Complex economic activities concentrate in large cities

Balland, Pierre Alexandre; Jara-Figueroa, Cristian; Petralia, Sergio G.; Steijn, Mathieu P.A.; Rigby, David L.; Hidalgo, César A.

Link to publication in VU Research Portal

Complex economic activities concentrate in large cities

Pierre-Alexandre Balland [1,2]*, Cristian Jara-Figueroa [2], Sergio G. Petralia [3], Mathieu P. A. Steijn [1,4], David L. Rigby [5] and César A. Hidalgo [2,6,7,8]

Human activities, such as research, innovation and industry, concentrate disproportionately in large cities. The ten most innovative cities in the United States account for 23% of the national population, but for 48% of its patents and 33% of its gross domestic product. But why has human activity become increasingly concentrated? Here we use data on scientific papers, patents, employment and gross domestic product, for 353 metropolitan areas in the United States, to show that the spatial concentration of productive activities increases with their complexity. Complex economic activities, such as biotechnology, neurobiology and semiconductors, concentrate disproportionately in a few large cities compared to less--complex activities, such as apparel or paper manufacturing. We use multiple proxies to measure the complexity of activities, finding that complexity explains from 40% to 80% of the variance in urban concentration of occupations, industries, scientific fields and technologies. Using historical patent data, we show that the spatial concentration of cutting-edge technologies has increased since 1850, suggesting a reinforcing cycle between the increase in the complexity of activities and urbanization. These findings suggest that the growth of spatial inequality may be connected to the increasing complexity of the economy.

With a gross domestic product (GDP) of US$1.4 trillion, the New York metro area generates more wealth than Australia, Spain or Mexico. With 1.39 patents per 1,000 people, the San Francisco Bay Area produced, in 2000, more than 12% of all of the patenting activity of the United States. Economic activities are known to concentrate in space, and that concentration appears to be increasing. In 15 years, the Bay Area more than doubled its rate of invention, growing to nearly 20% of all patents produced in the United States in 2015. But what factors explain this unprecedented concentration of knowledge and wealth in large cities? And why has the spatial concentration of activities increased in a world dominated by digital communications and international travel?

One factor may be the growth of complex economic activities: activities requiring a deep division of knowledge and labour. As an example, consider the division of labour involved in producing a single research paper in immunology. Immunology contributions usually require collaborations among people with narrow and complementary expertise. You may need experts in specific pathways and proteins, such as NF-κβ or Toll-like receptors, people experienced in in vivo murine biology and people with experience in a variety of laboratory techniques, such as flow cytometry. Depending on the nature of the contribution, you may also need to include people with clinical experience, which—once again—can be specific for each autoimmune disorder. This deep division of knowledge and labour is required in fields such as immunology or microbiology because it is not possible to accumulate all of that expertise in one or two people. In simple words, we can say that the complexity of this activity is large, not because each of the individuals involved is more skilled than people working in other activities, but because the activity requires a large network of people with deep expertise in complementary knowledge domains.

Our hypothesis is that complex industries, such as biotechnology and semiconductors, exhibit a much greater degree of spatial concentration than less complex industries, such as apparel and furniture manufacturing. This could help explain the rise in importance of superstar cities[1], and also contribute to our understanding of growing spatial inequality. In fact, as we show in this paper, the complexity of activities can account for approximately 40% to 80% of the variance in urban concentration across occupations, industries, scientific fields and technologies. This differs from literature in urban economics focused on the urban sorting of college graduates[2,3], instead of the complexity of innovative and productive activities.

We can draw a link between urban concentration and the complexity of economic activities by combining recent advances from development economics[4–7] and urban scaling[8–10]. On the one hand, scholars working on economic development have created methods to measure the complexity of economies (for example, countries and cities) and that of the activities present in them (for example, products and patents)[4,5]. On the other hand, scholars working in urban concentration have shown that output scales superlinearly with a city's population, which means that output per capita is larger in bigger cities[8–10]. This superlinear scaling is known to vary across activities[11], although it is unclear why. Here, we bring these two bodies of literature together by revealing that the urban concentration of activities increases with their complexity.

Why should we expect a link between complexity and spatial concentration? More complex activities require a deeper division of knowledge[12], compelling individuals to narrow down their expertise and specialize[13]. This division of knowledge creates high coordination costs, since specialized knowledge needs to be reconstituted to be put to work. Cities help solve the coordination problems created by the division of knowledge by creating multiple mixing and

[1]Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, the Netherlands. [2]Collective Learning Group, The MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Economic Geography, London School of Economics, London, UK. [4]Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [5]Department of Geography, University of California, Los Angeles, CA, USA. [6]ANITI Chair, University of Toulouse, Toulouse, France. [7]Manchester Institute of Innovation Research, Alliance Manchester Business School, Manchester, UK. [8]Datawheel, Cambridge, MA, USA. *e-mail: balland@mit.edu

matching opportunities[10,14,15]. In fact, economists have found that during the last decades college graduates have increasingly sorted themselves into high-wage, high-rent cities[2,3]. Cities are also home to a variety of knowledge spillover mechanisms[16–20], such as labour flows[21–23], spin-offs[24] and dense social networks[25]. Moreover, cities are also the preferred location of multiple private and public institutions focused on accumulating complex knowledge, such as research universities and private laboratories[26]. Together, these multiple reinforcing channels provide the increasing returns expressed in the superlinear scaling of output across cities. We expect this superlinear scaling to be more pronounced for complex activities, since increasing returns are stronger in sectors that are more intense in knowledge than in labour or capital[27].

In the following pages, we explore the link between complexity and spatial concentration by first measuring the urban scaling[8,9] of papers, patents, scientific papers, occupation and industries, and then exploring whether differences in the observed scaling exponents are explained by an activity's level of complexity[4,5]. Our findings show that complex activities concentrate more in large cities than less complex activities, and that the spatial concentration of complex activities has increased over time, contributing to our understanding of spatial inequality[28] and of the spatial organization of the economy.

Results

Figure 1 shows the urban concentration of research papers (Fig. 1a), patents (Fig. 1b), occupations (Fig. 1c) and industries (Fig. 1d) in the United States. Peaks are respectively proportional to the number of patents, the number of research papers, GDP and the total employment of each metro area. In all four examples, we find activities to be highly concentrated, especially in large cities. Figure 1e–h characterizes this urban concentration by showing the scaling laws followed by patents, research papers, industries and occupations. Scaling laws in cities follow power-law relationships of the form $y \approx x^{\beta}$, where x is the population of a city, y is a measure of output (patents, papers, GDP or jobs), and β is the scaling exponent. In the case of research papers (Fig. 1e), the number of papers published by authors in a metro area grows as the $\beta = 1.54$ power of the population. For patents, the patents granted to a city scale superlinearly with population with an exponent of $\beta = 1.26$. Similarly, total employment grows as the $\beta = 1.04$ power of the population in a metropolitan statistical area (MSA) and GDP scales as the $\beta = 1.11$ power of population.

We repeat this exercise by studying the scaling laws followed by specific research areas, technologies, occupations and industries. Figure 1j compares the scaling laws followed by patents in 'computer hardware and software' and 'pipes and joints'. Patents in 'computer hardware and software' concentrate more in large cities ($\beta = 1.57$) than patents in 'pipes and joints', which exhibit only a modest superlinear scaling ($\beta = 1.1$). Similarly, we observe large variations in the scaling coefficients of intuitively more and less knowledge intense research areas (Fig. 1i), occupations (Fig. 1l) and industries (Fig. 1k). Figures including each category of patents, papers, industries and occupations are available in Supplementary Section 2.

In Fig. 2, we explore the relationship between the urban concentration and the complexity of activities. For technologies, we proxy knowledge complexity using the age of the knowledge combined in patents, measured as the average year of appearance of the subclasses in which a patent makes a knowledge claim. Alternatively, we use the average number of inventors in a patent and the NK complexity measure of Fleming and Sorensen[29] (see Supplementary Information). The year of appearance of a subclass assumes that patents recombining more recent knowledge are, on average, more complex[30]. For scientific fields, we proxy knowledge complexity as the average size of the team involved in a scientific publication[31,32]. For occupations and industries, knowledge complexity is proxied

by the average years of education of the employees working in an occupation or industry. As we compare the spatial concentration of activities with their knowledge complexity, we avoid using measures of complexity that are derived from spatial information[4]. For more information about these definitions and robustness analyses, see Supplementary Section 3.

Figure 2a–d compares the urban concentration of activities with their respective scaling exponents. In all cases, we observe that the spatial concentration of activities increases with their complexity. For scientific fields, it increases with the average number of authors in a paper (Pearson's $r = 0.72$, $P < 1 \times 10^{-3}$), for technologies, it increases with the recency of the combined subclasses (Pearson's $r = 0.82$, $P < 1 \times 10^{-3}$) and with the average number of inventors in a patent ($r = 0.48$, $P < 1 \times 10^{-2}$), for occupations, it increases with the average years of education of the workers within that occupational category (Pearson's $r = 0.62$, $P < 1 \times 10^{-3}$), and for industries, it increases with the years of education of the workers employed in that industry (Pearson's $r = 0.70$, $P < 1 \times 10^{-3}$). In all four cases, the more complex the activity, the more superlinearly it scales with population, meaning that more complex activities concentrate more in large cities. We confirm the statistical significance of this relationship using regression analysis and a variety of alternative measures of spatial concentration and complexity (see Supplementary Section 3). For instance, we find that more complex patents of the same age still concentrate more in cities. In the Supplementary Information, we set up a model predicting the number of patents produced in a city in a technology as a function of a city's population, the NK complexity and recency of a technology, and an interaction term between a city's population and each measure of complexity (see Supplementary Information), finding that both measures of complexity are mutually significant.

Next, we examine historical patent data, spanning 150 years, to investigate whether the spatial concentration of activities has increased over time.

Figure 3a shows the scaling exponent observed for the top 25% most complex patents—those that recombine newer knowledge—granted each decade between 1850 and 2010 (red line). The figure reveals that the urban concentration of the most complex technologies has increased continuously for the past 150 years, accelerating with each industrial revolution. Starting with the second industrial revolution (1870), the urban scaling of complex technologies becomes increasingly superlinear, growing from a scaling exponent of $\beta \approx 1.15$ in 1870 to $\beta \approx 1.55$ by the 1930s. The urban concentration of the most complex patents then plateaus, increasing again after the third industrial revolution (1970s) and reaching a scaling exponent of almost 1.8 in 2010. The least complex patents (blue line) have always been less geographically concentrated than the most complex patents. After the 1970s, their urban concentration started to decrease, with the scaling exponent falling to less than 1.2. The information technology revolution has therefore been followed by an increasing concentration of the most complex technologies in cities, and a decreasing urban concentration of the least complex ones. Robustness analyses can be found in Supplementary Section 4.

We note that these results cannot be due to cities growing faster than rural areas, since cities becoming more populated relatively to rural areas would reduce the spatial concentration of the activities present in them (the β exponent). For the growth of urban areas to drive up the concentration of an activity, urban areas would need to generate employment in those activities faster than population growth.

We further our exploration of the evolution of the spatial concentration of patents by separating patents into the six main technological categories defined by the National Bureau of Economic Research: 'mechanical', 'chemical', 'electrical and electronic', 'computers and communication', 'drugs and medical' and 'others'. Figure 3b shows the scaling exponent observed for each of these technological
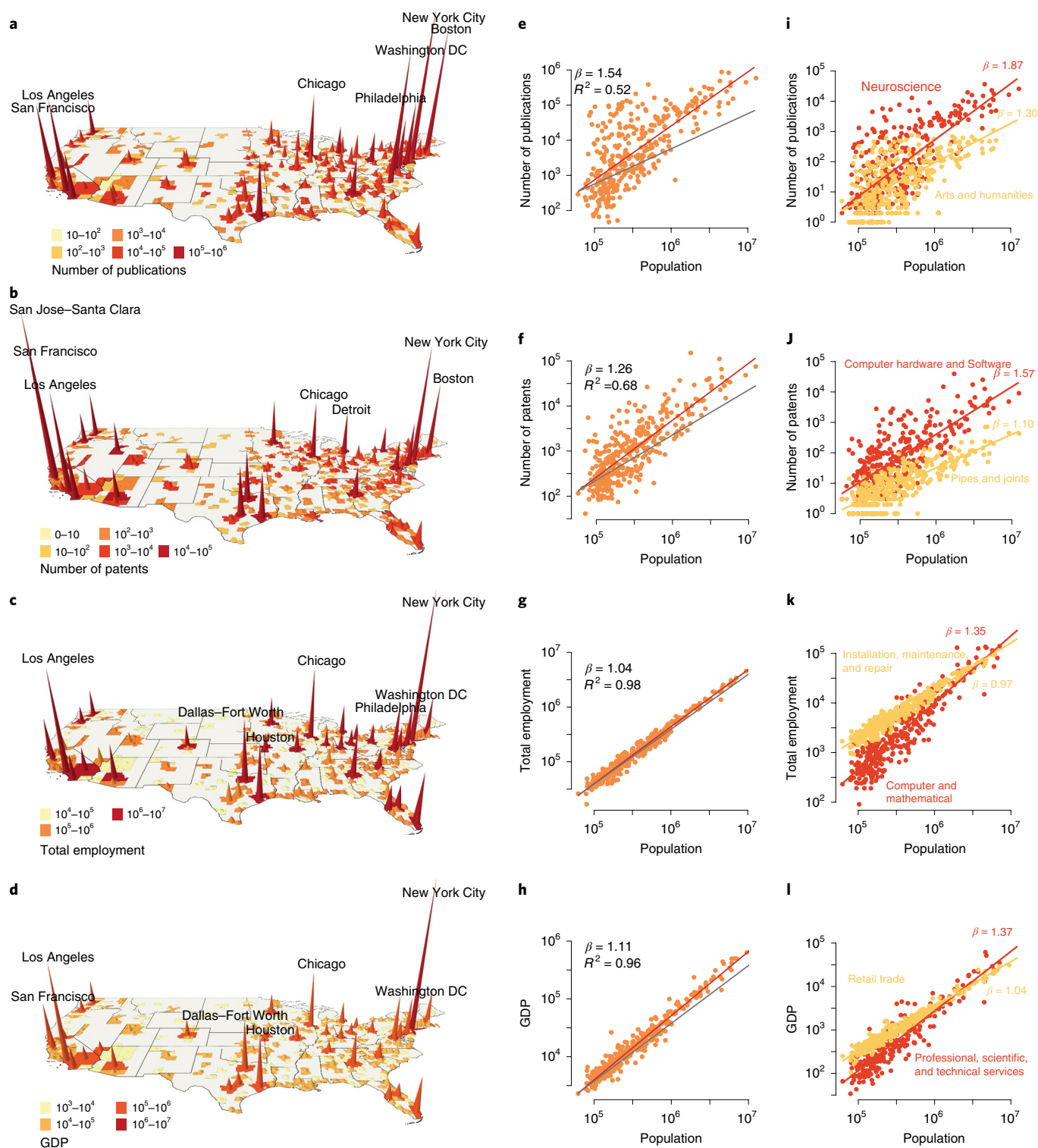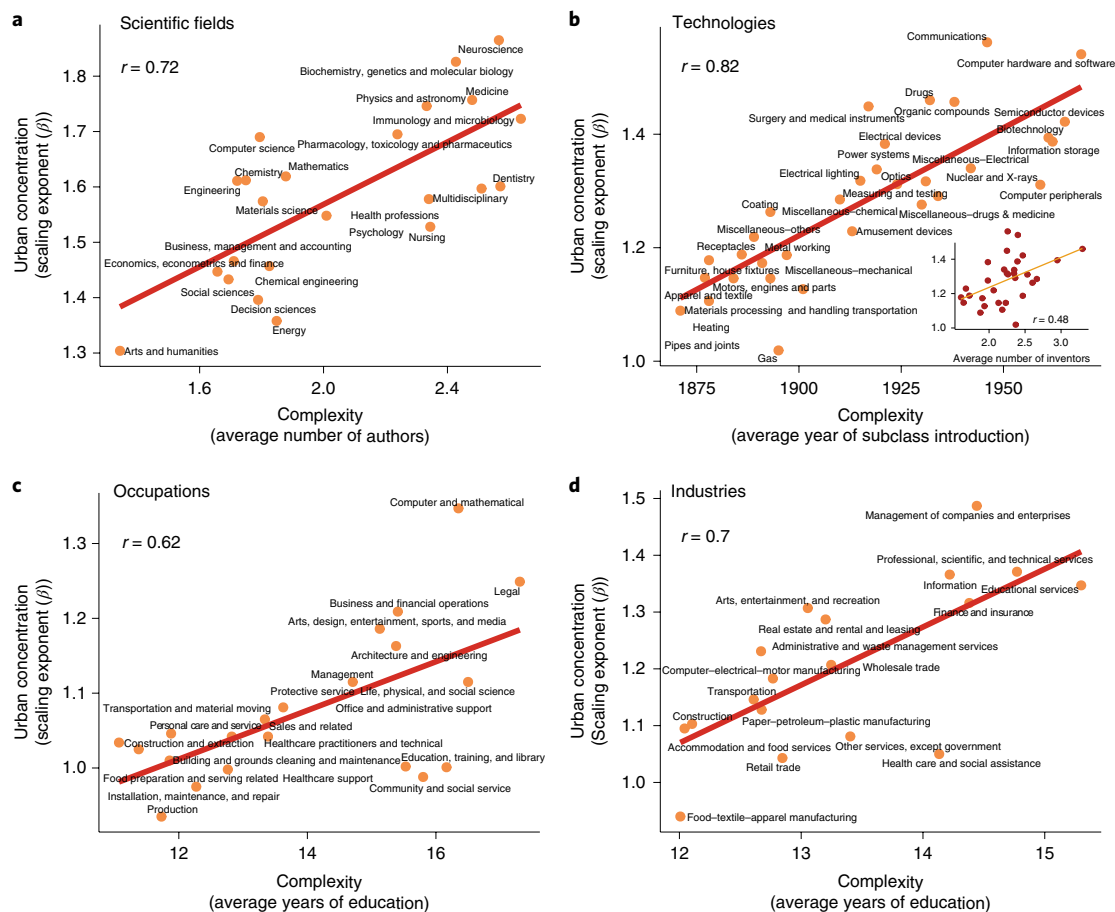
Fig. 1 | Spatial concentration of activities. a–d, Maps of the spatial concentration of research papers (**a**), patents (**b**), employment (**c**) and GDP (**d**) in the US MSAs. **e–h**, Scaling relations between the population of an MSA and the total number of research papers published from 1996 to 2008 (**e**), patents granted from 2000 to 2009 (**f**), the number of employees in 2015 (**g**) and the GDP of industries in 2015 (**h**). In **e–h** the grey line shows what would be expected if the scaling relationship was linear. **i–l**, Scaling relationships for pairs of activities with large differences in their scaling exponents: research papers in 'neuroscience' and 'arts and humanities' (**i**), patents in 'computer hardware and software' and 'pipes and joints' (**j**), employment in 'computer and mathematical' o-/ccupations and in 'installation, maintenance, and repair' (**k**) and output (GDP) of 'professional and scientific activities' and 'retail trade' (**l**).

categories by decade between 1850 and 2010. 'Mechanical' and 'others' are the technologies that exhibit the highest scaling in the mid-nineteenth century, meaning that they were the most concentrated

in large cities, with 'others' mostly composed of patents related to textiles during this period. However, the scaling exponent of these categories does not grow substantially during the following decades,

Fig. 2 | Urban concentration increases with knowledge complexity. a–d, Urban concentration of activities, as captured by the scaling exponent, increases with the average number of authors in a publication in a field (**a**), the average year of introduction of a patent's subclasses (and the average number of inventors in a patent (inset)) (**b**), the average years of education of workers within an occupational category (**c**) and the average years of education of the workers employed in an industry (**d**).

meaning that most of the rise in scaling observed after 1870 for all patents (Fig. 3a) can be attributed to an increase in the urban concentration of 'electrical and electronic' patents. Starting in 1950, 'computers and communications' and 'drugs and medical' become increasingly more concentrated, reaching the highest scaling exponents observed for all categories. Together, these results show that the urban concentration of patenting activity exhibits a long-term cycle, rising during the heyday of the technologies developed, and then declining as technologies mature.

Discussion

The core idea of this paper is that differences in the complexity of activities explain variations in the degree to which they agglomerate. We show this correlation to be true for the production of scientific papers and patents, for employment in multiple occupations, and for the output of industries. We argue that complex activities tend to be more concentrated in large urban areas because they require a deep division of knowledge that is distributed across many actors. This also suggests that much of the knowledge needed to perform complex activities—which may be tacit[33,34]—is subject to spillover mechanisms that are local, by virtue of being embodied in social networks[6], or because they flow together with labour[21,22], come from spin-offs[24] or require supporting local institutions, such as universities and private research centres[26]. Complex knowledge, therefore, does not travel well through digital communication channels and requires the richness of cities to be properly accumulated.

However, the dynamics of urban agglomeration are not static. When we look at the history of patenting activity in the US, we find that the concentration of patenting activity in urban areas has increased during most of the last century and a half, especially for more complex activities. During the third industrial revolution the concentration of the most complex and the least complex activities diverged. This could explain why the world has become flatter for some activities[35] and more spiky for others[36], and could be a reflection of an increase in the relative employment of more complex economic activities.

The idea that complex economic activities concentrate more in a few large cities poses a range of questions for planners and policymakers. It tells us that countries need to rethink their urbanization and transportation strategies. The economy of the future may well be dominated by those countries that succeed at developing the industrial policies[20,37–41] and megacities required to accumulate and process vast amounts of knowledge. Some countries in the East, especially China, have been successful at developing large megalopolises, such as the Yangtze River Delta Megalopolis (150 million people, including cities such as Shanghai, Nanjing and Hangzhou) or the Pearl River Delta Megalopolis (70 million people, including Hong Kong, Shenzhen and Guangdong). These megalopolises have been made possible by investments in transportation and housing that dwarf the efforts of the West. Of course, this does not mean that the West should copy the modernist efforts of China verbatim. However, it does suggest the need to think about the transportation,
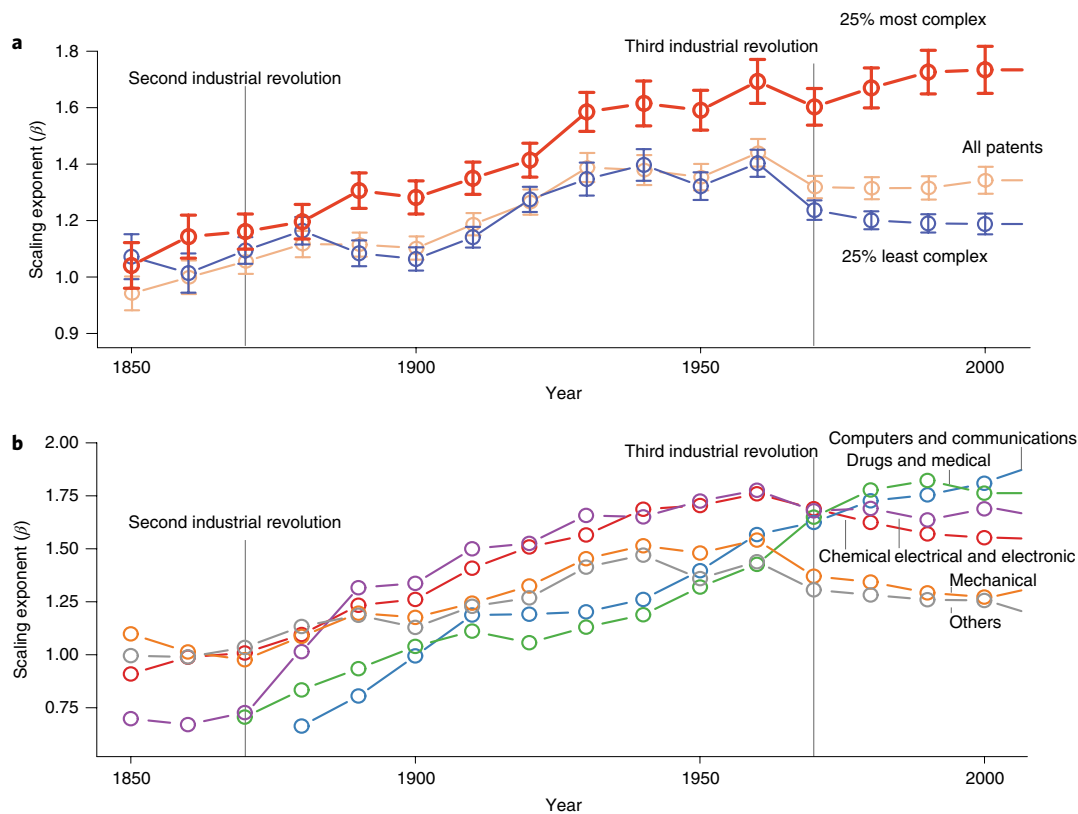
Fig. 3 | Evolution of the urban scaling of technologies. a, The scaling exponent of the top 25% most complex technologies increases throughout the observation period, while that of the bottom 25% of technologies based on complexity peaks in 1960 and then decreases. The scaling exponent for all patents increases from 1850 to 1930, and then remains relatively stable until 2000. **b**, The scaling exponent of the six main patent categories between 1850 and 2010.

housing and densification strategies that the West will need to develop to remain competitive in the high complexity activities of the future.

Of course, our study has limitations that need to be taken into consideration. The descriptive nature of our analysis does not provide a clear indication of the mechanisms connecting complexity and spatial concentration, or the causes leading to increases in both. For instance, a city's location within the global network of knowledge flows may provide an advantage to its ability of generate employment in complex activities. Furthermore, because we mostly use cross-sectional data, we cannot look at the dynamics of these relationships beyond the case of patenting activity. Moreover, United States Patent and Trademark Office (USPTO) classification recency is not a perfect measure of technological complexity, although it is correlated with other measures of technological complexity, such as NK and the average number of inventors in a patent.

If knowledge complexity and agglomeration cannot be divorced, the spatial inequality observed among large and small cities is likely to increase. Policymakers must recognize that the mechanisms generating growth and innovation may be the same as those that are contributing to the growth of inequality, both within and between cities. We face critical questions regarding the future of economic growth and the distribution of the returns and costs of economic activity within a world of growing unevenness[28] across multiple spatial scales.

Methods

We analyse the urban concentration of economic activities in the United States. We interchangeably refer to cities, metropolitan areas and urban areas. For the analysis, the spatial delimitation of these urban areas corresponds to the core-based statistical areas (CBSAs) defined by the United States Office of Management and Budget. CBSAs are statistical units, not administrative units. They are county aggregates, and refer to an urban area that includes a central city and the surrounding communities that are economically linked to this city (commuting patterns and shared labour market). Figures 1 and 2 report findings based on cities for which we have data on all four economic activities (patents, scientific publications, industries and occupations); that is, 353 MSAs. For Fig. 3, we analyse only historical patent data and are able to analyse patenting activity in all CBSAs (metropolitan and micropolitan statistical areas; that is, 923 observations). In the Supplementary Information, we present these three figures using alternative sets of cities. We use population data from the US Census (population in year 2000 for scientific publications and patents, estimated population in year 2015 for industries and occupations; for historical patents we use US Census population for each corresponding decade from 1850 to 2010). In the Supplementary Information, we also check the robustness of our results for occupations and industries using data for Brazilian cities (mesoregions).

We use patent data sourced from the USPTO, which provides inventor addresses for patents granted from 1975 onwards. For historical patents (1850–1974), we use HistPat. HistPat was built using optically recognized and publicly available documents from the USPTO, combining text-mining algorithms with statistical models to provide geographical information for older patents[42]. We disaggregate patents into 30 technologies as defined by the National Bureau of Economic Research (two-digit subcategories)[43]. For the 1975–2010 period, we use the Patent Network Dataverse. In total, we analyse 8,731,024 patents from 1850 to 2010. In the main text, we consider only technologies for which there are more than 200 cities with any recorded activity and we remove categories that are based on natural advantages: 'agriculture, husbandry, food', 'agriculture, food, textiles' and 'earth working and wells'. In the Supplementary Information, we also present results including these additional categories.

For scientific papers, we use publication data from Elsevier's Scopus database covering the time period 1996–2008[44,45]. Publications are disaggregated into 23 scientific disciplines as defined by the Scopus classification (two-digit major thematic categories). These data have kindly been provided by Ö. Nomaler, K. Frenken and G. Heimeriks. We analyse a total of 4,4000,000 scientific publications. The data include documents that have at least one author who has (at least) one affiliation to a US scientific organization. In the main text, we consider only

scientific fields for which there are more than 200 cities with any recorded activity and we remove categories that are based on natural advantages: 'agricultural and biological sciences', 'environmental science', 'Earth and planetary sciences' and 'veterinary'. In the Supplementary Information, we also present results including these categories.

For industries, we use 2015 GDP data from the Bureau of Economic Analysis to quantify the economic output of MSAs in 18 industries as defined by the North American Industry Classification System (two-digit NAICS). In the main text, we consider only industries for which there are more than 200 cities with any recorded activity and we remove categories that are based on natural advantages: 'agriculture, forestry, fishing, and hunting' and 'utilities'. In the Supplementary Information, we also present results including these categories.

For occupations, we use 2015 employment statistics from the Bureau of Labor Statistics disaggregated into 22 occupations according to the Standard Occupational Classification system (two-digit SOC). For occupations, we use 2015 employment statistics from the Bureau of Labor Statistics. In the main text, we consider only occupations for which there are more than 200 cities with any recorded activity and we remove categories that are based on natural advantages: 'farming, fishing, and forestry'. In the Supplementary Information, we also present results including this category. See Supplementary Section 1 for descriptive statistics on these data (Supplementary Figs. 1–4).

Data availability
The data that support the findings of this study are available from the corresponding author upon request.

Code availability
The code that supports the findings of this study is available from the corresponding author upon request.

References
1. Florida, R. *The New Urban Crisis: How Our Cities Are Increasing Inequality, Deepening Segregation, and Failing the Middle Class—and What We Can Do About It* (Basic Books, 2017).
2. Diamond, R. The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000. *Am. Econ. Rev.* **106**, 479–524 (2016).
3. Berry, C. R. & Glaeser, E. L. The divergence of human capital levels across cities. *Pap. Reg. Sci.* **84**, 407–444 (2005).
4. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl Acad. Sci. USA* **106**, 10570–10575 (2009).
5. Balland, P.-A. & Rigby, D. The geography of complex knowledge. *Econ. Geogr.* **93**, 1–23 (2017).
6. Hidalgo, C. *Why Information Grows: The Evolution of Order, from Atoms to Economies* (Basic Books, 2015).
7. Hausmann, R. et al. *The Atlas of Economic Complexity: Mapping Paths to Prosperity* (MIT Press, 2014).
8. 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. USA* **104**, 7301–7306 (2007).
9. Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–913 (2010).
10. Bettencourt, L. M. A. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
11. Youn, H. et al. Scaling and universality in urban economic diversification. *J. R. Soc. Interface* **13**, 20150937 (2016).
12. Warsh, D. *Knowledge and the Wealth of Nations: A Story of Economic Discovery* (W. W. Norton & Co., 2007).
13. Jones, B. F. The burden of knowledge and the "Death of the Renaissance Man": is innovation getting harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).
14. Scott, A. J. & Storper, M. The nature of cities: the scope and limits of urban theory. *Int. J. Urban Reg. Res.* **39**, 1–15 (2015).
15. Gans, J. S. & Stern, S. The product market and the market for "ideas": commercialization strategies for technology entrepreneurs. *Res. Policy* **32**, 333–350 (2003).
16. Feldman, M. P. & Audretsch, D. B. Innovation in cities: science-based diversity, specialization and localized competition. *Eur. Econ. Rev.* **43**, 409–429 (1999).
17. Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. & Shleifer, A. Growth in cities. *J. Political Econ.* **100**, 1126–1152 (1992).
18. Hidalgo, C. A. et al. in *Unifying Themes in Complex Systems IX* (eds Morales, A. J. et al.) 451–457 (Springer, 2018).
19. Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317**, 482–487 (2007).
20. Zheng, S., Sun, W., Wu, J. & Kahn, M. E. The birth of edge cities in China: measuring the effects of industrial parks policy. *J. Urban Econ.* **100**, 80–103 (2017).
21. Neffke, F. & Henning, M. Skill relatedness and firm diversification. *Strateg. Manag. J.* **34**, 297–316 (2013).
22. Jara-Figueroa, C., Jun, B., Glaeser, E. L. & Hidalgo, C. A. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proc. Natl Acad. Sci. USA* **115**, 12646–12653 (2018).
23. Park, J. et al. Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. *Nat. Commun.* **10**, 3449 (2019).
24. Klepper, S. Employee startups in high-tech industries. *Ind. Corp. Change* **10**, 639–674 (2001).
25. Breschi, S. & Lissoni, F. in *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (eds Moed, H. F. et al.) 613–643 (Springer, 2005).
26. Feldman, M. P. & Florida, R. The geographic sources of innovation: technological infrastructure and product innovation in the united states. *Ann. Assoc. Am. Geogr.* **84**, 210–229 (1994).
27. Romer, P. M. Endogenous technological change. *J. Political Econ.* **98**, S71–S102 (1990).
28. Moretti, E. *The New Geography of Jobs* (Houghton Mifflin Harcourt, 2012).
29. Fleming, L. & Sorenson, O. Technology as a complex adaptive system: evidence from patent data. *Res. Policy* **30**, 1019–1039 (2001).
30. Mukherjee, S., Romero, D. M., Jones, B. & Uzzi, B. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: the hotspot. *Sci. Adv.* **3**, e1601315 (2017).
31. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
32. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
33. Polanyi, M. *The Tacit Dimension* (Univ. Chicago Press, 2009).
34. Collins, H. *Tacit and Explicit Knowledge* (Univ. Chicago Press, 2010).
35. Friedman, T. L. *The World Is Flat 3.0: A Brief History of the Twenty-First Century* (Picador, 2007).
36. Florida, R. The world is spiky. *Atl. Mon.* **296**, 48–51 (2005).
37. He, C. & Zhu, S. *Evolutionary Economic Geography in China* (Springer, 2019).
38. Zhu, S., He, C. & Zhou, Y. How to jump further and catch up? Path-breaking in an uneven industry space. *J. Econ. Geogr.* **17**, 521–545 (2017).
39. Alshamsi, A., Pinheiro, F. L. & Hidalgo, C. A. Optimal diversification strategies in the networks of related products and of related research areas. *Nat. Commun.* **9**, 1328 (2018).
40. Lee, K. & Lim, C. Technological regimes, catching-up and leapfrogging: findings from the Korean industries. *Res. Policy* **30**, 459–483 (2001).
41. Lee, K. & Malerba, F. Catch-up cycles and changes in industrial leadership: windows of opportunity and responses of firms and countries in the evolution of sectoral systems. *Res. Policy* **46**, 338–351 (2017).
42. Petralia, S., Balland, P.-A. & Rigby, D. L. Unveiling the geography of historical patents in the United States from 1836 to 1975. *Sci. Data* **3**, 160074 (2016).
43. Hall, B. H., Jaffe, A. B. & Trajtenberg, M. *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools* (National Bureau of Economic Research, 2001).
44. Nomaler, Ö., Frenken, K. & Heimeriks, G. On scaling of scientific knowledge production in U.S. metropolitan areas. *PLoS One* **9**, e110805 (2014).
45. Patience, G. S., Patience, C. A., Blais, B. & Bertrand, F. Citation analysis of scientific categories. *Heliyon* **3**, e00300 (2017).

Author contributions
P.-A.B., C.J.-F., S.G.P., M.P.A.S., D.L.R. and C.A.H. all contributed equally to the work and have supervised it jointly.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Pierre-Alexandre Balland

Last updated by author(s): Sep 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Google Chrome used to download data
Data analysis	R and python used for data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data will be made available upon publication.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☒ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative analysis of the determinants of the scaling coefficient for economic output in four distinct domains: scientific publications, patents, industries, and occupations.
Research sample	353 largest U.S. metropolitan statistical areas. The spatial delimitation of these urban areas corresponds to the Core Based Statistical Areas (CBSAs) defined by the United States Office of Management and Budget (OMB)
Sampling strategy	We took the largest U.S. metropolitan areas where data was available on all four measures of economic output.
Data collection	No original data was collected for this publication. Patent data comes from the HistPat dataset. Data on publications comes from Elsevier's Scopus. Industry GDP data comes from the Bureau of Economic Analysis. Occupation data comes from the Bureau of Labor Statistics. Population data comes from the U.S. Census Bureau.
Timing	Patent data and population data were collected starting in 1850 until 2010. All other datasets are cross sectional and correspond to the decade of the 2000's.
Data exclusions	Economic activities related to natural resources were excluded from the results presented in the main text, but they were added to the supplementary material.
Non-participation	NA
Randomization	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging