



## OPEN

## Professional diversity and the productivity of cities

Luís M. A. Bettencourt<sup>1</sup>, Horacio Samaniego<sup>2,3</sup> & Hyejin Youn<sup>1,4</sup>

<sup>1</sup>Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA, <sup>2</sup>Center for Non Linear Studies, Theoretical Division MS B284, Los Alamos National Laboratory, Los Alamos NM 87545, USA, <sup>3</sup>Instituto de Conservación, Biodiversidad & Territorio, Universidad Austral de Chile, Valdivia, Chile, <sup>4</sup>Institute for New Economic Thinking, Oxford Martin School, Oxford, OX2 6ED, UK. Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK.

## SUBJECT AREAS:

STATISTICAL PHYSICS

THEORETICAL ECOLOGY

SUSTAINABILITY

SOCIOECONOMIC SCENARIOS

Received

25 February 2014

Accepted

2 June 2014

Published

23 June 2014

Correspondence and requests for materials should be addressed to

H.S. (horacio@ecoinformatica.cl)

Attempts to understand the relationship between diversity, productivity and scale have remained limited due to the scheme-dependent nature of the taxonomies describing complex systems. We analyze the diversity of US metropolitan areas in terms of profession diversity and employment to show how this frequency distribution takes a universal scale-invariant form, common to all cities, in the limit of infinite resolution of occupational taxonomies. We show that this limit is obtained under general conditions that follow from the analysis of the variation of the occupational frequency across taxonomies at different resolutions in a way analogous to finite-size scaling in statistical physical systems. We propose a theoretical framework that derives the form and parameters of the limiting distribution of professions based on the appearance, in urban social networks, of new occupations as the result of specialization and coordination of labor. By deriving classification scheme-independent measures of functional diversity and modeling cities as social networks embedded in infrastructural space, these results show how standard economic arguments of division and coordination of labor can be articulated in detail in cities and provide a microscopic basis for explaining increasing returns to population scale observed at the level of entire metropolitan areas.

A fundamental theme across the study of complex systems<sup>1–3</sup> -from ecosystems<sup>4,5</sup> to human behavior<sup>1,7</sup> and socioeconomic organization<sup>3,8,9</sup>- deals with the mechanisms by which diversity arises and is sustained.

Diversity means generally that component agents of a large system are heterogeneous so that they exhibit different traits and perform different functions. For all its ubiquity and importance, the functional role of heterogeneity in complex systems has remained poorly understood. General lines of argument for explaining diversity can be roughly divided into two classes: those that emphasize its protective role in uncertain environments and those that invoke direct gains in productivity as a result of functional heterogeneity<sup>5,6</sup>.

On the one hand, heterogeneity may confer collectives with resilience to selective shocks through a hedging effect that allows some components of the system to survive in times of crisis<sup>10</sup>. For example, cities that have remained very specialized (in manufacturing) have historically fared poorly over the long run. This effect already speaks of the importance of heterogeneity in enabling differential selection and therefore speeding up “evolution”<sup>11,12</sup>, whether in ecosystems or in markets. Selection, in this picture, remains a function of the external environment and confers no immediate incentive on particular individuals to differentiate: It develops because its effects, in retrospect, are necessary for the group’s survival.

On the other hand, a more direct and potentially more fundamental link between diversity, productivity and scale may exist. This connection has often been invoked, both in ecosystems and in economics, to explain why larger systems tend to be at once more diverse and more productive<sup>3–5,8–10,13</sup>. Nevertheless, the formalization of these ideas remains incomplete and does not yet explain quantitative patterns in observed levels of diversity<sup>3,14</sup>. The key concept is that diversity is both a consequence and an enabler of scale-dependent system organization and that it is directly associated with higher *individual* productivity. As a consequence, higher heterogeneity can be positively selected endogenously by agents and organizations that benefit directly from such strategies. Here, we show how diversity, productivity and social network structure can be combined at the microscopic level to generate an open-ended dynamics of growth, differentiation and complexification in cities.

In contemporary human societies socioeconomic diversity is associated primarily with large cities<sup>7,15</sup>. This reflects the primary role of urban environments in promoting the production of new ideas and stimulating socioeconomic development<sup>3,15,16</sup>. However, counter-arguments have also emerged noting that specialized cities are, at specific times, more productive<sup>16–19</sup>. Familiar examples are contemporary Silicon Valley or manufacturing cities in their heyday such as Manchester or Detroit. These questions remain far from settled, in part because of the difficulties inherent to measuring diversity in any complex system<sup>2,20</sup>.



Empirically, diversity has remained difficult to characterize unambiguously. Most measures account for the presence, and sometimes the relative proportion<sup>21–23</sup>, of different functional types: For example different professions or business types in cities or nations, or different species in an ecosystem<sup>24</sup>. Such measures of diversity, are inevitably linked to particular classification schemes or taxonomies. To appreciate this point, consider the question: How many different professions are there in a large city, like New York? In general, there is no objective answer as it depends on how finely one differentiates similar functions. However, under general conditions, a limit of infinite taxonomic resolution can be inferred through an analysis analogous to that of statistical physical quantities near phase transitions<sup>25</sup>. If such a limit exists, taxonomic scheme-independent answers emerge.

Here we show how addressing these two issues at once - measuring functional diversity and linking it to productivity by grounding it in scale-dependent urban social structure - can help shed light on the dynamics and structure of cities. As a result we are able to develop a predictive theoretical framework for the emergence of functional heterogeneity as a consequence of a process of division and coordination of functions in social networks, stimulated by productivity gains and made possible by increases in population, that is: system size.

## Results

We start by showing that the simplest measure of functional diversity in cities, the count of the number of distinct professions, is classification scheme dependent. We then show how a scheme-invariant measure can be obtained in the limit of infinite resolution via well-known methods of finite size scaling in statistical physics. This allows us to derive the general form of the distribution of professions in US urban areas, which we show applies to cities of all sizes and over time. We then develop a theoretical framework that derives the observed patterns of professional diversity from the maximization of economic productivity under social network constraints reflecting coordination costs at the metropolitan level.

### Classification resolution and the number of distinct professions.

The simplest measure of diversity,  $D(N)$ , counts the number of distinct professions in a city. Fig. 1A shows  $D(N)$ , for US metropolitan areas (MSAs) as a function of their total employment,  $N_e$ . Because  $N_e$  is, on average, proportional to population<sup>26</sup>,  $N$ , we use the two measures of scale interchangeably. We observe that  $D$  increases with  $N_e$  initially and then saturates for large cities and is well fit by

$$D(N_e) = d_0 \frac{\left(\frac{N_e}{N_0}\right)^\gamma}{1 + \left(\frac{N_e}{N_0}\right)^\gamma}. \quad (1)$$

Eq. 1 holds over time (Fig. S1) and for different levels of resolution,  $r$ , in the occupations hierarchical classification scheme (see Supplementary Information). The parameters in Eq. 1 are, in general, functions of  $r$ . The scale  $d_0(r)$  is the effective size of the classification scheme at resolution  $r$ ,  $N_0(r)$  is the characteristic size of the city at which saturation starts. The exponent (elasticity),  $\gamma$ , is empirically independent of  $r$ : It expresses the proportionality between the population growth rate and that for new occupations in the city, in the absence of saturation (see Supplementary Information).

Coarsening the hierarchical classification leads to similar saturation at each of the taxonomy's size  $d_0(r_6) > d_0(r_5) > d_0(r_4)$ , etc (Fig. 1B). This behavior is the hallmark of a finite resolution artifact, a phenomenon well understood through finite-size scaling analysis in statistical physical systems<sup>25</sup>. The explicit dependence of  $d_0$  on  $r$  suggests that existing classification schemes are too coarse to capture the true professional diversity of large US cities<sup>27</sup>, beyond an employment size of order  $N_e \sim N_0 \simeq 10^5$ . Nevertheless, we can use the variation of the statistics of occupations with  $r$  to derive classification

scheme independent results. We reconcile all curves for  $D(N)$  at different  $r$  and extract their limit as  $r \rightarrow \infty$ . We define a dimensionless function  $h\left(\frac{N}{N_0}, \gamma\right)$  such that

$$D(N) = d_0 \left(\frac{N}{N_0}\right)^\gamma h\left(\frac{N}{N_0}, \gamma\right) \rightarrow \begin{cases} D_0 N^\gamma, & N \ll N_0, \\ d_0(r), & N \gg N_0, \end{cases}$$

where  $D_0$  is a constant. Comparison with Eq. 1 tells us that in the limit  $\frac{N}{N_0} \rightarrow 0$ ,  $h \rightarrow 1$ ,  $D_0 = \frac{d_0}{N_0^\gamma}$ , and in the limit  $\frac{N}{N_0} \rightarrow \infty$ ,  $h \rightarrow \left(\frac{N_0}{N}\right)^\gamma$ . A

universal scaling regime exists if and only if the quantity  $D_0 = \frac{d_0(r)}{N_0(r)^\gamma}$  becomes a constant, independent of  $r$ , as  $r \rightarrow \infty$  (Fig. 1B). Fig. 1C shows  $d_0$  vs.  $N_0^\gamma$  across  $r$  and over time. The relationship is well described by a straight-line with slope  $D_0 = 0.05$  across all years.

These results suggest the existence of a resolution independent, scale-invariant limit for  $D(N)$  and show that the occupational diversity of cities is in fact open-ended: the number of distinct occupations in US cities increases by  $\gamma \sim 84\%$  with each doubling of its labor force, meaning that larger cities are at once more diverse in absolute terms and more specialized per capita. These insights can be proven as simple theorem, see SI Text.

**Universal Distribution of Urban Professions.** Beyond considering the presence or absence of professions, which gives only a crude measure of urban diversity, we can characterize their probability. To derive this distribution we identify  $D(N)$  in Eq. 1 with the maximum rank at each  $N$ , which has probability  $p(D) = \frac{1}{N}$ . Inverting this relation and generalizing it to all ranks,  $i$ , leads to the occupational frequency,  $f(i)$ :

$$f(i) = \frac{N_e}{N_0} \left(\frac{d_0 - i}{i}\right)^{1/\gamma} \quad (2)$$

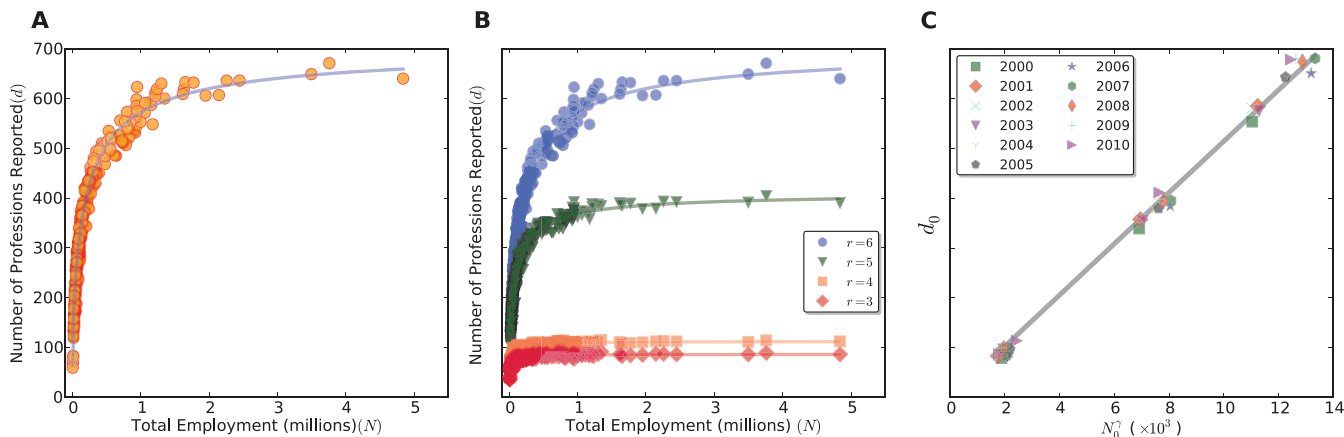
This is also scheme independent in the large resolution limit and can be used to derive the probability density,  $p(i)$ , as

$$p(i) = \frac{f(i)}{\sum_{j=1}^{D(N)} f(j)} = \frac{1 - \gamma}{\gamma} \frac{i^{-1/\gamma}}{1 - D(N)^{-\frac{1-\gamma}{\gamma}}}; \quad (3)$$

which is also independent of  $r$ . The occupational probability has a residual dependence on  $N$  through  $D(N)$  because the rarest professions in each city cannot have less than one person. Figs. 2A, B shows that the distribution of occupations for different cities is universal: When adjusted for scale,  $N_e$ , the frequency curves for every city collapse onto a single line, Fig. 2B. This shows that there is an expected nested sequence of occupations predicted by city size, as expected by the hierarchy principle of central place theory<sup>28</sup> and its extensions in regional economics<sup>23,29</sup>. This sequence of professions arising with growing city size is also analogous to the hierarchy of products vs. level of economic development observed at the national level<sup>8,9</sup> and shows how large cities play a fundamental role in embodying the economic diversity of nations.

**Indices of diversity.** The diversity associated with the distribution of types is commonly measured by functionals of the probability distribution,  $p(i)$ . Examples are the Herfindahl–Hirschman Index ( $HH$ ) and the Shannon entropy ( $S$ )<sup>21,22,24</sup>. Here we show that the residual dependence of  $p(i)$  on  $N$ , through  $D(N)$  in Eq. 3 is the only source of city size dependence of these measures of occupational diversity.

The Herfindahl–Hirschman index measures how concentrated a distribution is. For this reason it is often applied to economic sectors to measure their concentration in terms of firm's market share and assess the existence of monopolies. Given the asymptotic form of the distribution (Eq. 3) the  $HH(N)$  can be calculated analytically as



**Figure 1 | The number of distinct occupations in US Metropolitan Statistical Areas vs. total employment.** (A) The relationship between the number of professions present for each city (orange dots) and city size is well described by  $D(N_e) = d_0 \frac{(N_e/N_0)^\gamma}{1 + (N_e/N_0)^\gamma}$ , with  $d_0 = 686, \gamma = 0.84, N_0 = 1.48 \times 10^5$  (blue line). (B)  $D(N_e)$  at different labels of resolution of the occupational classification scheme,  $r_i$ , with  $i = 6$  the finest and  $i = 3$  the coarsest. (C)  $d_0$  is proportional to  $N_0^\gamma$  across levels of classification scheme resolution and time, suggesting the there is a  $r$ -independent limit to the form of the occupational diversity of cities and that  $D$  is open-ended. In this limit,  $D(N_e) = D_0 N_e^\gamma$  and larger cities are always more diverse as a whole, but more specialized per capita.

$$\begin{aligned}
 HH(N) &= \sum_{i=1}^{D(N)} p^2(i) = \frac{\delta^2}{1-\delta^2} \frac{1 - D_0^{-\frac{1+\delta}{1-\delta}} N^{-1-\delta}}{\left(1 - D_0^{-\frac{\delta}{1-\delta}} N^{-\delta}\right)^2} \\
 &\simeq \frac{\delta^2}{1-\delta^2} \left(1 + \frac{2}{D_0^{-\frac{\delta}{1-\delta}} N^\delta}\right),
 \end{aligned}
 \tag{4}$$

where the exponent  $\delta = 1 - \gamma$ , Consequently the  $HH$  index decreases with  $N$  towards a small constant, set by the exponent  $\delta$ , as cities grow. This expresses an increase in diversity with city size. Note that the asymptotic value for  $N \rightarrow \infty$ , with  $\delta \simeq 1/6^{30}$ , is  $HH \rightarrow 0.028$ , which is typical of highly diverse (and competitive) markets.

Similarly the Shannon entropy,  $S$ , measures the diversity of the occupational distribution as

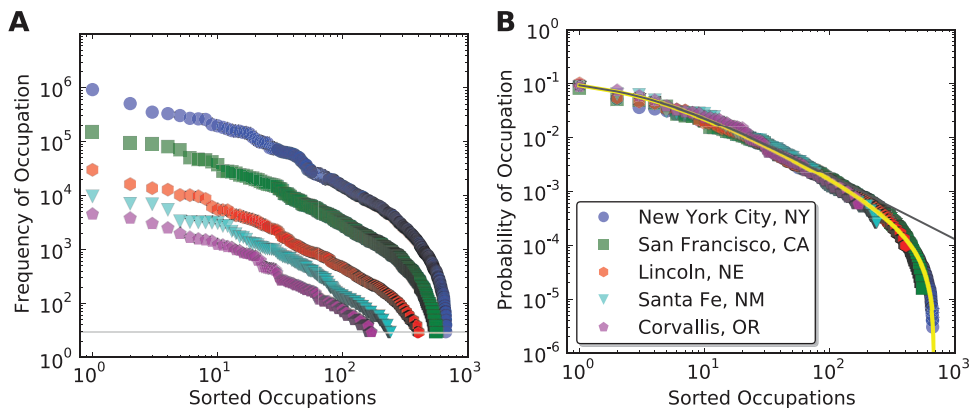
$$S = - \sum_{i=1}^{D(N)} p(i) \ln p(i) \simeq \frac{1}{\delta} - D_0^{-\delta/(1-\delta)} N^{-\delta} \ln\left(D_0^{1/\gamma} N\right), \tag{5}$$

which increases with  $N$  towards the Pareto distribution limit at infinite  $N$ . Thus, the increase in entropy signals the increase in diversity

of the occupational distribution in larger cities. Note that in both cases the increases in diversity are driven, at leading order, by a term of order  $N^{-\delta}$  (Fig. S3). As we shall see below, this term measures professional diversity per capita and is also implicated in the growth of labor productivity and social connectivity with city size.

**Diversity, Productivity and Urban Social Networks.** We now address the problem of how professional diversity, economic productivity and urban social networks are interconnected.

Given the empirical results so far, we should expect economic productivity to be inversely proportional to professional diversity. To see this more explicitly, consider that urban indicators of economic productivity (wages, GDP, etc) scale, on average, superlinearly with city size  $N$ ,  $W(N, t) = W_0(t)N(t)^\beta$ , with  $W_0(t)$  and  $\beta \simeq 1 + \delta > 1$  independent of  $N^{26}$ , see Fig. 3A. The average wage per capita is, then,  $w(N) = W_0 N^\delta$ , where  $\delta \sim 1/6 \simeq 1 - \gamma^{30}$ . This result has been derived from a general theoretical framework that defines cities as co-located social networks, subject to infrastructural efficiency constraints<sup>30</sup>, with  $w = Gk(N)$ , where  $G$  is a constant in  $N$ , involving a balance between mobility costs and infrastructural properties, and  $k(N) = k_0 N^\delta$  is the average social connectivity (network degree) per person.



**Figure 2 | The distribution of occupations in US metropolitan areas is universal.** (A) Frequency distribution for several cities with different population sizes only differ in their amplitude, which is set by city size and the extent to which they probe rare occupations. The horizontal grey line shows the minimum number of professions (thirty) reported. (B) The rank-probability distributions for different cities collapse on each other when adjusted for city size (total employment). The yellow line shows the fit of the universal form to  $f(i)/N_e = \frac{1}{N_0} \left(\frac{d_0 - (i + i_0)}{(i + i_0)}\right)^{1/\gamma}$ , where we introduce a scale  $i_0 \simeq 3$  at small ranks. The black line is the  $f$  of  $f(i)/N_e$  in the absence of saturation.



(This increase in social connectivity with city size has been observed directly in urban telecommunication networks<sup>31</sup>.) Similarly, diversity per capita,  $d(N) = D(N)/N = D_0 N^{-1} = D_0 N^{-\delta}$ . Hence, we conclude that  $w \propto 1/d$ . This relationship is an expression of the abundant evidence in economics for knowledge specialization – a decrease in  $d(N)$  – as the source of increases in (labor) productivity<sup>19,32–34</sup>. However, this view of productivity per se cannot be the target of optimization: no city has become rich by reducing its occupational diversity to a single activity! What then is the optimal level of diversity and productivity attainable at the city level?

Before we answer this question we note that the connection between the microscopic features of socioeconomic urban organization and the aggregate productivity of cities has remained a fundamental but difficult question in urban economics. A number of detailed socioeconomic mechanisms have been proposed to explain increases in productivity with city size<sup>16</sup>, all of which depend on the social coordination opportunities that larger cities enable<sup>36</sup>. A different, if not orthogonal, set of arguments seeks to explain the greater productivity of larger cities from the empirical observation of a disproportionate presence of economic sectors with greater value-added or of greater human capital<sup>35</sup>. Despite some important theoretical attempts<sup>33,34</sup>, what is still missing from these two lines of argument is an empirically based predictive formalization of urban diversity that ties together learning, via the specialization of knowledge (accumulation of human capital), to the coordination costs inherent to the division of labor in cities.

We now propose such a formalization by showing how knowledge specialization is set by the extent and costs of human social networks in cities. We implement these ideas in terms of the structural change of urban social networks with city size by assuming a process whereby individuals start as generalists in small social environments and are able to specialize in larger cities by externalizing some of their functions to others, while maintaining access to them through direct social links. Thus, we assume that only close coordination (at least in the initial stages of this transfer of functions) will be able to preserve these specialized functions suitably integrated and, as such, that there must be a conservation of the number of functions within the immediate network contacts (first neighbors) of each individual (Fig. 3B). This means that as an individual specializes by performing fewer functions, she maintains access to such functions via socioeconomic contacts, so that the rate of individual specialization with city size must equal in magnitude the rate of diversification of functions

across ones social ties, see below. Note that this process provides a (reversible) mechanism for the creation of human capital, through the gradual division of labor in larger social networks and is fundamentally different from arguments of comparative advantage where differential endowments of distinct individuals or organizations are taken as given.

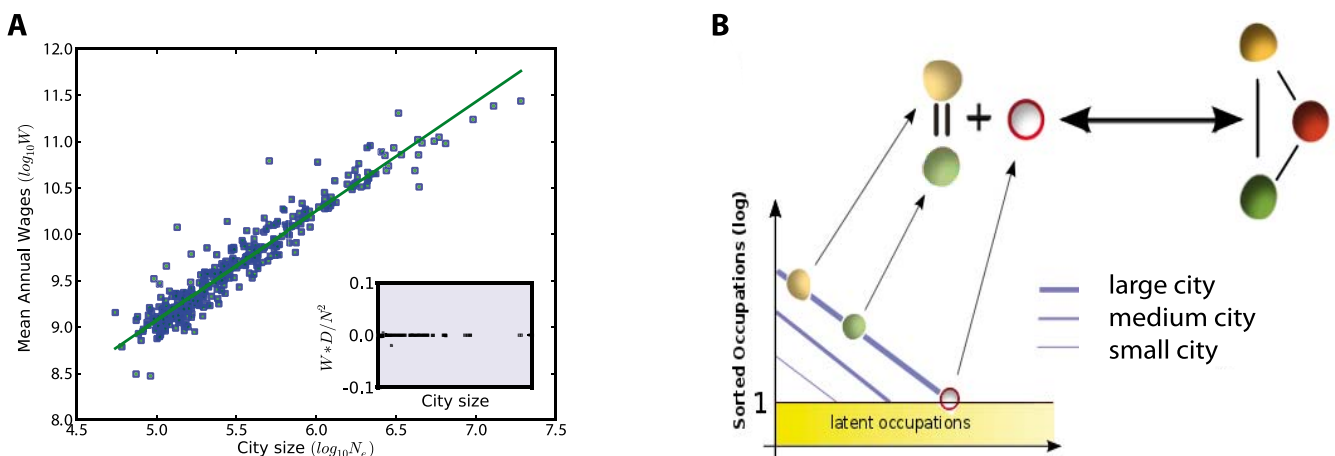
We formalize these ideas in terms of the condition that, on average, as  $k(N) \frac{D(n)}{N} \equiv k(N)d(N) = A$ , where  $A$  is a constant in  $N$  but that may vary over time, e.g. due to changes in communication and transportation technologies. This condition is equivalent to the conservation of the total rate of change of  $k$  plus  $d$  with  $N$ :  $\Delta \ln k + \Delta \ln d = 0$ , where  $\Delta$  denotes a population size variation. This conservation also enables each individual's capacity to innovate through processes of functional recombination resulting from re-wirings of ties in social networks. The main difference of this approach to that of Ref. 30, is to start from the more fundamental assumption that economic productivity is proportional to (labor) specialization that is  $w \sim 1/d$ , and not just to social connectivity,  $k$  and is subject to network constraints. To preserve physical dimensions we write  $w = g(kd)/d$ , where  $g$  is a function that transforms the index of specialization  $1/d$  into units of economic productivity (money/person/unit time). Its dependence on  $kd$  is necessary for consistency with observations.

With these definitions, we can formulate the problem of determining the optimal professional diversity of a city of size  $N$  in terms of the  $d$  that maximizes (labor) economic productivity,  $w$ , subject to the constraint that activities lost to an individual remain available through her neighbors in their social network. Note that because  $d = A/k$ , by assumption, that this is equivalent to finding the value of  $A$  and the function  $g(A)$  that maximize productivity. This can be written in terms of standard constrained optimization of a target function  $\mathcal{L}$ , as

$$\mathcal{L}(d; \lambda) = \frac{g(kd)}{d} - \lambda(kd - A), \quad (6)$$

where  $\lambda$  is a Lagrange multiplier. To proceed, we take  $k(N)$  as given and show how social connectivity and professional diversity must be related in order to maximize labor productivity. The solution follows by taking the variations of  $\mathcal{L}$  relative  $d$  and  $\lambda$  to zero:

$$d = \frac{A}{k} = \frac{A}{k_0 N^\delta}, \quad w = \frac{g(A)}{A} k = \frac{g(A)}{A} k_0 N^\delta, \quad (7)$$



**Figure 3 | The scaling of economic productivity with city size and the generation of professional diversity.** (A) Superlinear scaling relation between total wages in US metropolitan ( $W(N) = W_0 N^{1+\delta}$ , with  $\delta = 0.18 \pm 0.03$ , in agreement with theoretical expectations of  $\delta \simeq 1/6^{30}$ ). Inset shows the product  $A = wd$  versus city size, demonstrating that  $A$  it is on average independent of city size ( $R^2 = 0.91$ ). (B) The process of generation of new occupations as a function of city size: As the self-similar frequency distribution of occupations is pushed up by city size, latent occupations cross the lowest probability threshold and appear explicitly in  $D$ . The schematic shows how such a new explicit function (red circle) allow other occupations (yellow and green circles) to specialize further (acquire brighter colors), while remaining available to each other through network ties.



$$\frac{dg}{dA} - \frac{g}{A} - \lambda_1 A = 0 \rightarrow g(A) = \left[ C + \int^A dA' \lambda_1(A') \right] A, \quad (8)$$

where  $C$  is constant of integration;  $\lambda_1 \equiv \lambda/k$  and is a function of  $A$ . Both  $C$  and  $\lambda_1$  are to be set by boundary conditions on  $g(A)$ , the baseline for labor productivity across cities at a given time. From measurement of  $g(A)$ , we then obtain  $A$ , and the relationship between productivity, diversity and connectivity, since  $A = D_0 k_0$ .

These results derive the aggregate patterns of urban occupational diversity for all city sizes from microscopic considerations of the division and coordination of labor in cities and show the sense in which the patterns of diversity observed empirically above are optimal through their relation to individual productivity under network constraints.

## Discussion

We showed that US cities express a general statistical pattern of professions, where greater individual specialization coexists with increases in overall diversity as cities become larger. We demonstrated that these patterns of diversity are optimal, in the sense of providing the basis for open-ended increases in economic productivity per capita with city size while preserving the overall functionality available to each individual. This dynamics of differentiation and coordination at the individual level promotes increases in the quality of these functions available in larger cities and thus in urban systems by encouraging learning and the accumulation of human capital, enabled by the scaling properties of urban social networks. It also provides a general microscopic theory, based on fundamental principles of economic organization<sup>32,33</sup>, that derives observed increasing returns to scale in economic productivity at the metropolitan level consistent with data for thousands of cities worldwide<sup>30</sup>.

These ideas for the origin of functional diversity unify important concepts about the role of cities as generators of heterogeneity and innovation. The view of cities as “social reactors” from complex systems theory<sup>30</sup> provides the underpinnings for this dynamics by deriving the social and infrastructural conditions (coordination costs) under which the process of specialization can happen and remain open-ended. They also provide a context for ideas of innovation as a recombinatory process<sup>37</sup>, a framework often invoked to explain technological development: The growth in social connectivity with city size and the possibilities for the division of labor it affords, promote the appearance of explicit new functions, such as new professions. These functions can then be recombined in organizations and products, which in turn may stimulate the development of new functions and so on. Thus, the information about more elaborate products and services becomes increasingly encoded in social networks, rather than being contained in the knowledge of single individuals. Given inevitable individual cognitive constraints<sup>38</sup> this seems to be a necessary condition for socioeconomic complexity to develop as an open-ended process.

The framework developed here also unifies two distinct points of view that are often thought to be at odds: Growing diversity, in this picture, is the emergence of greater specialization. This demonstrates how diversity and specialization can be two facets of the same process for realizing increasing productivity gains through knowledge specialization in social organizations. In this picture, it is the transfer of already existing incipient functions from generalist individuals to growing social networks of interdependent specialists, see Fig. 3B, that supports their close integration and coordination. It also shows how system size (‘the extent of the market’<sup>32</sup> as the metropolitan area) is key in enabling these dynamics as it determines, indirectly<sup>30</sup>, both productivity gains and coordination costs<sup>33</sup>.

Where these two points of view - diversity and specialization - differ is in the unit of analysis at which they take place. Here, we showed that patterns of differentiation are predominately local in social networks and can be observed at the individual level. (The

original division of labor in human societies – by sex and by age – takes place in closely-knit cohabiting groups<sup>39</sup>). However, this does not exclude that processes of specialization and interdependence may also take place at higher levels of organization, such as firms, cities or even nations, whenever coordination costs are sufficiently low. In fact, a similar analysis at the establishment level (places of employment) reveals that there is less specialization in larger cities (a smaller increase in diversity) in terms of business types relative to professions<sup>40</sup>. This result implies that businesses classified as the same type in cities of different sizes contain, on the average, a more diverse set of occupations and realize greater productivity (e.g. in terms of value-added<sup>41</sup>) in larger cities. At the national level, scale (population) plays less of a role in determining economic diversity or productivity<sup>8</sup> (c.f. China or India vs. the Netherlands or Singapore), though these two characteristics remain firmly correlated, and intimately tied to levels of urbanization. In fact, urban hierarchy theories from central place theory<sup>28</sup> to more recent economic geography<sup>29</sup>, assume the existence of a general hierarchy of ‘services’ common to all cities that grows deeper with city size. In this picture, as we have demonstrated quantitatively here, the economic diversity,  $D$ , of a national economy is set by that of its largest city.

This evidence suggests that the development of fundamental theory linking diversity, productivity and scale in complex systems must be guided by the integrated analysis of microscopic productivity gains obtained via specialization and learning with associated coordination costs mediated by (social) networks of interdependence<sup>33</sup>. They also show how resilience deriving from greater diversity may be an unintended consequence of a general process of knowledge specialization in complex networks. Similar analysis based on different socioeconomic units (such as firms, cities or nations) may shed light on the necessary conditions for these entities to realize analogous patterns of growth and development and help establish a more general quantitative framework for this sort of analysis in other complex systems. In this respect, it is likely an important clue to theory that similar quantitative patterns of functional diversity characterize the technological complexity of at least some simpler (pre-urban) human societies<sup>42</sup> and may be a property of other networked systems that can experience increases in their productivity with scale. The reversibility of these processes, e.g. the absence of hysteresis in the externalization and reabsorption of functions by individuals interconnected by networks may also underlie the resilience of many complex systems<sup>5,20,43</sup> under functional change or population loss.

## Methods

**Urban Units of Analysis.** We adopt the standard definition of functional cities as metropolitan statistical areas (MSAs). MSAs are collections of political units (counties) aggregated by the US Census Bureau based on a set of criteria that includes population size, density and commuting flows. There were 403 MSAs in 2010, providing an ample basis for studying occupational diversity across city size (from 50K to 20M inhabitants) as well as other urban characteristics. MSAs are integrated labor markets and the best current official definition of functional cities in terms of a mixing population<sup>30</sup>.

**Data Sources.** Data on professional occupations in US MSAs used is the Occupational Employment Statistics (OES) obtained from the Bureau of Labor Statistics (BLS)<sup>44</sup> and is freely available online (<http://www.bls.gov/oes/>). Because the OES survey is a semi-annual mail survey of non-farm establishments it does not include self-employment. All occupations reported by employers in the US economy are hierarchically classified on the basis of their similarities at different levels of aggregation based on the BLS’s Standard Occupational Classification Scheme, which contains a total of 840 distinct professions at its finest ( $r=6$ -digit) level of resolution for 2010. It is important to notice that while this classification scheme is the finest occupation classification available at the MSA level, it does not actually represent a job that someone does, but instead a group of job titles performed on average by 7 specific jobs. Such even finer classification is only available at the national level (<http://www.bls.gov/soc/#match>). However, the inclusion of such definition do not alter the conclusions reached here as it only multiplies the diversity of occupation by a factor of 7 on average (See Supplementary Fig. S.4 and Table S.3).

**Fit methodology.** For each year, we estimated the parameters in Eq. 1 at  $r=6, 5, 4$ , and 3-digits level occupations (Fig. 1A–B) through ordinary least squares, using the



Gauss-Newton method, which relies on linear approximations to the nonlinear mean function<sup>45</sup>. We then used  $\gamma$ ,  $d_0$ ,  $N_0$  estimated for 2010 to fit Eq. 2. For visual purposes, a constant  $i_0 = 3$  is introduced to Eq. 2 to account for the initial curvature observed for most common occupations (Fig. 2B). Hence the fit in Fig. 2B corresponds to

$f(i; N) = NN_0^{-1} [(d_0 - i - i_0)(i + i_0)^{-1}]^{1/2}$ . Note that  $i_0$  is not determined by the process of analytic continuation, valid only at high ranks, used to obtain the form of the frequency distribution and constitutes from that point of view a functional freedom that can be set by the boundary condition of observing  $f(i; N)$  for sufficiently small cities, where  $D(N)$  is of order  $i_0$ .

- Page, S. E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies* (Princeton University Press, Princeton, NJ, 2007).
- May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, Princeton, NJ, 2001).
- Quigley, J. M. Urban Diversity and Economic Growth. *J Econ Perspect* **12**, 127–138 (1998).
- Bonner, J. T. *The Evolution of Complexity by Means of Natural Selection* (Princeton University Press, 1988).
- Tilman, D. et al. Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
- Dobson, A. P. et al. Sacred Cows and Sympathetic Squirrels: The Importance of Biological Diversity to Human Health. *PLoS Medicine* **3**, 714–718 (2006).
- Jacobs, J. *The Economy of Cities* (Random House, New York NY, 1969).
- Hidalgo, C. A., Klingler, B., Barabási, A.-L. & Hausmann, R. The Product Space Conditions the Development of Nations. *Science* **317**, 482–487 (2007).
- Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10570–5 (2009).
- Jacobs, J. *The Nature of Economies* (Random House, New York NY, 2000).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, UK, 1930).
- Frank, S. A. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**, 1712–1729 (1997).
- Jacobs, J. *Cities and the Wealth of Nations* (Random House, 1984).
- Dengler, J. Which function describes the species' area relationship best? A review and empirical evaluation. *J. Biogeogr.* **36**, 728–744 (2009).
- Jacobs, J. *The death and life of great American cities* (Random House, 1961).
- Durantón, G. & Puga, D. Micro-foundations of urban agglomeration economies. *Urban. Stud.* **37**, 533–555 (2000).
- Henderson, J. V. The sizes and types of cities. *Am. Econ. Rev.* **64**, 640–656 (1974).
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. & Shleifer, A. Growth in Cities. *J. Polit. Econ.* **100**, 1126–1152 (1992).
- Henderson, J. V., Kuncoro, A. & Turner, M. Industrial development in cities. *J. Pol. Econ.* **103**, 1067–1090 (1995).
- Marshall, C. R. Marine biodiversity dynamics over deep time. *Science* **329**, 1156–7 (2010).
- Hirschman, A. O. *National power and the structure of foreign trade* (University of California Press, 1980).
- Cover, T., Thomas, J. & Wiley, J. *Elements of information theory* (Wiley-Interscience, Hoboken NJ, 1991).
- Mori, T., Nishikimi, K. & Smith, T. E. The Number-average size rule: A new empirical relationship between industrial location and city size. *J. Reg. Sci.* **48**, 165–211 (2008).
- Magurran, A. *Measuring biological diversity* (Wiley-Blackwell, Hoboken NJ, 2004).
- Stanley, H. *Introduction to phase transitions and critical phenomena* (Oxford University Press, Oxford, UK, 1987).
- Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301–6 (2007).
- Mameli, F., Faggian, A. & McCann, P. Employment growth in Italian local labour systems: Issues of model specification and sectoral aggregation. *Spatial Economic Analysis* **3**, 343–360 (2008).
- Christaller, W. *Central places in southern Germany* (Prentice Hall, Upper Saddle River, NJ, 1966).
- Hsu, W.-T. Central Place Theory and City Size Distribution. *Econ. J.* **122**, 903–932 (2012).
- Bettencourt, L. M. A. The Origins of Scaling in Cities. *Science* **340**, 1438–1441 (2013).
- Schläpfer, M. et al. The Scaling of Human Interactions with City Size. *ArXiv preprint*. 1210.5215 (2012).
- Smith, A. *An Inquiry into the Nature and Causes of The Wealth of Nations* (W. Strahan, London UK, 1776).
- Becker, G. S. & Murphy, K. M. The Division of labor, coordination costs, and knowledge. *Q. J. Econ.* **CVII**, 1137–1160 (1992).
- Sun, G.-Z. & Yang, X. Agglomeration economies, division of labour and the urban land rent escalation: a general equilibrium analysis of urbanisation. *Aust. Econ. Pap.* **41**, 164–84 (2002).
- Glaeser, E. L. & Resseger, M. G. The Complementarity between Cities and Skills. *J. Reg. Sci.* **50**, 221–244 (2010).
- Lucas Jr, R. E. On the mechanisms of economic development. *J. Monetary Economics* **22**, 3–42 (1988).
- Arthur, W. B. *The Nature of Technology: What it is and How it Evolves* (The Free Press and Penguin Books, New York NY, 2009).
- Simon, H. Bounded Rationality and Organizational Learning. *Organization Science* **2**, 125–134 (1991).
- Johnson, A. W. & Earle, T. *The Evolution of Human Societies* (Stanford Univ. Press, Stanford CA, 2000).
- Youn, H.-J. et al. The systematic structure and predictability of urban business diversity. *ArXiv preprint*. 1405.3202 (2013).
- Sveikauskas, L. The productivity of cities. *Q. J. Econ.* **89**, 393–413 (1975).
- Kline, M. A. & Boyd, R. Population size predicts technological complexity in Oceania. *P. Roy. Soc.-Biol. Sci.* **277**, 2559–64 (2010).
- Scheffer, M. et al. Anticipating Critical Transitions. *Science* **338**, 344–8 (2012).
- OES. *Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics* (2011).
- Bates, D. M. & Watts, D. G. *Nonlinear Regression Analysis and Its Applications*. Wiley Series in Probability and Statistics eds. Bates, D. M. & Watts, D. G. (John Wiley & Sons, Inc., Hoboken, NJ, USA, 1988).

## Acknowledgments

We thank Doug Erwin, Ricardo Hausmann, Cesar Hidalgo, José Lobo and Geoffrey West for discussions. This research was partially supported by the Rockefeller Foundation (grant no. 2011 SRC 108), the James S. McDonnell Foundation (grant no. 220020195), the John Templeton Foundation (grant no. 15705), the Army Research Office Minerva Program (grant no. W911NF1210097), the Bill and Melinda Gates Foundation (grant OPP1076282), the U.S. Department of Energy through the LANL/LDRD Program (contract no. DE-AC52-06NA25396), and a FONDEF grant (D10I1038).

## Author contributions

L.B. and H.S. designed the research; H.S. performed the analysis; L.B., H.S. and H.Y. discussed the results; L.B. and H.S. wrote the paper.

## Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Bettencourt, L.M.A., Samaniego, H. & Youn, H. Professional diversity and the productivity of cities. *Sci. Rep.* **4**, 5393; DOI:10.1038/srep05393 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

# Supporting Information for: Professional diversity and the productivity of cities

Luís M. A. Bettencourt<sup>1</sup>, Horacio Samaniego<sup>2,3\*</sup>, HyeJin Youn<sup>1</sup>

<sup>1</sup> Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA,

<sup>2</sup> Center for Non Linear Studies, Theoretical Division MS B284,

Los Alamos National Laboratory, NM 87545, USA,

<sup>3</sup>Instituto de Conservación, Biodiversidad y Territorio,

Universidad Austral de Chile, Valdivia, Chile

## Asymptotic relation between diversity and city size holds over time

The relation between diversity of employment and the size of urban area is maintained across time (Fig. S.1). The diversity of professions in the urban area is well described by the relation:

$$d(N) = d_0 \frac{\left(\frac{N}{N_0}\right)^\gamma}{1 + \left(\frac{N}{N_0}\right)^\gamma}, \quad (\text{S.1})$$

where  $d_0$  is the effective size of the classification scheme (maximum number of occupations),  $N_0$  is the size of the city where saturation starts to occur and  $\gamma$  the characteristic exponent describing how  $d$  increases with city size, in the absence of saturation.

---

\*To whom correspondence should be addressed; E-mail: horacio@ecoinformatica.cl

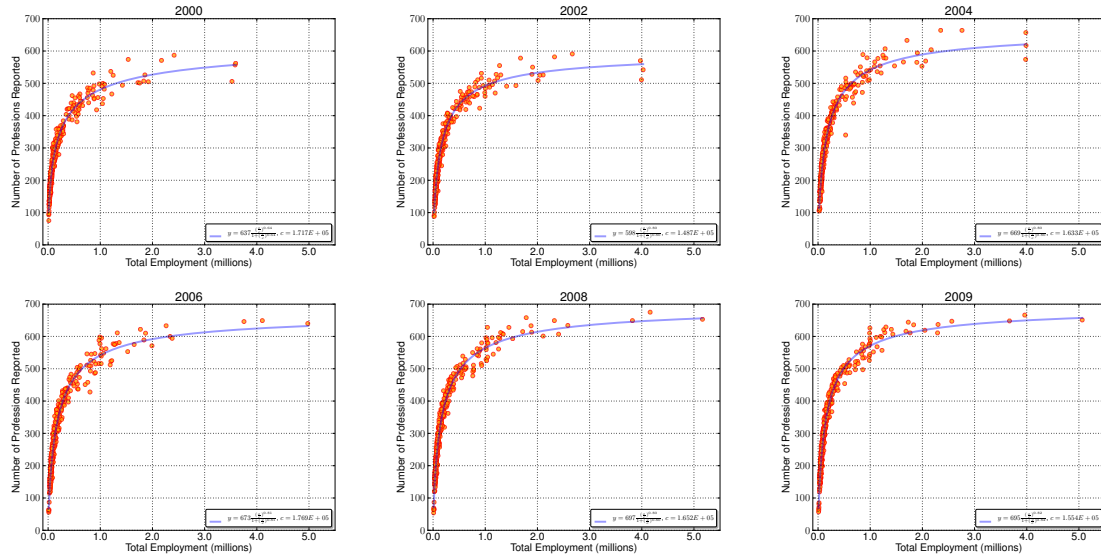


Figure S.1: The functional relationship between diversity of professions,  $d$ , and city size holds over time

## Variation of $\gamma$ over time

Variation of  $\gamma$  over time is fairly stable. The employment classification scheme has been refined over time with an important redefinition in 2002, which may explain larger fluctuations (and lower values) of  $\gamma$  prior to such date.

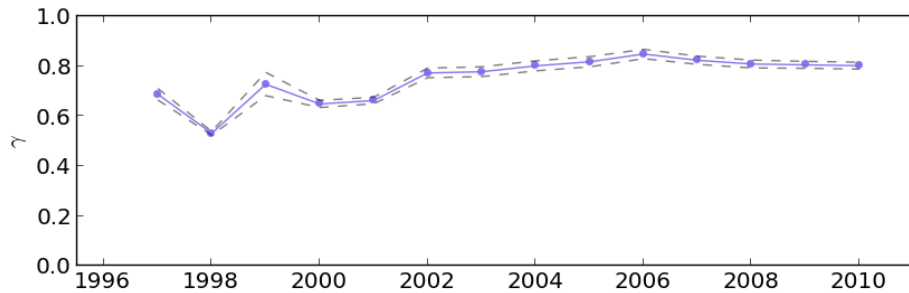


Figure S.2: Changes in the best fit value of the exponent  $\gamma$  across time.



## The number of distinct professions scales sub linearly with city size

**Theorem** *The number of distinct professions scales sublinearly with city population size.*

**Proposition:** For any not fully specialized city with  $N$  inhabitants (or workers), if professions are a property of individuals that cannot be accumulated and the total number of distinct professions  $D$  is a scale invariant function of  $N$  then its exponent  $\gamma < 1$  (sublinear scaling).

**Proof:** We take a profession as a property of a person that cannot be accumulated. As such, the total number of distinct professions in a city  $D$ ,  $\frac{D}{N} \leq 1$ . From this it follows that

$$\ln D \leq \ln N \rightarrow \frac{\Delta D}{D} \leq \frac{\Delta N}{N}, \quad (\text{S.2})$$

where  $\Delta$  denotes a variation. Because  $D$  is a scale invariant function of  $N$ ,  $D = D_0 N^\gamma$ , then

$$\frac{\Delta D}{D} = \gamma \frac{\Delta N}{N}, \quad (\text{S.3})$$

for some real number  $\gamma$ , independent of  $N$ . Then, from Eq. S.2 and Eq. S.3, it follows that  $\gamma \leq 1$ . Because no city is fully specialized, the inequality holds strictly and  $\gamma < 1$ , as stated.

## Matched jobs to occupation

While each occupation provided by BLS is only an approximation of the diversity of job titles in each cities, it accurately represent the functional relation between occupation/job

diversity with city size (Fig. 1a). The Standard of Occupation Classification (SOC) provides an equivalence of the actual job titles matched within each occupation at the national level. We use this 2010 data to estimate the difference of simple occupation and job-adjusted occupation to evaluate how job titles scales with the size of the employment per msa.

Table S.1: Number of job titles per occupation declared at the national level for 2010. The 15 out of 840 occupations with the largest number of matched jobs titles is shown.

ID	Jobs	Occupation Title
840	109	Military Officer Special and Tactical Operations Leaders, All Other
839	71	Command and Control Center Officers
838	68	Military Enlisted Tactical Operations and Air/Weapons Specialists and Crew Members, All Other
837	65	Radar and Sonar Technicians
836	54	Artillery and Missile Crew Members
835	53	Air Crew Members
834	43	Command and Control Center Specialists
833	41	Athletes and Sports Competitors
832	36	Musicians and Singers
831	35	First-Line Supervisors of All Other Tactical Operations Specialists
830	34	Lawyers
829	30	Artillery and Missile Officers
828	28	Fine Artists, Including Painters, Sculptors, and Illustrators
827	27	Captains, Mates, and Pilots of Water Vessels
826	25	Vocational Education Teachers, Postsecondary

The median and average of jobs per matched occupation is 7 and 7.8 respectively with only 12 occupations (out of 840) having over 30 jobs (mostly military occupations). Hence, while the diversity of jobs is certainly higher than the number of occupations in the dataset, the differences do not seem to affect the analysis portrayed here.

Table S.2: Descriptive statistics for the number of matched jobs per occupation declared at the national level for 2010.

Statistics	count	min	25%	50%	mean	75%	max
Value	840	1	4	7	7.85	9	109

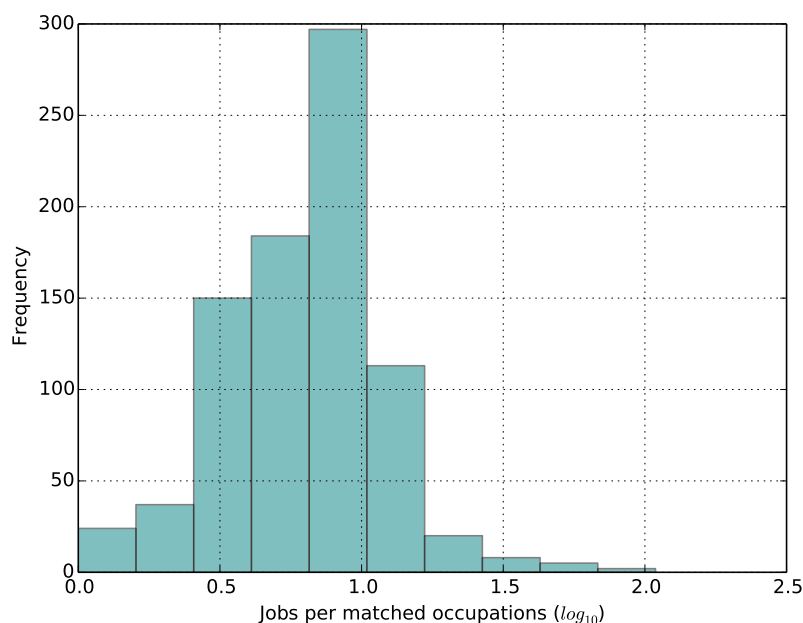


Figure S.3: Distribution of number of matched jobs per occupation declared at the national level for 2010

To further understand if this would affect our estimation of occupation diversity across cities ( $d$ ), we additionally plotted the diversity of occupations per city multiplied by the number of matched jobs for each occupation in 2012 (Fig. S.4). This graph shows that while there is indeed a larger diversity for each city, the correction to  $d$  does not affect the parameters value. of gamma and, most importantly, the data still fits Eqn. 1, suggesting

that our analysis remains insensitive even if this correction could be applied at the city level.

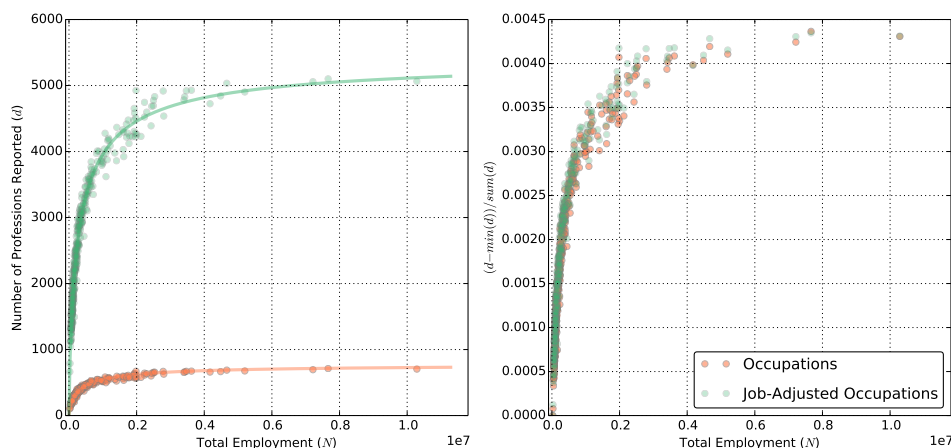


Figure S.4: Number of Occupations ( $d$ ) and job-adjusted occupations do not show significant differences with respect to city size. (See Table S.3 for parameter estimation values)

Table S.3: Fitted parameter estimation and 95% confidence interval for job-adjusted occupation diversity. (See Fig. S.4)

	$d_0$	$N_0 \times 10^5$	$\gamma$
Occupations	760 [740,780]	2.8 [2.7,3.2]	0.74 [0.71,0.77]
Adjusted Occupations	5442 [5301,5582]	2.9 [2.5,3.0]	0.77 [0.73,0.79]

## Indices of occupation diversity

Both Herfindahl–Hirschman (HH) index and Shannon entropy show consistent behavior to describe the diversity of occupations among cities of increasing sizes. Note that HH index decreases with  $N$  towards a small constant, set by the exponent  $\delta$ , as cities grow. Similarly, the increase in Shannon entropy signals the increase in diversity of the occu-

pational distribution in larger cities. Note that in both cases the increases in diversity are driven, at leading order, by a term of order  $N^{-\delta}$  (see main text).

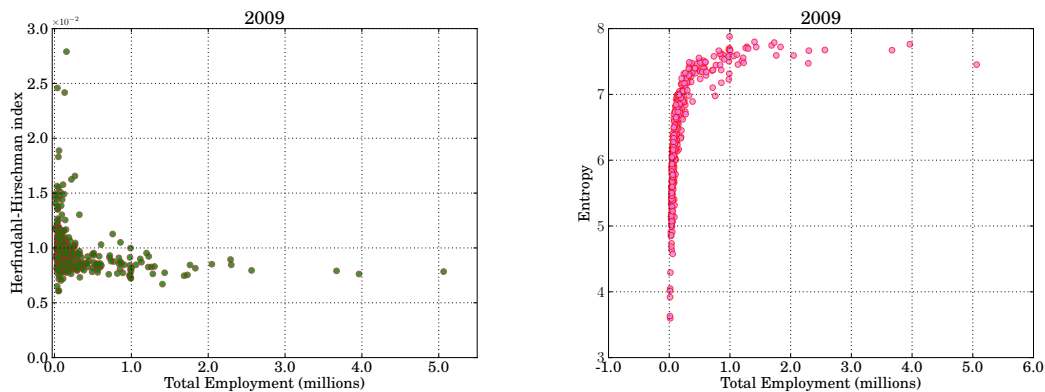


Figure S.5: Measures of occupation diversity in U.S. urban systems. (a) Herfindahl–Hirschman Index; (b) Shannon entropy. Both plots show the results of evaluating these indices at  $r = r_6$ , thus showing saturation in these measures of diversity for large cities.