Research

Author for correspondence:
Alex Rutherford
e-mail: alexisadams@gmail.com

The dynamic resilience of urban labour networks

Xiangnan Feng and Alex Rutherford

Centre for Humans and Machines, Max Planck Institute for Human Development,
Lentzeallee 94, Berlin 14195, Germany

XF, 0000-0001-8736-5018

Both cities and markets are well understood as complex systems which are amenable to analysis using physically inspired methods. Cities have shown fascinating universality with size, while labour markets modelled as networks have considerable explanatory power. Labour markets are a particularly attractive domain of study in this context due to societal importance, the influx of high-resolution data as well as exogenous influence of automation. While much previous work has studied the economic characteristics of cities as a function of size and examined the exposure of urban economies to automation, this has often been from a static perspective. In this work, we examine the diffusive properties of labour markets and examine their variance across cities. More specifically, we identify the occupations which are most important in promoting the diffusion of beneficial or deleterious properties. To this end, we propose a new measure of node centrality empSI. We find that these properties of influence vary considerably with city size.

1. Introduction

Both cities and markets have a long history of study within complex systems. Cities have been described as agglomerative processes [1] and as universally scaling systems [2]. Labour markets, the allocation of workers with skills to tasks required to be performed by organizations, has led to the modelling of labour markets, skills and jobs as complex networks [3–9].

These two classes of system are, however, closely intertwined, suggesting a rich set of diffusive behaviours varying on urban substrates. This is attributable to locations advantageous for commerce, leading to economies of scale and increasing returns [10,11], and more recently encouraging the urban migration of workers to convene in centralized workplaces [12,13]. However, setting labour policies within urban levels (or indeed nationally or regionally) in an optimal way is both critical for prosperity yet challenging in the technical perspective [14].

With the development of complexity system, various network models have been applied on large datasets across various

domains. Among these new trends, economies have been widely studied involving geographical features, skill demands, labour markets and economic profiles. Hosseinioun *et al.* analysed the skill portfolios to build the skill hierarchy network to study the latent structure behind the skill dependencies, diversification and specialization [15]. Neffke *et al.* built labour flow network to analyse the role of firms and entrepreneurs and their structure changes in different geographic and industrial divisions [16]. Hidalgo *et al.* used the trading data among countries to build the product space to study the patterns of products and adaption in different countries [17]. These frameworks are meaningful for various levels of economic agents: for policy-makers, finding out how to develop competitive products and services is one of their most crucial tasks; for firms, realizing their roles in local and global regions could help them extend their business both in geographical and industrial sectors; for employees, understanding the skill demands and industrial structure shifting will help make successful career plans and transitions.

Dynamic models and spreading patterns on networks have been a classic and important research topic in complexity science. Opinion dynamics, which represent the phenomenon of system dynamics created by local interaction among individuals, including group decision-making, marketing and spread of fake news, have received a lot of attention, with many concepts and models emerging [18]. With new datasets of human mobility collected from sensory devices and surveys [19], various models have been developed to track and predict the people's movement and contagion events, including the spread of COVID-19 in recent years [20]. From the perspective of theory, to avoid the limits of classical networks, dynamical processes and percolation transitions on more complex systems like higher-order networks [21] and multiplex networks [22] have been studied to explore the phenomena like synchronization and consensus formation. These studies have pushed the boundary of research on network dynamics with many applications.

Although networks have been widely applied on economic datasets to analyse the labour markets, much previous work has focused on static properties of labour networks. However, labour markets are dynamic with flows of workers between jobs, of working practices between organizations and automation within firms and industrial sectors. Meanwhile, networks are natural representations for, not only how workers move between jobs that are mutually accessible based on skills, but also for diffusive processes by which technology spreads [23]. Little research work has attempted to apply the spreading models and theories on the economic data to study the labour market landscapes. Therefore, in this work, we focus on the resilience in an urban job network based on its network structure amenable to diffusion. This view is consistent with a large body of work examining the role of relatedness in the study of economic complexity [24]

We wish to make a strong link from our physically motivated modelling to implications that might be relevant for policy-makers. From this perspective, there are a number of attributes that might be expected to diffuse on a network of occupations. Some of these might help build an optimal environment, e.g. a gender-balanced workforce or safer and healthier working conditions, while some might be unfavourable, e.g. skill-based technological changes leading to the displacement of human labour. However, with knowledge of the degree to which a local job network will be able to promote or constrain diffusion and the nodes which are most influential in this network, policy-makers can more effectively drive the labour market to a desired state. For example, the adoption of company-wide policies protective of children and young people has been found to preferentially occur along network connections defined by supply chains [25].

In this manuscript, we focus on the mechanisms by which new technology diffuses (although this is amenable to the diffusion of other occupation-based norms). As shown in figure 1, when the workplace tasks of truck drivers are exposed to technologies such as autonomous-driving systems, this automation could spread preferentially to a related job, e.g. tractor operators. Subsequently other occupations similar to tractor operators with respect to skills could also be affected. However, this spreading phenomenon would not be limited to physical jobs. Image recognition software can assist radiologists diagnosing diseases in medical imagery, while the same technology could possibly be adopted by nuclear engineers to detect the operation of nuclear power plants; then engineering technologists could apply similar software onto other instruments to gain more powerful tools.

We wish to emphasize that the spread of technology might be beneficial or deleterious depending on the technology itself, as well as the stakeholders. A piece of technology might displace human labour completely leading to unemployment, but benefit a firm which is able to lower its wage bill. Likewise, a new piece of technology might well enable increased productivity in occupations depending on the efficiency of the technology in question [26].

Our contributions in this work are focused on providing a novel understanding of the structure and resilience of urban labour markets. More specifically, we (i) propose the employment-weighted spreading
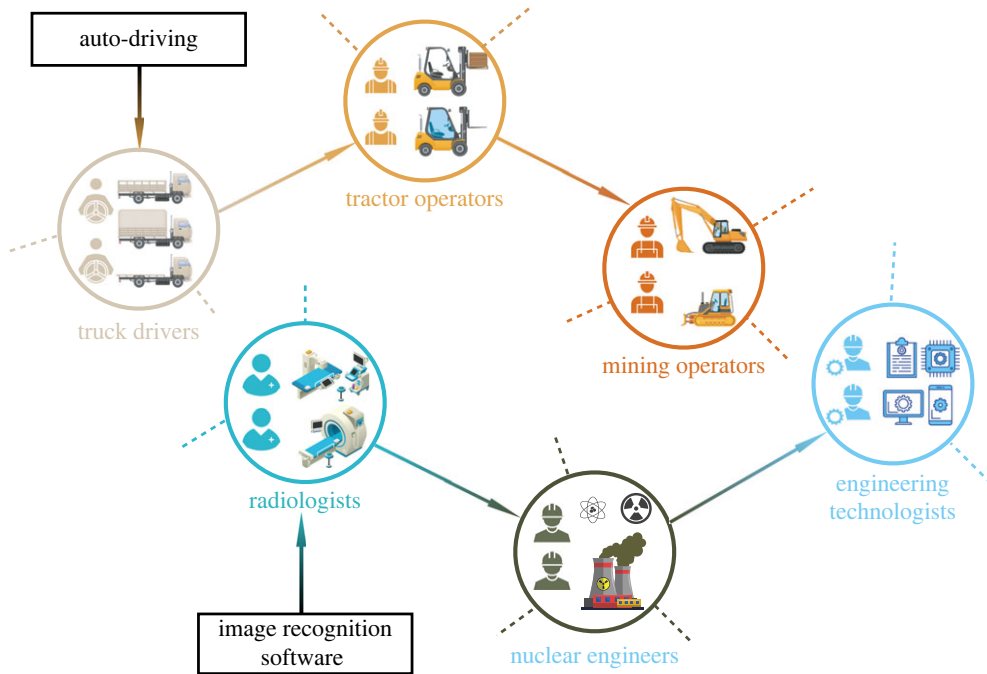
Figure 1. An illustration of workplace technologies spreading on an occupational network. Autonomous driving capability is first introduced for truck drivers, which subsequently spreads to impact tractor operators and then mining operators due to the similarities between these occupations. Likewise image recognition software introduced for radiologists spreads to nuclear engineers and engineering technologists. Not all occupations are drawn in this diagram; only those lying on one illustrative diffusive pathway.

influence, a new measure of a node's influence in a job network, (ii) we quantify the efficiency of diffusion based on a seeding strategy on this basis, (iii) we investigate how the cognitive and automatable nature of influential jobs changes with the size of the urban economy, and (iv) we uncover the most influential occupations, those that would optimize the efficacy of targeted policy interventions, across national urban centres.

2. Methods and data

2.1. Network construction

Firstly, occupation networks in each city are built. In the network, each node represents one occupation and each edge represents the similarity between the two occupations whose weight is given by skill similarity. We use the Jaccard similarity [27] to quantify the similarity between occupation i and i',

$$\text{SkillSimilarity}(i, i') = \frac{\sum_s \min(\text{onet}_{i,s}, \text{onet}_{i',s})}{\sum_s \max(\text{onet}_{i,s}, \text{onet}_{i',s})}, \tag{2.1}$$

where $\text{onet}_{i,s}$ is the importance of skill s to occupation i. High skill similarity between jobs suggests workers make transitions more easily between them.

Since this similarity value between any occupation pair is not zero, a complete network will be built containing 774 nodes and 299 151 edges, which is the global network composed of all occupations in the USA. In each city, its occupation network is built in the same way with occupations whose employment is non-zero locally, which means each local occupation network is a complete sub-network of the global one.

2.2. Employment spreading influence

With a view of understanding the dynamic behaviours among various individuals on networks, a number of theories studying the interactions between nodes have been considered. These operate globally and locally to model the spreading processes, which usually focus on information [28] and epidemic spread [29]. A key issue is to find the 'super-spreader', namely the nodes with highest

spreading abilities. Controlling these super-spreaders, that is to seed with them or isolate them purposefully, could lead to optimal spreading or immunization, respectively [30].

The spreading also happens on the labour markets and occupation networks. New technologies and skills from one occupation can potentially diffuse into other occupations, since new use cases beyond the original promote the introduction and adaptation of existing inventions. For example, the recording of speech was firstly invented by Edison to record the last words before people dying and books for blind persons, then it was adapted for office work. But even though Edison himself did argue against the use of this invention to record music, after around 20 years the recorders have totally revolutionized the whole music industry. If we view the skill similarity-based network from the perspective of spreading, the super-spreaders, namely the most influential occupations, would present the greatest ability to spread technologies or norms to the whole labour market. These occupations play crucial roles in leading and shaping the landscape of labour market. On one hand, if starting with them, new technologies improving efficiency could be adopted faster, which increases the productivity and economic growth. On the other hand, automation of skills and jobs could spread, leading to potentially negative effects such as short-term displacement of workers [31].

From this perspective, obtaining the most influential occupations could help regulate the whole labour market. For policy-makers, understanding which occupations are influential and making policies targeting these occupations and related skills could lead the labour market to more robust states with favourable economic outcomes.

Finding the optimal spreaders belongs to the non-deterministic polynomial (NP)-hard [32,33] problems generally. Interactions at different topological levels need to be studied [34,35] and multi-scale features tangling together increases the difficulty. It was studied that high-influence nodes are not always those whose degrees are highest [36]. By applying the cavity method [37] on this weighted network case [38], we define the spreading influence of occupation i in city c

$$S_i^c = \sum_{j \in N(i)} a_{ij} \sum_{k \in N(i), k \neq j} a_{ik}(d_j - 1), \tag{2.2}$$

where $a_{ij} = \text{SkillSimilarity}(i, j)$ is the edge weight between node i and j, d_j is the node degree. The summary and derivation with cavity method can be seen in appendix A.

Besides the network structure and skill similarity, the number of workers at each occupational node in the network is highly inhomogeneous, which has a strong influence on the labour market as a whole. Occupations with more employees should always be preferentially targeted by policies, since when new technologies are applied, they will have a large impact on the local or global labour market directly, either in a good way or in a bad way. Thus, when measuring the influence of occupations, we take into account the employment of occupation i in city c, denoted as emp_i^c, to get the employment spreading influence

$$\begin{aligned} \text{empSI}_i^c &= \log(\text{emp}_i^c) S_i^c \\ &= \log(\text{emp}_i^c) \sum_{j \in N(i)} a_{ij} \sum_{k \in N(i), k \neq j} a_{ik}(d_j - 1). \end{aligned} \tag{2.3}$$

For each city, the labour market is described by its occupation network and the empSI of each occupation in that city is calculated. Occupations with high empSI suggest high influence on the local market in the perspective of spreading, which has a contribution due to both their structural spreading abilities or large employment numbers. Although the spreading models and the 'super-spreader' searching have been widely applied on various fields like epidemic and information spreading, it is the first time to consider and apply the percolation theory on the occupation networks. It is the first time to view the labour market in the respect of spreading to track and understand the dynamics of human resource and technologies. Comparing with other labour market research by statistical inference, this fresh perspective could help understand the complexity of labour market networks and other crucial ideas like regional relatedness [39]. Various stakeholders could benefit from these findings.

2.3. Data

In this research, we consider data from O*NET [40] and US Bureau of Labor Statistics [41]. Although similar structured data are increasingly available, e.g. in the EU [42], the use of O*NET is common in network studies of labour, and given the high quality of this data broken down by urban area, we
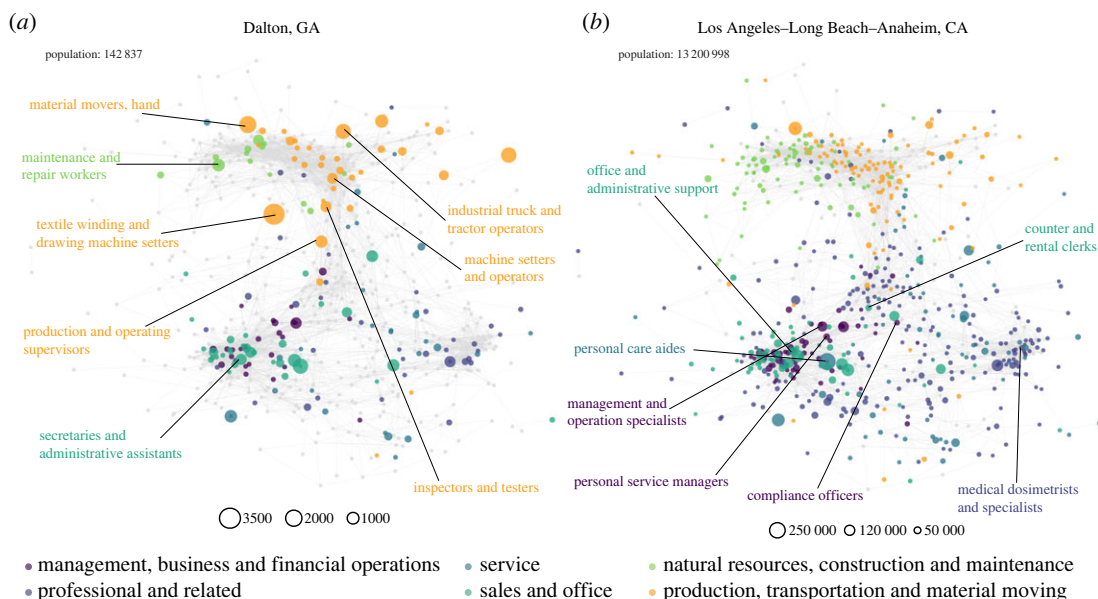
Figure 2. Two occupation networks of (*a*) Dalton, GA and (*b*) Los Angeles–Long Beach–Anaheim, CA. Nodes stand for occupations and edges stand for skill similarity between occupations. Node colours stand for sectors, and sizes indicate the employment numbers. The grey node indicates that this occupation does not appear in this city. Edges with weights higher than 0.7 are drawn, while the complete graph is applied in calculating. The labelled occupations are the different occupations in the two cities' top-20 highest empSI occupations.

proceed with this data. From this structured data, we build networks for each urban area as well as the whole nation. All the analyses are based on the built networks.

The O*NET data contain a set of standardized occupations and, for each occupation, a weighted importance against a set of standard skills. These data are compiled manually from expert interviews based on the skills that workers report. We consider the data from 2011 to 2020 as the standard occupational classification (SOC) system is consistent during this period. Thus the data we get includes 774 occupations globally and 120 items referred to as 'skills' including ability, knowledge and skill sections in the datasets are used to build networks. We use two-digit and six-digit SOC classifications for analysis.

The US Bureau of Labor Statistics provides employment data including occupations, employment numbers and salaries in each Metropolitan Statistical Area, commonly referred as a 'city'. Around 380 cities are selected in the analysis. Not all occupations appear in every city, ranging from 93 to 733 occupations in each city. We build networks for each city based on the occupations found in it (as described in more detail below).

Other data include population data from US census [43] and occupation automation probability data from Frey & Osborne's and Webb's research [44,45]. Details about the automation data are introduced in appendix D. The occupations in different data sources are indexed by SOC codes.

3. Results

In this manuscript, we investigate the spreading processes of the labour market in each city. More specifically, we wish to identify the most influential occupations in each city network and determine how this relates to city size. For each city *c*, the $empSI_i^c$ values of all occupations that are represented in the city workforce are calculated and those occupations with highest values are considered the most influential.

To begin, we compare the job networks from two cities with very different sizes: Dalton, GA (low population) and Los Angeles, CA (high population) in figure 2. The nodes (occupations) are linked by skill similarity (edges). We consider only the jobs that are found in each city, therefore the set of occupations (number of nodes) is typically different among cities. Each occupation has variable numbers of workers on that occupation that is represented by the size of the corresponding node. A selection of the most influential occupations are marked in each case. A distinction can be made between the two networks: in Dalton, the most influential occupations concentrate on the production and construction sectors, while in LA, the most influential ones belong to services, sales and clerical sectors.

To verify whether empSI meaningfully represents influential occupations, a simple spreading simulation within the susceptible-infected framework is conducted on the local networks of the two
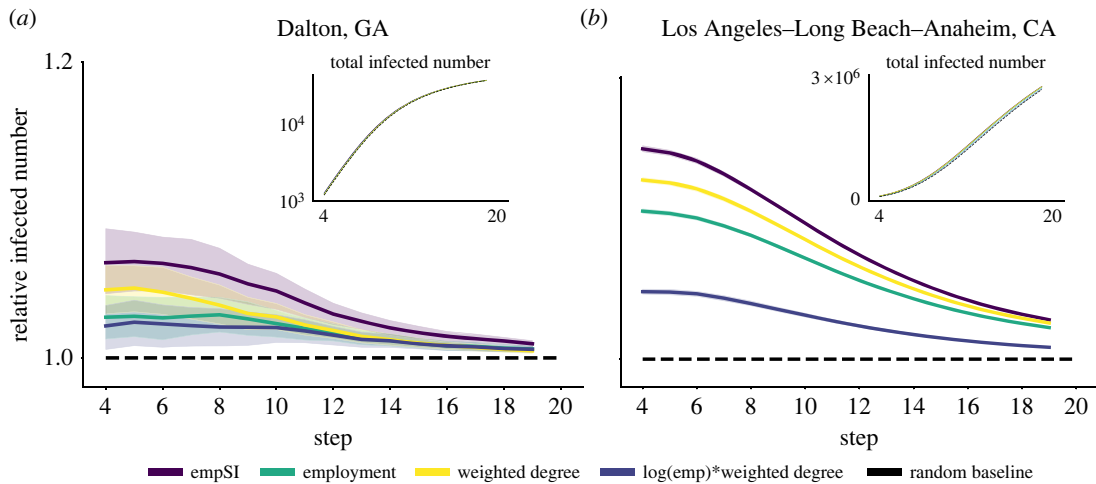
Figure 3. Simulation of a spreading process on the occupation networks of (*a*) Dalton, GA and (*b*) Los Angeles–Long Beach–Anaheim, CA. All results are the averaged values of 20 trials. In each simulation, 0.2% random agents of employment are selected by different strategies including highest empSI, highest employment numbers, highest weighted degrees and highest log(emp) × weighted degree as the starting seeds for spreading. Each node in the networks contains the same agent number as employment. At each step, each agent infects one of its neighbours with the probability given by their skill similarity. A strategy of random seed selection is applied as a baseline. The curves are the relative infected number and inside is the total infected number. The confidence intervals in Los Angeles are very small to distinguish, due to its large population.

cities above (see model details in appendix B). Figure 3 presents the spreading processes starting with seed agents selected by different strategies. The random selection of the seeds is applied as a null model. As shown in the figure, seeding based on the empSI value leads to a faster spread than seeding at random or based only on employment or degree. These results show that with the proposed index, the most influential jobs could be found more accurately. These occupations work as the spreading hubs in the networks and could transit human sources and technologies faster to the whole labour market. They are functionally crucial for a healthy economic landscape and city development, which deserve more attention from various levels of stakeholders and policy-makers. As a verification, we made another spreading model analysis with susceptible-infectious-susceptible (SIS), which is given in appendix B.

3.1. Diversity and automation

Using the proposed empSI, a number of influential occupations in each city could be found, which usually are regarded as the most important and significant ones according to network centrality studies. In this section, based on the findings from proposed index, we investigate the relationship between the characteristics of an occupation: the cognitive nature and exposure to automation, and the occupation's importance within a city job network.

For each city, we compute the correlation between all occupations' empSI values and cognitive scores, which is defined as the fraction of cognitive skills for each occupation [3], namely it is calculated by $\sum_{s \in \text{cognitive}} \text{onet}_{i,s} / \sum_{s} \text{onet}_{i,s}$. In smaller cities, we find little correspondence between the two measures; in this case, the more influential occupations in small cities are just as likely to be cognitive or non-cognitive, whereas in larger cities, we find a reasonable correlation ($\rho \approx 0.3$) between these measures. This is demonstrated by a clear positive trend between city size and the job-wise correlation in cognitive score and empSI ((r, p) = (0.32, <10^{-9})). Thus we expect that any occupational characteristic presented in more cognitive jobs, whether positive or negative, will spread more efficiently to other occupations in larger cities. Interestingly, in contrast to many economic indicators [10], we found no significant correspondence to a super- or sub-linear scaling with population.

We note that this trend of increasing influence of cognitive jobs with city size also coincides with an increasing diversity of job influence, both with respect to city size and time. We calculate the Shannon entropy [46] of empSI in each city. Figure 4*b* demonstrates that the entropy of node influence across all occupations found in a city increases with population. Further, the level of diversity appears to be generally increasing between 2011 and 2020. Taken together, these results suggest that in larger cities, the correlation between the cognitive nature of jobs and their influence is not driven by a small number of highly influential cognitive jobs. Rather, the increasing uniformity in empSI, both over time and with city
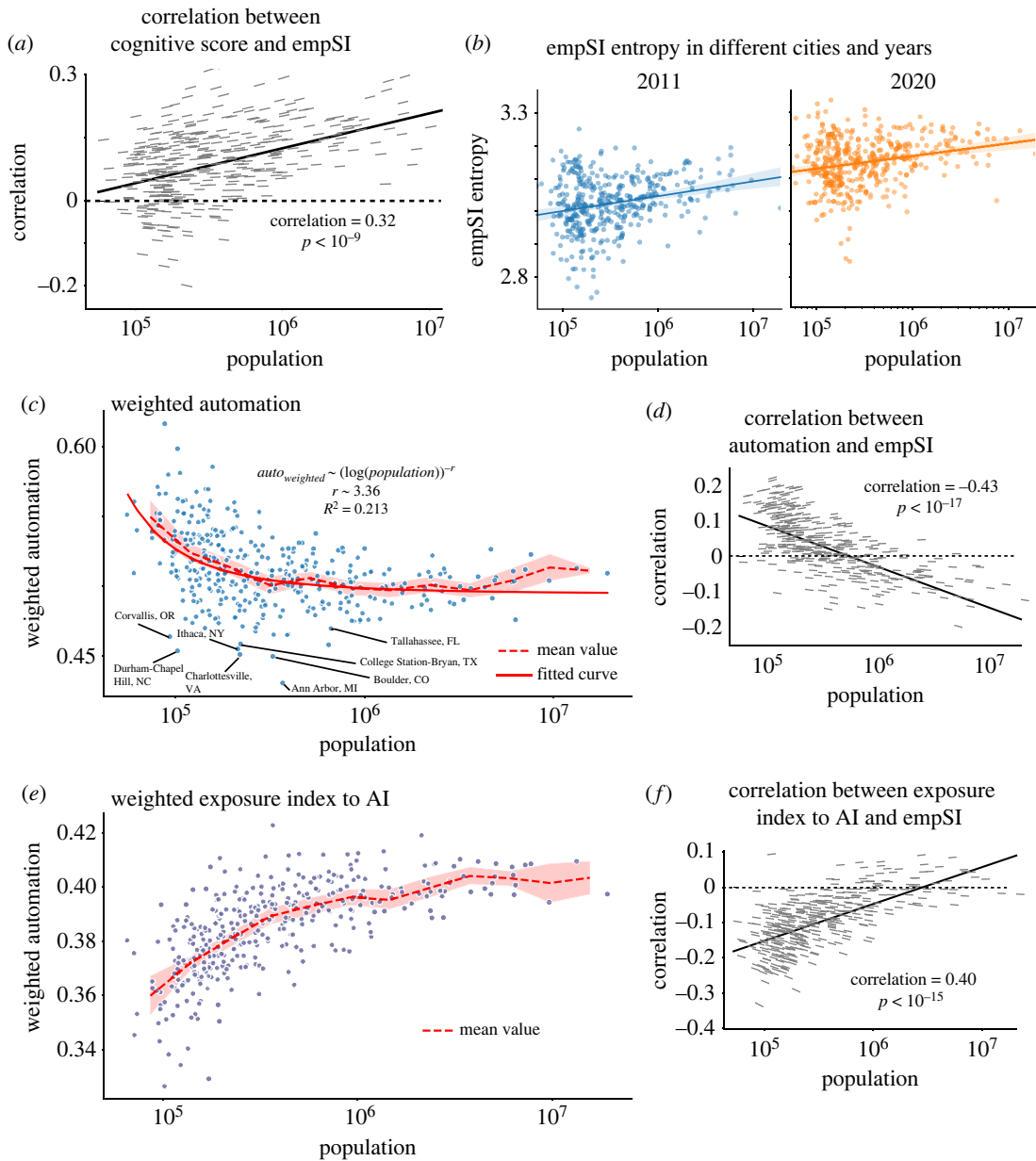
Figure 4. (*a*) Correlations between cognitive score values and empSI against populations in each city. The Pearson correlation coefficient is marked. Different slopes stand for different correlation values. (*b*) The empSI entropy of years 2011 and 2020 against populations. For each city, occupation empSI values are divided into same number of bins and their entropy is calculated. (*c*) Weighted automation probability by empSI against populations. The mean value is the average value in each bin, which is log-scale by populations. A log-power fitting is implemented on the data. On the left bottom, a group of college towns are labelled, featured by small-scale population and low-weighted automation probability. (*d*) Correlations between automation and empSI against population. (*e*) Weighted exposure index to AI from Webb's research by empSI against populations. (*f*) Correlations between exposure to AI and empSI against population.

size, suggests that the choice of occupations for effective targeted policy interventions is not trivial (see a similar result related to spreading influence S and cognitive score in the appendix).

Next, we consider whether these trends in increasing diversity and more influential cognitive jobs in larger cities lead to increased *resilience* to automation specifically. In figure 4*c*, we evaluate the city-wise exposure to automation as measured by Frey & Osborne [44], weighted by empSI as below

$$\text{auto}^c_{\text{weighted}} = \sum_i \frac{\text{empSI}^c_i}{\sum \text{empSI}} \times P_{\text{auto}}(i), \tag{3.1}$$

where $P_{auto}(i)$ is the automation probability of occupation i from Frey & Osborne [44]. This value gives the probability of a job being susceptible to computerization, namely the risk of replacement from machines (see details in appendix D). The higher the value, the higher the replacement risk.

In common with previous findings [47], we see a decreasing exposure to automation with city size increasing when weighted by empSI: smaller cities face greater risks of automation. Meanwhile, at the left-bottom part of the figure, we observe a group of outliers, mostly corresponding to college towns including Ithaca, NY (Cornell University), Ann Arbor, MI (University of Michigan), Charlottesville, VA (University of Virginia) and Durham–Chapel Hill, NC (Duke University and North Carolina Central University). This phenomenon is consistent with the intuition that these college towns should be under low automation risk.

Further, the occupations that are most exposed to automation are less influential in larger cities. This result goes beyond a static picture of a city's overall automation risk as measured by its present workforce. Rather, this index measures the degree to which an occupation is able to spread its own exposure to automation to other related occupations. Although this trend is relatively weak (from $\rho \approx 0.2$ in the smallest cities to $\rho \approx -0.2$ in the largest cities), it is consistent across all cities $((\rho, p) = (-0.425, <10^{-17}))$.

However, we find that this result is extremely sensitive to the exact measure of automation exposure. In figure 4d, we present the same weighted automation measure using data from [45], an index of exposure to AI specifically, which measure the overlap between job descriptions and pattern description related to AI. By contrast, we find the opposite trend in the relationship between the urban exposure to automation and population. Specifically, the weighted automation exposure has a positive correlation with city size $((\rho, p) = (0.40, <10^{-17}))$. Similar calculations are repeated based on other automation data [45,48], see appendix D.

The exposure of an occupation to automation is notoriously difficult to quantify, and the measures presented here all differ in what is measured. Here, we principally compare Frey & Osbourne's [44] results as the most established measure on the one hand, and Webb's [45] data as a more recent measure based on the similarity between occupation task descriptions and patent documents. Webb's measure offers the key advantage of being validated on historical changes in workforce numbers, and is also broken down into three categories of AI, software and robotics.

Our results suggest that the cognitive jobs found preferentially in larger cities are more amenable to automation through the deployment of AI technology (as opposed to robots or software). AI technology can be more easily deployed in work environments where computers and data infrastructures are already common as well as through flexible cloud computing resources. The fact that no correspondence is found between city size and equivalent measures of automation exposure through robots and software support this.

In conclusion, results from our analysis suggest that in large cities, high-cognitive jobs tend to have higher influence on the labour markets. This conclusion agrees with the finding that larger cities tend to rely more strongly on high-cognitive skills with higher household income and annual wage [3], meanwhile larger cities always lead the development of economy and smaller cities follow the same pathway [49]. Also in larger cities, the labour markets present higher diversity, which is in agreement with the finding that there is less specialization in larger cities [50]. In respect to automation trend, urban labour market resilience has a nuanced relationship with city size and depends sensitively on the nature of the occupational automation risk. This relation is complicated since the influence of automation on labour market has been widely studied yet is still sophisticated [44]. Recent research suggests that in larger cities their higher job connectivity and diversity may bring higher resilience than in smaller ones [51]. We can conclude that larger cities are more exposed to AI-based automation technologies and that cognitive jobs will be able to diffuse occupational characteristics, possibly through targeted policy interventions, more efficiently. Still, AI-driven technologies are new trends [52] and history data may not help us capture and predict the future of labour markets accurately [53].

3.2. Landscape of US labour markets

Previous work has shown that the industry structure varies geographically [8,54,55]. For different cities, is there any pattern or particular combination of the influential occupations? What do the most common influential occupations and uncommon ones look like across cities?

To find the common pattern of influential occupations, in each city, we sort all active occupations by their local empSI and for each occupation its average rank is calculated across the whole nation.
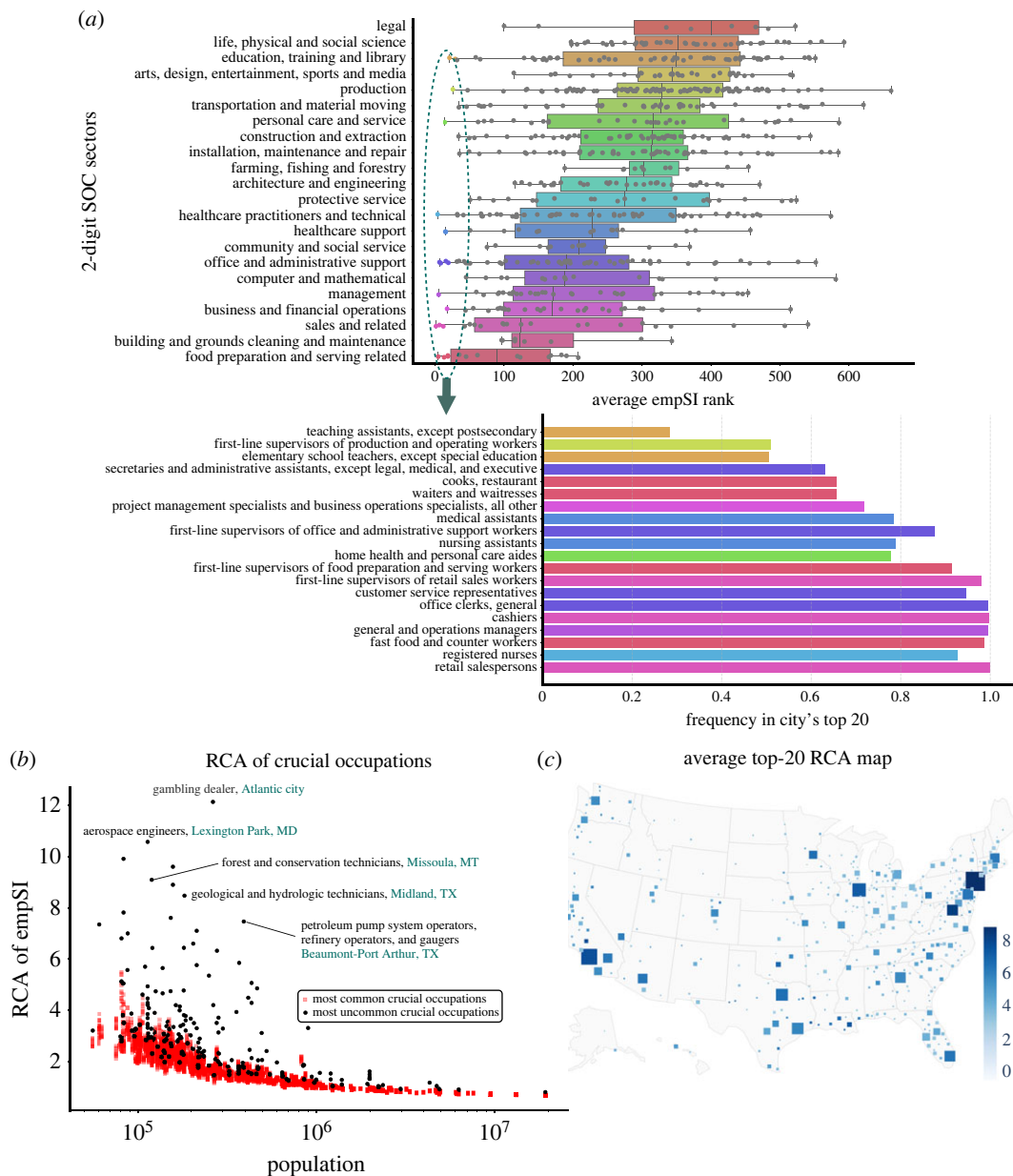
Figure 5. (a) Box plot of average empSI ranks of occupations in each sector. The two-digit SOC code is applied to divide occupations into different sectors, and different colours stand for different sectors. For each occupation, its empSI in every city is calculated and the average empSI rank is calculated. We select the 20 highest average rank occupations (marked by colourful nodes in the box plot) to count their frequencies of appearing in each city's top-20 empSI. (b) Revealed comparative advantage (RCA) values of most common crucial occupations and uncommon crucial occupations against populations. The crucial occupations are defined as the top-20 empSI ones in each city. Several occupations are labelled. (c) The map of top-20 empSI RCA occupations in all cities. Square size stands for populations and colour stands for the value.

We investigate the occupations with highest average empSI ranks by counting their frequency of ranking top 20 in local labour markets. As shown in figure 5a, we find that occupations in different sectors demonstrate different rank patterns. There are some occupations ranking among the highest in almost all the cities, suggesting that they are crucial and influential everywhere, which concentrate in sales, services, health and office sectors, like retail salespersons, registered nurses, cashiers and office clerks. Most of these occupations are regarded as basic life-supporting and economic-activity ones. Although some of them do not have a high essential score [56,57] and their automation risks vary [44,45,48], these occupations, due to both large employment numbers or high spreading abilities based on skill similarities, play influential roles across all cities. For policy-makers, these occupations are crucial since they are fundamental and prevailing for local economy, and global policies targeting these occupations may cause similar effects around the whole country.

While there are a group of common influential occupations across the whole nation, one might wonder what makes the labour market particular to the cities? Previous work has shown that features vary with the scale of cities, including wages, patent activity and characteristic industries [49,58,59]. Are the existing influential occupations specific to particular cities? To verify this, we calculate the revealed comparative advantage (RCA) [24] of occupation empSI in each city

$$\text{RCA}_i^c = \frac{\text{empSI}_i^c / \sum_i \text{empSI}_i^c}{\sum_c \text{empSI}_i^c / \sum_i \sum_c \text{empSI}_i^c}. \tag{3.2}$$

This index could help find the relative advantage of an occupation, namely occupations with high RCA_i^c will present high influence on local labour market and this influence is identical to this city—in cities other than c, these occupations are either playing negligible roles or not active.

We check the occupation frequencies in every city and obtain the most common and uncommon crucial occupations. In figure 5b, we present their RCA_i^c values. It could be observed that some occupations, featured by high RCA_i^c, are influential in only one or several cities, like gambling dealer in Atlantic City which is famous for gambling industry, aerospace engineers in Lexington Park in Maryland, where the NASA Goddard Space Flight Center is close, and geological and petroleum-related work in Midland and Port Arthur in Texas, where the oil sources were discovered. These high RCA_i^c occupations present strong geographical identities (see in figure 5c). Similar identities of which could be found in the map of job embeddedness [51], distribution of general skill scores [15], and expected job impact from automation in every city [47]. These geographically identical occupations deserve more attention from policy-makers, since they could play special roles across the whole nation, e.g. output special productions and deal with particular issues.

Our findings, in respect to higher diversity in larger cities, agree with conclusions from recent literature. Bettencourt $et\ al.$ claimed that the coexistence of greater individual specialization and increases in overall diversity is the optimal [50]. In our findings, both larger and smaller cities present specialization abilities in different ways: while in smaller cities geographical specialization features are performed by some specific high RCA_i^c value occupations, in large cities like New York and Los Angeles, high average RCA_i^c values of top influential occupations suggest there are also rare but crucial occupations. Combining the high exposure to AI in larger cities in our previous results, the AI may be the next great opportunity for them to keep the increase of diversity and economic vigour. Nevertheless, elucidating the intrinsic mechanisms and dynamics underneath the labour market diversity and specialization still requires further research. Some researcher found that the job connectivity is crucial for diversity and resilience [51], while network indices like nestedness is applied to study the dynamics of industrial ecosystem [60]. More research with all-around models and detailed analysis on new data could be expected in the future.

In summary, it has been shown from this empSI perspective that there are some occupations that appear and play crucial roles in every city, which have high average empSI ranks and usually focus on life and health sectors, like sales, services, health and office. They are supporting the everyday routine life in the global nation and prevailing in many professions. At the same time, there are some unique occupations in some cities, which are deeply related to the local economy structures and city geographical features like college towns and nature resource centres, being crucial and influential locally. They tend to have high empSI RCA values and deserve specialized attention and policies. These two aspects of crucial occupations, the globally crucial ones and locally crucial ones, make up the backbones of US labour market landscape.

4. Conclusion and discussion

The adoption of workplace behaviours, like many other social norms, can have positive and negative effects. In the context of workplace norms, the adoption of safer working conditions would be beneficial for workers whereas the adoption of new technology could potentially be a negative development for both workers and policy-makers. Given the many successes of describing complex social systems, such as labour markets, as complex networks, it is natural to consider the diffusive properties of labour markets. This provides an attractive opportunity for policy-makers to take advantage of this diffusive property to encourage (inhibit) the faster (slower) adoption of positive (negative) attributes on the network of occupations.

From the perspective of spreading, we have proposed an employment spreading index to measure the occupation influence on occupation networks. For each city, an occupation network is built based

on skill similarities between jobs and the most influential occupations are obtained by the index. We investigate the systematic effect of city size on the susceptibility of jobs. We find that cognitive jobs are consistently more influential as city population increases. Regarding the effect of automation specifically, we find that the relationship with occupational influence is more nuanced. The Frey & Osbourne [44] measure of exposure to automation suggests that larger cities will be more resilient to automation as the influential occupations are those with a lower exposure. Conversely, Webb's [45] measure of exposure to AI specifically suggests the opposite trend; larger cities have their more influential occupations more exposed.

Considering influential occupations on a global level, by aggregating the most influential occupations across all cities, we are able to rank occupations. We find that the top jobs tend to include an element of physical and socio-cognitive work e.g. nurses, cashiers and customer service representatives. These results suggest that these mixed-nature jobs should be given special attention by policy-makers when considering how to manage labour market change, whether technological in nature or not.

In this work, we have considered the occupation networks linked by the similarity of the tasks required. However, behaviours such as the adoption of automation technologies are able to diffuse on many network substrates; between the successive jobs of individual workers, between firms in common sectors in physical space or along supply chains. Considerable explanatory power has been found in novel combinations of these, e.g. geo-industrial networks [8] and supply chain networks [61]. It is likely that the true dynamics involve some combination of these into a multiplex network [62], on which the cooperative spreading of more than one social norm would diffuse [63,64]. More related studies of the diffusion mechanism on labour market networks could bring deeper understanding on them. A key challenge will be how to apply new strategies to capture these patterns, like research on high-order interactions [65,66] and structures including motif [67,68] and graphlet [69]. The occupation network structure deserves more explorations in the future.

Our study is constrained by the availability of public data. More research into the spreading processes and models deserve further attention. If the flows among occupations of specific technology or human resource could be accurately tracked, the trends of automation and skill demands would be studied much more clearly, which would tremendously help depict the landscape of labour market. This study relies on much more detailed data. Higher fidelity data would also allow for more careful validation of these findings. In our research, we make conceptual simplifications that the occupational network is static for the timescale for diffusion to take place. Implicitly, this means that we ignore the fact that the labour market is dynamic: new occupations are generated every year and for many occupations new skills and working contents are introduced [70,71].

Appendix A. Derivation of spreading influence

This result by 1RSB is assumed on sparse network [37,38], especially on locally tree-like networks where there are not many loops, yet it still could work on more dense networks. In our research, we applied the weighted networks to model the occupation networks and the conclusion still works. In many studies related to cavity method and message passing, it is common to extend the conclusions from locally tree-like networks to more complicated ones [73–75]. Usually, it is too complicated to do the derivations on networks with many loops, while results from tree-like structure networks provide a rigorous bound to results from networks containing loops [76]. Recently, there are

some studies attempting to remedy and rectify this problem [77,78], which deserve further exploration in future research.

By using message passing formula, the dynamical systems of spreading patterns on networks could be built. To guarantee a stable solution of this system, the largest eigenvalue of the corresponding linear operator which is represented using non-backtracking matrix [79] with directed links should be less than one. In this way identifying super-spreaders problem is transferred into solving the stable solution problems. To estimate the largest eigenvalue, the power method from numerical analysis [80] is applied to make calculations. After derivation, the largest eigenvalue could be estimated by each node contribution to it, which is regarded as the influence of node on the network. The larger its contribution, the greater its influence in spreading.

Under the sparse network assumption, for a network G with node set V and edge set E, its adjacency matrix A is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } v_j \text{ and are connected} \\ 0 & \text{otherwise.} \end{cases} \tag{A 1}$$

In a weighted network case, we use a_{ij} to stand for the weight of edge between node v_i and v_j.

The maximum spreading problem equals the optimal percolation problem [36], which is how to break down the major component in networks with minimal set of nodes removed. Let $c_{i \rightarrow j}$ denote the probability of node v_i belonging to the major component with node v_j removed. For a tree-like weighted network, this relation could be formulated as

$$c_{i \rightarrow j} = n_i [1 - \prod_{h \in \partial i, h \neq j} (1 - c_{h \rightarrow i} a_{ih} A_{ih})], \tag{A 2}$$

where $n_i = 0$ stands for node v_i being an optimal spreader and not if $n_i = 1$ and $h \in \partial i$ means the neighbours of node v_i.

Based on knowledge from dynamic system, the above system will have a stable solution if the largest eigenvalue of linear operation R is smaller than 1 [81]. Here, the R is a $2|E| \times 2|E|$ matrix and each row and column corresponds to a one-direction edge. The matrix R is defined as

$$R_{i \rightarrow j, k \rightarrow l} = \frac{\partial c_{i \rightarrow j}}{\partial c_{k \rightarrow l}} \bigg|_{c_{i \rightarrow j} = 0}, \tag{A 3}$$

where $i \rightarrow j$ and $k \rightarrow l$ stand for the edges between v_i and v_j and v_k and v_l with directions. R could be calculated by non-backtracking matrix M [79]

$$R_{i \rightarrow j, k \rightarrow l} = n_k a_{ij} M_{i \rightarrow j, k \rightarrow l}, \tag{A 4}$$

where

$$M_{i \rightarrow j, k \rightarrow l} = \begin{cases} 1 & \text{if } i = j, k \neq l \\ 0 & \text{otherwise.} \end{cases} \tag{A 5}$$

Thus, we have

$$R_{i \rightarrow j, k \rightarrow l} = n_k a_{ij} A_{ij} A_{kl} \delta_{jk} (1 - \delta_{il}), \tag{A 6}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{A 7}$$

In the equation above, A_{ij}, A_{kl} and δ_{jk} suggest that there is a path $v_i \rightarrow v_j \rightarrow v_k \rightarrow v_l$, and δ_{il} makes sure that it is a non-backtracking path.

Let $\lambda(\mathbf{n})$ be the largest eigenvalue of R where $\mathbf{n} = (n_1, n_2, ..., n_{|V|})$ indicates which nodes are selected as optimal spreader. We apply the power method [80] from numerical analysis to approximate $\lambda(\mathbf{n})$. Let $\mathbf{w}_0$ be a vector satisfying that it has non-zero projection on the direction of $\lambda(\mathbf{n})$'s eigenvector and let $\mathbf{w}_r(\mathbf{n})$ be

$$\mathbf{w}_r(\mathbf{n}) = R^r \mathbf{w}_0. \tag{A 8}$$

Then, according to power method, we could get

$$\lambda(\mathbf{n}) = \lim_{r \to \infty} \lambda_r(\mathbf{n}) = \lim_{r \to \infty} \left(\frac{|\mathbf{w}_r(\mathbf{n})|}{|\mathbf{w}_0|} \right)^{1/r}, \tag{A 9}$$

where

$$|\mathbf{w}_r(\mathbf{n})|^2 = \langle \mathbf{w}_r(\mathbf{n}) | \mathbf{w}_r(\mathbf{n}) \rangle = \langle \mathbf{w}_0 | (R^r)^T R^r | \mathbf{w}_0 \rangle. \tag{A 10}$$

When $r = 1$, the approximated eigenvector $\mathbf{w}_1$ is

$$|\mathbf{w}_1(\mathbf{n})\rangle = R |\mathbf{w}_0\rangle. \tag{A 11}$$

Let $|\mathbf{w}_0\rangle = |\mathbf{1}\rangle$, the left vector could be calculated as

$$\begin{aligned}
_{i \to j}\langle \mathbf{w}_1(\mathbf{n})| &= \sum_{k \to l} {}_{k \to l}\langle \mathbf{w}_0 | R_{k \to l, i \to j} \\
&= \sum_{k \to l} n_i a_{kl} A_{kl} A_{ij} \delta_{li} (1 - \delta_{kj}) \\
&= n_i A_{ij} \sum_{k \to l} a_{kl} A_{kl} \delta_{li} (1 - \delta_{kj}),
\end{aligned} \tag{A 12}$$

and the right vector

$$\begin{aligned}
|\mathbf{w}_1(\mathbf{n})\rangle_{i \to j} &= \sum_{k \to l} R_{i \to j, k \to l} |\mathbf{w}_0\rangle_{k \to l} \\
&= \sum_{k \to l} n_j a_{ij} A_{ij} A_{kl} \delta_{jk} (1 - \delta_{il}) \\
&= n_j a_{ij} A_{ij} \sum_{k \to l} A_{kl} \delta_{jk} (1 - \delta_{il}).
\end{aligned} \tag{A 13}$$

So the norm of $\mathbf{w}_1(\mathbf{n})$ is

$$\begin{aligned}
|\mathbf{w}_1(\mathbf{n})|^2 &= \sum_{i \to j} {}_{i \to j}\langle \mathbf{w}_1(\mathbf{n}) | \mathbf{w}_1(\mathbf{n}) \rangle_{i \to j} \\
&= \sum_{i \to j} n_i n_j a_{ij} A_{ij} \times \sum_{k \to l} a_{kl} A_{kl} \delta_{li} (1 - \delta_{kj}) \sum_{k \to l} A_{kl} \delta_{li} (1 - \delta_{kj}).
\end{aligned}$$

Let $M_{i \to j} = \sum_{k \to l} a_{kl} A_{kl} \delta_{li} (1 - \delta_{kj})$ and $N_{i \to j} = \sum_{k \to l} A_{kl} \delta_{jk} (1 - \delta_{il})$, then

$$|\mathbf{w}_1(\mathbf{n})|^2 = \sum_{i \to j} n_i n_j a_{ij} A_{ij} M_{i \to j} N_{i \to j}, \tag{A 14}$$

and the approximation to largest eigenvalue is

$$\lambda_1(\mathbf{n}) = \left(\frac{1}{2|E|} \sum_{i \to j} n_i n_j a_{ij} A_{ij} M_{i \to j} N_{i \to j} \right)^{1/2}. \tag{A 15}$$

For $r = 2$, the left vector $_{i \to j}\langle \mathbf{w}_2(\mathbf{n})|$ and right vector $|\mathbf{w}_2(\mathbf{n})\rangle_{i \to x=j}$ could also be calculated

$$\begin{aligned}
_{i \to j}\langle \mathbf{w}_2(\mathbf{n})| &= \sum_{k \to l} {}_{k \to l}\langle \mathbf{w}_1 | R_{k \to l, i \to j} \\
&= \sum_{k \to l} n_i n_k a_{kl} A_{kl} A_{ij} \delta_{li} (1 - \delta_{kj}) M_{k \to l} \\
&= n_i A_{ij} \sum_{k \to l} n_k a_{kl} A_{kl} \delta_{li} (1 - \delta_{kj}) M_{k \to l};
\end{aligned}$$

and

$$|\mathbf{w}_2(\mathbf{n})\rangle_{i \to j} = \sum_{k \to l} R_{i \to j, k \to l} |\mathbf{w}_1\rangle_{k \to l}$$

$$= \sum_{k \to l} n_l a_{kl} A_{kl} N_{k \to l} n_k a_{ij} A_{ij} \delta_{jk} (1 - \delta_{li})$$

$$= a_{ij} A_{ij} \sum_{k \to l} n_k n_l a_{kl} A_{kl} \delta_{jk} (1 - \delta_{li}) N_{k \to l}.$$

Thus, the second-order eigenvector is

$$|\mathbf{w}_2(\mathbf{n})|^2 = \sum_{i \to j} {}_{i \to j}\langle \mathbf{w}_1(\mathbf{n}) | \mathbf{w}_1(\mathbf{n}) \rangle_{i \to j}$$

$$= \sum_{i \to j, k \to l} n_i n_j n_k n_l a_{ij} a_{ki} a_{lj} A_{ij} A_{ki} A_{lj} M_{k \to i} N_{j \to l},$$

and the approximated eigenvalue is

$$\lambda_2(\mathbf{n}) = \left(\frac{1}{2|E|} \sum_{i \to j, k \to l} n_i n_j n_k n_l a_{ij} a_{ki} a_{lj} \times A_{ij} A_{ki} A_{lj} M_{k \to i} N_{j \to l} \right)^{1/4}. \tag{A 16}$$

A more generalized form of $|\mathbf{w}_r|$ and $\lambda_r(\mathbf{n})$ could be got

$$|\mathbf{w}_r(\mathbf{n})| = \sum_{i_1 \to j_1, i_2 \to j_2, \dots, i_r \to j_r} a_{i_1 j_1} M_{i_2 \to i_1} N_{j_{r-1} \to j_r}$$

$$\times \prod_{k=1}^{r} n_{i_r} n_{j_r} \prod_{k=1}^{r-1} a_{i_k i_{k+1}} a_{j_k j_{k+1}} A_{i_k i_{k+1}} A_{j_k j_{k+1}}$$

and

$$\lambda_r(\mathbf{n}) = \left(\frac{1}{2|E|} \sum_{i_1 \to j_1, i_2 \to j_2, \dots, i_r \to j_r} a_{i_1 j_1} M_{i_2 \to i_1} N_{j_{r-1} \to j_r} \right.$$

$$\left. \times \prod_{k=1}^{r} n_{i_r} n_{j_r} \prod_{k=1}^{r-1} a_{i_k i_{k+1}} a_{j_k j_{k+1}} A_{i_k i_{k+1}} A_{j_k j_{k+1}} \right)^{1/2r}.$$

In our research, we use the $r = 1$ case. The spreading influence of node v_i is regarded as the contribution of all value of edge between v_i and v_j contained in $\lambda_1(\mathbf{n})$, namely

$$S_i = \sum_{j \in N(i)} n_i n_j a_{ij} A_{ij} M_{i \to j} N_{i \to j}$$

$$= \sum_{j \in N(i)} a_{ij} \sum_{k \in N(i), k \neq j} a_{ik} (d_j - 1). \tag{A 17}$$

Appendix B. Spreading simulation

To examine the validation of employment spreading influence, we apply an agent-based spreading simulation. We want to verify if the infection starts with high empSI occupations will it spread through the whole network faster.

The complete occupation networks in Dalton, GA and Los Angeles, CA are used to simulate the infection process. In the networks, nodes stand for occupations and edges are weighted by skill similarity. In each node, there are agents with the same number to employment of corresponding occupation.

We select 0.2% agents in Los Angeles and Dalton network scattered uniformly into top-20 occupations as initial infected agents. If the employment of one selected job is lower than the infection number, we will take the employment number as the infected one, corresponding to the case of occupations with high similar to others but low employment numbers. For each step, each infected
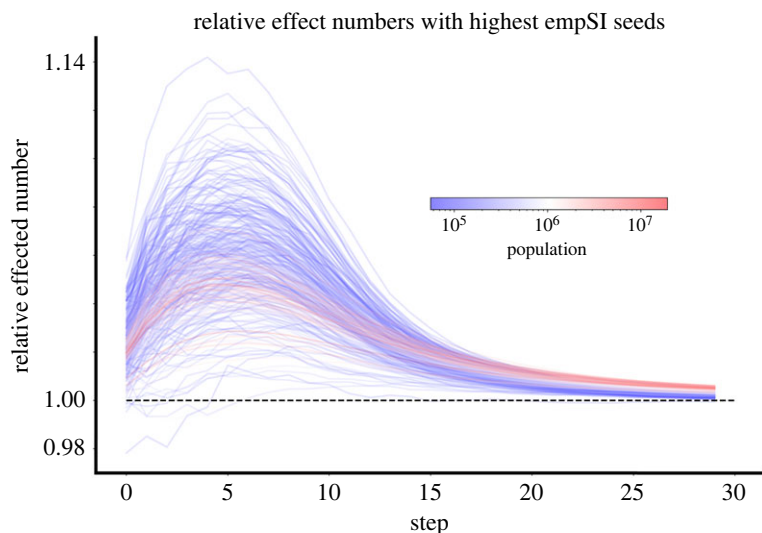
Figure 6. Spreading simulation with high empSI seeds against random results. Cities with different populations are marked by different colours, with blue standing for small cities and red large cities. Each curve is the average results of 20 simulation trials.

agent (source agent) selected one agent from its neighbour occupations, which we call 'target agent'. If the target agent is not infected, the source agent infects the target with the probability of skill similarity; if the target agent is already infected, the source agent will do nothing. After each step, all the infected agents will become source agents and could start to infect others.

In different simulations, the initial infected agents are selected by highest empSI, highest employment and highest weighted degree. At the same time, a simulation with random selected initial infected agents is implemented to work as the baseline. For each simulation, 20 trials are implemented and average results are calculated. We present the relative infected numbers in each step against the baseline in the main paper.

The simulation process is described in algorithm 1.

Algorithm 1. Infection Simulation Process.

Input: Graph G, Seed Infection Set I

for Each step **do**
    for Any agent i in I **do**
        Selected one agent j from the neighbours of i
        if j is not infected **then**
            With probability $SkillSimilarity(i, j)$, infected agent j
        end if
    end for
    Update I
end for

Output: Infected Sets I in Each Step

Results of spreading simulation with highest empSI jobs as seeds against random results in all cities are presented in figure 6. A subtle negative correlation between spreading speed and city scale could be observed: the spreading starting with high empSI seeds tends to be slower than in smaller cities. We infer that this is due to the higher complexity and more diverse occupation structure in larger cities. Further exploration in the spreading dynamics geo-variation could be expected in the future.

To examine the spreading patterns in the case of more complex contagions, we conducted a similar spreading processes with SIS model, namely during the spreading progress, the agent will not transfer the skill to others with a certain possibility at next step. We introduced a parameter *ResetProb* to stand for the probability that agent will 'recover' temporarily from the skill and not infect others. Meanwhile, we introduced another parameter *SpreadProb* to tune the spreading probability. The specific process is described in algorithm 2.
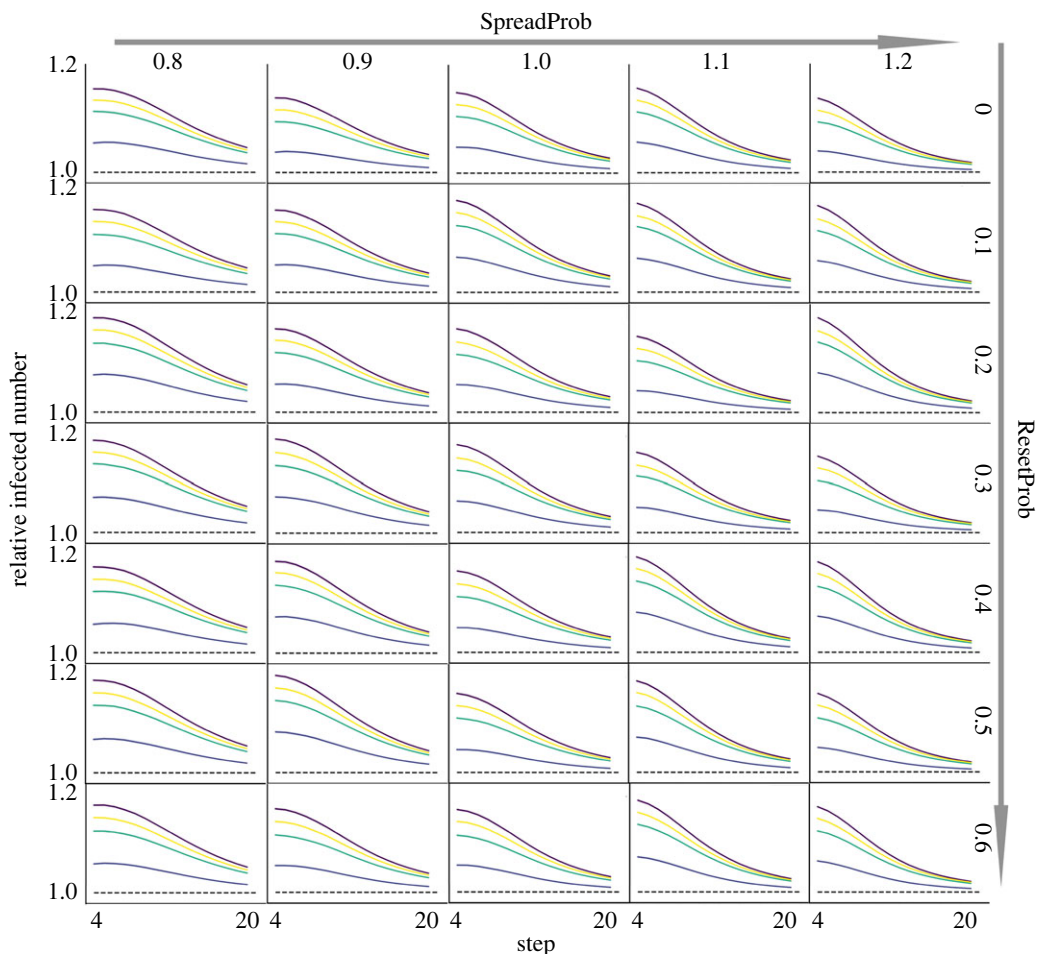
Figure 7. Simulation of SIS process on the occupation networks of Los Angeles–Long Beach–Anaheim, CA. All results are the averaged values of 20 trials. In each simulation, 0.2% random agents of employment are selected by different strategies including highest empSI, highest employment numbers, highest weighted degrees and highest log(emp) $\times$ weighted degree as the starting seeds for spreading. Each node in the networks contains the same agent number as employment. At each step, each agent infects one of its neighbours with the probability given by their skill similarity. A strategy of random seed selection is applied as a baseline. The curves are the relative infected number and inside is the total infected number. The confidence intervals in Los Angeles are very small to distinguish, due to its large population. Cases with different values of *ResetProb* and *SpreadProb* mentioned in algorithm 2 are presented in the simulations.

Algorithm 2. Susceptible-Infectious-Susceptible Process.

Input: Graph G, Seed Infection Set I

for Each step **do**
    for Any agent i in I **do**
        Selected one agent j from the neighbours of i
        if j is not infected **then**
            With the probability of *SkillSimilarity* $(i, j) \times$ *SpreadProb*, infected agent j
        end if
    end for
    Update I
end for
for Any agent i in I **do**
    With probability *ResetProb* remove i from I
end for

Output: Infected Sets I in Each Step

For various values of *ResetProb* and *SpreadProb*, we conducted the spreading simulation on LA occupation network (see figure 7). We find that although the derivation of empSI is based on SI model,
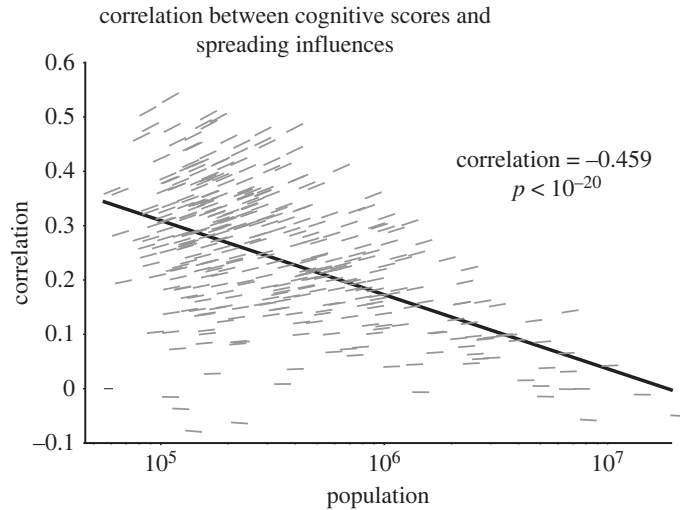
Figure 8. Correlations between cognitive score values and spreading influence S against populations in each city. The Pearson correlation coefficient is marked. Different slopes stand for different correlation values.

this index still performs well in SIS cases. In real world, the spreading process could be much more complex, which deserves further research to build models with more precise details and dynamical patterns.

Appendix C. Correlation between cognitive scores and spreading influences

In the main text, we presented relationships between empSI and cognitives in different cities. Here, a similar work is done for spreading influence S and cognitive scores (see figure 8). We find that there is a clear negative trend between the job-wise correlation in cognitive score and S. This is despite smaller cities generally having fewer workers in cognitive jobs, suggesting that this is driven by the more numerous *non-cognitive* jobs being relatively *un-influential*.

Appendix D. Weighted automation

Automation is a widely studied topic in labour market research. It measures the risk of jobs being substituted by machine, which is a crucial issue especially in this AI era. In the main text, we calculate the weighted automation by empSI with occupation data from Frey and Osborne's research work [44]. In their work, automation possibility of 70 occupations are selected to be hand-labelled by a group of machine learning researchers as 0 for impossible to be replaced by machine and 1 for will be replaced by machine, which are used as the labels in training set. Then nine variables-related computerization bottleneck from O*NET [40] describing the occupations are selected as features to be trained by a Gaussian process classifier. In this way, the automation and computerization probability of other occupations from O*NET could be estimated by the trained classifier.

There are some other research works about automation probability estimation. Brynjolfsson *et al.* designed a rubric with 23 distinct statements to evaluate the 'suitability for machine learning' (SML) of 2069 direct work activities which the occupations could be mapped to, to estimate the SML scores of occupations [48]. Webb calculated the overlaps between job task description texts and patent texts to measure the exposure of occupations to automation, in the perspectives of software, industrial robots and artificial intelligence. These indices are measures of the dependence of jobs to corresponding technologies. They both got meaningful data and conclusions [45].

Based on the automation data from this research, we also calculate their corresponding weighted automation by empSI, and results are presented in figure 9. A similar conclusion could be seen from the SML data and larger cities perform higher resilience against automation. The curves from Webb's data look opposite and larger cities present higher weighted values.

Appendix E. Notes on RCA values

When we calculate the RCA values of empSI, there are some occupations with very high RCA values, mostly in the natural sources and production sectors. These high RCA values are mostly because of
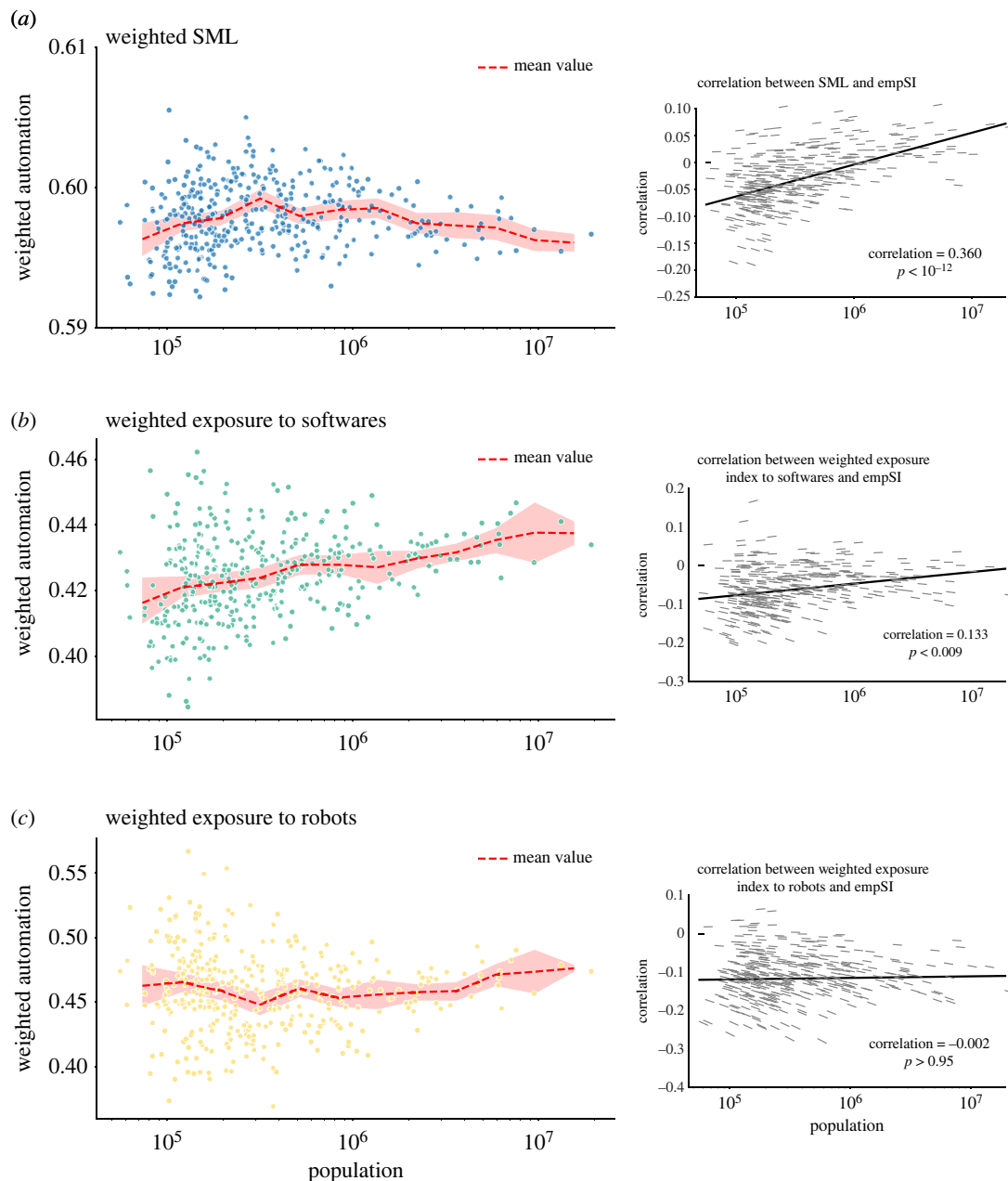
Figure 9. (a) Weighted suitability for machine learning (SML) from Brynjolfsson and Mitchell's [52] research work with empSI. The correlation between SML and empSI is drawn on the right side. (b) Weighted exposure score to software patents. (c) Weighted exposure score to robot patents.

insufficient data, since these occupations do appear in other cities yet their employment numbers are not available. It is hard to decide whether the RCA values follow some certain distributions (like power law [82], which will make the average meaningless). So in our research, we remove some extreme high RCA values.

References

1. Batty M, Longley PA. 1994 *Fractal cities: a geometry of form and function*. New York, NY: Academic press.

2. Bettencourt LM, Cintrón-Arias A, Kaiser DI, Castillo-Chávez C. 2006 The power of a good idea:

quantitative modeling of the spread of ideas from epidemiological models. *Physica A* **364**, 513–536. (doi:10.1016/j.physa.2005.08.083)

3. Alabdulkareem A, Frank MR, Sun L, AlShebli B, Hidalgo C, Rahwan I. 2018 Unpacking the

polarization of workplace skills. *Sci. Adv.* **4**, eaao6030. (doi:10.1126/sciadv.aao6030)

4. del Rio-Chanona RM, Mealy P, Beguerisse-Diaz M, Lafond F, Farmer JD. 2020 Occupational mobility and automation: a data- driven

network model. *J. R. Soc. Interface* **18**, 20200898. (doi:10.5281/zenodo.4453162)

5. Dworkin JD. 2019 Network-driven differences in mobility and optimal transitions among automatable jobs. *R. Soc. Open Sci.* **6**, 182124. (doi:10.1098/rsos.182124)

6. Guerrero OA, Axtell RL. 2013 Employment growth through labor flow networks. *PLoS ONE* **8**, e60808. (doi:10.1371/journal.pone.0060808)

7. Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, White DR. 2009 Economic networks: what do we know and what do we need to know? *Adv. Complex Syst.* **12**, 407–422. (doi:10.1142/S0219525909002337)

8. Park J, Wood IB, Jing E, Nematzadeh A, Ghosh S, Conover MD, Ahn YY. 2019 Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. *Nat. Commun.* **10**, 1–10. (doi:10.1038/s41467-019-11380-w)

9. Meindl B, Ayala NF, Mendonça J, Frank AG. 2021 The four smarts of Industry 4.0: evolution of ten years of research and future perspectives. *Technol. Forecast. Soc. Change* **168**, 120784. (doi:10.1016/j.techfore.2021.120784)

10. Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB. 2007 Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306. (doi:10.1073/pnas.0610172104)

11. Pinheiro FL, Hartmann D, Boschma R, Hidalgo CA. 2021 The time and frequency of unrelated diversification. *Res. Policy* **51**, 104323. (doi:10.1016/j.respol.2021.104323)

12. Glaeser E. 2012 Viewpoint: triumph of the city. *J. Transp. Land Use* **5**, 1–4. (doi:10.5198/jtlu.v5i2.371)

13. Tóth G, Wachs J, Di Clemente R, Kertész J, Lengyel B. 2021 Inequality is rising where social network segregation interacts with urban topology. *Nat. Commun.* **12**, 1–9. (doi:10.1038/s41467-021-21465-0)

14. Ferragina E, Filetti FD. 2022 Labour market protection across space and time: a revised typology and a taxonomy of countries' trajectories of change. *J. Eur. Soc. Policy* **32**, 09589287211056222. (doi:10.1177/09589287211056222)

15. Hosseinioun M, Neffke F, Youn H. 2023 Deconstructing human capital to construct hierarchical nestedness. *arXiv*. (http://arxiv.org/abs/2303.15629)

16. Neffke F, Hartog M, Boschma R, Henning M. 2018 Agents of structural change: the role of firms and entrepreneurs in regional diversification. *Econ. Geogr.* **94**, 23–48. (doi:10.1080/00130095.2017.1391691)

17. Hidalgo CA, Klinger B, Barabási AL, Hausmann R. 2007 The product space conditions the development of nations. *Science* **317**, 482–487. (doi:10.1126/science.1144581)

18. Noorazar H. 2020 Recent advances in opinion propagation dynamics: a 2020 survey. *Eur. Phys. J. Plus* **135**, 1–20. (doi:10.1140/epjp/s13360-020-00541-2)

19. Zhang J, Litvinova M, Liang Y, Zheng W, Shi H, Vespignani A, Viboud C, Ajelli M, Yu H. 2021 The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. *Sci. Adv.* **7**, eabe2584. (doi:10.1126/sciadv.abe2584)

20. Alessandretti L. 2022 What human mobility data tell us about COVID-19 spread. *Nat. Rev. Phys.* **4**, 12–13. (doi:10.1038/s42254-021-00407-1)

21. Majhi S, Perc M, Ghosh D. 2022 Dynamics on higher-order networks: a review. *J. R. Soc. Interface* **19**, 20220043. (doi:10.1098/rsif.2022.0043)

22. Li M, Lü L, Deng Y, Hu MB, Wang H, Medo M, Stanley HE. 2020 History-dependent percolation on multiplex networks. *Natl Sci. Rev.* **7**, 1296–1305. (doi:10.1093/nsr/nwaa029)

23. Rogers EM. 2003 *Diffusion of innovations*. New York, NY: Free Press.

24. Hidalgo CA. 2021 Economic complexity theory and applications. *Nat. Rev. Phys.* **3**, 92–113. (doi:10.1038/s42254-020-00275-1)

25. Sekara V, Rutherford A, Mann G, Dredze M, Adler N, Garcia-Herranz M. 2018 Trends in the adoption of corporate child labor policies: an analysis with bloomberg terminal ESG data. *Proc. Data Good Exchange* **3**.

26. Goldfarb A, Gans J, Agrawal A. 2019 *The economics of artificial intelligence: an agenda*. Chicago, IL: University of Chicago Press.

27. Jaccard P. 1912 The distribution of the flora in the alpine zone, 1. *N. Phytol.* **11**, 37–50. (doi:10.1111/j.1469-8137.1912.tb05611.x)

28. Pei S, Muchnik L, Andrade Jr JS, Zheng Z, Makse HA. 2014 Searching for superspreaders of information in real-world social media. *Sci. Rep.* **4**, 1–12. (doi:10.1038/srep05547)

29. Kang D, Choi H, Kim JH, Choi J. 2020 Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int. J. Infect. Dis.* **94**, 96–102. (doi:10.1016/j.ijid.2020.03.076)

30. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA. 2010 Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893. (doi:10.1038/nphys1746)

31. Acemoglu D, Restrepo P. 2019 *8. Artificial intelligence, automation, and work*. Chicago, IL: University of Chicago Press.

32. Kempe D, Kleinberg J, Tardos É. 2003 Maximizing the spread of influence through a social network. In *Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Washington, DC, 24–27 August*, pp. 137–146. Association for Computing Machinery.

33. Karp RM. 1972 Reducibility among combinatorial problems. In *Complexity of computer computations*, pp. 85–103. New York, NY: Springer.

34. Rosvall M, Esquivel AV, Lancichinetti A, West JD, Lambiotte R. 2014 Memory in network flows and its effects on spreading dynamics and community detection. *Nat. Commun.* **5**, 1–13. (doi:10.1038/ncomms5630)

35. Holland PW, Leinhardt S. 1970 A method for detecting structure in sociometric data. *Am. J. Sociol.* **76**, 492–513. (doi:10.1086/224954)

36. Morone F, Makse HA. 2015 Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65–68. (doi:10.1038/nature14604)

37. Mézard M, Parisi G. 2003 The cavity method at zero temperature. *J. Stat. Phys.* **111**, 1–34. (doi:10.1023/A:1022221005097)

38. Zhang R, Pei S. 2018 Dynamic range maximization in excitable networks. *Chaos* **28**, 013103. (doi:10.1063/1.4997254)

39. Neffke FM, Otto A, Weyh A. 2017 Inter-industry labor flows. *J. Econ. Behav. Organ.* **142**, 275–292. (doi:10.1016/j.jebo.2017.07.003)

40. National Center for O*NET Development. 2022 O*NET OnLine. See https://www.onetonline.org/.

41. US Department of Labor. 2021 US Bureau of Labor Statistics. See https://www.bls.gov/.

42. European Commission. 2022 European Skills/Competences, Qualifications and Occupations. See https://ec.europa.eu/social/main.jsp?catId=1326&langId=en.

43. US Department of Commerce. 2021 US Census Bureau. See https://www.census.gov/.

44. Frey CB, Osborne MA. 2017 The future of employment: how susceptible are jobs to computerisation? *Technol. Forecast. Soc. Change* **114**, 254–280. (doi:10.1016/j.techfore.2016.08.019)

45. Webb M. 2019 The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.

46. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Techn. J.* **27**, 379–423. (doi:10.1002/j.1538-7305.1948.tb01338.x)

47. Frank MR, Sun L, Cebrian M, Youn H, Rahwan I. 2018 Small cities face greater impact from automation. *J. R. Soc. Interface* **15**, 20170946. (doi:10.1098/rsif.2017.0946)

48. Brynjolfsson E, Mitchell T, Rock D. 2018 What can machines learn, and what does it mean for occupations and the economy? *AEA Papers & Proc.* **108**, 43–47. (doi:10.1257/pandp.20181019)

49. Hong I, Frank MR, Rahwan I, Jung WS, Youn H. 2020 The universal pathway to innovative urban economies. *Sci. Adv.* **6**, eaba4934. (doi:10.1126/sciadv.aba4934)

50. Bettencourt LM, Samaniego H, Youn H. 2014 Professional diversity and the productivity of cities. *Sci. Rep.* **4**, 5393. (doi:10.1038/srep05393)

51. Moro E, Frank MR, Pentland A, Rutherford A, Cebrian M, Rahwan I. 2021 Universal resilience patterns in labor markets. *Nat. Commun.* **12**, 1–8. (doi:10.1038/s41467-021-22086-3)

52. Brynjolfsson E, Mitchell T. 2017 What can machine learning do? Workforce implications. *Science* **358**, 1530–1534. (doi:10.1126/science.aap8062)

53. Frank MR *et al.* 2019 Toward understanding the impact of artificial intelligence on labor. *Proc. Natl Acad. Sci. USA* **116**, 6531–6539. (doi:10.1073/pnas.1900949116)

54. Molinero C, Thurner S. 2021 How the geometry of cities determines urban scaling laws. *J. R. Soc. Interface* **18**, 20200705. (doi:10.1098/rsif.2020.0705)

55. Chen HC, Li X, Frank M, Qin X, Xu W, Cebrian M, Rahwan I. 2021 Automation impacts on China's polarized job market. *J. Comput. Soc. Sci.* **5**, 517–535. (doi:10.1007/s42001-021-00134-8)

56. Pichler A, Pangallo M, Lafond F, Farmer JD. 2021 In and out of lockdown: propagation of supply and demand shocks in a dynamic input-output model. (http://arxiv.org/abs/2102.09608)

57. Fana M, Tolan S, Torrejón S, Brancati CU, Fernández-Macías E. 2020 *The COVID confinement measures and EU labour markets*. Publications office of the European Union Luxembourg.

58. David H. 2015 Why are there still so many jobs? The history and future of workplace automation. *J. Econ. Perspect.* **29**, 3–30. (doi:10.1257/jep.29.3.3)

59. Acemoglu D, Restrepo P. 2018 The race between man and machine: implications of technology for growth, factor shares, and employment. *Am. Econ. Rev.* **108**, 1488–1542. (doi:10.1257/aer.20160696)

60. Bustos S, Gomez C, Hausmann R, Hidalgo CA. 2012 The dynamics of nestedness predicts the evolution of industrial ecosystems. *PLoS ONE* **7**, e49393. (doi:10.1371/journal.pone.0049393)

61. Lafond F, Farmer JD, Mungo L, Astudillo-Estévez P. 2022 *Reconstructing production networks using machine learning*. Technical report. Oxford, UK: Institute for New Economic Thinking at the Oxford Martin School, University of Oxford.

62. Tuninetti M, Aleta A, Paolotti D, Moreno Y, Starnini M. 2021 Prediction of new scientific collaborations through multiplex networks. *EPJ Data Sci.* **10**, 25. (doi:10.1140/epjds/s13688-021-00282-x)

63. Mišić B, Betzel RF, Nematzadeh A, Goni J, Griffa A, Hagmann P, Flammini A, Ahn YY, Sporns O. 2015 Cooperative and competitive spreading dynamics on the human connectome. *Neuron* **86**, 1518–1529. (doi:10.1016/j.neuron.2015.05.035)

64. Hyland CC, Tao Y, Azizi L, Gerlach M, Peixoto TP, Altmann EG. 2021 Multilayer networks for text analysis with multiple data types. *EPJ Data Sci.* **10**, 33. (doi:10.1140/epjds/s13688-021-00288-5)

65. Battiston F et al. 2021 The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093–1098. (doi:10.1038/s41567-021-01371-4)

66. Millán AP, Ghorbanchian R, Defenu N, Battiston F, Bianconi G. 2021 Local topological moves determine global diffusion properties of hyperbolic higher-order networks. *Phys. Rev. E* **104**, 054302. (doi:10.1103/PhysRevE.104.054302)

67. Benson AR, Gleich DF, Leskovec J. 2016 Higher-order organization of complex networks. *Science* **353**, 163–166. (doi:10.1126/science.aad9029)

68. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002 Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827. (doi:10.1126/science.298.5594.824)

69. Sarajlić A, Malod-Dognin N, Yaveroğlu ÖN, Pržulj N. 2016 Graphlet-based characterization of directed networks. *Sci. Rep.* **6**, 1–14. (doi:10.1038/srep35098)

70. Acemoglu D, Restrepo P. 2019 Automation and new tasks: how technology displaces and reinstates labor. *J. Econ. Perspect.* **33**, 3–30. (doi:10.1257/jep.33.2.3)

71. Börner K, Scrivner O, Gallant M, Ma S, Liu X, Chewning K, Wu L, Evans JA. 2018 Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proc. Natl Acad. Sci. USA* **115**, 12 630–12 637. (doi:10.1073/pnas.1804247115)

72. Barabási AL, Albert R. 2023 The dynamic resilience of urban labour networks. *Zenodo*. (https://doi.org/10.5281/zenodo.8026260)

73. Pei S, Wang J, Morone F, Makse HA. 2020 Influencer identification in dynamical complex systems. *J. Complex Netw.* **8**, cnz029. (doi:10.1093/comnet/cnz029)

74. Yoon S, Goltsev AV, Dorogovtsev SN, Mendes J. 2011 Belief-propagation algorithm and the Ising model on networks with arbitrary distributions of motifs. *Phys. Rev. E* **84**, 041144. (doi:10.1103/PhysRevE.84.041144)

75. Wang J, Zhang R, Wei W, Pei S, Zheng Z. 2019 On the stability of multilayer Boolean networks under targeted immunization. *Chaos* **29**, 013133. (doi:10.1063/1.5053820)

76. Karrer B, Newman ME. 2010 Message passing approach for general epidemic models. *Phys. Rev. E* **82**, 016101. (doi:10.1103/PhysRevE.82.016101)

77. Kirkley A, Cantwell GT, Newman M. 2021 Belief propagation for networks with loops. *Sci. Adv.* **7**, eabf1211. (doi:10.1126/sciadv.abf1211)

78. Cantwell GT, Newman ME. 2019 Message passing on networks with loops. *Proc. Natl Acad. Sci. USA* **116**, 23 398–23 403. (doi:10.1073/pnas.1914893116)

79. Hashimoto K. 1989 Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties*, pp. 211–280. Orlando, FL: Elsevier.

80. Atkinson KE. 2008 *An introduction to numerical analysis*. Hoboken, NJ: John Wiley & Sons.

81. Bhatia NP, Szegö GP. 2002 *Stability theory of dynamical systems*. Berlin, Germany: Springer Science & Business Media.

82. Barabási AL, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)