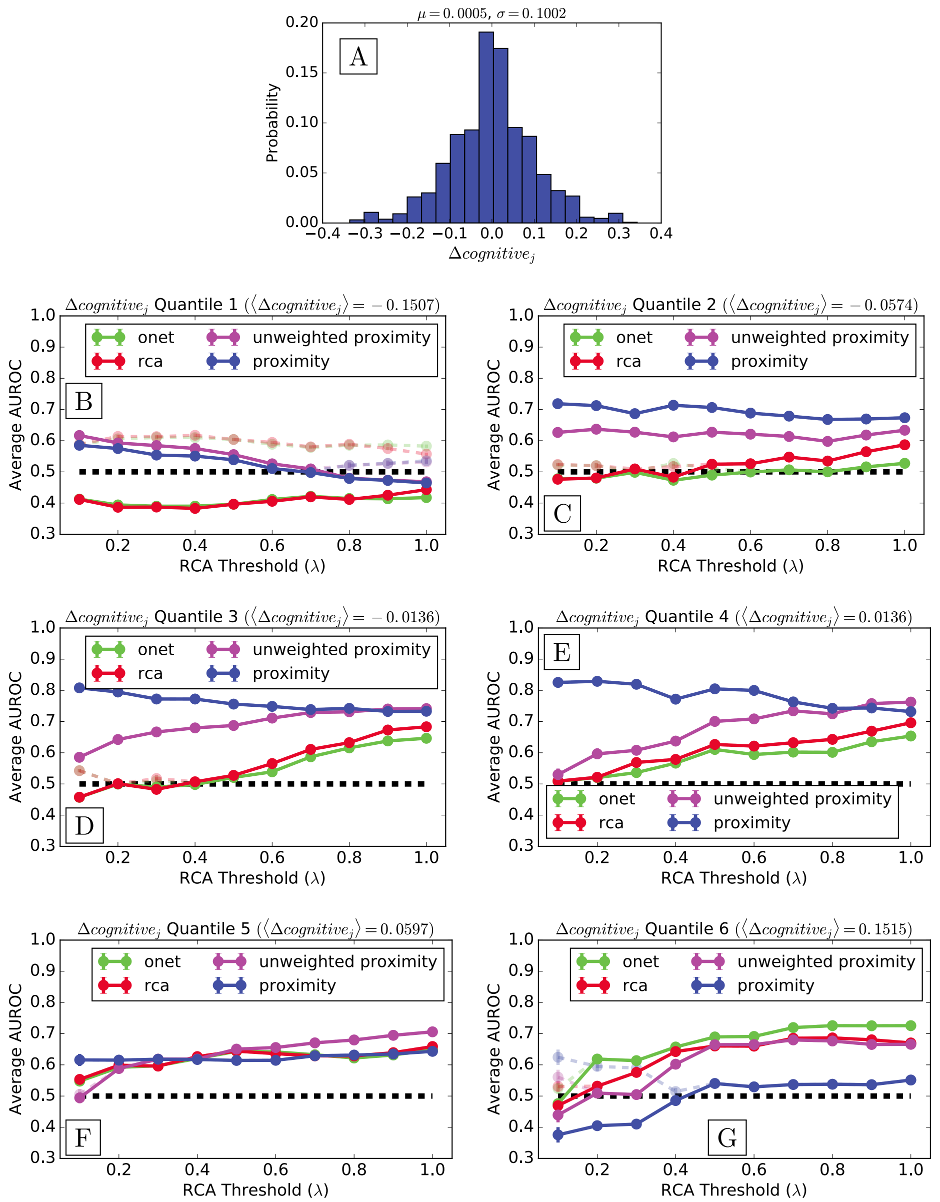

Figure 11: **(A)** The distribution of changes in cognitive skill fraction (denoted $\Delta cognitive_j$) associated with occupation transitions of individual workers from the CPS data. **(B)-(G)** The performance of *onet*, *rca*, and Skillscape proximity for predicting changes in effectively used skills after binning CPS occupation transitions into six quantiles ($N = 900$ transitions per bin) according to the change in cognitive skill fraction.

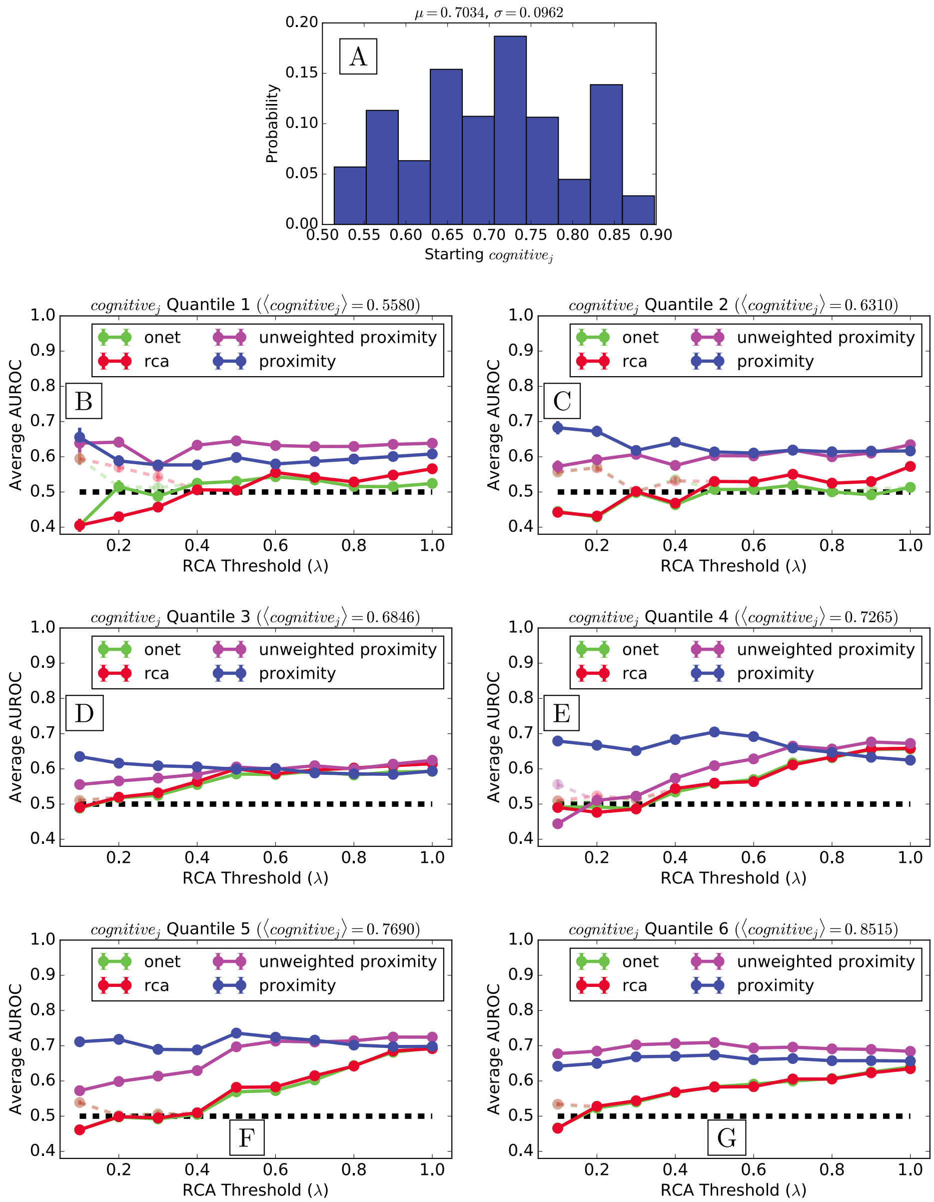

Figure 12: **(A)** The distribution of starting cognitive skill fraction associated with occupation transitions of individual workers from the CPS data. **(B)-(G)** The performance of *onet*, *rca*, and Skillscape proximity for predicting changes in effectively used skills after binning CPS occupation transitions into six quantiles ($N = 900$ transitions per bin) according to the cognitive skill fraction for the worker's original occupation.

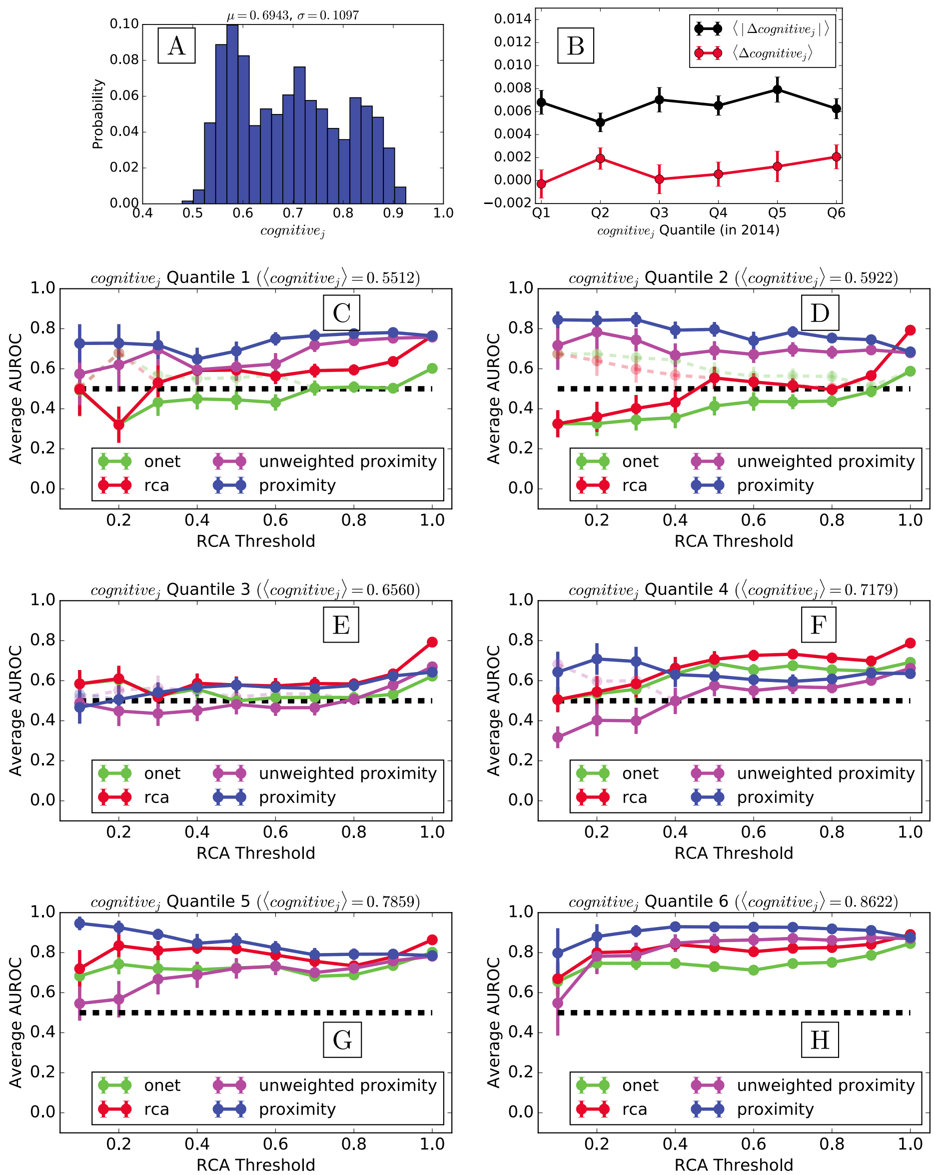

Figure 13: **(A)** The distribution of starting cognitive skill fraction associated with each occupation in 2014. **(B)** The average change (red) and average magnitude of change (black) in cognitive skill fraction for occupations from 2014 to 2015 for each quantile of occupations according to $cognitive_j$ in 2014 ($N = 100$ occupations per bin). **(C)-(H)** The performance of *onet*, *rca*, and Skillscape proximity for predicting changes in effectively used skills after binning occupations into six quantiles according to the cognitive skill fraction of the occupation in 2014.

5.1 Worker Occupation Mobility is Constrained by Skill Polarization

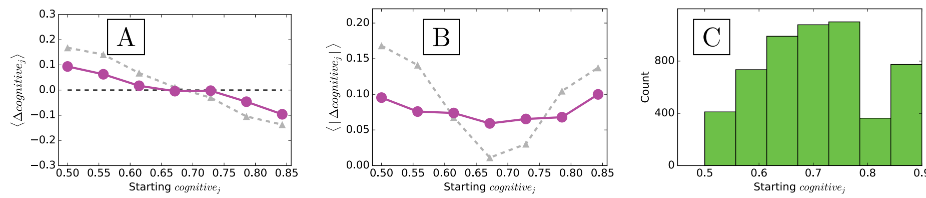

Figure 14: Binning individual worker occupation transitions into evenly spaced bins according to starting $cognitive_j$, we provide the **(A)** expected change in cognitive skill fraction and **(B)** the expected magnitude of change in cognitive skill fraction resulting from the occupation change. For both (A) & (B), standard error bars are plotted, but are negligible. **(C)** The number of observations in each bin.

Here, we investigate the changes in cognitive skill fraction (i.e. $cognitive_j$) associated with each occupation transition in the CPS dataset. Recall that occupations with higher $cognitive_j$ tend to have higher annual salaries (see Fig. 2B of the main text), but we do not find that workers necessarily increase $cognitive_j$ when they transition between occupations. Clearly, workers do not have complete freedom in selecting their next occupation.

Therefore, the task is to identify feasible null models that might explain the trends we observe in occupation transitions. First, we consider the case where individual workers select their new occupation at random with probability proportional to the number of workers of each occupation according to national labor statistics. In this case, the expected cognitive skill fraction of the newly obtained occupation is $cognitive_j = 0.70$. Figure 3 C&D in the main text demonstrate the average change and the average magnitude of change in cognitive skill fraction after binning occupation transitions into quantiles (i.e. evenly populated bins) in comparison to the null model. Figure 14 is the analogous plot using evenly spaced bins for the $cognitive_j$ of the original occupation in each occupation transition. In both cases, we find that workers transitioning away from occupations with low cognitive skill fraction tend to transition into occupations with higher $cognitive_j$, and workers transitioning away from occupations with high cognitive skill fraction tend to transition into occupations with lower $cognitive_j$, but these transitions appear to be smaller than we would expect given random occupation selection. On the other hand, we find that the magnitude of change in cognitive

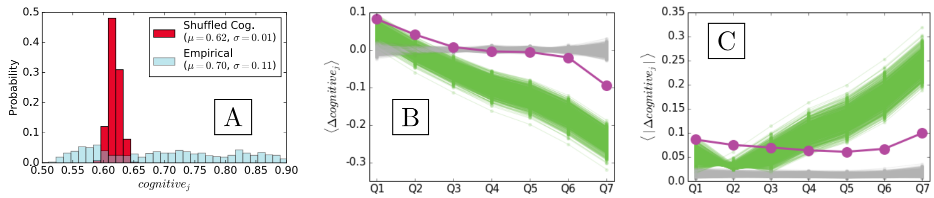

Figure 15: The effects of randomly selecting "cognitive skills" as a null model alternative to Louvain community detection. For 1,000 trials, we randomly select skills to be considered as cognitive skills (denoted C^*) and calculate the updated cogntive skill fraction, $cognitive^*_j$, for each occupation. **(A)** The $cognitive^*_j$ of one randomization (red) is provided for comparison to the empirical distribution of $cognitive_j$ (blue). For each cognitive skill randomization, we calculate the **(B)** expected change in $cognitive_j$ and **(C)** the expected change in $|cognitive_j|$ using C^* for the cognitive skill fraction of just new occupations (green) and for both new and old occupations (grey) from CPS data, and compare to the empirical results (purple).

skill fraction (i.e. $\langle | \Delta cognitive_j | \rangle$. See Fig. 14B) is maximized relative to the null model for workers transitioning away from middle-quantile occupations.

Combined, these results suggest constraints on worker mobility that we would expect given the polarized skill network. Specifically, we expect workers transitioning away from low (high) $cognitive_j$ occupations to obtain occupations with higher (lower) $cognitive_j$ just by chance due to the bounded nature of $cognitive_j$. However, our results demonstrate that the mobility for low and high $cognitive_j$ workers is actually more constrained than we would expect from pure randomness. This constraint is a result of the divide between sensory-physical skills and socio-cognitive skills we observe in the skill network. Conversely, workers transitioning away from occupations relying on skills that straddle this skills divide (i.e. middle $cognitive_j$ occupations) have greater ability to transition to either low or high $cognitive_j$ occupations because of their privileged starting location according to the Skillscape structure. Thus, we conclude that occupation transitions are not achieved through randomly selecting a worker's next occupation, but, rather, these transitions are constrained in a way that makes sense given the Skillscape structure

It remains to show that our results about occupation transitions are not an artifact of how cognitive skills are selected empirically. We carry out 1,000 trials of randomly selecting skills to be "cognitive skills", and we measure the change in cognitive skill fraction when the new occupation's $cognitive_j$ is calculated from the randomized skill selection and when both the old and new occupations' $cognitive_j$ are calculated from the randomized skills. Figure 15A demonstrates that the resulting distribution of $cognitive_j$ across occupations after randomization is very narrow by comparison to the empirical distribution of $cognitive_j$. Also, Fig. 15 B&C demonstrate that the changes in $cognitive_j$ and $|cognitive_j|$ that we observe empirically for occupation transitions is not well-captured by this null model either. These results suggest that the selection of cognitive skills is not a result of randomness and reveals something intrinsic about how individual workers move about the Skillscape. We conclude that the Skillscape and $cognitive_j$ captured something meaningful about workers' abilities and their freedom to change occupations.

5.2 The Trimodal Employment Distribution over $cognitive_j$ is Robust

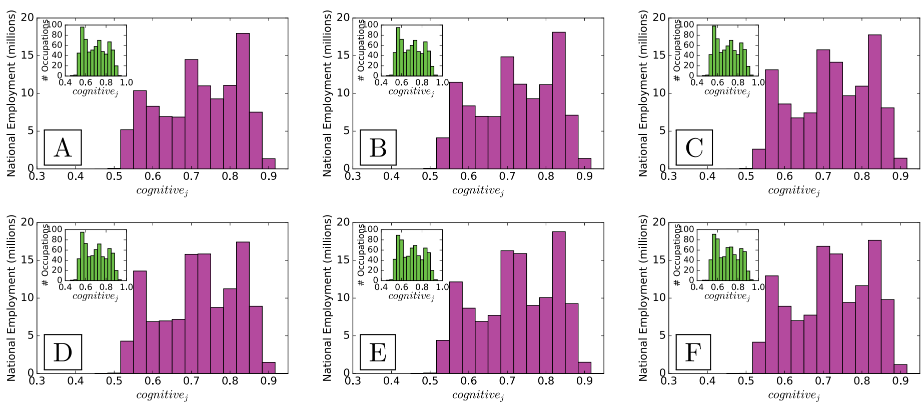

Figure 16: The distribution of national employment, and of individual occupations as an inset, after binning by $cognitive_j$ in **(A)** 2010, **(B)** 2011, **(C)** 2012, **(D)** 2013, **(E)** 2014, and **(F)** 2015.

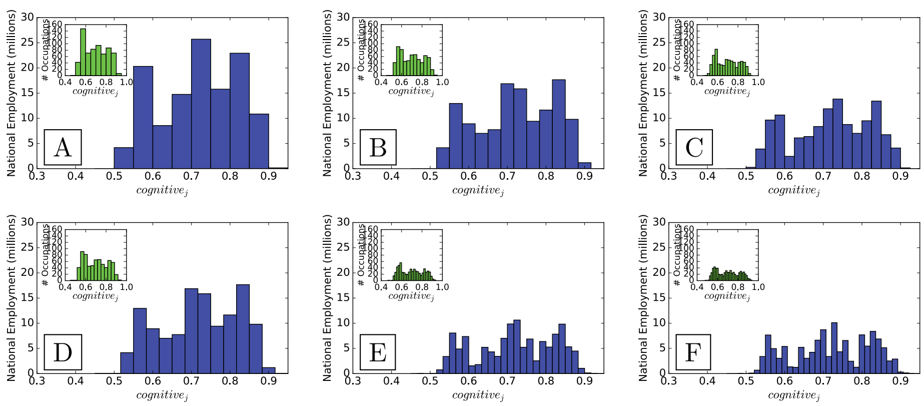

Figure 17: The distribution of national employment in 2015, and of individual occupations as an inset, after binning by $cognitivce_j$ while varying the number of bins between **(A)** 10, **(B)** 15, **(C)** 20, **(D)** 25, **(E)** 30, and **(F)** 35 bins.

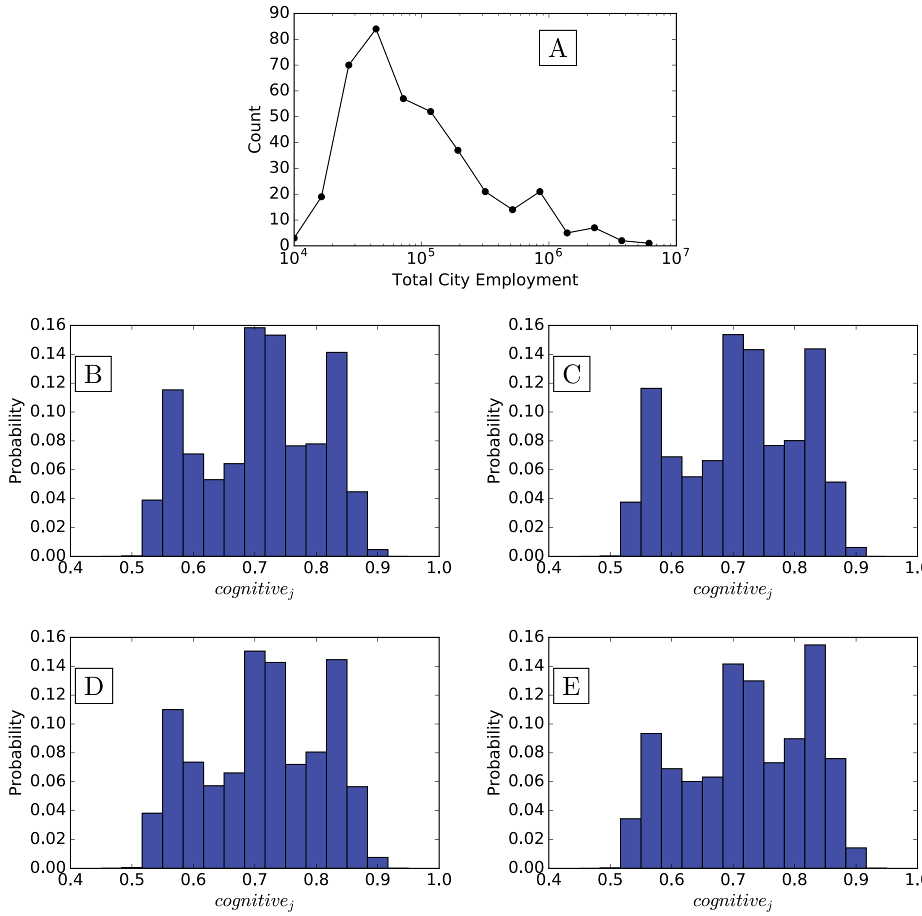

Figure 18: **(A)** The distribution of total employment in cites (i.e. metropolitan statistical areas). The distribution of city-level employment binned by $cognitive_j$ after dividing cities into quartiles according to total employment in the city: **(B)** total employment between 7,000 and 45,000, **(C)** total employment between 45,000 and 84,000, **(D)** total employment between 84,000 and 2,000,000, and **(E)** total employment between 2,000,000 and 8,900,000.

5.3 Occupation Transitions of Urban Workforces

Analogous to our investigation of individual worker occupation transitions and occupation redefinition, we ask if changes in the effectively used skills of urban workforces is predicted by complementarity of skills. To this end, we identify the effectively used skills of each city in 2010 and predict from the remaining skills which ones will become effectively used in 2015. Again, we compare *onet*, *rca*, and Skillscape proximity measures while varying the RCA threshold (see Fig. 19A). The first observation is that the RCA values of skills in cities do not change much from 2010 to 2015 (see Fig. 19B). This is not to say that urban workforces do not change their constituent workplace skills, but, rather, it may take longer time-scales (i.e. longer than 5-years) for appreciable labor dynamics to present themselves in urban workforces. All the same, Skillscape proximity narrowly outperforms the other measures for most choices of RCA threshold; however, if you allow for the $1 - onet$ predictor (indicated by pale green in Fig. 19A), then this method tends to outperform Skillscape proximity. Although the $1 - onet$ method may be reasonably predictive, we suspect that all results presented in Figure 19 are not conclusive due to the small amounts of observable dynamics in the constituent skills of urban workforces at the five year time-scale.

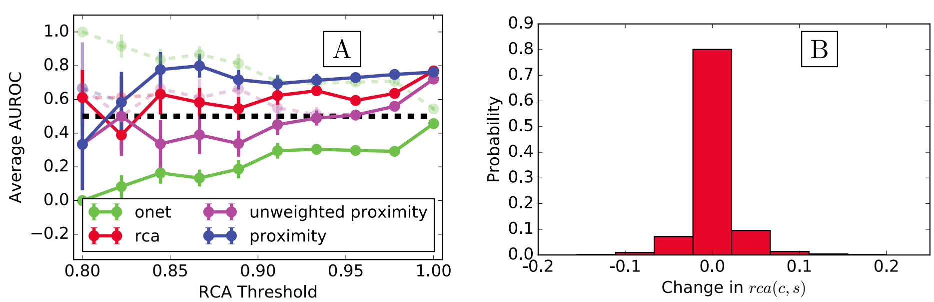

Figure 19: **(A)** Predicting the effectively used skills of urban workforces in 2015 from the effectively used skills of urban workforces in 2010. **(B)** The distribution of changes in $rca(c, s)$ comparing 2010 to 2015.